

RESEARCH ARTICLE

Communication-Efficient Federated Learning for Hybrid VLC/RF Indoor Systems

SEYED MOHAMMAD SHEIKHOESLAMI¹, ARASH RASTI-MEYMANDI¹,
SEYED JAMAL SEYED-MOHAMMADI², JAMSHID ABOUEI³, (Senior Member, IEEE),
AND KONSTANTINOS N. PLATANIOTIS¹, (Fellow, IEEE)

¹The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada

²School of Electrical Engineering, Iran University of Science and Technology, Tehran 13114-16846, Iran

³Department of Electrical Engineering, Yazd University, Yazd 89195-741, Iran

Corresponding author: Jamshid Abouei (abouei@yazd.ac.ir)

The work of Jamshid Abouei and Konstantinos N. Plataniotis was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

ABSTRACT Federated Learning (FL) enables smart devices to collaboratively train Machine Learning (ML) models in a distributed manner without sharing their private data with a central server. However, the disparity between the communication and computation capabilities, and the heterogeneity of local datasets of smart devices degrades the performance of FL in terms of latency and accuracy. To mitigate this effect, we address the problems of device selection and resource allocation in an indoor environment where multiple smart devices participate in the FL process. To further reduce the communication latency, we use Visible Light Communication (VLC) for the downlink transmission while a Radio Frequency (RF) access point supports the uplink transmission in the proposed system. Accordingly, we formulate a multi-objective optimization problem for joint device selection and resource allocation in a hybrid VLC/RF system. Then, using the weight methods, the problem is converted to a single-objective optimization which is solved by incrementally selecting devices in each iteration. The embedded device selection scheme in the proposed algorithm is based on the significance of candidate devices' local gradients and their alignment with the global tendency in order to intelligently prioritize the candidates in the training procedure. Simulation results show that the joint device selection and resource allocation scheme improves the accuracy of the ML model and reduces the average delay in presence of both system and data heterogeneity. Additionally, the proposed hybrid VLC/RF system decreases the latency of the FL process in the downlink mode compared to conventional RF systems.

INDEX TERMS Visible light communication federated learning, device selection, resource allocation, Internet of Things.

I. INTRODUCTION

Machine Learning (ML) has become an undistinguished part of Internet of Things (IoT) networks which employs the massive amount of data generated by the IoT smart devices to enable indoor services, such as localization and activity recognition. Recent ML models require a large dataset to reach a generalization in performance before their employment in real-world applications. However, such a dataset is often distributed among smart devices in a decentralized manner. On the other hand, conventional ML methods only perform the training procedure when the local datasets are

collected in one location; as a result, they cannot guarantee the privacy of clients. Additionally, the transmission of datasets from smart devices to data centers in wireless networks is not feasible due to communication limitations. In this regard, Federated Learning (FL) has been introduced as a ML-based revolutionary solution that can preserve the privacy of clients and reduce the communication burden on the network by applying distributed machine learning [1]. Typically, FL involves several iterations to train an ML model. In each iteration, referred to as a communication round, the selected devices update their local models using the local datasets available at the edge. The updated models are further transmitted to a central server where they are aggregated to construct a new global model. Finally, the

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

aggregated model is distributed among all the smart devices. This process is repeated until the model is converged.

Implementation of FL in 5G wireless networks and beyond comes with various challenges including delay, limited communication resources, restricted energy consumption in smart devices, and particularly the heterogeneity challenge. Generally speaking, there exist two types of heterogeneity for the smart devices in an indoor environment: *i*) data heterogeneity, and *ii*) system heterogeneity. The first one refers to the statistical difference among the local datasets stored at the edge side. The latter is due to the performance variability of smart devices in computation and communication. Both statistical and system heterogeneity can severely affect the performance of the FL algorithm. Statistical heterogeneity has more impact on the convergence and accuracy of the trained ML model, while the problems caused by system heterogeneity are primarily concerned with the delay and energy consumption in each communication round. Efficient device selection and resource allocation are regarded as essential techniques that can improve the performance of FL in the current communication systems [2]. Accordingly, it is imperative to select devices that can not only contribute more to the FL process but also have adequate communication channels. Typically, resource allocation techniques in wireless networks are designed to maximize the average throughput of the network. However, in an FL process, the objective of resource allocation problems is to minimize the delay in each communication round while meeting energy constraints.

Although applying resource allocation techniques in Radio Frequency (RF) systems can improve the FL process in terms of latency and energy consumption, it cannot easily handle the transmission of huge ML models between smart devices and the central server. In recent years, Visible Light Communication (VLC) has been recognized as a complementary technology for indoor RF systems. Compared to RF, VLC provides higher data rates, energy efficiency, and security, however, it requires Line-of-Sight (LoS) link for reliable transmission mode. In most indoor applications, hybrid VLC/RF systems are used for improving the performance of communication, where VLC is utilized for downlink transmission and RF is employed for uplink case. Motivated by the aforementioned challenges, in this paper, we study the problem of device selection and bandwidth allocation for federated learning in a hybrid VLC/RF indoor system. We formulate an optimization problem for joint device selection and resource allocation under delay and energy constraints.

A. RELATED WORKS

The problem of client selection [3], [4], [5], [6], [7], [8] and resource allocation [9], [10], [11], [12], [13], [14] for federated learning in wireless networks has been widely studied in the literature. While some works have addressed the two problems separately, some literature considers joint client selection and resource allocation [15], [16]. In [9], an optimization problem was formulated for joint learning and communication in an FL process, whose goal was to

minimize the total energy consumption under delay constraints. Authors in [10] proposed a joint device scheduling and resource allocation to maximize the model accuracy in a given training time with latency constraints. [11] presented a Hierarchical Federated Edge Learning (HFEL) framework in which model aggregation is partially migrated to edge servers from the cloud. Additionally, it formulated a joint computation and communication resource allocation and edge association problem under the proposed HFEL framework. [12] developed an importance-aware joint data selection and resource allocation algorithm to maximize learning efficiency. The closed-form results for optimal communication resource allocation and data selection were both developed. Reference [13] studied adaptive power allocation for distributed gradient descent in wireless FL within both Orthogonal Multiple Access (OMA) and Non-Orthogonal Multiple Access (NOMA) transmission with “over-the-air-computing”. The work in [14] aimed to accelerate the Deep Neural Network (DNN) training task by jointly optimizing the local training batch size and communication resource allocation to achieve fast training speed while maintaining learning accuracy for both Central Processing Unit (CPU) and Graphics Processing Unit (GPU) scenarios. Reference [17] proposes a probabilistic device selection framework where the candidates with larger gradient absolute have a higher chance of being selected. Authors in [18] develop a device selection scheme that identifies irrelevant updates made by the clients and stops them from being uploaded to the server. To this end, each client checks whether its update is in alignment with the global model.

Authors in [3] considered a multi-criteria FL approach for client selection and maximized the number of users in each round. An online heuristic FL approach was presented in [4] to choose the best candidates based on test accuracy. Authors in [5] proposed a client selection for the FL algorithm, named Client Selection Federated Averaging (CSFedAvg), to mitigate the biases in model training caused by non-independent identically distributed clients. The work in [6] presented a dynamic client selection scheme in a power grid mobile edge computing environment for the FL problem. In [7] an algorithm based on reinforcement learning was proposed for client selection to minimize the energy consumption and the training delay such that users are encouraged to participate in the FL process. Reference [8] modeled the client selection for an FL algorithm with a fairness guarantee as a Lyapunov optimization problem. Reference [15] formulated an optimization problem to minimize the FL loss by considering joint learning, resource allocation, and user selection. Reference [16] follows a later-is-better policy to maximize the weighted sum of selected clients in the FL algorithm for a fixed number of communication rounds in the FL process, while satisfying a long-term energy budget.

There has been limited work on the application of VLC indoor systems for federated learning. Reference [19] formulated a problem for user selection and bandwidth allocation in a hybrid VLC/RF system. The problem was first

TABLE 1. A comparison of the state-of-the-art works on federated learning in wireless systems.

Work	Objective	DS	RA	Comm.
[9]	Energy consumption	✓	×	RF
[10]	Test accuracy	✓	✓	RF
[11]	Energy & delay	×	✓	RF
[12]	Learning efficiency	✓	✓	RF
[13]	Power allocation	×	✓	RF
[14]	Training acceleration	×	✓	RF
[15]	Training loss	✓	✓	RF
[16]	No. of devices & energy	✓	×	RF
[17]	Convergence time & training loss	✓	✓	RF
[18]	Alignment with global tendency	✓	×	RF
[19]	No. of samples	✓	✓	VLC/RF
[20]	No. of samples & model compression	✓	✓	VLC/RF
Proposed Method	Device contribution & delay	✓	✓	VLC/RF

Note- DS: Device Selection, RA: Resource Allocation

separated into two sub-problems: a user selection problem with a known bandwidth allocation solved by a traversal algorithm, and a bandwidth allocation problem with a given user selection, which was solved by a numerical method. The final user selection and bandwidth allocation were obtained by iteratively solving the two sub-problems. The proposed algorithm in [20] first used a model compression method to reduce the size of the FL model in a hybrid VLC/RF system. Then, similar to [19], it solved the user selection and bandwidth allocation problems as two separate problems. In addition, Table 1 compares the existing works with the proposed algorithm in this paper.

B. MAIN CONTRIBUTIONS

In this paper, we study the performance of FL in a hybrid VLC/RF system. In this regard, we consider an indoor environment with multiple users which have several smart devices which generate their particular local datasets used for training an ML model. Regarding the communications, we consider multiple VLC transmitters on the ceiling used for illumination and downlink transmission. We assume that smart devices are equipped with Photodiodes (PD) to receive data from VLC transmitters. For the uplink transmission, it is presumed that each smart device has RF antennas to upload its data. The main contributions of this paper are summarized as the following items:

- In order to improve the performance of the FL algorithm in training DNN models, we formulate an optimization problem for joint device selection and bandwidth allocation in presence of multiple VLC transmitters for downlink and RF transmission for uplink. We also consider both data and system heterogeneity among devices in order to create a more realistic scenario.
- Due to the limited bandwidth in the communication system, only a subset of devices can be selected in each communication round. In addition, due to

the data heterogeneity in the local datasets, different devices contribute non-identically to the training process. To improve the convergence of the FL process under the communication limitations, we propose a novel device selection technique in which smart devices with the most contribution to the training process have a higher chance to be selected in each communication round. To this end, different from [17] and [18] which consider only the absolute or angle of local gradients, the proposed device selection scheme takes into account both the absolute and angle of the local gradients based on the inner product of gradients (GradInn). Accordingly, GradInn prioritizes devices based on the significance and mutuality of local gradients while allowing all to participate in the training procedure which potentially increases the accuracy and decreases the convergence time.

