IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 January 2023, accepted 25 January 2023, date of publication 10 February 2023, date of current version 22 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3244099

RESEARCH ARTICLE

Energy-Efficient Federated Learning With Resource Allocation for Green IoT Edge Intelligence in B5G

ADEB SALH[1,7], RAZALI NGAH[1], (Member, IEEE), LUKMAN AUDAH[2], (Member, IEEE), KWANG SOON KIM[3], (Senior Member, IEEE), QAZWAN ABDULLAH[2,6], (Senior Member, IEEE), YAHYA M. AL-MOLIKI[4], (Member, IEEE), KHALED A. ALJALOUD[5], (Member, IEEE), AND HAIRUL NIZAM TALIB[6], (Member, IEEE)

[1]Wireless Communication Centre, Faculty of Electrical Engineering, Universiti Teknologi Malaysia (UTM), Skudai, Johor Bahru 81310, Malaysia
[2]Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Parit Raja, Johor 86400, Malaysia
[3]School of Electrical and Electronics Engineering, Yonsei University, Seodaemun-gu, Seoul 03277, South Korea
[4]Faculty of Electrical Engineering, King Saud University, Riyadh 11421, Saudi Arabia
[5]College of Engineering, Muzahimiyah Branch, King Saud University, Riyadh 11451, Saudi Arabia
[6]Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka (UTeM), Durian Tunggal 76100, Malaysia
[7]Faculty of Information and Communication Technology, University Tunku Abdul Rahman (UTAR), Petaling Jaya, Perak 31900, Malaysia

Corresponding authors: Razali Ngah (razalingah@utm.my) and Qazwan Abdullah (gazwan20062015@gmail.com)

This work was supported in part by the Ministry of Higher Education (MoHE) Malaysia through the Higher Institution Centres of Excellence (HICoE) Grant; in part by the Universiti Teknologi Malaysia (UTM) through the Professional Development Research University Grant Q. J130000.21A2.06E40; in part by the King Saud University, Riyadh, Saudi Arabia, through the Researchers Supporting Project under Grant RSP2023R474; and in part by the MoHE Malaysia through the Fundamental Research Grant Scheme under Grant FRGS/1/2022/TK07/UTHM/02/25.

ABSTRACT An edge intelligence-aided Internet-of-Things (IoT) network has been proposed to accelerate the response of IoT services by deploying edge intelligence near IoT devices. The transmission of data from IoT devices to the edge nodes leads to large network traffic in the wireless connections. Federated Learning (FL) is proposed to solve the high computational complexity by training the model locally on IoT devices and sharing the model parameters in the edge nodes. This paper focuses on developing an efficient integration of joint edge intelligence nodes depending on investigating an energy-efficient bandwidth allocation, computing Central Processing Unit (CPU) frequency, optimization transmission power, and the desired level of learning accuracy to minimize the energy consumption and satisfy the FL time requirement for all IoT devices. The proposal efficiently optimized the computation frequency allocation and reduced energy consumption in IoT devices by solving the bandwidth optimization problem in closed form. The remaining computational frequency allocation, transmission power allocation, and loss could be resolved with an Alternative Direction Algorithm (ADA) to reduce energy consumption and complexity at every iteration of FL time from IoT devices to edge intelligence nodes. The simulation results indicated that the proposed ADA can adapt the central processing unit frequency and power transmission control to reduce energy consumption at the cost of a small growth of FL time.

INDEX TERMS Internet-of-things, federated learning, energy consumption, edge nodes, central processing unit.

I. INTRODUCTION

The Internet-of-Things (IoT) is a critical technology in integrating heterogeneous electronic devices. These IoT devices is used to generate a significant amount of data that can be used to enable machine learning. The IoT networks connect billions of IoT devices such as smartphones, personal computers, wearable devices, and vehicles in the smart environment [1]. To increase the system's reliability and promote green IoT, Federated Learning (FL) in the edge should focus

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

on decreasing latency and energy consumption of resources without affecting the convergence rate of the global model. Regarding energy-efficient FL: Since mobile devices are battery-powered, it is essential to lower edge device energy consumption to ensure the sustainable viability of the FL framework. A green FL or energy-efficient should consider special attention to reducing communication and computation energy. Energy-efficient radio resource allocation for delay-constrained FL was proposed in [1] and [2]. Moreover, the authors in [2] ignored the computation energy and only reduced the communication energy.

Edge Nodes (ENs) utilize intelligence features to alleviate the network burdens at the core network by reducing energy and latency requirements for time critical IoTs such as communication, storage, and analytics [1], [2]. In addition, edge intelligence aided IoT networks have been utilized near physical IoT devices to help execute applications by storing large volumes of data on ENs with low latency and in real time [3]. The edge intelligence assisted IoT networks collect data from IoT devices and perform the processing and training at the edge intelligence nodes to reduce bandwidth and latency. The challenge arises from the limitations of machine learning. With the growing number of IoT services, the training data from ENs creates a large traffic burden between the IoT devices and ENs. The FL is proposed to train with high quality using relatively few rounds of communication by allowing IoT devices to train learning models locally and the EN to learn a shared model by aggregating locally computed updates [4]. This involves downloading the global model parameters from the EN, updating the local data, and sending the updated parameters to ENs. The performance of FL depends on updated local data aggregated as the global model parameter. Then, the local and global parameters are updated iteratively until an accuracy level of the learning model is reached and sent to all IoT devices. By classifying the learning tasks between the IoT device and ENs, FL not only decreases the large traffic burden over a wireless channel but also keeps the privacy and securing IoT data.

A. RELATED WORK

Edge intelligence IoT networks have been extensively studied, such as the reduction of training losses, and the distribution of high bandwidth to devices with poor channel conditions [5], [6], [7], [8], [9]. However, some major problems are not yet fully resolved with the FL model such as reduced energy consumption while satisfying the FL time required on all IoT devices. This is because of the local loss of IoT devices, it becomes a challenge to analyze the total time consumption of a wireless FL. The authors in [5] proposed saddle point approximation to obtain the approximate delay distribution. Also, the authors in [5] study only the distribution delay to provide resource efficiency and avoid intractable computations to provide high accuracy guarantees in delay distribution for FL. Optimizing a minimum loss to perform local updates was studied in [6], where the authors proposed Federated Delayed Averaging (FedDelAvg) technique that joins communication delay between the edge nodes and the aggregator on FL. In addition, the bandwidth allocation should be a guarantee on learning speed to all devices by adapting both channel states and more powerful computation capacities to minimize energy consumption [2], [7]. The improved convergence rate and minimized loss of the trained model depend on a derivative of a convergence bound for bandwidth allocation and scheduling policy [8]. The performance convergence of the proposed joint scheduling in [8] cannot be guaranteed to schedule more devices with the highest level of accuracy, and it is only evaluated by experiments. The joint scheduling and resource block allocation policy was studied using Lyapunov optimization to minimize the training loss [9]. Some of the challenges of FL over wireless networks have been studied in [10], [11], [12], [13], [14]. The effectiveness of federated training over wireless networks is limited by the computing resource and energy consumption of the IoT device, which requires computational tasks of different data sizes over time [10]. To reduce the computational time of the IoT device and communication and accelerate the FL training process, the Lyapunov optimization-based dynamic computation offloading was proposed. This controls the Central Processing Unit (CPU)-cycle frequencies by allowing the CPU to operate at different frequencies to adjust the power consumption [11]. Other authors studied energy-aware task allocation in IoT networks to reduce the completion time based on reinforcement learning [12]. The authors in [12] proposed reinforcement learning to adapt a varying in wireless channel and storage capacity to optimize the task completion delay without considering the device battery. Moreover, some authors evaluated the reduction of service delays for IoT based on collaboration and unloading policies [13]. In addition, minimizing the total energy consumption and computation resource allocation for FL have been studied under joint learning and latency constraints based on a proposed iterative algorithm for time allocation, power control, and computation frequency [14]. To improve the efficient learning of IoT in edge computing, the time complexity of the bandwidth allocations, computational frequency, and joint optimization of power should be factored in. The proposed strategy in [14] ignores CPU frequency to avoid long waiting times reduce energy consumption and enhance energy efficiency in an FL system. A mixed-integer nonlinear programming problem was proposed to reduce the weighted sum of the system and learning costs by shortening the time spent on FL due to the univariate non-convexity [15]. A study proposed a novel energy-efficient FL approach that utilizes a novel fine-grained training technique that efficiently and specifically selects the training samples to enhance energy efficiency [16]. This performance depends on exploiting the round deadline constraint to maximize the uploading time by avoiding long waiting times and decreasing the expended energy. This proposed strategy optimizes the transmission power and local CPU speed of workers in a federated edge learning system to increase energy efficiency.

However, another study focused on the minimization of transmission completion time for uplink transmissions of a given number of bits per node and jointly optimized the locations based on analyzing the data gathering capacity to minimize the transmission completion time and enable nodes to harvest energy [17]. Moreover, enhancing FL in edge intelligence IoT to reduce the energy consumption of all IoT devices was studied by considering both CPU frequency and power control to reduce energy consumption [3]. According to IoT devices and big data sizes, the training of the local models was investigated to provide more energy efficiency [1]. Reducing the energy consumption of FL was investigated by applying joint scheduling based on the proposed two scenarios; the first scenario depends on the training of fog access points to collect data from the IoT device, and the second scenario depends on training local models at the IoT devices to upload these local model parameters to the cloud server [1]. Also, from [10], [11], [12], [13], [14], [15], [16], [17], [18] both CPU frequency and power control still require further investigation to increase the number of IoT devices. To develop resource allocation mechanisms to facilitate energy-efficient FL in an edge intelligence and improve the number of connected devices. The authors in [19], [20], [21], [22] focused on addressing the training of the local models to provide more energy efficiency and reduce computation time and wireless data on IoT devices with minimal CPU frequency allocation. For a sustainable operation of an FL framework, it is imperative to reduce energy consumption and satisfy the FL time of edge intelligence nodes. The challenge of energy-efficient FL is the trade-off between FL time and energy consumption [23], [24], [25], [26]. In particular, the energy budget is required to finish the updates and prevent long wait times due to poor channels, a slow CPU, or insufficient energy to minimize computation and communication. Accordingly, it is essential to concurrently optimize the CPU frequency to satisfy a particular FL time and decrease energy consumption. This work is different from the previously existing one [19], [20], [21], [22], [23], [24], [25], [26]. This work focused on jointly optimizing resource and learning performance to decrease communication costs and enhance learning performance in wireless FL systems. To fill this gap, this work focused on the joint optimization of CPU frequency, optimal transmission power, and learning accuracy for IoT devices to minimize energy consumption, and completion time and determined the proper amount of data needed to meet the FL time required on all IoT devices as shown in the new subsection (I.B).

B. MOTIVATION AND CONTRIBUTIONS

To further increase the performance of IoT devices, reduce the completion time, and provision of future event prediction by training data collected from IoT devices to ENs, we developed an efficient integration of joint edge intelligence nodes to minimize the energy consumption. We established an energy-efficient bandwidth allocation, joint with computing CPU frequency, transmission optimization power, and the desired level of learning accuracy to decrease the total energy consumption by satisfying the FL time requirement for all IoT devices. Besides, idle power consumption was not considered in the optimization process of this study. The main contributions of this paper are:

- To formulate the data transmission and the transmission time of finding the optimal bandwidth allocation based on the proposed Alternative Direction Algorithm (ADA) and apply the optimal data receiving time on FL and the optimal channel gain from all IoT devices of the trained model to minimize the energy consumption and satisfy the FL time.
- To improve the computation and transmission to meet the FL time and reduce energy consumption by jointly optimizing the CPU frequency. We apply the closed-form solution of minimum computation frequency allocation and optimal transmission power with Lagrangian multiplier methods of each IoT device to edge intelligence nodes.
- The local computation time can be higher for the convex function of the learning loss to obtain optimal learning accuracy with the shortest completion time and minimizing loss. Therefore, we verified the optimal time allocation by minimizing the training learning accuracy based on a repeated number of local iterations until the global model is achieved. We also validated the level of learning accuracy through the first and second derivatives of the convex function to reduce the complexity at every iteration and meet the optimal learning accuracy of FL.

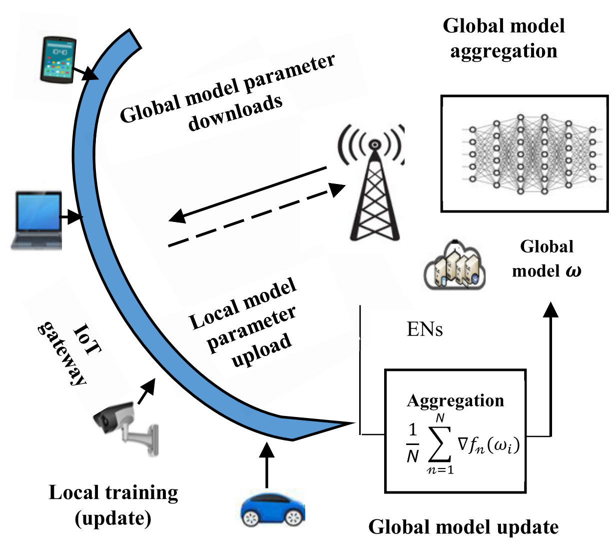

FIGURE 1. Framework of integrated FL in edge intelligence IoT networks.

II. SYSTEM MODEL

A. FEDERATED LEARNING MODEL

We proposed an edge FL-supported wireless IoT network for selecting the global aggregator EN at each round, for $\boldsymbol{N}$ intelligent IoT devices as $\mathcal{N} = \{\mathbf{1}, \ldots, \boldsymbol{N}\}$. The $\boldsymbol{N}$ IoT devices were joined to the IoT gateway and connected with ENs as shown in Fig. 1. The local training data in IoT devices were independently trained based on locally collected data. The optimization of model parameters was obtained from parameters uploaded via wireless connections from devices n. We introduced the vector ω to obtain the related parameters of the global FL model. Each of the IoT devices has an amount of data $\mathcal{Z}_n$ in the local dataset $\mathbb{Z}_n$. For each access to a dataset $\mathbb{Z}_n = \{\varkappa_{nl}, \Upsilon_{nl}\}_l^{\mathcal{Z}_n}$ of device n that does not share any dataset with the model manager, the device is called the server. Where $\varkappa_{nl}$ represents the input vector $\varkappa_{nl} \in \mathbb{R}^z$, Υ_{nl} represents the equivalent output and $l-th$ training data. The IoT devices train local models from the global model using their respective private datasets $\mathcal{Z}_n$ and send the updates to the model manager. In each iteration, the global model aggregated all local model parameters from the IoT devices(see Fig. 1). In addition, Fig. 1, shows the steps in each round of FL training: Step 1 – Server initializes the parameters of the global model and sends to each device, Step 2 - Each device completes training on its local dataset and sends the local model to the server, and Step 3 – Server aggregates local models to generate a new global model. The FL trained for all IoT devices dataset $\mathcal{F}_n(\omega, \varkappa_{n1}, \Upsilon_{n1}, \ldots, \varkappa_{n\mathcal{Z}_n}, \Upsilon_{n\mathcal{Z}_n})$ is called the total local function with the local FL. The whole dataset is called the global FL model of device n [10]. The loss function of device n with $\mathcal{Z}_n$, is determined from:

$$\mathcal{F}_n(\omega, \varkappa_{n1}, \Upsilon_{n1}, \ldots, \varkappa_{n\mathcal{Z}_n}, \Upsilon_{n\mathcal{Z}_n}) = \frac{1}{\mathcal{Z}_n}\sum_{l\in\mathcal{Z}_n} f_l(\omega, \varkappa_{nl}, \Upsilon_{nl}) \quad \forall n \in \mathcal{N}, \tag{1}$$

where $f_l(\omega, \varkappa_{nl}, \Upsilon_{nl})$ represents the loss function of the IoT device n with a single data sample. The EN aggregates the data from all IoT devices and estimates the new global model ω_i to solve the local optimization problem that reduces the loss function. Every device n computes $\nabla\mathcal{F}_n(\omega_i) = \frac{1}{N}\sum_{l=1}^{\mathcal{Z}_n}\nabla f_n(\omega_i)$ and sends it to the EN, where i is the iteration number with a vector ω_i until the global model achieves learning accuracy. It is necessary to train the implicit model to deploy an FL [10], which can be written as:

$$\min_{\omega} \mathcal{F}(\omega) = \sum_{n=1}^{N}\frac{\mathcal{Z}_n}{\mathcal{Z}}\mathcal{F}_n(\omega) = \frac{1}{\mathcal{Z}}\sum_{n=1}^{N}\sum_{l=1}^{\mathcal{Z}_n} f_l(\omega, \varkappa_{nl}, \Upsilon_{nl}), \tag{2}$$

where $\mathcal{Z} = \sum_{l=1}^{N}\mathcal{Z}_n$ represents the total data for all IoT devices. In the FL setting, a global model is copied from a server to many of the IoT devices and establishes a connection to be used for their local inference. A subset of the IoT devices is chosen to contribute to the update of the global model by sending the local training results and data to the server. The server aggregates all the local training results to update the global model at the end of an FL round. Subsequently, the updated global model is distributed to IoT devices to start a new round. This training process works well when most IoT devices have similar computing capabilities. In each iteration, all IoT devices downloaded the global FL model for local computing from ENs and informed the local model of its local training data via transmission of the local model $\frac{1}{N}\sum_{n=1}^{N}\nabla f_n(\omega_i)$ to the ENs for aggregation. The updated value of local FL is an estimate of the gradient to train its model according to the gradient descent method to solve the local optimization problem:

$$\min_{\mathcal{G}_n\in\mathbb{R}} \mathscr{b}(\omega_i, \mathcal{G}_n) \triangleq \mathcal{F}(\omega_i + \mathcal{G}_n) - \nabla\mathcal{F}_n(\omega_i) - \lambda\nabla\mathcal{F}_n(\omega_i)^T\mathcal{G}_n, \tag{3}$$

where $\mathcal{G}_n$ represents the difference between the global and local FL for IoT device n, and λ is a constant value. The term $\omega_i + \mathcal{G}_n$ represents the local FL model of IoT device n at iteration i with vector $(\omega_i)^T$, which depends on the gradient method with a given minimum global loss and local data. Due to the difficulty to solve the local optimization problem in (3), the IoT device n obtains the optimum $\mathcal{G}_{i,n}^*$ that reduces $\min_{\mathcal{G}_n\in\mathbb{R}} \mathscr{b}(\omega_i, \mathcal{G}_n)$ and uploads it to the suitable ENs to be aggregated. So, the optimal solution $\mathcal{G}_n^*$ to solve (3) with a global loss F at i th iteration number can be written as:

$$\mathscr{b}(\omega_i, \mathcal{G}_{i,n}) - \mathscr{b}(\omega_i + \mathcal{G}_{i,n}^*) \le \Psi\left(\mathscr{b}_n(\omega_i, \mathcal{G}_{i,n}) - \mathscr{b}_n(\omega_i), \mathcal{G}_{i,n}^*\right). \tag{4}$$

The global loss depending on the solution ω_i of the problem (2) with the minimum global loss Φ is

$$\mathcal{F}(\omega_i) - \mathcal{F}(\omega^*) \le \Phi\left(\mathcal{F}(\omega_0) - \mathcal{F}(\omega^*)\right), \tag{5}$$

where ω^* represents the actual optimal solution of a problem in (2). The loss function analyzed based on (3) and (4) can be written as:

$$\Gamma\,\mathfrak{l} \le \nabla^2\mathcal{F}_n(\omega_i) \le L\,\mathfrak{l}, \ \forall n \in \mathcal{N}, \tag{6}$$

where $\mathfrak{l}$ denotes the identity matrix, the values of Γ and L are determined by the loss function. Under assumption (6) to achieve a learning accuracy depends on a selection a very small value of λ satisfying $0 < \lambda \le \Gamma/L$, with the learning accuracy Ψ at the rate of $i \ge a/(1-\Psi)$, where $a = 2L^2/\lambda\Gamma^2\ln(1/\Phi)$. The overall convergence rate is achieved with an arbitrary Ψ by solving the local learning problem in (3) and step size $\mathfrak{q}$ based on using the gradient method for $(j+1)$ local iteration for $j \ge \sigma\log_2(1/\Psi)$ at every IoT device [27]. This new global model repeats iteration until reaching a confirmed accuracy level $\mathcal{G}_{i,n}^{(j+1)} = \mathcal{G}_{i,n}^{(j)} - \mathfrak{q}\nabla\mathscr{b}_n\left(\omega_i, \mathcal{G}_{i,n}^{(j)}\right)$. The next section describes the task of IoT devices for conducting local training and uploading local parameters to the ENs, divided into FL time and energy consumption.