- To evaluate the performance of the proposed FL algorithm, we define separate metrics for accuracy and latency in the FL process. In addition, we demonstrate the overall performance based on a unified parameter from the mentioned metrics. Simulation results show that the proposed GradInn selection scheme increases the convergence speed of the training process compared to the alternatives where the absolute value and angle of gradients are used. In addition, using VLC for downlink in a realistic environment reduces the average delay of the FL algorithm in presence of different levels of system heterogeneity.

II. SYSTEM MODEL AND PROBLEM DESCRIPTION

In this section, we first describe the hybrid VLC/RF indoor environment where several clients possess multiple smart devices equipped with both VLC and RF capabilities. Then, a summarized background of FL is presented in order to get more insight into the proposed framework. Moreover,

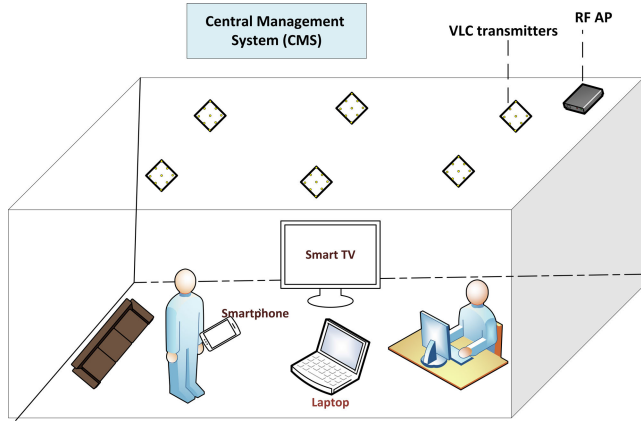


FIGURE 1. Snapshot of the smart hybrid VLC/RF indoor environment.

we elaborate on the communication and computation models of smart devices in the FL process as we further use them to formulate the proposed joint device selection and resource allocation scheme. Table 2 summarizes the used notations in this paper.

A. INDOOR ENVIRONMENT CONFIGURATION

This article considers a hybrid VLC/RF system that employs FL in an indoor environment, such as a smart living room, with multiple RF transceivers, e.g., Wi-Fi modems. It is assumed that N LEDs are installed on the ceiling of the indoor area, known as VLC transmitters in the downlink communication mode, indexed by $\mathbb{T}_v = \{T_1, \dots, T_N\}$. It is worthy to mention that the number and position of the VLC transmitters are set such that the illumination requirements of the indoor environment are satisfied. In this work, we differentiate between users (clients or persons) and devices in the indoor environment. In this regard, we consider N_u users in the set $\mathbb{U} = \{u_1, \dots, u_{N_u}\}$, where each user u_i has N_u^d smart devices, e.g., smart TV, laptop, and smartphone, indexed by $\mathbb{U}_i = \{u_{i1}, \dots, u_{iN_u^d}\}$. Moreover, each smart device $u_{ij} \in \mathbb{U}_i$ has access to a dataset $\mathcal{D}_{u_{ij}} = (\mathbf{x}_{u_{ij}}, \mathbf{y}_{u_{ij}})$, where $\mathbf{x}_{u_{ij}}$ and $\mathbf{y}_{u_{ij}}$ denote the inputs and labels of the data set, respectively. We also presume that each device is equipped with Photodiodes (PDs) as VLC receivers, and mmWave antennas to communicate RF signals. The LoS channel coefficient between smart device u_{ij} and VLC transmitter T_v , $v = 1, \dots, N$, is calculated as follows:

$$h_{T_v, u_{ij}}^{VLC} = \frac{(m + 1)g(\phi_{T_v, u_{ij}})A_r \cos^m(\theta_{T_v, u_{ij}}) \cos(\phi_{T_v, u_{ij}})}{2\pi d_{T_v, u_{ij}}^2}, \quad (1)$$

where m , A_r , $\theta_{T_v, u_{ij}}$, and $\phi_{T_v, u_{ij}}$ denote the Lambertian factor, the area of the receiver, the irradiance angle, and the incident angle, respectively [21]. In addition, $g(\phi_{T_v, u_{ij}}) = (\frac{n_r}{\sin \Phi_c})^2 \Pi(\frac{\phi_{T_v, u_{ij}}}{2\Phi_c})$ represents the gain of the optical concentrator [21], n_r denotes the refractive index, and Φ_c indicates the receiver Field of Vision (FoV) semiangle. The VLC transmitters are connected to a Central Management System

TABLE 2. List of notations used in this paper.

Notation	Definition
$h_{t, u_{ij}}^{ul}$	uplink channel of device u_{ij} at round t
$h_{T_v, u_{ij}}^{VLC}$	channel between VLC T_v and device u_{ij}
$b_{t, u_{ij}}^{ul}$	bandwidth allocated to device u_{ij} at round t
$c_{u_{ij}}$	number of CPU cycles to process one bit of data at device u_{ij}
d_t	total delay in communication round t .
$d_{t, u_{ij}}^{dl}$	downlink delay at device u_{ij} and communication round t .
$d_{t, u_{ij}}^{ul}$	uplink delay at device u_{ij} and communication round t .
$d_{t, u_{ij}}^{comp}$	computation delay at device u_{ij} and communication round t .
$\mathcal{D}_{u_{ij}}$	Local dataset at smart device u_{ij}
$e_{t, u_{ij}}^{ul}$	energy consumption for uplink transmission at device u_{ij} and round t .
$e_{t, u_{ij}}^{comp}$	energy consumption for processing at device u_{ij} and communication round t .
$L_{t, u_{ij}}^{ul}$	number of bits required for local update at device u_{ij} in communication round t
N_u^d	number of devices related to each user
N_u	number of total users
$p_{t, u_{ij}}^{ul}$	power transmission of device u_{ij} at round t
$r_{t, u_{ij}}^{ul}$	uplink transmission rate of device u_{ij} at round t
$S_{t, u_{ij}}$	number of bits required for device u_{ij} to upload its local model at round t
\mathbb{T}_v	set of VLC transmitters
\mathbb{U}_s	set of VLC selected devices
\mathbb{U}_c	set of VLC candidate devices
\mathbb{U}	set of users
ω_t	global model weights at round t .
$\omega_{t, u_{ij}}$	local model weights at device u_{ij} and communication round t .
$\mathbf{x}_{u_{ij}}$	feature set at device u_{ij}
$\mathbf{y}_{u_{ij}}$	label set at device u_{ij}

(CMS) using fast communication tools, like fiber optics. The CMS can be any high performance device in the environment, such as smart TV, or a separate device with sufficient

processing capabilities, e.g., Raspberry Pi. It should be noted that CMS can also act as the central server in the FL algorithm which is responsible for selecting devices, distributing the updated global model among the selected devices, and performing communication tasks such as horizontal/vertical handovers.

B. FL PRELIMINARY

For the sake of completeness and ease to understand the proposed Fed-VL process, we briefly explain the basics of FL. The aim of federated learning is to train a global model using the local data distributed among smart devices while maintaining the privacy of clients. In the FL process, firstly, an ML model (e.g., a deep neural network) is initialized in the central server. The rest of the FL procedure consists of several rounds of updating the ML model and communicating it between smart devices and the CMS. In each round, a pre-defined loss function is firstly minimized over the local datasets at each selected device. Then, the local models are combined in the CMS to minimize the loss function over the collection of all local datasets. This process continues until the global model converges to the desired model that reaches sufficient accuracy for all devices. It is worthy of mention that we consider label heterogeneity among the devices of different users (e.g., $\mathcal{D}_{u_{11}}$, $\mathcal{D}_{u_{21}}$, and $\mathcal{D}_{u_{32}}$ cover different classes). While, the datasets on the devices of the same user (e.g., $\mathcal{D}_{u_{11}}$, $\mathcal{D}_{u_{21}}$, $\mathcal{D}_{u_{32}}$) have distribution heterogeneity, i.e., they have the same classes with the different distributions. We describe the six stages of a single communication round in a typical FL algorithm as follows:

1) DEVICE REQUEST

In this step, a set of \mathbb{U}_c smart devices candidates to join the FL process based on their remaining energy, CPU usage, and data availability. Thus, they transmit the requested information, e.g., channel state information, energy capabilities, and their locations, to the CMS and volunteer to participate in the next communication round of the FL process.

2) DEVICE SELECTION

The CMS chooses $\mathbb{U}_s \subset \mathbb{U}_c$ of the candidates based on the received information to actively participate in the FL process while the other devices remain silent in the current round. Note that selecting all the candidates may increase the convergence speed of FL algorithm because more data samples contribute to the learning process assuming there is no malicious users in the network. However, due to the limited bandwidth and energy capabilities, selecting all the clients is not feasible. Thus, it is important to maximize the number of selected devices in each communication round such that the delay and energy constraints are satisfied. Moreover, based on their datasets, smart devices have different contributions to the learning process. Thus, selecting devices that follow the global tendency of the network can improve the convergence of the ML model.

3) LOCAL UPDATE

The objective of each device $u_{ij} \in \mathbb{U}_s$ is to update the local model parameters, represented by a vector $\omega_{t,u_{ij}}$, by minimizing the local objective function $f(\omega_{t,u_{ij}}, \mathcal{D}_{u_{ij}})$ over all the samples in its local dataset. Accordingly, the new local model parameters, $\omega_{t+1,u_{ij}}$, at client u_{ij} is obtained as follows:

$$\omega_{t+1,u_{ij}} = \underset{\omega_{t,u_{ij}}}{\operatorname{argmin}} f(\omega_{t,u_{ij}}, \mathcal{D}_{u_{ij}}). \quad (2)$$

4) UPLINK TRANSMISSION

The selected smart devices should transmit their updated models, $\omega_{t+1,u_{ij}}$, to the CMS through uplink RF transmission. Since gradient can have information leakage, usually the difference between the gradient in two consequence communication rounds are shared with the CMS.

5) MODEL AGGREGATION

In the communication round t , the CMS is responsible for aggregating the local models and generating a global model. Denoting ω_t as the parameter weights of the aggregated model in communication round t , the global loss, denoted by $F(\omega_t)$, is obtained as follows:

$$F(\omega_t) = \sum_{u_{ij} \in \mathbb{U}_s} \pi_{u_{ij}} f(\omega_t, \mathcal{D}_{u_{ij}}), \quad (3)$$

where $\pi_{u_{ij}} > 0$ denotes the global aggregation weight for device $u_{ij} \in \mathbb{U}_s$ which satisfies $\sum_{u_{ij} \in \mathbb{U}_s} \pi_{u_{ij}} = 1$. In order to minimize the objective function in (3), different techniques have been proposed. In this paper, we apply the well-known FedAvg approach, which uses the weighted average of the parameters in the local models to obtain the global model as follows:

$$\omega_{t+1} = \sum_{u_{ij} \in \mathbb{U}_p} \frac{n_{u_{ij}}}{n} \omega_{t+1,u_{ij}}, \quad (4)$$

where n is the total number of data samples in the collection of local datasets $\mathcal{D} = \cup_{u_{ij} \in \mathbb{U}_p} \mathcal{D}_{u_{ij}}$.