B. TRANSMISSION IN FL TIME

In local IoT devices, the local computing model trains and updates the model, and each device uploads its local model to the EN [28], [29]. The transmit rate of the n th IoT device is

$$\mathcal{R}_n = \mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N\mathcal{B}_n}\right), \tag{7}$$

where $\mathcal{B}_n$ represents the bandwidth allocated to device n, and the wireless network has a total bandwidth of B. Due to the limited bandwidthof the system $\sum_{n=1}^{N} \mathcal{B}_n = B$, the Shannon capacity achieves the data transmission rate in the upper bound, as shown in (7). $\mathcal{P}_n$ is the average transmitted power of device n, k_n represents the channel gain available to the N IoT devices, and N is the power spectral density of the Gaussian noise. Let τ_n be the transmission time for uploading its parameters to ENs. Hence, the τ_n can be characterized by

$$\tau_n = \frac{\mathcal{Z}_n}{\mathcal{R}_n}, \tag{8}$$

where $\mathcal{Z}_n$ represents a transmit data size. Let f_n denote the computation speed of the CPU. Therefore, the computation time τ_n^c needed for data processing is:

$$\tau_n^c = \frac{\varepsilon_n \mathbb{Z}_n}{f_n} \quad \forall n \in \mathcal{N}, \tag{9}$$

where ε_n (cycles/bit) is the number of CPU cycles needed for computing one sample of data of IoT device n, and consequently the number of CPU cycles needed for one local iteration on all data samples is $\varepsilon_n\ \mathbb{Z}_n$. To synchronize the updates and minimize waiting times, the edge nodes should use an FL round deadline restriction $\mathcal{T}$ during every FL training cycle. As a result, the computation and communication stages of the training period in each iteration must be finished within $\mathcal{T}$. Formally, each training computation and communication time must meet the Quality of Service (QoS) requirement, and can be expressed as:

$$\max_{n\in\mathcal{N}} \left\{ \frac{\varepsilon_n \mathbb{Z}_n}{f_n} + \frac{\mathcal{Z}_n}{\mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N_0 \mathcal{B}_n}\right)} \right\} \leq \mathcal{T}, \quad \forall n \in \mathcal{N}. \tag{10}$$

Every IoT device must meet the learning time $\tau \leq \mathcal{T}$, where $\mathcal{T}$ is the maximum completion time of FL. From (8) and (9), the training time in every iteration is established by the total time τ of the IoT device n. To lower the CPU speed and transmission power to conserve energy during the computation and communication phases [16], the total time of receiving between all IoT devices can be calculated as:

$$\begin{aligned} \tau &= \max_{n\in\mathcal{N}} \left\{ \tau_n^c + \tau_n \right\} \\ &= \max_{n\in\mathcal{N}} \left\{ \frac{\varepsilon_n \mathbb{Z}_n}{f_n} + \frac{\mathcal{Z}_n}{\mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N_0 \mathcal{B}_n}\right)} \right\}. \end{aligned} \tag{11}$$

To support IoT device$n$ with a diverse QoS depends on when the learning time τ should meet the maximum FL time $\mathcal{T}$, i.e., $\tau = \mathcal{T}$.

C. ENERGY CONSUMPTION IN IoT DEVICE

In this section, the energy consumption mainly occurs in two stages: local training and data transmission. Let f_n be the computation speed which represents the CPU frequency of IoT device n [30], [31], [32], [33]. The energy consumption of device$n$ for local computation can be written as:

$$e_n^c = \delta \varepsilon_n \mathbb{Z}_n (f_n)^2 \ , \tag{12}$$

where $\delta = \sigma \log_2(\frac{1}{\psi})$ is effective switched capacitance, and σ is a constant related to the data size. The energy consumption for uploading the local model is

$$e_n = \mathcal{P}_n \tau_n = \frac{\mathcal{P}_n \mathcal{Z}_n}{\mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N_0 \mathcal{B}_n}\right)}. \tag{13}$$

The total energy consumption of all IoT devices that join in FL will be:

$$e = \sum_{n\in\mathcal{N}} \left(e_n^c + e_n\right) = \sum_{n\in\mathcal{N}} \left(\delta \varepsilon_n \mathbb{Z}_n (f_n)^2 + \frac{\mathcal{P}_n \mathcal{Z}_n}{\mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N_0 \mathcal{B}_n}\right)}\right). \tag{14}$$

III. PROBLEM FORMULATION

The performance of the FL depends on applying local learning and assigning active IoT devices to appropriate ENs to select the active IoT devices intelligently. Our goal is to minimize the total energy consumption under an optimized bandwidth allocation, power allocation, and desired level of learning accuracy to decrease total energy consumption for FL in ENs networks. The energy-minimization problem can be formulated as follows:

$$\min_{\tau,\, \mathcal{B}, f, \mathcal{P},\, \psi} \quad e, \tag{15}$$

$$\text{S.t } \sigma \log_2(\frac{1}{\psi}) \varepsilon_n \mathbb{Z}_n (f_n)^2 + \frac{\mathcal{P}_n \mathcal{Z}_n}{\mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N_0 \mathcal{B}_n}\right)} \leq \varrho \tag{15a}$$

$$\tau_n \mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N_0 \mathcal{B}_n}\right) \geq \mathcal{Z}_n, \ \forall n \in \mathcal{N}, \tag{15b}$$

$$\frac{a}{1-\psi} \left(\frac{\sigma \varepsilon_n \mathbb{Z}_n \log_2(\frac{1}{\psi})}{f_n} + \frac{\mathcal{Z}_n}{\mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n k_n}{N_0 \mathcal{B}_n}\right)} \right) \leq \mathcal{T}, \ \forall n \in \mathcal{N}, \tag{15c}$$

$$\sum_{n=1}^{N} \mathcal{B}_n \leq B, \tag{15d}$$

$$\mathcal{B}_n \geq 0, \quad \tau_n \geq 0, \quad \forall n \in \mathcal{N}, \tag{15e}$$

$$0 \leq \mathcal{P}_n \leq \mathcal{P}_n^{max}, \quad \forall n \in \mathcal{N}, \tag{15f}$$
$$f_n^{min} \leq f_n \leq f_n^{max}, \quad \forall n \in \mathcal{N}, \tag{15g}$$
$$0 \leq \Psi \leq 1, \tag{15h}$$

where $\tau = [\tau_1, \ldots, \tau_N]$, $\mathcal{B} = [\mathcal{B}_1, \ldots, \mathcal{B}_N]$, $f = [f_1, \ldots, f_N]$, and $\mathcal{P} = [\mathcal{P}_1, \ldots, \mathcal{P}_N]$. $\mathcal{B}_n, f_n^{max}$ and $\mathcal{P}_n^{max}$ are the bandwidths for IoT devices, maximum local CPU frequency in ENs (in cycles per second), and maximum value of the average transmit power to all IoT devices, respectively. Constraint (15a) reduces the energy consumption of each IoT device. The energy budget $\mathcal{E}$ for FL training was not exceeded by any IoT devices that were subjected to minimizing computation and communication. From (15a), the energy budget $\mathcal{E}$ is required to finish the updates and prevent long wait times due to poor channels, a slow CPU, or insufficient energy. This constraint ensures that the assigned IoT devices have enough energy to complete the update [31]. Constraint (15b) represents the data transmission rate, whereas constraint (15c) represents the time of local task execution. The time of all IoT devices should not exceed the total time in the whole FL ($\tau \leq \mathcal{T}$). In addition, the number of iterations is the same for each IoT device, and the constraint (15c) ensures the maximum time constraint for all IoT devices in each iteration $\sigma \log_2(\frac{1}{\Psi})$ if we divide the number of local iterations on both sides of the constraint (15c). Constraints (15d) and (15e) represent the bandwidth and completion time, respectively. Constraints (15f) and (15g) are bandwidth and CPU frequency ranges. The model accuracy constraint is given in (15h). Therefore, the joint energy efficiency can be optimized in the next section to reduce the energy consumption of all devices.

IV. PROBLEM SOLUTION

This work aims to obtain the optimal solution to all IoT devices based on the proposed ADA until convergence. ADA is frequently operated at the IoT gateway to optimize energy-efficient bandwidth allocation, computation frequency allocation, and optimal transmission power allocation until the global model fulfils a learning accuracy.

A. ENERGY EFFICIENT BANDWIDTH ALLOCATION

According to (8), the bandwidth allocation to all IoT devices would only affect the data-receiving time. The bandwidth problem can be formulated as

$$\min_{\mathcal{B}} \tau_n \tag{16}$$
$$\sum_{n=1}^{N} \mathcal{B}_n \leq B, \tag{16a}$$
$$\mathcal{B}_n \geq 0, \quad \forall n \in \mathcal{N}. \tag{16b}$$

By analyzing the time taken for the learning process from data collection τ to reach the optimal value of τ^* as shown in (16), where τ^* represents the shortest time to complete a round of FL. Each EN maintains reachability with the shortest estimated waiting time, by adjusting the power allocation, preventing long wait times due to poor channels, and satisfying the FL time requirement when $\mathcal{O}_n(\Psi) = \tau_n^*$ for all IoT devices (Fig. 2). If $\tau = \tau^*$, the wireless transmission time τ_n decreases to receive data from all IoT devices when its range is equal, and the time goes down if the bandwidth allocation increases. From (16), the optimal bandwidth allocation $\mathcal{B}_n^*$ can be obtained if τ_n is minimized by making small adjustments to the $\mathcal{B}_n$ as $\tau = \max_{n \in \mathcal{N}} \{\tau_n^c + \tau_n\} < \tau^*$. The optimal received data satisfy when the small-time $\tau = \tau^*$, and also when the $\mathcal{B}_n$ meet the following requirements:

$$\tau_n = \begin{cases} \dfrac{\mathcal{Z}_n}{\mathcal{B}_n \log_2\left(1 + \frac{\mathcal{P}_n \hbar_n}{N_0 \mathcal{B}_n}\right)} = \Omega_n^*, & \forall n \in \mathcal{N}, \\ \sum_{n=1}^{N} \mathcal{B}_n^* = B, \end{cases} \tag{17}$$

where Ω_n^* represents the positive constant for receiving optimal data from all IoT devices $\tau_n^* = \Omega_n^*$, which can be formulated as

$$\frac{\mathcal{Z}_n}{\mathcal{B}_n^* \log_2\left(1 + \frac{\mathcal{P}_n \hbar_n}{N_0 \mathcal{B}_n}\right)} = \Omega_n^* \Longrightarrow \mathcal{B}_n^* = \frac{1}{\Omega_n^*}\left(\frac{\mathcal{Z}_n}{\log_2\left(1 + \frac{\mathcal{P}_n \hbar_n}{N_0 \mathcal{B}_n}\right)}\right). \tag{18}$$

From the (17) and (18), the optimal receiving data for n *th* IoT device depends on a better channel gain, can be written as:

$$\Omega_n^* = \frac{\mathcal{Z}_n / \log_2\left(1 + \frac{\mathcal{P}_n \hbar_n^*}{N_0 \mathcal{B}_n}\right)}{\mathcal{B}_n}. \tag{19}$$

From (18) and (19), the ENs eliminate the interference from the n *th* device by using self-interference cancellation.The optimal bandwidth allocation can be formulated as

$$\mathcal{B}_n^* = \frac{\mathcal{B}_n \mathcal{Z}_n / \log_2\left(1 + \frac{\mathcal{P}_n \hbar_n^*}{N_0 \mathcal{B}_n}\right)}{\sum_{n \in \mathcal{N}} \left(\mathcal{Z}_n / \log_2\left(1 + \frac{\mathcal{P}_n \hbar_n^*}{N_0 \mathcal{B}_n}\right)\right)}, \ \forall n \in \mathcal{N}. \tag{20}$$

When the minimum time $\tau_n^* = \Omega_n^*$, the received data is optimized and can be simplified as:

$$\min_{f, \mathcal{P}, e} \Psi a \max_n (\Omega_n^* + \tau_n^c + \tau_n) + \Psi(1-a) e + (1-\Psi)\mathcal{B}_n, \tag{21}$$
$$\text{S. t. } (15e), (15f), (15g), (15h). \tag{21a}$$

The optimization problem in (21) is a mixed-integer nonlinear programming problem, which can be decomposed into three subproblems and solved by the ADA to optimize variable time $\mathcal{T} = \Omega_n^* + \tau_n^c + \tau_n$ in (21) and can be rewritten as:

$$\min_{f, \mathcal{P}, e, \mathcal{T}} \Psi a \mathcal{T} + \Psi(1-a) e + (1-\Psi)\mathcal{B}_n, \tag{22}$$

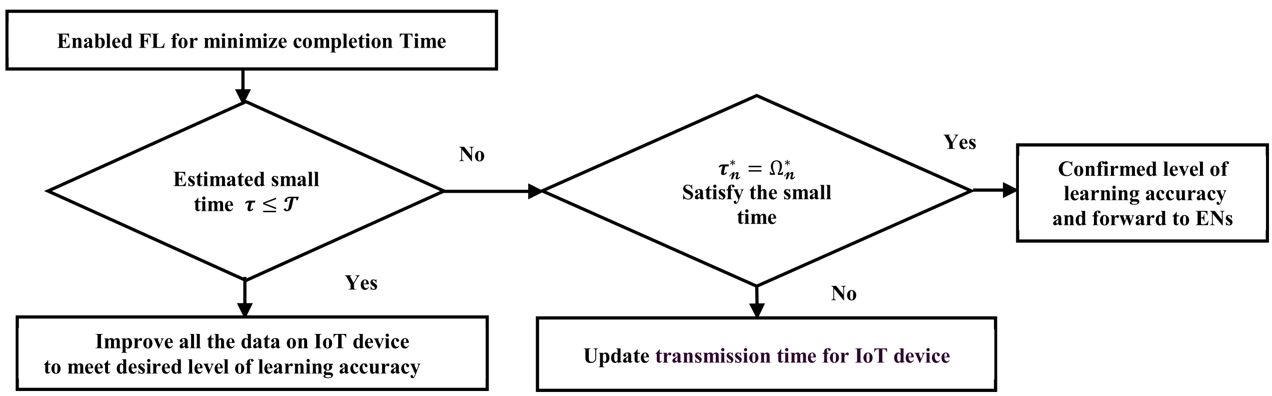

FIGURE 2. Reducing the completion time based on intelligently selecting the active IoT devices.

$$\text{S.t. (15e), (15f), (15g), (15h),} \tag{22a}$$

$$\Omega_n^* + \tau_n^c + \tau_n \leq \mathcal{T}, \quad \forall n \in \mathcal{N}. \tag{22b}$$

By analyzing (22), the optimization problem is still challenging because of its non-convexity. To find a good solution, we decomposed the three subproblems by using the ADA to process the IoT. We then established CPU frequency allocation iteratively f_n, optimized transmission power $\mathcal{P}_n^*$, and minimized completion time until the global model meets the optimal learning accuracyΨ, as shown in the next subsection:

B. OPTIMIZE COMPUTATION FREQUENCY ALLOCATION

When CPU frequency and time of execution of local tasks are given $(f, \mathcal{T})$ with fixed $(\mathcal{P}, e)$, in this case, the power allocation and energy consumption can be minimized by substituting the problem (21) into:

$$\min_{f,\mathcal{T}} \Psi\, a\mathcal{T} + \Psi\,(1-a) \sum\nolimits_{n\in\mathcal{N}} \varepsilon_n \mathbb{Z}_n (f_n)^2 , \tag{23}$$

$$\text{S.t}\ \mathcal{T} - \frac{\varepsilon_n \mathbb{Z}_n}{f_n} \geq \Omega_n^* + \tau_n,\ \forall n \in \mathcal{N}, \tag{23a}$$

$$f_n^{min} \leq f_n \leq f_n^{max},\ \forall n \in \mathcal{N}. \tag{23b}$$

From the constraint in (23a) all the data on device n will have optimal computing f_n.

The constraint in (23b) is most effective to assign active IoT devices with minimum f_n^{min}. Based on the minimization problem in the second term in (23), the optimal f_n^* can obtain from (23a), which leads to optimizing energy consumption minimization based on integrated EN-enabled IoT networks as:

$$\min_{f} \sum\nolimits_{n\in\mathcal{N}} \phi\,(1-a)\, \varepsilon_n \mathbb{Z}_n (f_n)^2. \tag{24}$$

Constraint (23a) can be transformed into $\max_{n\in\mathcal{N}} \left\{\tau_n^c + \tau_n\right\} = \max_{n\in\mathcal{N}} \left\{\frac{\varepsilon_n \mathbb{Z}_n}{f_n} + \frac{\mathcal{Z}_n}{\mathcal{R}_n}\right\} \leq \mathcal{T},\ \forall n \in \mathcal{N}$. Hence, the CPU frequency f_n is bounded between f_n^{min} and f_n^{max}, and the optimal computation frequency allocation f_n^* can be obtained from (23b) as:

$$f_n^* = \frac{\varepsilon_n \mathbb{Z}_n}{\mathcal{T} - \Omega_n^* - \tau_n} \quad \forall n \in \mathcal{N}, \tag{25}$$

Then, f_n satisfies $f_n \geq \max\left\{f_n^{min}, f_n^*\right\}$, and accordingly, (23a) and (23b) can be joined as $\max\left\{f_n^{min}, f_n^*\right\} \leq f_n \leq f_n^{max},\ \forall n \in \mathcal{N}$. Therefore, the closed form solution of the minimum f_n is obtained as:

$$f_n = \begin{cases} f_n^{min}, & \text{if } f_n^* \leq f_n^{min} \\ f_n^*, & \text{if } f_k^{min} < f_n^* < f_n^{max} \\ f_n^{max}, & \text{if } f_n^* \geq f_n^{max}, \end{cases} \tag{26}$$

The optimal CPU frequency is found as soon as we solve for $\mathcal{T}$, by substituting the optimal f_n^* in (25) into constraint (23a). The optimization problem concerning $\mathcal{T}$ is transformed into a univariate optimization problem $\min_{\mathcal{T}} \Psi\, a\, \mathcal{T} + \Psi\,(1-a)\,\varepsilon_n{}^3 \sum_{n\in\mathcal{N}} (\mathbb{Z}_n{}^3 / (\mathcal{T} - \Omega_n^* - \tau_n)^2)$, and by setting $\mathcal{T} \geq \mathcal{T}^{min}$ the minimum time can define according to the constraints in (22b) as $\mathcal{T}^{min} = \max_{n\in\mathcal{N}} (\Omega_n^* + \tau_n + \tau_n^c)$, where $\tau_n^c = \max_{n\in\mathcal{N}} \frac{\varepsilon_n \mathbb{Z}_n}{f_n^{min}}$ [32]. The optimal time $\mathcal{T}^*$ is obtained when the time it takes to receive data from all IoT devices according the constraint in (23a) and (25) is $\mathcal{T}^* = \max_{n\in\mathcal{N}} \left(\Omega_n^* + \tau_n + \frac{\varepsilon_n \mathbb{Z}_n}{f_n^{max}}\right) \hat{\mathcal{T}} \geq 0$, where $\hat{\mathcal{T}}$ represents the solution when $\hat{\mathcal{T}} = 0$.

C. OPTIMIZE TRANSMISSION POWER

To enhance $(\boldsymbol{\tau}, \mathcal{B}, f, \mathcal{P}, \Psi)$ in problem (15) depends on firstly optimizing $(\mathcal{P}, \mathcal{B})$ with fixed $(\boldsymbol{\tau}, f, \Psi)$. To update the nth IoT device's optimal power allocation $\mathcal{P}_n^*$ depends on providing efficient $\mathcal{P}_n$ levels for the IoT devices, which can be written as:

$$\min_{\mathcal{P}}, \mathcal{B} \frac{a}{1-\Psi} \sum\nolimits_{n\in\mathcal{N}} \mathcal{P}_n \tau_n, \tag{27}$$

$$\text{S.t. (15b), (15d), (15f).} \tag{28}$$

Reducing transmit power$\mathcal{P}_n$ to all IoT devices depends on avoiding weak channels by selecting the optimal $\mathcal{B}_n^*$. To solve a problem in (27) depends on selecting the optimal solution

$(\mathcal{B}_n^*, \mathcal{P}_n^*)$. From (18), the optimal bandwidth is satisfied by defining $\mathcal{B}_n^* = \max\left\{\mathcal{B}_{n,}(\mathrm{V}), \mathcal{B}_n^{min}\right\}$. From (27), the transmission power should be reduced as:

$$\mathcal{P}_n^* = \frac{N_0 \mathcal{B}_n}{k_n}\left(2^{\mathcal{Z}_n/\mathcal{B}_n^*\tau_n} - 1\right), \tag{29}$$

The transmission power (29) is a convex function. The optimal power $\mathcal{P}_n^*$ in (29) is a decreasing function of bandwidth. The conventional successive convex function in (29) was obtained by taking the first- and second-order derivatives of $\mathcal{P}_n^*$, which can be written as:

$$\frac{\partial \mathcal{P}_n^*}{\partial \mathcal{B}_n} = \frac{N_0}{k_n}\left(e^{\frac{\mathcal{Z}_n(\ln 2)}{\mathcal{B}_n\tau_n}} - \frac{\mathcal{Z}_n(\ln 2)}{\mathcal{B}_n\tau_n}e^{\frac{\mathcal{Z}_n(\ln 2)}{\mathcal{B}_n\tau_n}} - 1\right), \tag{30}$$

and

$$\frac{\partial^2 \mathcal{P}_n^*}{\partial \mathcal{B}_n^2} = \frac{N_0 \mathcal{Z}_n^2 (\ln 2)^2}{k_n \tau_n^2 \mathcal{B}_n^3} e^{\frac{\mathcal{Z}_n(\ln 2)}{\mathcal{B}_n\tau_n}} \geq 0. \tag{31}$$

From the first derivative in (30), it is a decreasing function of bandwidth $\mathcal{B}_n$. The second derivative in (31) is an increasing function of $\mathcal{B}_n$. The convex function for $\min_{\mathcal{B}} \mathcal{P}_n^*$ is solved if $\frac{\partial \mathcal{P}_n^*}{\partial \mathcal{B}_n} < 0$, for $0 < \mathcal{B}_n < \infty$. Therefore, according to (29) and (31) $\min_{\mathcal{B}} \mathcal{P}_n^*$ is a convex function that can be solved by using the Karush-Kuhn-Tucker conditions [7], [33], [34], [39]. By introducing Lagrange multipliers V in (29), this is:

$$L(\mathcal{B}, \mathrm{V}) = \sum_{n=1}^{N} \frac{N_0 \mathcal{B}_n \tau_n}{k_n}\left(2^{\mathcal{Z}_n/\mathcal{B}_n\tau_n} - 1\right) + \mathrm{V}\sum_{n=1}^{N}(\mathcal{B}_n - \mathrm{B}). \tag{32}$$