6) MODEL DISTRIBUTION

Finally, after aggregating the local updates in the communication round t , the CMS sends the vector of the model weights, denoted by ω_t , to all the smart devices with the downlink VLC transmission.

C. COMMUNICATION AND COMPUTATION MODEL

To get more insight into the FL process, we formulate the latency and energy models for each communication round.

1) DELAY

The delay of each smart device $u_{ij} \in \mathbb{U}_i$ in communication round t is calculated as follows:

$$d_{t,u_{ij}} = d_{t,u_{ij}}^{\text{dl}} + d_{t,u_{ij}}^{\text{comp}} + d_{t,u_{ij}}^{\text{ul}}, \quad (5)$$

where $d_{t,u_{ij}}^{dl}$, $d_{t,u_{ij}}^{comp}$, and $d_{t,u_{ij}}^{ul}$ represent the delay associated with the downlink VLC transmission, local updates, and uplink RF transmission. Note that the total delay of communication round t is obtained as follows:

$$d_t = \max_{u_{ij} \in \mathbb{U}_i} d_{t,u_{ij}}. \quad (6)$$

Let $b_{t,u_{ij}}^{ul}$ denote the allocated bandwidth to j^{th} device from i^{th} client in the uplink transmission mode. In addition, let $p_{t,u_{ij}}^{ul}$ and $h_{t,u_{ij}}^{ul}$ represent the transmission power and the channel coefficient of the uplink channel between device u_{ij} and the CMS at communication round t . The uplink transmission rate of client u_{ij} at time t , denoted by $r_{t,u_{ij}}^{ul}$, is achieved as

$$r_{t,u_{ij}}^{ul} = b_{t,u_{ij}}^{ul} \log \left(1 + \frac{h_{t,u_{ij}}^{ul} p_{t,u_{ij}}^{ul}}{N_0^{ul} + I_{t,u_{ij}}^{ul}} \right), \quad (7)$$

where N_0^{ul} and $I_{t,u_{ij}}^{ul}$ denote the powers of the white Gaussian noise and the interference related to user $u_{ij} \in \mathbb{U}$ at round t , respectively. Let $S_{t,u_{ij}}$ denote the number of bits required for device u_{ij} to transmit its local model to the CMS. The delay of uplink transmission, $d_{t,u_{ij}}^{ul}$, in device u_{ij} at time t is achieved as

$$d_{t,u_{ij}}^{ul} = \frac{S_{t,u_{ij}}}{r_{t,u_{ij}}^{ul}}. \quad (8)$$

Furthermore, denoting $b_{t,u_{ij}}^{dl}$, $p_{t,u_{ij}}^{dl}$, and N_0^{dl} as the bandwidth, transmission power, and the noise power related to device u_{ij} in the downlink transmission mode, the rate and delay of downlink transmission can be calculated similar to (7) and (8). The delay of computational tasks in communication round t at device u_{ij} is obtained as follows:

$$d_{t,u_{ij}}^{comp} = \frac{c_{u_{ij}} L_{t,u_{ij}}}{f_{u_{ij}}}, \quad (9)$$

where $c_{u_{ij}}$ is the number of required CPU cycles for processing one bit of data, and $L_{t,u_{ij}}$ is the number of bits required for the local update at device u_{ij} and round t . In addition, $f_{u_{ij}}$ is the frequency of the processing unit for client u_{ij} allocated to the FL task.

2) ENERGY CONSUMPTION

The total energy consumption of each participating device $u_{ij} \in \mathbb{U}$ at communication round t is the summation of the energy consumed for computational updates and the uplink RF transmission which is calculated as follows:

$$e_{t,u_{ij}} = e_{t,u_{ij}}^{comp} + e_{t,u_{ij}}^{ul}, \quad (10)$$

where $e_{t,u_{ij}}^{comp}$ and $e_{t,u_{ij}}^{ul}$ represent the energy consumption associated with the local updates and uplink RF transmission of device u_{ij} at communication round t . Note that $e_{t,u_{ij}}^{comp}$

depends on the processing capabilities of client u_{ij} , and can be obtained as follows:

$$e_{t,u_{ij}}^{comp} = \epsilon_{u_{ij}} c_{u_{ij}} f_{u_{ij}}^2 L_{t,u_{ij}}, \quad (11)$$

where ϵ_{ij} is the energy consumption for one CPU cycle. Regarding the energy consumption for the uplink RF transmission, we have

$$e_{t,u_{ij}}^{ul} = d_{t,u_{ij}}^{ul} \times p_{t,u_{ij}}^{ul}. \quad (12)$$

Note that the energy consumption of each client $u_{ij} \in \mathbb{U}$ is bounded by $e_{u_{ij}}^{max}$, i.e., $e_{t,u_{ij}} \leq e_{u_{ij}}^{max}$. In the FL process, we aim to reduce the energy consumption of smart devices which have energy limitation, while reaching a high accuracy for the trained model.

III. PROPOSED DEVICE SELECTION AND RESOURCE ALLOCATION

This section addresses two crucial problems in the FL process, namely, device selection and resource allocation in the proposed hybrid VLC/RF environment. From a machine learning perspective, device selection has a significant effect on the convergence of the FL process. Accordingly, increasing the number of data samples, or equivalently, the number of selected devices, can improve the performance of the FL algorithm [22]. However, from a communication aspect, a larger number of participants in each communication round leads to a higher delay for a fixed bandwidth. Additionally, despite having a major learning contribution, some smart devices might have a poor communication channel with the CMS which can severely degrade the performance of the FL process by slowing down the other participants. It is, thus, necessary to make a trade-off between the selected devices and the bandwidth assigned to each participating device by jointly considering both the problems of device selection and resource allocation. Limited by the communication constraints, the objective of the FL algorithm is to select smart devices which make a higher contribution to training the shared model. In the meantime, it is aimed to maximize the number of selected devices, while minimizing the delay in each communication round. To this end, we propose a device selection scheme named GradInn, which is based on the local gradient of the candidates and the global gradient. Furthermore, we define an optimization problem to jointly optimize the selected devices and the delay of each communication round.

In the following, we explicitly describe our model and formulation for device selection and bandwidth allocation problems. Afterward, a multi-objective optimization problem is formulated for joint device selection and resource allocation with delay and energy constraints. The problem is then converted to a single-objective optimization problem which is shown to be non-convex. Finally, we propose an algorithm for solving the problem by incrementally adding devices to the set of participating devices, \mathbb{U}_s , in the FL algorithm.

A. PROBLEM FORMULATION

The main objective of the proposed FL algorithm is two folds; *i*) maximizing the number of participating devices considering the limited communication resources such that the devices with higher contribution to the training process and with better communication conditions are selected, and *ii*) synchronizing the selected smart devices and minimizing the delay at each communication round by performing bandwidth allocation.

1) PROPOSED GradInn METHOD

The convergence of the global model in the FL framework is heavily dependent upon the data on which the smart devices are trained. Consequently, the more statistical heterogeneity among $\mathcal{D}_{u_{ij}}, \forall u_{ij} \in \mathbb{U}_p$, the slower the global model converges. As a result, the whole training procedure takes longer to complete. On the other hand, since only a limited number of devices are allowed to participate in the aggregation phase, it is imperative to select the ones that could contribute the most to reducing the global loss function $F(\omega_t)$. In this regard, we propose a device selection scheme based on the information extracted from local gradient updates of each device $u_{ij} \in \mathbb{U}_c$ in the framework.

For each candidate $u_{ij} \in \mathbb{U}_c$, we define an indicator variable $v_{t,u_{ij}}$ to indicate whether the smart device is selected in the communication round t or not, given as

$$v_{t,u_{ij}} = \begin{cases} 1, & u_{ij} \text{ is selected in round } t, \\ 0, & u_{ij} \text{ is not selected in round } t. \end{cases} \quad (13)$$

Although increasing the number of participants can improve the performance of the FL algorithm, selecting all candidates is not always feasible due to communication limitations. In this regard, some devices might make a higher contribution to the FL algorithm due to their data set characteristics, e.g., the number of samples or the mutuality of their data distributions. Thus, we define the weighted sum of the selected users as the objective of the device selection problem as follows:

$$\mathcal{F}^{\text{DS}} = \sum_{u_{ij} \in \mathbb{U}_c} \alpha_{t,u_{ij}} v_{t,u_{ij}}, \quad (14)$$

where $\alpha_{t,u_{ij}}$ determines the importance of smart device u_{ij} in communication round t , and is adjusted based on the local updates in each round.

To obtain a decent value for coefficients $\alpha_{t,u_{ij}}$, we assume that each device u_{ij} runs κ local update iterations at each round t , where $\kappa \leq K$ and K represents the number of local iterations at each round. We denote $\omega_{t^+,u_{ij}}$ as the updated weights of smart device u_{ij} after κ local iterations, where $t \leq t^+ \leq t+1$. The corresponding gradient at each device u_{ij} at time step t^+ with respect to ω_t is given by

$$\mathbf{g}_{t^+,u_{ij}} = \omega_{t^+,u_{ij}} - \omega_t. \quad (15)$$

The main idea of the proposed device selection scheme is to choose candidates which contribute more to the training process. In this regard, we prioritize the candidates which

have a higher value of gradient and follow the global tendency of all devices. To measure the similarity of the local updates at round t with the global tendency, the local gradients $\mathbf{g}_{t,u_{ij}}$ are compared with the global gradient \mathbf{g}_t . However, since \mathbf{g}_t is not available at the communication round t , we approximate it with the previously aggregated gradient, i.e., \mathbf{g}_{t-1} given as

$$\mathbf{g}_t \approx \mathbf{g}_{t-1} = \omega_t - \omega_{t-1}. \quad (16)$$

Additionally, to avoid redundant computations in the smart devices, we use the local gradients obtained after $\kappa \leq K$ local iterations, $\mathbf{g}_{t^+,u_{ij}}$. We thus define the following score based on the inner product of \mathbf{g}_{t-1} and $\mathbf{g}_{t^+,u_{ij}}$ for each device $u_{ij} \in \mathbb{U}_c$ such that the ones with higher scores get more chance to participate in the aggregation at communication round t .

$$\alpha_{t,u_{ij}} = \frac{\langle \mathbf{g}_{t^+,u_{ij}}, \mathbf{g}_{t-1} \rangle}{\sum_{u_{ij} \in \mathbb{U}} \|\mathbf{g}_{t^+,u_{ij}}\| \|\mathbf{g}_{t-1}\|}. \quad (17)$$

The proposed GradInn criterion in (17) implies that the local gradients that are more aligned to the previous global gradient should be allowed to engage more in the training procedure. This is because the mutuality of the selected gradients results in faster convergence. In addition, the gradients' amplitude is also informative since larger gradients tend to contribute more to changing the loss function. Note that the alignment and the impact of the local gradients' amplitude are both embedded in the inner product in (17).

2) RESOURCE ALLOCATION

The performance of the FL algorithm relies on the fast and robust communications between the devices and the CMS. Due to the system heterogeneity, i.e., the difference in the communication and computation characteristics of smart devices, it takes different times to perform local updates and upload the new models to the CMS. In a synchronized FL algorithm, the model aggregation step in each round, i.e., equation (4), requires gathering the local updates from all the selected devices. Consequently, the speed of each communication round is limited by the slowest participating device. The aim of FL is to minimize the delay of each communication round, or equivalently, the maximum delay of participating devices while satisfying the energy constraints. Efficient bandwidth allocation in FL enables the CMS to synchronize the selected candidates and minimize the delay at each communication round. In this regard, we define the maximum delay of the participating devices at the communication round t as the objective of the resource allocation problem as follows:

$$\mathcal{F}^{\text{RA}} = \max_{u_{ij} \in \mathbb{U}_c} \{d_{t,u_{ij}} v_{t,u_{ij}}\}. \quad (18)$$

In such a communication system, the total available bandwidth B should be optimized for the devices in each communication round such that none of the devices have a delay and energy consumption higher than thresholds d^{th} and $e_{u_{ij}}^{\text{max}}$, respectively.

By considering the aforementioned analytic results, we define the following multi-objective optimization problem for each communication round of the FL process:

$$\mathcal{P1) \quad \min_{v_{t,u_{ij}}, b_{t,u_{ij}}^{\text{ul}}} \{ \mathcal{F}^{\text{RA}}, \frac{1}{\mathcal{F}^{\text{DS}}} \} \quad (19)$$

$$\text{s.t. C1. } v_{t,u_{ij}} \in \{0, 1\}, \forall u_{ij} \in \mathbb{U}_c, \quad (20)$$

$$\text{C2. } d_{t,u_{ij}} \leq d^{\text{th}}, \forall u_{ij} \in \mathbb{U}_c, \quad (21)$$

$$\text{C3. } e_{t,u_{ij}} \leq e_{u_{ij}}^{\text{max}}, \forall u_{ij} \in \mathbb{U}_c, \quad (22)$$

$$\text{C4. } 0 \leq b_{t,u_{ij}}^{\text{ul}} \leq B^{\text{ul}}, \forall u_{ij} \in \mathbb{U}_c, \quad (23)$$

$$\text{C5. } \sum_{u_{ij} \in \mathbb{U}_c} b_{t,u_{ij}}^{\text{ul}} v_{t,u_{ij}} = B^{\text{ul}}, \forall u_{ij} \in \mathbb{U}_c, \quad (24)$$

where (21) indicates the delay constraint for the current communication round, while (22) guarantees that the energy consumption of none of the smart devices exceeds its limitation. In addition, (23) determines the lower and upper bounds for the allocated bandwidth $b_{t,u_{ij}}^{\text{ul}}$, and (24) guarantees that the total available bandwidth, B^{ul} , is utilized. Note that the downlink VLC bandwidth is also limited by B^{dl} . However, the above joint device selection and resource allocation problem is solved after the distribution of the updated global model among smart devices in downlink VLC transmission, and thus, the optimization variables $\mathcal{P1}$ are the uplink bandwidth, $b_{t,u_{ij}}^{\text{ul}}$, and selection variables $v_{t,u_{ij}}$.

B. SOLUTION OF PROBLEM $\mathcal{P1}$

The optimization problem $\mathcal{P1}$ has two objective functions, namely, \mathcal{F}^{RA} and $\frac{1}{\mathcal{F}^{\text{DS}}}$, which correspond to the resource allocation and device selection objectives, respectively. Additionally, the optimization variables include the integer device selection parameter, $v_{t,u_{ij}}$, and the bandwidth allocated to each device, $b_{t,u_{ij}}^{\text{ul}}$. Hence, according to [23], the optimization problem $\mathcal{P1}$ can be considered a multi-objective optimization problem which is non-convex. The weighted method is one of the popular algorithms to solve multi-objective optimization problems. In this technique, the first step is to make a dimensionless objective function. Since the objective function in $\mathcal{P1}$ has two parameters with different dimensionalities, we divide these parameters by their nominal values. In this paper, we denote N_c and d^{th} as the nominal value for the number of selected devices and the latency term. To show the trade-off, or equivalently, the relative importance between the two parameters, we present w^{DS} and w^{RA} as weight factors for the number of selected devices, and the maximum delay, such that $w^{\text{US}} + w^{\text{RA}} = 1$. Therefore, the optimization problem $\mathcal{P1}$ is transformed to a single-objective optimization problem as follows:

$$\mathcal{P2) \quad \min_{v_{t,u_{ij}}, b_{t,u_{ij}}^{\text{ul}}} \mathcal{F} \triangleq \frac{w^{\text{RA}}}{d^{\text{th}}} \mathcal{F}^{\text{RA}} - \frac{w^{\text{DS}}}{N_c} \mathcal{F}^{\text{DS}} \quad (25)$$

$$\text{s.t. (20) - (24).}$$

It can be shown that the objective function of problem $\mathcal{P2}$ is non-convex. More precisely, the second term of the objective function is an integer programming, while the first term is integer and non-convex. Overall, problem $\mathcal{P2}$ is a mixed-integer non-linear problem, and finding its optimal solution is difficult.

In order to solve $\mathcal{P2}$, we present an iterative algorithm where in each iteration a new candidate device is added to the selection set based on a priority criterion until the total bandwidth B is filled for a specific latency threshold. Let $\mathbb{U}_{s,t}$ denote the set of selected devices at communication round t which is empty at the beginning. We first obtain a minimum bandwidth that satisfies the delay and energy constraints of the candidate devices in \mathbb{U}_c . Substituting (5), (7), and (8) in constraint (21), we obtain

$$d_{t,u_{ij}}^{\text{dl}} + d_{t,u_{ij}}^{\text{comp}} + \frac{S_{t,u_{ij}}}{b_{t,u_{ij}}^{\text{ul}} \log(1 + \frac{h_{t,u_{ij}}^{\text{ul}} p_{t,u_{ij}}^{\text{ul}}}{N_0 + I_{t,u_{ij}}^{\text{ul}}})} \leq d^{\text{th}}. \quad (26)$$

Defining $\mathcal{A}_{t,u_{ij}} = d_{t,u_{ij}}^{\text{dl}} + d_{t,u_{ij}}^{\text{comp}}$ and $\mathcal{B}_{t,u_{ij}} = \frac{S_{t,u_{ij}}}{\log(1 + \frac{h_{t,u_{ij}}^{\text{ul}} p_{t,u_{ij}}^{\text{ul}}}{N_0 + I_{t,u_{ij}}^{\text{ul}}})}$,

inequality (26) can be reformulated into

$$b_{t,u_{ij}}^{\text{ul}} \geq \frac{\mathcal{B}_{t,u_{ij}}}{d^{\text{th}} - \mathcal{A}_{t,u_{ij}}}. \quad (27)$$

Additionally, substituting (10) and (12) in **C3**, and using (7), we obtain

$$e_{t,u_{ij}}^{\text{comp}} + \frac{p_{t,u_{ij}}^{\text{ul}} S_{t,u_{ij}}}{b_{t,u_{ij}}^{\text{ul}} \log(1 + \frac{h_{t,u_{ij}}^{\text{ul}} p_{t,u_{ij}}^{\text{ul}}}{N_0 + I_{t,u_{ij}}^{\text{ul}}})} \leq e_{u_{ij}}^{\text{max}}. \quad (28)$$

Similarly, setting $\mathcal{C}_{t,u_{ij}} = e_{t,u_{ij}}^{\text{comp}}$ and $\mathcal{D}_{t,u_{ij}} = \frac{p_{t,u_{ij}}^{\text{ul}} S_{t,u_{ij}}}{\log(1 + \frac{h_{t,u_{ij}}^{\text{ul}} p_{t,u_{ij}}^{\text{ul}}}{N_0 + I_{t,u_{ij}}^{\text{ul}}})}$,

we have

$$b_{t,u_{ij}}^{\text{ul}} \geq \frac{\mathcal{D}_{t,u_{ij}}}{e_{ij}^{\text{max}} - \mathcal{C}_{t,u_{ij}}}. \quad (29)$$

Hence, constraints **C2** and **C3** are satisfied if (27) and (29) hold, or equivalently,

$$b_{t,u_{ij}}^{\text{ul}} \geq \max\{ \frac{\mathcal{B}_{t,u_{ij}}}{d^{\text{th}} - \mathcal{A}_{t,u_{ij}}}, \frac{\mathcal{D}_{t,u_{ij}}}{e_{ij}^{\text{max}} - \mathcal{C}_{t,u_{ij}}} \}. \quad (30)$$

Without loss of generality, we assume that the right side of (30) satisfies **C4**, i.e, it is in range $[0, B^{\text{ul}}]$, otherwise, we remove the corresponding device from the selection process. We define a priority metric as follows:

$$\gamma_{u_{ij}} = w^{\text{DS}} \alpha_{t,u_{ij}} - w^{\text{RA}} \left(d_{t,u_{ij}}^{\text{dl}} + d_{t,u_{ij}}^{\text{comp}} + \tilde{d}_{t,u_{ij}}^{\text{ul}} \right), \quad (31)$$

where $\tilde{d}_{t,u_{ij}}^{\text{ul}}$ is the maximum uplink delay for device u_{ij} which is obtained with the minimum bandwidth $b_{t,u_{ij}}^{\text{ul}} =$

$\max\{\frac{B_{t,u_{ij}}}{d^{\text{th}}-\mathcal{A}_{t,u_{ij}}}, \frac{D_{t,u_{ij}}}{e_{t,u_{ij}}^{\text{max}}-\mathcal{C}_{t,u_{ij}}}\}$. In each iteration, the smart device with the highest priority metric is selected. In other words,

$$v_{u_{ij}^*} = 1, \text{ where } u_{ij}^* = \arg \max_{u_{ij} \in \mathbb{U}_c} \gamma_{u_{ij}}. \quad (32)$$

Therefore, the set of selected devices, \mathbb{U}_s and integer variable v_{t,u_{ij}^*} in optimization problem $\mathcal{P2}$ are known and updated in each iteration of Algorithm 1. Moreover, since constraint **C1** is related to the selection of the candidate devices, we remove it from the optimization problem. The modified version of $\mathcal{P2}$ is given as

$$\mathcal{P3) \quad \frac{w^{\text{RA}}}{d^{\text{th}}} \max_{u_{ij} \in \mathbb{U}_s} \{d_{t,u_{ij}}\} - \frac{w^{\text{DS}}}{n_s} \sum_{u_{ij} \in \mathbb{U}_s} \alpha_{t,u_{ij}} \quad (33)$$

$$\text{s.t. (21) - (24),}$$

where n_s is the number of selected devices until the current iteration.