Given the constraints in (28), more bandwidth should be allocated with weaker channels within a short time to reduce $\mathcal{P}_n$, the Lagrange multipliers associated with $\min_{\mathcal{B}} \mathcal{P}_n^* = \min_{\mathcal{B}} \sum_{n=1}^{N} \frac{N_0 \mathcal{B}_n}{k_n}\left(2^{\mathcal{Z}_n/\mathcal{B}_n\tau_n} - 1\right)$. The $\mathcal{B}_n^*$ can be obtained by taking the first derivative of the objective function with respect to $\mathcal{B}_n$:

$$\frac{\partial L(\mathcal{B}, \mathrm{V})}{\partial \mathcal{B}_n} = \frac{N_0}{k_n \tau_n}\left(e^{\frac{\mathcal{Z}_n(\ln 2)}{\mathcal{B}_n\tau_n}} - \frac{\mathcal{Z}_n(\ln 2)}{\mathcal{B}_n\tau_n}e^{\frac{\mathcal{Z}_n(\ln 2)}{\mathcal{B}_n\tau_n}} - 1\right) + \mathrm{V}. \tag{33}$$

From the constraint (15d), the optimal solution becomes continuously controlled with $\mathcal{B}_n^* = \max\left\{\mathcal{B}_n(\mathrm{V}), \mathcal{B}_n^{min}\right\}$. The maximum power constraint $\mathcal{P}_n^* \leq \mathcal{P}_n^{max}$ is equivalent to $\mathcal{B}_n^{min}$ and satisfied when $\mathcal{B}_n(\mathrm{V}) \geq \max\left\{\mathcal{B}_n^{min}, \mathcal{B}_n^*\right\}$. Accordingly, (33) can be joined as $\max\left\{\mathcal{B}_n^{min}, \mathcal{B}_n^*\right\} \leq \mathcal{B}_n(\mathrm{V}) \leq \mathcal{B}_n^{max}$, $\forall n \in \mathcal{N}$. Then, the closed-form solution of the optimal $\mathcal{B}_n^*$ is obtained as:

$$\mathcal{B}_n(\mathrm{V}) = \begin{cases} \mathcal{B}_n^{min}, & \text{if } \mathcal{B}_n^* \leq \mathcal{B}_n^{min} \\ \mathcal{B}_n^*, & \text{if } \mathcal{B}_n^{min} < \mathcal{B}_n^* < \mathcal{B}_n^{max} \\ \mathcal{B}_n^{max}, & \text{if } \mathcal{B}_n^* \geq \mathcal{B}_n^{max} \end{cases}. \tag{34}$$

The algorithm significantly reduced the overall completion time and data exchange between IoT devices and ENs. Due to the local loss function, the analysis of the total time consumption of a wireless FL became a challenge. Solving these challenges would depend on finding a feasible solution to the problem (15) by selecting the desired level of the learning accuracy parameter, as shown in the following subsection.

Algorithm 1 Enable FL for Optimizing Computation Frequency Allocation and Power Allocation to Minimize Energy Consumption

[1-]
Initialization k_n, Ψ, f_n, and $\mathcal{P}_n^{max}$
repeat
for $n = 1 : \mathcal{N}$
Compute τ_n^c and Ω_n^* for their corresponding frequencies using (23a) with fixed $(\mathcal{P}, e)$
Improve all the data on IoT devices with minimal CPU frequency allocation with a range $f_n \geq \max\left\{f_n^{min}, f_n^*\right\}$
Arrange the devices in appearing order as per local computation time τ_n^c
Assign active IoT devices with minimum energy consumption (24)
if $\max_{n\in\mathcal{N}}\left\{\tau_n^c + \tau_n\right\} \leq \mathcal{T}$ **then**
Obtain a minimum f_n as (29),
Join $\max\left\{f_n^{min}, f_n^*\right\} \leq f_n \leq f_n^{max}$, $\forall n \in \mathcal{N}$.
end if
If $\mathcal{P}_n^* \leq \mathcal{P}_n^{max}$ **then**
Apply the first and second derivatives in (30) and (31)
Assign device n with $\min_{\mathcal{B}} \mathcal{P}_n^*$
else
Efficiently allocates power $\mathcal{P}_n^*$ based on optimizing $(f, \mathcal{T})$ with fixed $(\mathcal{P}, e)$.
end if
Compute $\mathcal{P}_n^*$ at an increasing function of $\mathcal{B}_n(\mathrm{V})$ as (34)
Update transmission time for IoT device τ_n
Update the corresponding $\mathcal{P}_n^*$ according to e in (15)
end for
Compute $\min_{\mathcal{B}} \mathcal{P}_n^* = \min_{\mathcal{B}} \sum_{n=1}^{N} \frac{N_0 \mathcal{B}_n}{k_n}\left(2^{\mathcal{Z}_n/\mathcal{B}_n^*\tau_n} - 1\right)$
update $n = n + 1$,
Until $\mathcal{P}_n^*, f_n^*, \mathcal{B}_n^*$ converge.

D. COMPLETION TIME MINIMIZATION

To minimize the total energy consumption, prevent long wait times due to poor channels, and satisfy the FL time requirement for all IoT devices. This subsection formulated the convex problem by selecting the desired level of learning accuracy based on a repeated number of local iterations until the global model is achieved. Likewise, the time consumed was also minimized in each round of FL and can be presented as:

$$\min_{\mathcal{T},f,\tau,\Psi,\mathcal{B},\mathcal{P}} \quad \mathcal{T}, \tag{35}$$

$$\text{S.t. (15a)}, -\text{(15h)}. \tag{35a}$$

A set of conditions is required to minimize completion time if $\mathcal{T}^*\leq \mathcal{T}$, such that $\mathcal{T}^*, f^*, \tau^*, \Psi^*, \mathcal{B}^*, \mathcal{P}^*$ is the optimal solution of (35). From the constraints (15b) and (15c), the completion time in (35) was still non-convex. The solution to the convex optimization problem would depend on achieving learning accuracy when the loss does not decrease over time if $\mathcal{T}^*\leq \mathcal{T}$. By setting $\mathcal{T}$, the completion time in (35) can determine a better solution that satisfies the constraints (15a)–(15h). The highest computation frequency is always efficient if $f_n^* = f_n^{max}$. From the constraints (15g), (15e) and (15h), the minimized completion time was attained when $\mathcal{P}_n^* = \mathcal{P}_n^{max}$. Thus, by substituting (15f) and (15g) into (35), the solution to jointly minimize the $\mathcal{T}$ problem can be expressed as:

$$\min_{\mathcal{T},\tau,\Psi,\mathcal{B}} \quad \mathcal{T}, \tag{36}$$

$$\text{s.t. } \tau_n \leq \frac{(1-\Psi)\,\mathcal{T}}{a} + \frac{\sigma\varepsilon_n\mathbb{Z}_n\log_2(\Psi)}{f_n^{max}}, \forall n\in\mathcal{N}, \tag{36a}$$

$$\frac{\mathcal{Z}_n}{\tau_n} \leq \mathcal{B}_n\log_2\left(1+\frac{\mathcal{P}_n^{max}\hbar_n}{\mathrm{N}_0\mathcal{B}_n}\right), \quad \forall n\in\mathcal{N}, \tag{36b}$$

$$\sum\nolimits_{n=1}^{\mathcal{N}} \mathcal{B}_n \leq B, \tag{36c}$$

$$\text{S. t. (15d), (15e), (15h).} \tag{36d}$$

The sufficient and necessary conditions were obtained when the loss significantly decreased, which enabled IoT devices to save energy and optimize the training time. We designed the learning accuracy for FL to solve the convex optimization problem by selecting the desired level of learning accuracy Ψ with a set $\mathcal{T}$ in (36) and finding the optimal solution from (36a)–(36d), as follows:

$$B \geq \min_{0\leq\Psi\leq 1} \sum\nolimits_{n=1}^{N} \vartheta_n\left(\sigma_n(\Psi)\right), \tag{37}$$

$$\sigma_n(\Psi) = -\frac{(\ln 2)\,\Psi}{W\left(-\frac{(\ln 2)\Psi\mathrm{N}_0}{\mathcal{P}_n^{max}\hbar_n} exp^{-\frac{(\ln 2)\Psi\mathrm{N}_0}{\mathcal{P}_n^{max}\hbar_n}}\right) + \frac{(\ln 2)\Psi\mathrm{N}_0}{\mathcal{P}_n^{max}\hbar_n}}, \tag{38}$$

$$\sigma_n(\Psi) = \frac{\mathcal{Z}_n}{\frac{(1-\Psi)\mathcal{T}}{a} + \frac{\sigma\varepsilon_k\mathbb{Z}_n\log_2(\Psi)}{f_n^{max}}}. \tag{39}$$

Based on the analysis from (37)-(39), the maximum B is defined by $z = L\ln(1+1/L)$ with $L > 0$. Function z is proportional to the constraint of $\mathcal{B}_n$ on the right side of (15d). So, by taking the first-order derivative of z with respect to L, the target function $z' = \ln(1+1/L) - (1+1/L)$, where z' is the decreasing function, such that $z' > 0$ for $0 < L < \infty$. From the constraints (36b) and (36c), the left side should be small, and the communication time τ_n in (36a) should be as long as possible to satisfy the maximum bandwidth as shown in (37a), the optimal time allocation can be written as:

$$\tau^*{}_n = \frac{(1-\Psi)\,\mathcal{T}}{a} + \frac{\sigma\varepsilon_n\mathbb{Z}_n\log_2(\Psi)}{f_n^{max}}, \quad \forall n\in\mathcal{N}. \tag{40}$$

By substituting (40) into (36b), the major complexity at every iteration depends on the transmission time, which can be written as:

$$\min_{\Psi,\mathcal{B}} \sum\nolimits_{n=1}^{\mathcal{N}} \mathcal{B}_n, \tag{41}$$

$$\text{S.t.} \begin{cases} \sigma_n(\Psi) \leq \mathcal{B}_n\log_2\left(1+\frac{\mathcal{P}_n^{max}\hbar_n}{\mathcal{N}_0\mathcal{B}_n}\right), \forall n\in\mathcal{N}, \\ \text{(15e)} \\ \text{(15h)} \end{cases} \tag{41a}$$

where $\sigma_n(\Psi)$ is the learning accuracy. The learning accuracy increases with the number of global iterations because more iterations are required if the efficiency of the local computation is low. The optimal solution is obtained when (42) is less than B. Then, the constraints on the right side in (36b) should balance the optimal solution to increase the function. The convex problem in (37) can be reformulated to obtain the optimal solution of the learning accuracy Ψ^* by addressing sets (36a) and (36b), and can be presented as:

$$B \geq \min_{0\leq\Psi\leq 1} \sum\nolimits_{n=1}^{N} \vartheta_n\left(\sigma_n\left(\Psi^*\right)\right). \tag{42}$$

Due to the convexity of function $\vartheta_n(\sigma_n(\Psi))$ is an increasing function of Ψ^*. The learning accuracy Ψ^* is the unique solution to $\sum_{n=1}^{N}\vartheta'_n\left(\sigma_n\left(\Psi^*\right)\right)\sigma'_n\left(\Psi^*\right) = 0$, which can be solved by taking the first-order derivative $\sum_{n=1}^{N}\vartheta'_n\left(\sigma_n\left(\Psi^*\right)\right)\sigma'_n\left(\Psi^*\right)$. The optimal solution of the learning accuracy should not be too small or too large to meet the optimal Ψ^* of FL because the local computation time will be higher for small learning Ψ. While using many global iterations for transmission time, the learning accuracy Ψ is large. To confirm that $\sigma_n(\Psi)$ is a convex function, the learning process is repeated continuously until the desired learning of optimal Ψ^* is achieved. The global accuracy function meets the minimum requirements, which can be defined as:

$$\mathcal{Y}(\Psi) = \frac{\mathcal{Z}_n}{\phi, 0\leq\Psi\leq 1}. \tag{43}$$

By substituting the optimal $\tau^*{}_n$ in terms of an accuracy as shown in (40), (40) can fulfil a better performance learning in terms of accuracy Ψ with optimal $\tau^*{}_n$ by reformulating (40) as:

$$\mathcal{O}_n(\Psi) = \tau^*{}_n = \frac{(1-\Psi)\,\mathcal{T}}{a} + \frac{\sigma\varepsilon_n\mathbb{Z}_n\log_2(\Psi)}{f_n^{max}}, \quad 0\leq\Psi\leq 1. \tag{44}$$

From (39), the convex function $\sigma_n(\Psi) = \vartheta_n(\mathcal{O}_n(\Psi))$ is obtained by taking the first and second derivatives of (43) and (44). From (43), the optimal Ψ^* is obtained by taking the first derivative $\mathcal{Y}'(\Psi) = -\mathcal{Z}_n/\Psi^2 \leq 0$, and also the second derivative is $\mathcal{Y}''(\Psi) = 2\mathcal{Z}_n/\Psi^3 \geq 0$. From (44), the optimal time allocation is obtained by applying the second derivative as $\mathcal{O}_n''(\Psi) = -\sigma\varepsilon_n\mathbb{Z}_n/\ln 2f_n^{max}\Psi^2$, $\forall n \in \mathcal{N}$. By combining $\mathcal{Y}''(\Psi)$ for (44) and$\mathcal{O}_n''(\Psi)$ in (44) and substituting for $\sigma_n(\Psi)$ in (42), the second derivative is $\sigma_n''(\Psi) = \mathcal{Y}''(\mathcal{O}_n(\Psi))(\mathcal{O}_n'(\Psi))^2 + \mathcal{Y}'(\mathcal{O}_n(\Psi))\mathcal{O}_n''(\Psi)$. With reference to $\mathcal{Y}''(\mathcal{O}_n(\Psi)) \geq 0$ and $\mathcal{Y}'(\mathcal{O}_n(\Psi))$, $\vartheta_n(\mathcal{O}_n(\Psi))$ is a convex function.

Algorithm 2 Proposed FL To Achieve an Optimal Learning Accuracy Ψ With Minimized Completion Time

1- **Initialize** $\mathcal{T}_{min}$ minimal time allocation, $\mathcal{T}_{max}$ maximum allowed number of iteratio ns, and $\mathcal{E}$ threshold value for judging the convergence
2- **repeat**
3- Compute the completion time $\mathcal{T} = (\mathcal{T}_{min} + \mathcal{T}_{max})/2$
4- Achieve the minimum $\mathcal{T}$ as shown (36)
5- **while**$(\mathcal{T}_{min} - \mathcal{T}_{max}) > \mathcal{E}$ **do**
6- **if** (36a)- ((36d) has a sufficient solution,
7- **Set** $\mathcal{T}_{max} = \mathcal{T}$.
8- **else**
9- Check a better performance in (42) to get the desired level of learning accuracy Ψ^*,
10- Set $\mathcal{T}_{min} = \mathcal{T}$
11- Confirm that $\sigma_n\left(\Psi^*\right)$ is a convex function
12- **end if**
13- **end while**
14- **until** $\mathcal{T} = \frac{\mathcal{T}_{max} - \mathcal{T}_{min}}{\mathcal{T}_{max}} \leq \mathcal{E}$

The convergence on a stable solution that reduces the overall completion time can be easily obtained by applying the closed-form solution of the minimum computation frequency allocation and optimal transmission power with the Lagrangian multiplier methods. Therefore, the convergence of our algorithm mainly depended on the properties of ADA for solving the problem (15). In one iteration, subproblems (21), (24), and (27) were successively solved. The learning accuracy for FL was proposed to solve the convex optimization problem from (36a)–(36d). The desired level of learning accuracy ϕ depended on addressing the subproblem (36) and reformulating to convexity in (40). Thus, the optimal time allocation $\tau^*{}_n$ in terms of accuracy Ψ^* met the minimum requirements as shown in (44).

V. SIMULATION RESULTS

Our proposed schemes ADA were compared with three algorithms from our simulation: Power-only, CPU-only, and Fixed. Power-only optimizes the transmission of power control to all IoT devices and sets f_n to its highest value to minimize energy consumption, which is inspired by existing work [35]. CPU-only optimizes f_n and sets the power allocation to its highest value to minimize energy consumption [36]. Meanwhile, Fixed sets both f_n and transmission power to their maximum values [10]. We considered the square area of 1000 m $\times$1000 m, where IoT devices were randomly distributed in this region. The IoT gateway to which the EN is connected was located at its centre, and the path loss model was $32.44 + 20\log_{10}(d)$, where d is the distance in meters [37]. The standard deviation of shadow fading was 8 dB for IoT devices, and the effective switched capacitance and processing density of learning was represented by the noise power density $\aleph = -174$ dB/Hz. The maximum IoT device transmission power was 3 W. The FL time was 1 second [38], and bandwidth allocation was 20 MHz. The IoT device data size $\mathcal{Z}_n$ was randomly selected from 5 to 10 Kb. The number of training data samples $\mathbb{Z}_n$ was randomly selected from 100 to 800, whereas the range of CPU frequency f_n was set from 0.3 to 10^9, and the CPU was programmed based on the parameter $\delta = 10^{-28}$ [11]. The baseline in [3], [14] is considered to compare with our solution. The suggested ADA schemes balance the aforementioned factors by intelligently scheduling devices, customizing the device's transmission rate, and maximizing each IoT device's transmission power. Our suggested strategies also effectively reduce energy usage by altering CPU frequencies while optimizing the computation frequency distribution of IoT devices. To save energy, the performance of energy consumption for the proposed ADA does not significantly change, and only minor adjustments occur in energy consumption starting from 21 J to 38 J as shown in Table 1.

A. ENERGY CONSUMPTION

From Fig. 3, the performance of each algorithm, the ADA provides less energy consumption by adjustable CPU frequencies and is followed by power-only, and CPU-only. Whereas the power-only consumes more power because of selecting a maximum computation frequency allocation for local training. However, the CPU-only leads to more consumption of energy because the CPU-only ignores the power optimization at transmitting a signal to IoT devices which leads to additional interference occurring and becoming inefficient to offload data. Finally, fixed has the highest energy consumption compared to ADA proposed, power-only and CPU- only because it chooses the maximum CPU frequency and maximum transmission power to reduce the FL time. From Fig. 3, the impact of a number of CPU cycles to be treated ranging from 4×10^7 to 12×10^7, the energy consumption greatly increases with more a number of IoT devices, which need to fix the CPU frequency at the greatest value in both of fixed and CPU only. While proposed ADA offered the minimum energy consumption for both small and large numbers of CPU cycles. Fig.3, show that the energy consumption of both CPU-only and proposed ADA with adjustable CPU frequencies does not change much for large numbers of CPU cycles or if the number of CPU cycles varied

TABLE 1. Simulation setting for ADA.

	Energy consumption											
CPU cycles × 10^7	20.44487	23.51804	24.94486	27.90826	30.98142	34.71312	36.13994	37.67652	39.2131	40.63993	43.60333	45.13991
Number of Data samples	9.2152	13.97571	19.90204	23.39956	24.56539	28.16006	30.49173	32.82341	35.25224	35.35224	37.58391	38.74975
	Completion time											
CPU cycles× 10^7	0.6873	0.71471	0.74241	0.76981	0.79427	0.83111	0.84938	0.85851	0.8647	0.87383	0.86765	0.87059
Number of Data samples	0.54907	0.64151	0.73855	0.71667	0.70457	0.69737	0.68758	0.75785	0.81861	0.86728	0.90126	0.94964

beyond 12. In comparison to the baselines, the proposed ADA exhibits a significant decrease in energy consumption.

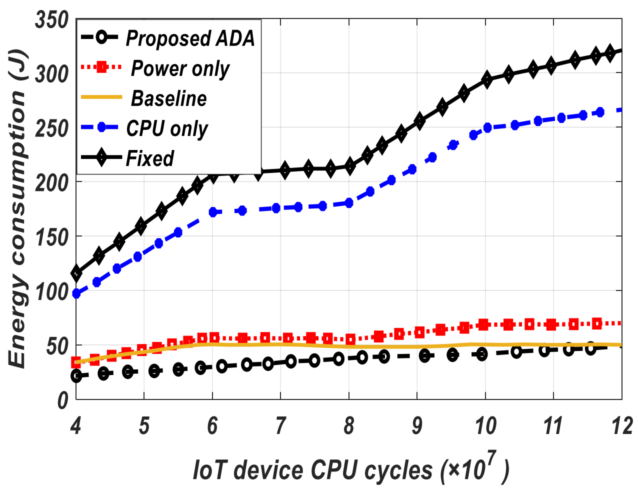

FIGURE 3. Energy consumption related to CPU cycles.

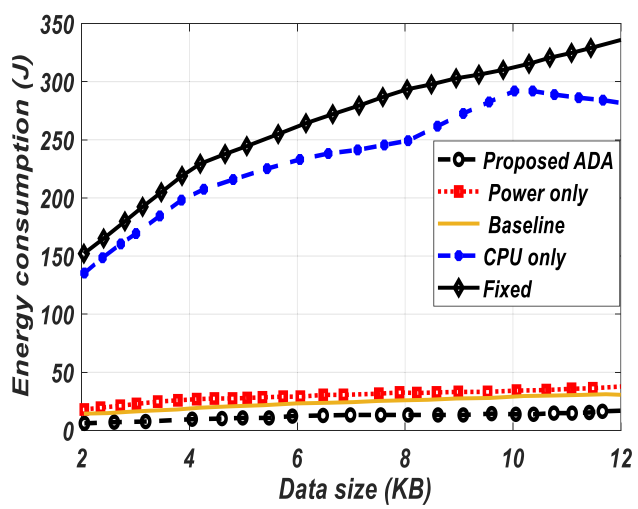

FIGURE 4. Energy consumption related to data size.