In order to solve optimization problem $\mathcal{P3}$, we set $Z = \max_{u_{ij} \in \mathbb{N}_c} d_{t,u_{ij}}$, and add constraint $d_{t,u_{ij}} \leq Z$. Since the first term in the optimization problem $\mathcal{P3}$ is fixed and does not affect the solution, we remove it from the objective function. Finally, the maximization problem is converted to the following minimization problem:

$$\mathcal{P4) \quad \min_{b_{t,u_{ij}}^{\text{ul}}} Z \quad (34)$$

$$\text{s.t. (21) - (24)}$$

$$\mathbf{C6.} \quad Z \geq d_{t,u_{ij}}. \quad (35)$$

Problem $\mathcal{P4}$ is a convex optimization problem and can be solved by standard convex optimization techniques, such as Barrier methods [23].

C. PROPOSED FEDERATED LEARNING FRAMEWORK

In the proposed FL framework, the joint device selection and resource allocation scheme has the following six steps; *i*) each device $u_{ij} \in \mathbb{U}_c$ performs κ local update iterations and calculates the gradient $\mathbf{g}_{t^+,u_{ij}}$, *ii*) the inner product $\langle \mathbf{g}_{t^+,u_{ij}}, \mathbf{g}_{t-1} \rangle$ is calculated in each device and then the results are sent to the CMS, *iii*) the CMS assigns $\alpha_{t,u_{ij}}$ according to (17) to each device, *iv*) the CMS solves the optimization problem $\mathcal{P1}$ using Algorithm 1 and determines which device is qualified to participate in the current FL round, *v*) the selected devices update their local weights $\omega_{t^+,u_{ij}}$ for $K - \kappa$ more iterations, and *vi*) finally, the updated weights, $\omega_{t+1,u_{ij}}$, are sent to the CMS for the aggregation.

D. COMPUTATIONAL COMPLEXITY OF ALGORITHM 1

In this subsection, we present the computational complexity analysis of Algorithm 1 by the following proposition.

Proposition 1: The computational complexity of Algorithm 1 is of order $O(N_c N_s)$, in which N_c is the number of candidates and N_s is the total number of selected devices.

Algorithm 1 Proposed Algorithm to Solve $\mathcal{P2}$

Input: The set of candidate clients, \mathbb{U}_c , the number of bits required for device, $S_{t,u_{ij}}$, the total bandwidth, B^{ul} .

- 1: $n_s = 0$
- 2: $v_{u_{ij}} = 0, \forall u_{ij} \in \mathbb{U}_c$.
- 3: **while** $\sum_{u_{ij} \in \mathbb{U}_c} b_{t,u_{ij}}^{\text{ul}} v_{t,u_{ij}} \leq B^{\text{ul}}$ **do**
- 4: $r_{t,u_{ij}}^{\text{dl}} \leftarrow b_{t,u_{ij}}^{\text{dl}} \log\left(1 + \frac{h_{t,u_{ij}}^{\text{dl}} p_{t,u_{ij}}^{\text{dl}}}{N_0^{\text{dl}} + I_{t,u_{ij}}^{\text{dl}}}\right), \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 5: $d_{t,u_{ij}}^{\text{dl}} \leftarrow \frac{S_{t,u_{ij}}}{r_{t,u_{ij}}^{\text{dl}}}, \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 6: Determine $d_{t,u_{ij}}^{\text{comp}}$ as in (9), $\forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 7: $\mathcal{A}_{t,u_{ij}} \leftarrow d_{t,u_{ij}}^{\text{dl}} + d_{t,u_{ij}}^{\text{comp}}, \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 8: $\mathcal{B}_{t,u_{ij}} \leftarrow \frac{S_{t,u_{ij}}}{\log\left(1 + \frac{h_{t,u_{ij}}^{\text{ul}} p_{t,u_{ij}}^{\text{ul}}}{N_0^{\text{ul}} + I_{t,u_{ij}}^{\text{ul}}}\right)}, \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 9: $\mathcal{C}_{t,u_{ij}} \leftarrow e_{t,u_{ij}}^{\text{comp}}, \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 10: $\mathcal{D}_{t,u_{ij}} \leftarrow \frac{p_{t,u_{ij}}^{\text{ul}} S_{t,u_{ij}}}{\log\left(1 + \frac{h_{t,u_{ij}}^{\text{ul}} p_{t,u_{ij}}^{\text{ul}}}{N_0^{\text{ul}} + I_{t,u_{ij}}^{\text{ul}}}\right)}, \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 11: $b_{t,u_{ij}}^{\text{ul}} \leftarrow \max\left\{\frac{B_{t,u_{ij}}}{d^{\text{th}}-\mathcal{A}_{t,u_{ij}}}, \frac{D_{t,u_{ij}}}{e_{t,u_{ij}}^{\text{max}}-\mathcal{C}_{t,u_{ij}}}\right\}, \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 12: Obtain uplink rate using (7), $\forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 13: Determine uplink delay, $\tilde{d}_{t,u_{ij}}^{\text{ul}}$, using (8), $\forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 14: $\gamma_{u_{ij}} = w^{\text{DS}} \alpha_{t,u_{ij}} - w^{\text{RA}} \left(d_{t,u_{ij}}^{\text{dl}} + d_{t,u_{ij}}^{\text{comp}} + \tilde{d}_{t,u_{ij}}^{\text{ul}}\right), \forall u_{ij} \in \mathbb{N}_c, v_{u_{ij}} \neq 1$.
- 15: $u_{ij}^* = \arg \max_{u_{ij} \in \mathbb{U}_c} \gamma_{u_{ij}}$.
- 16: $v_{u_{ij}^*} = 1$
- 17: $n_s \leftarrow n_s + 1$
- 18: Solve optimization problem $\mathcal{P4}$.
- 19: **end while**

Output: The set of selected devices \mathbb{U}_s , bandwidth $b_{t,u_{ij}}^{\text{ul}}, \forall u_{ij} \in \mathbb{U}_c$.

Proof: Algorithm 1 contains one “while” loop in lines 3-19. In each iteration of the loop, a series of computations are performed (lines 4-17), and an optimization problem is solved (line 18). The computation complexity of each line for an iteration of the loop, from line 4 to line 17, equals the number of candidates which are not selected yet, i.e., $N_c - n_s$, where n_s is the counter variable in Algorithm 1. Additionally, in line 17 of Algorithm 1, a convex optimization problem with some constraints is solved using barrier methods. According to subsection 11.5 of the Boyd optimization book [23], the complexity of barrier methods is of order $O(n_s)$. Thus, the overall complexity of a single iteration of the “while” loop

Algorithm 2 Pseudocode of the Proposed FL Algorithm

Input: The number of users, N_s .

Initialization:

- 1: Randomly set ω_t .
- 2: **while** convergence **do**

Model distribution:

- 3: Send the model weights ω_t to all the clients using VLC links.

Device Selection and Bandwidth Allocation:

- 4: Obtain ω_{t+} by performing κ local update iterations.
- 5: $\mathbf{g}_{t+}^i \leftarrow \omega_t - \omega_{t+}^i$.
- 6: Calculate scores $\alpha_{t,i,j}$ according to (17)
- 7: Obtain $v_{t,u_{ij}}$ and bandwidth $b_{t,u_{ij}}$ by solving optimization problem **P2** using Algorithm 1.

Local update:

- 8: Perform $K - \kappa$ more local iterations in the selected devices $v_{t,u_{ij}}$.

Uplink transmission:

- 9: Upload the updated local weights $\omega_{t,i,j}$ to the central server.

Model aggregation:

- 10: Obtain ω_t according to (4).
- 11: $t \leftarrow t + 1$

12: **end while**

Output: Weights ω_t .

is N_c . The loop is repeated for N_s iterations, hence, the computation complexity of the Algorithm 1 is of order $O(N_c N_s)$.

IV. EXPERIMENTATION

In this section, we conduct several experiments to evaluate the proposed joint device selection and resource allocation FL algorithm in a hybrid VLC/RF system and compare it with other alternatives proposed in [24] and [17]. First, the simulation setup of the smart indoor environment is described and summarized in Table 3 [25]. We then introduce multiple metrics in terms of the model classification accuracy and latency to evaluate the performance of the proposed algorithm against the baselines. Finally, the simulation results are reported.

A. SIMULATION SETUP

1) COMMUNICATION SETUP

We consider a typical indoor environment with size $9 \times 6 \times 3 \text{ m}^3$ and $N_v = 6$ VLC transmitters placed on the ceiling. The downlink total bandwidth and the corresponding downlink transmission power are set to $B^{\text{dl}} = 20 \text{ MHz}$ and $p_{t,u_{ij}}^{\text{dl}} = 1.3 \text{ watts}$, respectively. We consider $N_u = 5$ users in the smart environment, where each user owns $N_{u_i}^d = 6$ smart devices, each equipped with $P = 4$ PDs as VLC receivers. It is assumed that $N_c = 15$ devices randomly candidate to participate in the FL process at each communication round. We opt to select an uplink RF transmission with total band-

TABLE 3. Simulation parameters [25].

Parameter	value
Room size	$6 \times 4 \times 3 \text{ m}^3$
Downlink VLC bandwidth (B^{dl})	20 MHz
Downlink transmit power ($p_{t,u_{ij}}^{\text{dl}}$)	1.3 W
Noise power for downlink (N_0^{dl})	10^{-21} W
Semi-angle of LEDs at half power ($\theta_{1/2}$)	60°
FoV of each PD (Φ_c)	120°
Gain of optical filter	1
Gain of optical concentrator	1
Area of PDs (A_r)	1 cm^2
PD responsivity (R)	0.8 A/W
Uplink RF bandwidth (B^{ul})	10 MHz
Uplink transmit power ($p_{t,u_{ij}}^{\text{ul}}$)	1 W
Noise power for uplink (N_0^{ul})	$3.89 \times 10^{-21} \text{ W}$
CPU energy consumption coefficient (η_i)	10^{-24}
CPU cycles for processing one bit data (c_i)	40 Hz
CPU frequency (f_i)	10^9

width $B^{\text{ul}} = 10 \text{ MHz}$ and complex Gaussian noise with power $N_0^{\text{ul}} = 3.89 \times 10^{-21} \text{ watts}$.