From Fig. 4, the energy consumption with the local data size $\mathcal{Z}_n$ is high in Fixed and CPU-only compared to Power-only and the proposed ADA. From Fig. 4, the proposed

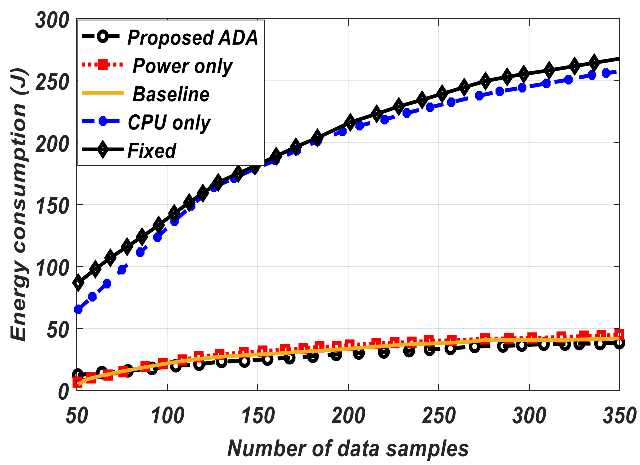

FIGURE 5. Energy consumption vs a number of data samples.

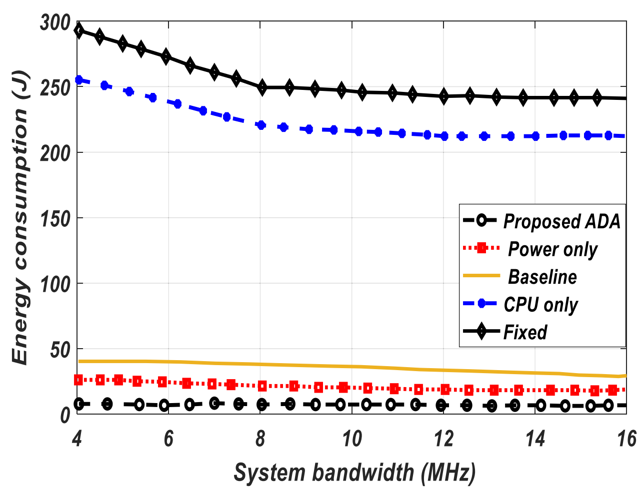

FIGURE 6. Energy consumption related to system bandwidth.

ADA provides the best energy consumption, followed by Power-only, while in the case of CPU-only and Fixed, the energy consumptions greatly increase, and the worst performance will occur at the highest value because Fixed and CPU-only fix transmission power. From (15a), the local training minimizes the consumed energy of IoT devices by adjusting the CPU frequency and power transmission. The proposed ADA consumes less energy of the data size $\mathcal{Z}_n$ because the IoT devices achieve local learning on the data. When the data size $\mathcal{Z}_n$ changes from 2 to 12, the performance of energy consumption for the proposed ADA does not change much and provides the changes from 1 J to 15 J. Fig. 5 shows the energy consumption versus the number of data samples $\mathbb{Z}_n$. The small and large size of the proposed ADA data samples $\mathbb{Z}_n$ provides less energy consumption. Moreover, Power-only does not significantly change with the proposed ADA at the increased number of data samples. However, when the number of data samples increases, energy consumption greatly increases and provides the worst performance for CPU-only and fixed effects. When the number of data samples $\mathbb{Z}_n$ changes from 50 to 350, the performance of energy consumption for the proposed ADA does not change much in order to save energy, and only small changes occur in energy consumption starting from 25 J to 40 J. For Fixed and CPU-only, energy consumption increases and changes more. In Fig. 6, we investigate the impact of energy consumption with a different system bandwidth with a range from 4 to 16 MHz. The energy consumption is reduced if the allowed transmission time is larger. From (8), the bandwidth allocation to all IoT devices will only affect the shortest estimated waiting time. The proposed ADA and Power-only can keep energy consumption steady and adjust the power allocation to guarantee learning speed to all IoT devices by enhancing channel states and more powerful computation capacities. Therefore, the proposed ADA provides less energy consumption compared to the three algorithms. In the case of Fixed and CPU-only, energy consumption greatly increases because a larger bandwidth requires a higher processing speed. These figures show that the proposed ADA still performs better energy consumption than the baselines based on guaranteed learning speed to all IoT devices by enhancing channel states to prevent long wait times due to poor channels, and a slow CPU to provide efficient $\mathcal{P}_n$ levels for the IoT devices.

B. FL TIME

From Fig. 7, both the Fixed and Power-only algorithms selected the highest CPU frequency. This effectively reduced the FL time due to maximum power consumption. Our proposed ADA increased the CPU when more data samples were input, in terms of FL time. Moreover, the Fixed algorithm performed much better in optimizing the computation frequency allocation f_n of the IoT devices, in terms of FL time. The computation learning time and the transmission time of IoT devices depended on FL time. From Fig. 8, the Fixed algorithm optimized the computation frequency allocation f_n by selecting the highest CPU frequency and transmission power to reduce the FL time based on the intelligent selection of active IoT devices that implement local learning to the appropriate ENs. The other algorithms did not change with the number of devices or completion time. This was because the FL time was generally controlled by

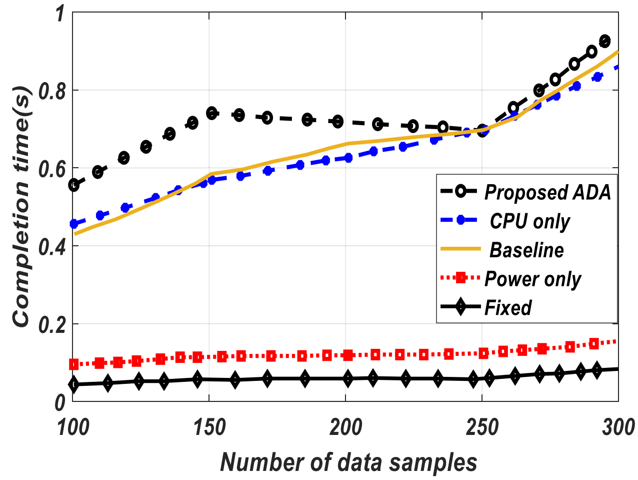

FIGURE 7. Completion time related to number of data samples.

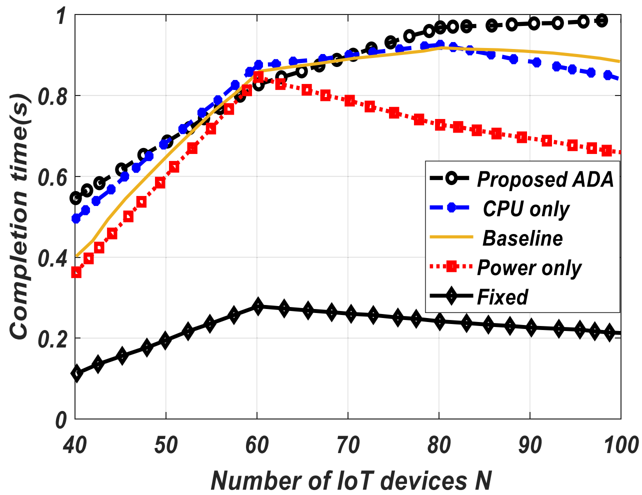

FIGURE 8. Completion time related to number of IoT devices.

one IoT device's extended local learning time. The proposed ADA maintained a steady energy consumption, adjusted the power allocation, prevented long wait times due to poor channels, and met the FL time requirement for all IoT devices. The proposed ADA utilized the first and second derivative $\sum_{n=1}^{N} \vartheta_n' \left(\sigma_n \left(\Psi^*\right)\right) \sigma_n' \left(\Psi^*\right)$ with the desired level of training learning accuracy based on a repeated number of local iterations to ensure the global model is achieved and minimize the time consumed in each round of FL.The completion times for all algorithms do not change much at the number of IoT devices increases as shown in Fig. 8. This is because the FL time is primarily determined by the IoT device with the longest local training time, which does not significantly change when the number of IoT device is increased. From Fig. 9, the duration of local model training depended on the number of CPU cycles, as shown in (23). Hence, Fixed and Power-only algorithms would increase with more IoT device CPU cycles, setting the CPU frequency at the largest value. However, the proposed ADA and CPU-only algorithms would be adaptable to CPU frequencies and not change much with the number of CPU cycles.

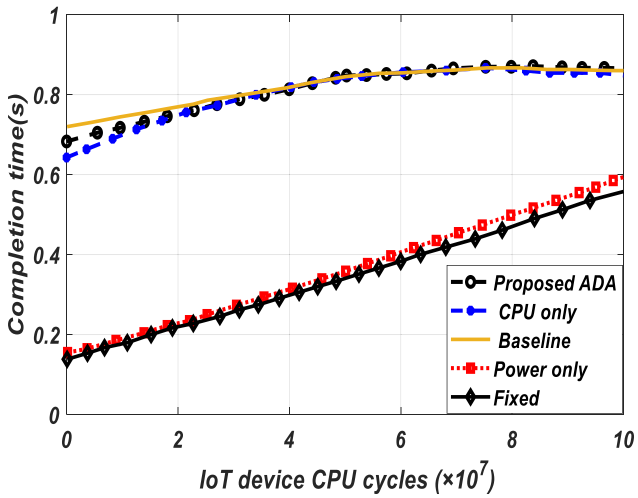

FIGURE 9. Completion time vs CPU cycles.

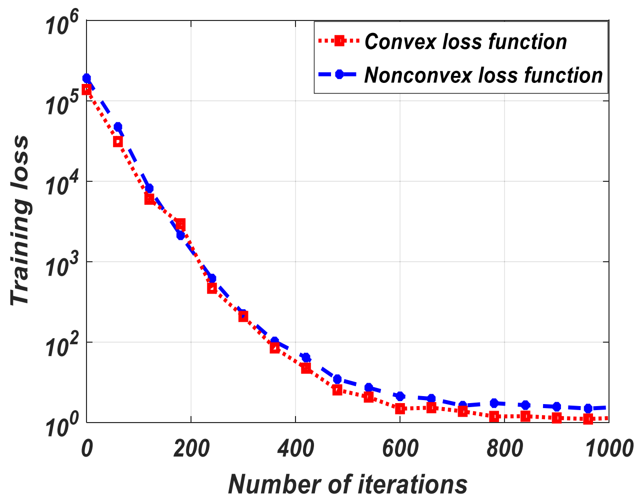

FIGURE 10. Training loss vs number of iterations.

C. CONVERGENCE BEHAVIOR

From Fig. 10, the value of training loss varied with the number of iterations for convex and non-convex loss functions. As the number of iterations increased, the training loss decreased rapidly. This was followed by a gradual decrease in both convex and non-convex loss functions. The completion time in (35) was still non-convex, which reduced the overall completion time and data exchange between IoT devices and ENs. The total time consumption was challenging to analyze with the local loss function, which depended on the desired level of learning accuracy to solve the convex optimization problem. The learning accuracy determines the trade-off between computation learning time and transmission time by establishing the total time as a convex function of learning accuracy. From Fig. 10, the initial value of the training loss $\mathcal{F}(\omega_i) = 10^5$, and the value of the loss function reduced to $\mathcal{F}(\omega_{1000}) = 1$ for the convex loss function. The convex loss function approaches were faster than the non-convex loss function. The FL can be applied to the non-convex loss function as it changes with the number of iterations.