2) MACHINE LEARNING SETUP

We use the MNIST dataset and its extension, EMNIST, for the commonly-used handwritten image classification task. MNIST contains 70, 000 data samples in the form of 28×28 images. There are 10 classes in the dataset corresponding to handwritten digits from 0 to 9. We also evaluate the algorithms on the heterogeneous dataset EMNIST where each device has $|D_{u_{ij}}| = 450$ data samples of 90% of the classes.

B. PERFORMANCE METRICS

The performance of an FL framework can be examined from the perspectives of classification accuracy, and latency. We, thus, define the following metrics to evaluate the proposed algorithm for each of the mentioned aspects of:

1) TEST ACCURACY

For an ML classification model, the test accuracy is defined as the number of correctly labeled samples in the test set, denoted by n^{corr} , to the total number of test samples. In our work, we assume that a balanced test set, denoted by D^{test} , is available in the CMS and can be used to evaluate the trained model. In order to make a realistic evaluation of the model accuracy in the FL algorithm, we define metric η_1 as the average accuracy over multiple runs given as

$$\eta_1 = \mathbb{E} \left[\frac{n^{\text{corr}}}{n^{\text{test}}} \right], \tag{36}$$

where $\mathbb{E}[\cdot]$ presents the expectation operator.

2) LATENCY

To acquire a fair understanding of the latency between the devices and CMS in each communication round of the FL algorithms, the cumulative latency of the FL process is

defined as follows:

$$\eta_2 = \sum_{t=1}^R d_t, \quad (37)$$

where d_t is the delay in communication round t .

3) UNIFIED METRIC

The goal of FL is to maximize the accuracy while minimizing the communication and computation costs. To reach an overall understanding of the FL algorithms, we define a unified metric by combining the aforementioned parameters as follows:

$$\eta = \frac{\eta_1}{\eta_1^{\max}} \exp\left(-\frac{\eta_2}{\eta_2^{\max}}\right), \quad (38)$$

where $\eta_1^{\max} = 100$ and η_2^{\max} indicate the nominal values of the test accuracy and the latency, respectively.

4) SYSTEM HETEROGENEITY

To better understand the effect of system heterogeneity on the average delay of each communication round, we define the following metric to represent the system heterogeneity:

$$H^{\text{SYS}} = 1 - \frac{\min_{u_{ij}} \{\hat{d}_{u_{ij}}\}}{\max_{u_{ij}} \{\hat{d}_{u_{ij}}\}}, \quad (39)$$

where $\hat{d}_{u_{ij}}$ represents the delay of each device $u_{ij} \in \mathbb{U}_s$ for a given bandwidth $b_{u_{ij}} < B^{\text{ul}}$ in each communication round. Note that $H^{\text{SYS}} \rightarrow 0$, when there is no system variability, i.e., when all devices have the same delay, and $H^{\text{SYS}} \rightarrow 1$ for high system heterogeneity.

C. PERFORMANCE EVALUATION

In this subsection, we evaluate the performance of the proposed FL algorithm in three scenarios. In the first one, we present the simulation results of the proposed GradInn device selection scheme and compare it with the alternatives [17] in terms of test accuracy η_1 . In the second scenario, we analyze the effect of the proposed resource allocation algorithms in both hybrid VLC/RF and conventional RF systems in terms of delay for various levels of system heterogeneity. Finally, in the third scenario, the simulation results of the proposed joint device selection and resource allocation algorithm are presented for both hybrid VLC/RF and conventional RF systems.

1) SCENARIO I

In order to evaluate the performance of the proposed device selection algorithm, we set $w^{\text{RA}} = 0$ and $w^{\text{DS}} = 1$ to eliminate the effect of the resource allocation term in the objective function of problem $\mathcal{P}2$, i.e., (25). We compare the proposed GradInn algorithm addressed in (17) with the following baselines in terms of accuracy:

i) The case where the angle between the local gradients with the previously aggregated gradient is used (namely

GradAng) [18]. In this scheme, the parameter $\alpha_{t,u_{ij}}$, formulated in (17), is obtained as

$$\alpha_{t,u_{ij}} = \arccos\left(\frac{\langle \mathbf{g}_{t^+,u_{ij}}, \mathbf{g}_{t-1} \rangle}{\|\mathbf{g}_{t^+,u_{ij}}\| \|\mathbf{g}_{t-1}\|}\right). \quad (40)$$

ii) The absolute value of local gradients is used for device selection (GradAbs). We use this scheme similar to the work proposed in [17] with a minor modification, where

$$\alpha_{t,u_{ij}} = \frac{|\mathbf{g}_{t^+,u_{ij}}|}{\sum_{u_i \in \mathbb{U}} |\mathbf{g}_{t^+,u_i}|}. \quad (41)$$

Fig. 2 compares the test accuracy of the proposed GradInn algorithm and baselines with respect to communication rounds for different number of selected devices, namely $N_s = 6, 9,$ and 12 . Overall, it is clear that the proposed GradInn algorithm performs better than the GradAng and GradAbs methods. For the GradAbs [17] method, the devices with bigger local gradients are selected. In this situation, local gradients of the selected devices might not be in the same direction, and thus mitigate each other's effects. Although the GradAng method selects devices that are in the same direction of the previously aggregated gradient, it does not take into account the absolute value of the local gradients. Thus, the selected devices do not necessarily have large local gradient, which leads to slow convergence of the algorithm. The proposed GradInn technique, however, takes into account both the absolute value and the direction of the local gradient. More precisely, it chooses the devices that have a larger value of local gradient applied in the direction of the globally aggregated gradient. In addition, as the number of selected devices increases from $N_s = 6$ to $N_s = 12$, the convergence speed of the device selection schemes reduces. This is due to the fact that the device selection methods have to pick a larger portion of the randomly candidate devices. Thus, they have a smaller degree of freedom in the selection process, and hence they behave similar more similarly compared to the case where they have a larger degree of freedom.

2) SCENARIO II

We evaluate the performance of the proposed resource allocation algorithm in the hybrid VLC/RF systems and compare it with the alternative, where only RF is used for both downlink and uplink transmission modes [24]. To this end, we set $w^{\text{DS}} = 0$ and $w^{\text{RA}} = 1$ in the objective function of the optimization problem formulated in (25). Fig. 3 compares the average delay of the communication rounds in the FL algorithm for both hybrid VLC/RF and conventional RF systems, described in (6), while popular neural networks, namely, AlexNet, VGG-M, VGG-S, GoogleNet, and MobileNet, are used as the global model of the FL framework.

The system heterogeneity, imposed by the difference in communication and computation capabilities of devices, leads to discrepancy in delay of the participating devices. In this situation, the delay in each communication round, which is limited to the slowest device, can be reduced by

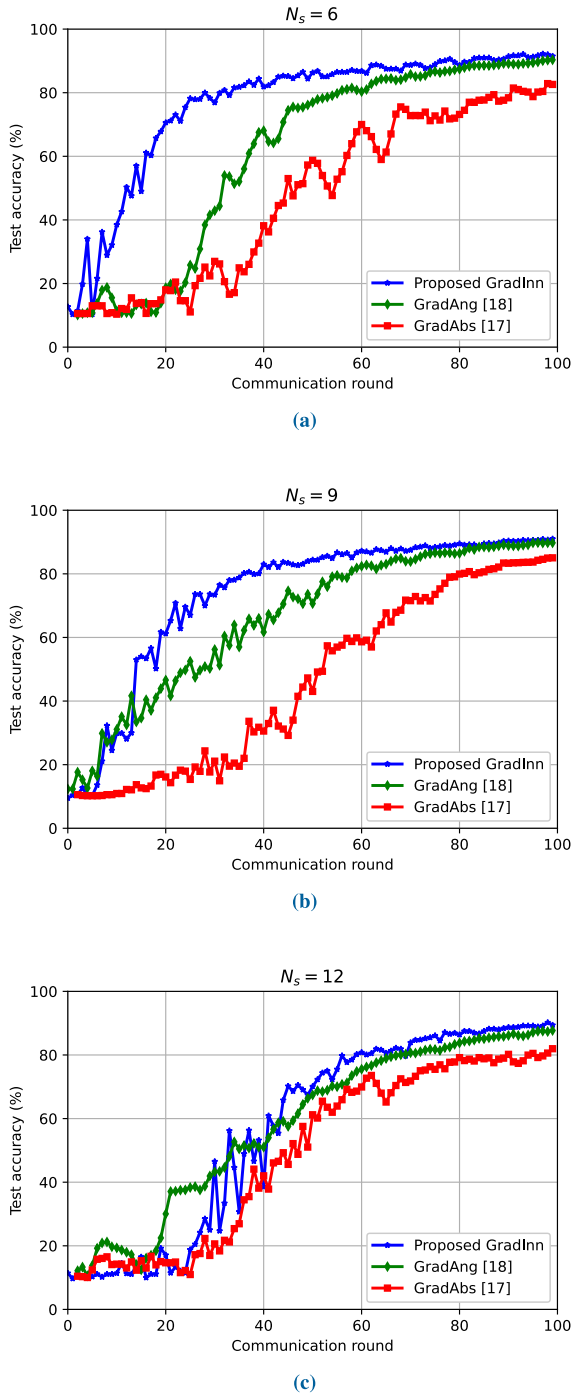


FIGURE 2. Comparing the performance of the proposed device selection algorithm (GradInn) with other techniques, namely GradAbs [17] and GradAng [18] for (a) $N_s = 6$, (b) $N_s = 9$, and (c) $N_s = 12$ selected devices.

optimally allocating bandwidth to the participating devices such that the slow participants are given more bandwidth compared to fast devices.

Fig. 4 compares the average delay per communication round for different values of H^{SYS} in (39) for both the proposed hybrid VLC/RF system and the conventional RF model similar to the work in [24]. In addition, we consider two cases, where a) the proposed resource allocation algorithm

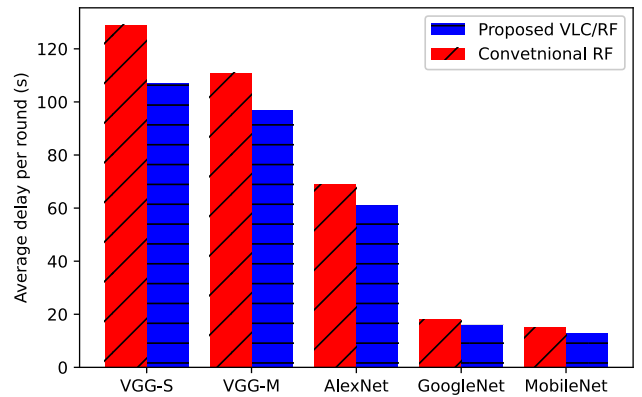


FIGURE 3. Comparison of the average delay of a communication round for popular neural networks with the proposed hybrid VLC/RF and conventional RF systems.

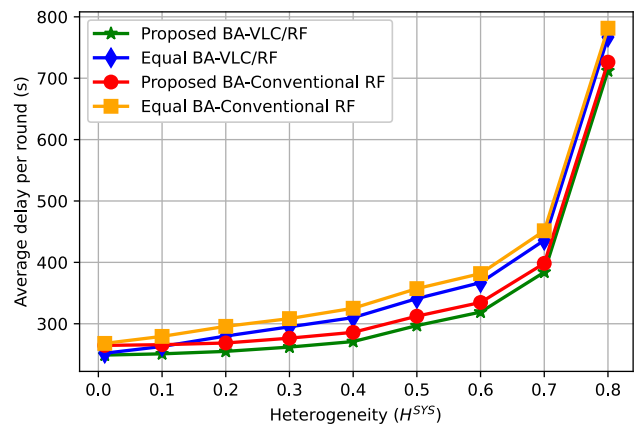


FIGURE 4. Average delay per communication round versus heterogeneity with and without applying the proposed resource allocation for the proposed hybrid VLC/RF system and the conventional RF system used in [24].

is applied, and b) an equal bandwidth allocation among the selected candidates ($b_{u_{ij}}^{ul} = \frac{B^{ul}}{N_s}$) is considered. As seen from Fig. 4, both hybrid VLC/RF and conventional RF schemes have almost the same average delay for low values of H^{SYS} . However, as the system heterogeneity increases, the proposed resource allocation algorithm achieves lower average delay than the equally partitioned bandwidth allocation method. In addition, in the case of the extreme system heterogeneity (i.e., $H^{SYS} > 0.7$), the four schemes start to perform similarly. This behavior stems from the fact that the variation between the delay of participants is too high that cannot be compensated by the proposed bandwidth allocation scheme. Moreover, a sudden increase is observed in the average delay after $H^{SYS} = 0.7$ which is due to the high variability of communication and computation capabilities between the devices, or equivalently, the high variation in the delay of different devices. Furthermore, in both the proposed and equal bandwidth allocation methods, the VLC/RF system has lower average delay per communication round since it provides faster downlink communication compared to the conventional RF system.

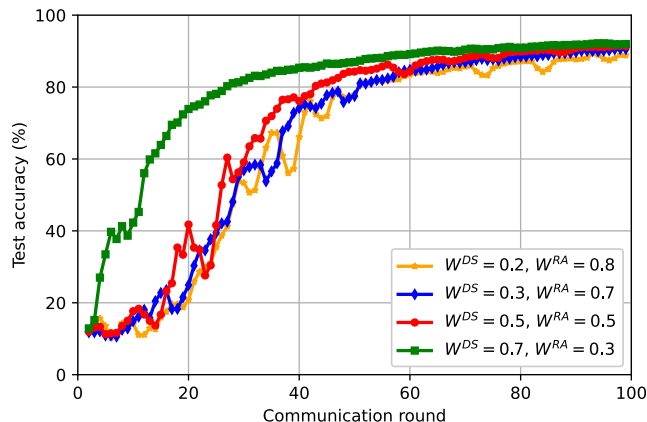


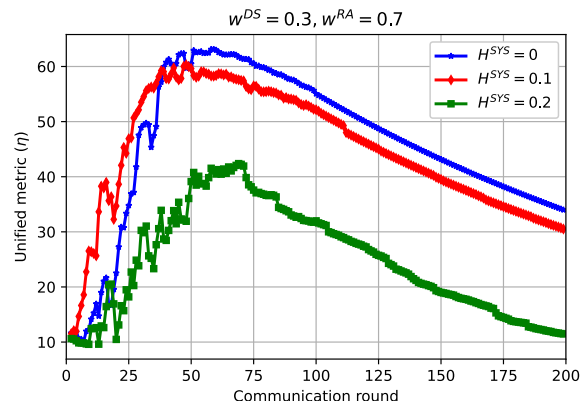
FIGURE 5. Test accuracy versus communication rounds for different values of w^{DS} and w^{RA} in the hybrid VLC/RF system.

3) SCENARIO III

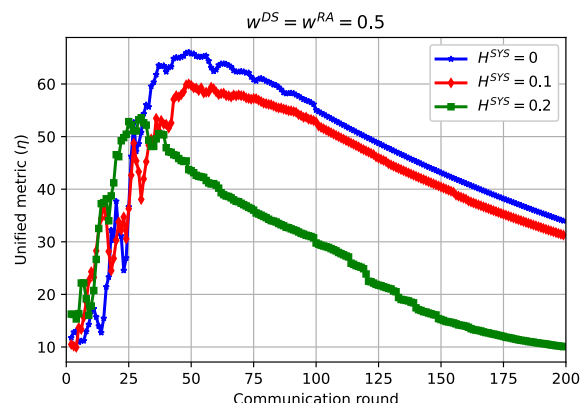
As shown in Scenarios I and II, the proposed device selection and resource allocation techniques can improve the accuracy and latency in the FL algorithm. Accordingly, the weighting coefficients w^{DS} and w^{RA} in (25) have a direct impact on both test accuracy and the delay of the proposed FL algorithm. To better understand this affect, we alternative the importance weights and evaluate the proposed FL framework in Algorithm 2 in terms of test accuracy and latency. In this regard, Fig. 5 depicts the test accuracy for different values of w^{DS} and w^{RA} . As observed, increasing the weight values of the device selection term in the joint device selection and resource allocation algorithm gives rise to the test accuracy and convergence speed of the learning curve. However, this leads to a raise in the completion time and the average delay at each round.

To have a general understanding of the proposed algorithm with respect to the test accuracy and delay, Fig. 6 illustrates the unified metric η , formulated in (38), versus communication rounds and for different values of system heterogeneity H^{SYS} . We consider three cases where (a) $w^{DS} = 0.3, w^{RA} = 0.7$, (b) $w^{DS} = 0.5, w^{RA} = 0.5$, and (c) $w^{DS} = 0.7, w^{RA} = 0.3$. As it can be seen from the figure, a larger system heterogeneity in the setup results in a lower η . This comes from the fact that higher system heterogeneity leads to higher delay in each round and thus a lower η . Moreover, as the importance weight of the devices selection term, w^{DS} , in (25) increases, the curve gets a sharper ramp with respect to communication rounds to reach its peak. The reason is that a higher priority for selection has been given to the smart devices with more learning contributions. However, it is seen that the curves reduce gradually after reaching their peaks. This is because the test accuracy does not improve as much as for the initial communication rounds.

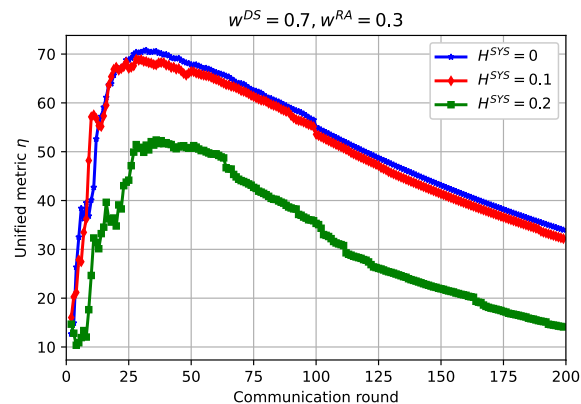
Finally, Table 4 shows the quantitative results of the unified metric η for both the hybrid VLC/RF and conventional RF systems after the completion of the FL process. It can be concluded from the table that the VLC/RF system reaches a higher value of η for the proposed VLC/RF system compared to the conventional RF communication. The main reason



(a)



(b)



(c)

FIGURE 6. Comparing the unified metric η for different values of heterogeneity in (a) $w^{DS} = 0.3, w^{RA} = 0.7$, (b) $w^{DS} = 0.5, w^{RA} = 0.5$, and (c) $w^{DS} = 0.7, w^{RA} = 0.3$.

behind this phenomenon is the lower latency of the VLC transmission in downlink which results in a lower cumulative delay, and thus a higher η . Additionally, as the system heterogeneity increases, the value of η reduces after the FL process completion since the average delay per communication round becomes larger with respect to the increase in system heterogeneity.

TABLE 4. Comparison between the unified metric η obtained in the hybrid VLC/RF and conventional RF systems for different heterogeneity levels.

Algorithms	(w^{DS}, w^{RA})	$H^{SYS} = 0$			$H^{SYS} = 0.1$			$H^{SYS} = 0.2$		
		(0.7, 0.3)	(0.5, 0.5)	(0.3, 0.7)	(0.7, 0.3)	(0.5, 0.5)	(0.3, 0.7)	(0.7, 0.3)	(0.5, 0.5)	(0.3, 0.7)
Conventional RF		68.61	63.08	60.59	67.20	57.32	57.77	50.56	51.29	39.39
Proposed hybrid VLC/RF		70.79	66.03	63.12	68.94	60.01	60.29	52.39	53.58	42.37

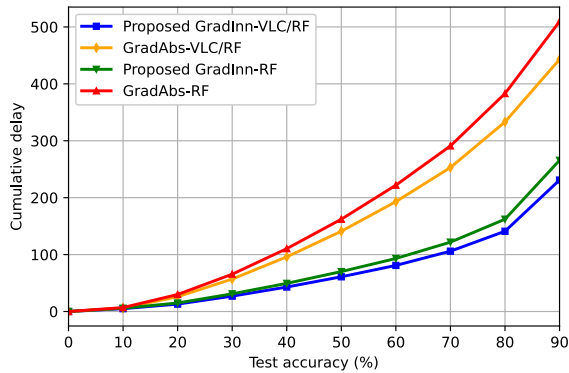


FIGURE 7. Cumulative delay versus test accuracy for the proposed GradInn and GradAbs in both conventional RF and hybrid VLC/RF systems.

The communication efficiency in the FL algorithm depends on *i*) which devices are selected and *ii*) how the communication resources are assigned to the selected devices. To make a better understanding of the communication efficiency of the proposed hybrid VLC/RF system model, Fig. 7 compares the cumulative delay required to reach a specified accuracy, e.g., 80% with the alternatives. Our proposed algorithm selects the devices with a higher contribution to the training process, thus, in long term, less number of communication rounds, and hence a lower cumulative delay, is required for the convergence of the FL algorithm resulting in a fewer model exchange between devices and the CMS. Additionally, in each communication round, the optimal resource allocation is applied which makes optimal usage of the available bandwidth. The proposed hybrid VLC/RF system slightly reduces the delay due to faster communication in the downlink mode.

V. CONCLUSION

This paper put forward a communication-efficient FL in a smart indoor environment with multiple VLC transmitters for the downlink transmission and an RF access point for the uplink transmission. We assumed that several users (clients or persons) are present in the indoor environment where each user has multiple smart devices equipped with PDs to receive VLC signals. We proposed an FL framework to train a deep learning models in which devices with the highest contribution to the learning process and low latency were prioritized in the selection scheme. Towards this goal, we formulated the problems of device selection and resource allocation as a multi-objective optimization problem which was then converted to a single-objective problem using weighting methods. Furthermore, the optimization problem was solved by incrementally adding the candidates to the set of selected devices in each round of FL process. We conducted several experiments to evaluate the performance of the proposed

joint device selection and resource allocation algorithm in a hybrid VLC/RF indoor system. A comprehensive simulation was conducted to show that the proposed GradInn scheme not only reduces the average delay but also increases the convergence speed of the FL algorithm in presence of both system and data heterogeneity. In addition to that, the proposed hybrid VLC/RF system could further reduce the average delay in downlink transmission mode compared to the conventional systems based on RF. For the future works, we suggest to use Intelligent Reflecting Surfaces (IRS) to improve the uplink transmission in FL.

ACKNOWLEDGMENT

The author Jamshid Abouei’s work was performed when he was a visiting professor in the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada.

REFERENCES

- [1] T. Gafni, N. Shlezinger, K. Cohen, Y. C. Eldar, and H. V. Poor, “Federated learning: A signal processing perspective,” 2021, *arXiv:2103.17150*.
- [2] L. Yu, R. Albelaihi, X. Sun, N. Ansari, and M. Devetsikiotis, “Jointly optimizing client selection and resource management in wireless federated learning for Internet of Things,” *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4385–4395, Mar. 2022.
- [3] S. Abdulrahman, H. Tout, A. Mourad, and C. Talhi, “FedMCCS: Multi-criteria client selection model for optimal IoT federated learning,” *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4723–4735, Mar. 2021.
- [4] I. Mohammed, S. Tabatabai, A. Al-Fuqaha, F. E. Bouanani, J. Qadir, B. Qolomany, and M. Guizani, “Budgeted online selection of candidate IoT clients to participate in federated learning,” *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5938–5952, Apr. 2021.
- [5] W. Zhang, X. Wang, P. Zhou, W. Wu, and X. Zhang, “Client selection for federated learning with non-IID data in mobile edge computing,” *IEEE Access*, vol. 9, pp. 24462–24474, 2021.
- [6] S. Zhai, X. Jin, L. Wei, H. Luo, and M. Cao, “Dynamic federated learning for GMEC with time-varying wireless link,” *IEEE Access*, vol. 9, pp. 10400–10412, 2021.
- [7] H. Zhang, Z. Xie, R. Zarei, T. Wu, and K. Chen, “Adaptive client selection in resource constrained federated learning systems: A deep reinforcement learning approach,” *IEEE Access*, vol. 9, pp. 98423–98432, 2021.
- [8] T. Huang, W. Lin, W. Wu, L. He, K. Li, and A. Y. Zomaya, “An efficiency-boosting client selection scheme for federated learning with fairness guarantee,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 7, pp. 1552–1564, Jul. 2021.
- [9] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, “Energy efficient federated learning over wireless communication networks,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935–1949, Mar. 2021.
- [10] W. Shi, S. Zhou, Z. Niu, M. Jiang, and L. Geng, “Joint device scheduling and resource allocation for latency constrained wireless federated learning,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 453–467, Jan. 2021.
- [11] S. Luo, X. Chen, Q. Wu, Z. Zhou, and S. Yu, “HFEL: Joint edge association and resource allocation for cost-efficient hierarchical federated edge learning,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6535–6548, Oct. 2020.
- [12] Y. He, J. Ren, G. Yu, and J. Yuan, “Importance-aware data selection and resource allocation in federated edge learning system,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13593–13605, Nov. 2020.

- [13] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 170–185, Jan. 2021.
- [14] J. Ren, G. Yu, and G. Ding, "Accelerating DNN training in wireless federated edge learning systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 219–232, Jan. 2021.
- [15] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Oct. 2021.
- [16] J. Xu and H. Wang, "Client selection and bandwidth allocation in wireless federated learning networks: A long-term perspective," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1188–1200, Feb. 2021.
- [17] M. Chen, N. Shlezinger, H. V. Poor, Y. C. Eldar, and S. Cui, "Communication-efficient federated learning," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 17, 2021, Art. no. e2024789118. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.2024789118>
- [18] L. Wang, W. Wang, and B. Li, "CMFL: Mitigating communication overhead for federated learning," in *Proc. IEEE 39th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jul. 2019, pp. 954–964.
- [19] C. Liu, C. Guo, Y. Yang, M. Chen, H. V. Poor, and S. Cui, "Optimization of user selection and bandwidth allocation for federated learning in VLC/RF systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–6.
- [20] W. Huang, Y. Yang, M. Chen, C. Liu, C. Feng, and H. V. Poor, "Wireless network optimization for federated learning with model compression in hybrid VLC/RF systems," *Entropy*, vol. 23, no. 11, p. 1413, Oct. 2021. [Online]. Available: <https://www.mdpi.com/1099-4300/23/11/1413>
- [21] F. Miramirkhani and M. Uysal, "Channel modelling for indoor visible light communications," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 378, no. 2169, Mar. 2020, Art. no. 20190187.
- [22] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.
- [23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [24] C. T. Dinh, N. H. Tran, M. N. H. Nguyen, C. S. Hong, W. Bao, A. Y. Zomaya, and V. Gramoli, "Federated learning over wireless networks: Convergence analysis and resource allocation," *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 398–409, Feb. 2021.
- [25] M. Kafafy, Y. Fahmy, M. Abdallah, and M. Khairy, "A novel bandwidth and power allocation scheme for power efficient hybrid RF/VLC indoor systems," *Phys. Commun.*, vol. 31, pp. 187–195, Dec. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S187449071730589X>



SEYED JAMAL SEYED-MOHAMMADI received the B.Sc. degree in power engineering from Bu-Ali Sina University, Iran, in 2017, and the M.Sc. degree in telecommunication systems engineering from the Iran University of Science and Technology, Iran, in 2021. His research interests include signal processing, machine learning, federated learning, human sensing, and human–computer interaction.



JAMSHID ABOUEI (Senior Member, IEEE) received the B.Sc. degree (Hons.) in electronics engineering and the M.Sc. degree (Hons.) in communication systems engineering from the Isfahan University of Technology (IUT), Iran, in 1993 and 1996, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Canada, in 2009. In 1996, he joined the Department of Electrical Engineering, Yazd University, Iran, as a Lecturer, and was promoted to an Assistant Professor, in 2010, and an Associate Professor, in 2015. From 1998 to 2004, he served as a Technical Advisor and a Design Engineer at the Research and Development Center, Cable Design Department, SGCC, Iran. From 2009 to 2010, he was a Postdoctoral Fellow at the Multimedia Laboratory, Department of Electrical and Computer Engineering, University of Toronto, Canada, and worked as a Research Fellow at the Self-Powered Sensor Networks (ORF-SPSN) Consortium. During his sabbatical, he was an Associate Researcher at the Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Toronto, ON, Canada. He currently directs the Research Group with the Wireless Networking Laboratory (WINEL), Yazd University. His research interests include next generation of wireless networks (5G) and wireless sensor networks (WSNs), with a particular emphasis on PHY/MAC layer designs including the energy efficiency and optimal resource allocation in cognitive cell-free massive MIMO networks, multi-user information theory, mobile edge computing, and femtocaching. He is a member of the IEEE Information Theory. He has received several awards and scholarships, including FOE and IGSA Awards for Excellence in Research at the University of Waterloo; the MSRT Ph.D. Scholarship from the Ministry of Science, Research and Technology, Iran, in 2004; the Distinguished Researcher Award in Province of Yazd, Iran, in 2011; and the Distinguished Researcher Award in Electrical Engineering Department, Yazd University, in 2013. He was a recipient the Best Paper Award from the IEEE Iranian Conference on Electrical Engineering (ICEE 2018). He was the International Relations Chair in 27th ICEE 2019 Conference, Iran, in 2019.



light communication, and smart grid.

SEYED MOHAMMAD SHEIKHOESLAMI received the B.Sc. degree in power electrical engineering from Bu-Ali Sina University, Iran, in 2017, and the M.Sc. degree in communication systems engineering from the University of Tehran, Iran, in 2020. He is currently pursuing the Ph.D. degree with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto. His main research interests include indoor localization, federated learning, visible



with a particular emphasis on distributed learning, graph signal processing, and graph neural networks.

ARASH RASTI-MEYMANDI received the B.Sc. degree (Hons.) in electronics and biomedical engineering from Yazd University and the M.Sc. degree (Hons.) in electronics and biomedical engineering from the Iran University of Science and Technology. He is currently pursuing the Ph.D. degree with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto. His research interests include signal processing and machine learning approaches



KONSTANTINOS (KOSTAS) N. PLATANIOTIS (Fellow, IEEE) received the B.Eng. degree in computer engineering from the University of Patras, Greece, and the M.S. and Ph.D. degrees in electrical engineering from the Florida Institute of Technology Melbourne, Florida. He holds the Bell Canada Endowed Chair in Multimedia, since 2014. He is currently a Professor with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where he directs the Multimedia Laboratory. His research interests include image/signal processing, machine learning and adaptive learning systems, visual data analysis, multimedia and knowledge media, and affective computing. He is a fellow of the Engineering Institute of Canada and a fellow of the Canadian Academy of Engineering. He was the Technical Co-Chair of the IEEE 2013 International Conference in Acoustics, Speech and Signal Processing, the General Co-Chair of the 2017 IEEE GLOBALSIP, the General Co-Chair of the 2018 IEEE International Conference on Image Processing (ICIP 2018), and the IEEE International Acoustics, Speech and Signal Processing (ICASSP 2021), Toronto. He served as the Inaugural IEEE Signal Processing Society Vice President for Membership, from 2014 to 2016. He has served as the Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS. He is a Registered Professional Engineer in Ontario.