VI. CONCLUSION

In this paper, we investigated the optimization of computation frequency allocation, optimal transmission power, and the desired level of learning accuracy to reduce energy consumption completion time and improve FL performance in edge intelligence for green IoT networks beyond fifth generation. Also, we propose ADA by applying the closed-form solution and Lagrangian multiplier methods of each IoT device to develop a computationally efficient resource allocation that satisfies the FL time and reduces energy consumption based on joint optimization of the CPU frequency and optimal transmission power. The simulation results indicated that the proposed ADA can adjust the CPU frequency and power transmissions to reduce the energy consumption of IoT devices at the cost of FL computing/training time. Finally, the proposed ADA performed better than all other algorithms by incurring the smallest energy consumption for a large number of IoT devices and large data sizes to train the local models. From the results, at increasing the amount of data samples from 50 to 350, the performance of energy consumption for the proposed ADA does not vary much in order to save energy. By scheduling IoT devices, adapting the transmission rate, and maximizing the transmission power of each IoT device, our proposed schemes illustrate the balance between the aforementioned factors. The ADA achieves the best performance among all schemes by jointly optimizing bandwidth and desired level of learning accuracy to cancel self-interference and minimize the completion time until the global model meets the optimal learning accuracy. In future works, we will concentrate on designing efficient and robust double deep Q-learning algorithms to provide smart packet transmission scheduling in channel state information evolution over time in large-cognitive IoT networks.

REFERENCES

[1] M. S. Al-Abiad, M. Z. Hassan, and M. J. Hossain, "Energy efficient federated learning in integrated fog-cloud computing enabled Internet-of-Things networks," 2021, *arXiv:2107.03520*.

[2] Q. Zeng, Y. Du, K. Huang, and K. K. Leung, "Energy-efficient radio resource allocation for federated edge learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[3] J. Yao and N. Ansari, "Enhancing federated learning in fog-aided IoT by CPU frequency and wireless power control," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3438–3445, Mar. 2021.

[4] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[5] L. Li, "Delay analysis of wireless federated learning based on saddle point approximation and large deviation theory," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3772–3789, Dec. 2021.

[6] F. P.-C. Lin, C. G. Brinton, and N. Michelusi, "Federated learning with communication delay in edge networks," in *Proc. GLOBECOM IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.
[7] R. Saha, S. Misra, and P. K. Deb, "FogFL: Fog-assisted federated learning for resource-constrained IoT devices," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 8456–8463, May 2021.
[8] 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.
[9] M. M. Wadu, S. Samarakoon, and M. Bennis, "Joint client scheduling and resource allocation under channel uncertainty in federated learning," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5962–5974, Sep. 2021.
[10] S. Wang, "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 3, pp. 1205–1221, Jun. 2019.
[11] Y. Mao, J. Zhang, Z. Chen, and K. B. Letaief, "Dynamic computation offloading for mobile-edge computing with energy harvesting devices," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3590–3605, Dec. 2016.
[12] J. Yao and N. Ansari, "Task allocation in fog-aided mobile IoT by Lyapunov online reinforcement learning," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 556–565, Jun. 2020.
[13] A. Yousefpour, G. Ishigaki, R. Gour, and J. P. Jue, "On reducing IoT service delay via fog offloading," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 998–1010, Apr. 2018.
[14] 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.
[15] J. Ren, W. Ni, G. Nie, and H. Tian, "Research on resource allocation for efficient federated learning," 2021, *arXiv:2104.09177*.
[16] A. Albaseer, M. Abdallah, A. Al-Fuqaha, and A. Erbad, "Fine-grained data selection for improved energy efficiency of federated edge learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3258–3271, Sep. 2022.
[17] P. Santi, "On the data gathering capacity and latency in wireless sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1211–1221, Sep. 2010.
[18] T. T. Anh, N. C. Luong, D. Niyato, D. I. Kim, and L.-C. Wang, "Efficient training management for mobile crowd-machine learning: A deep reinforcement learning approach," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1345–1348, Oct. 2019.
[19] A. Brecko, E. Kajati, J. Koziorek, and I. Zolotova, "Federated learning for edge computing: A survey," *Appl. Sci.*, vol. 12, no. 18, pp. 9124–9137, 2022.
[20] H. G. Abreha, M. Hayajneh, and M. A. Serhani, "Federated learning in edge computing: A systematic survey," *Sensors*, vol. 22, no. 2, pp. 450–464, 2022.
[21] F. Foukalas and A. Tziouvaras, "Federated learning protocols for IoT edge computing," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13570–13581, Aug. 2022.
[22] J. Ren, J. Sun, H. Tian, W. Ni, G. Nie, and Y. Wang, "Joint resource allocation for efficient federated learning in Internet of Things supported by edge computing," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6.
[23] H. Chen, S. Huang, D. Zhang, M. Xiao, M. Skoglund, and H. V. Poor, "Federated learning over wireless IoT networks with optimized communication and resources," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16592–16605, Sep. 2022.
[24] R. Yu and P. Li, "Toward resource-efficient federated learning in mobile edge computing," *IEEE Netw.*, vol. 35, no. 1, pp. 148–155, Jan. 2021.
[25] Z. Chen, W. Liao, K. Hua, C. Lu, and W. Yu, "Towards asynchronous federated learning for heterogeneous edge-powered Internet of Things," *Digit. Commun. Netw.*, vol. 7, no. 3, pp. 317–326, Aug. 2021.
[26] J. Xu, H. Wang, and L. Chen, "Bandwidth allocation for multiple federated learning services in wireless edge networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2534–2546, Apr. 2022.
[27] H. H. Yang, Z. Liu, T. Q. S. Quek, and H. V. Poor, "Scheduling policies for federated learning in wireless networks," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 317–333, Jan. 2020.
[28] J. Yao and N. Ansari, "Caching in energy harvesting aided Internet of Things: A game-theoretic approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3194–3201, Apr. 2019.
[29] J. Yao and N. Ansari, "Fog resource provisioning in reliability-aware IoT networks," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8262–8269, Oct. 2019.
[30] T. D. Burd and R. W. Brodersen, "Processor design for portable systems," *J. VLSI Signal Process. Syst.*, vol. 13, nos. 2–3, pp. 203–221, 1996.
[31] A. Albaseer, M. Abdallah, A. Al-Fuqaha, and A. Erbad, "Threshold-based data exclusion approach for energy-efficient federated edge learning," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6.
[32] C. T. Dinh, "Federated learning over wireless networks: Convergence analysis and resource allocation," *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 398–409, Feb. 2021.
[33] H. Hult, F. Lindskog, O. Hammarlid, and C. J. Rehn, "Convex optimization," in *Risk and Portfolio Analysis: Principles and Methods* (Springer Series in Operations Research and Financial Engineering). New York, NY, USA: Springer, 2012, pp. 33–38.
[34] J. Zhao, Y. Feng, X. Chang, and C. H. Liu, "Energy-efficient client selection in federated learning with heterogeneous data on edge," *Peer Peer Netw. Appl.*, vol. 15, no. 2, pp. 1139–1151, Mar. 2022.
[35] J. Yao and N. Ansari, "QoS-aware power control in Internet of Drones for data collection service," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6649–6656, Jul. 2019.
[36] J. Kwak, Y. Kim, J. Lee, and S. Chong, "DREAM: Dynamic resource and task allocation for energy minimization in mobile cloud systems," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2510–2523, Dec. 2015.
[37] L. Klozar and J. Prokopec, "Propagation path loss models for mobile communication," in *Proc. 21st Int. Conf. Radioelektronika*, Apr. 2011, pp. 1–4.
[38] J. Yao and N. Ansari, "Secure federated learning by power control for Internet of Drones," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 4, pp. 1021–1031, Dec. 2021.
[39] N. A. Khan, A. Awang, and S. A. A. Karim, "Security in Internet of Things: A review," *IEEE Access*, vol. 10, pp. 104649–104670, 2022.

ADEB SALH received the Bachelor of Electrical and Electronic Engineering degree from IBB University, Ibb, Yemen, in 2007, and the master's and Ph.D. degrees in electrical and electronic engineering from the University Tun Hussein Onn Malaysia, in 2015 and 2020, respectively. From 2007 to 2012, he worked as a Lecturer Assistant with the Yareem Community College. He is currently a Postdoctoral Researcher with the Faculty of Electrical and Electronic Engineering, Universiti Technology Malaysia. His research interests include 5G, 6G wireless communications, massive MIMO, artificial intelligence (AI), and the Internet of Things (IoT).

RAZALI NGAH (Member, IEEE) received the bachelor's degree in electrical engineering (communication) from Universiti Teknologi Malaysia (UTM), Skudai, in 1989, the M.Sc. degree in RF communication engineering from the University of Bradford, U.K., in 1996, and the Ph.D. degree in photonics from Northumbria University, U.K., in 2005. Since 1989, he has been with the Faculty of Electrical Engineering, UTM. Currently, he is an Associate Professor with the Wireless Communication Centre (WCC), Faculty of Electrical Engineering, UTM. He has published more than 150 technical papers for journals and international conferences. His research interests include mobile radio propagation, antenna and RF design, photonics networks, wireless communication systems, and radio over fiber (RoF).

LUKMAN AUDAH (Member, IEEE) received the Bachelor of Engineering degree in telecommunications from Universiti Teknologi Malaysia, in 2005, and the M.Sc. degree in communication networks and software and the Ph.D. degree in electronic engineering from the University of Surrey, U.K. He is currently a Senior Lecturer with the Communication Engineering Department, Universiti Tun Hussein Onn Malaysia. His research interests include wireless and mobile communications, internet traffic engineering, network system management, data security, and satellite communications.

KWANG SOON KIM (Senior Member, IEEE) received the B.S. (summa cum laude), M.S.E., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in February 1994, February 1996, and February 1999, respectively. From March 1999 to March 2000, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA, USA. From April 2000 to February 2004, he was a Senior Member of Research Staff with the Mobile Telecommunication Research Laboratory, Electronics and Telecommunication Research Institute (ETRI), Daejeon. Since March 2004, he has been with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea, where he is currently a Professor. His research interests include signal processing, communication theory, information theory, and stochastic geometry applied to wireless heterogeneous cellular networks, wireless local area networks, wireless D2D networks, wireless ad hoc networks, and new radio access technologies for 5G. He was a recipient of the Postdoctoral Fellowship from the Korea Science and Engineering Foundation (KOSEF), in 1999. He received the Outstanding Researcher Award from ETRI, in 2002, the Jack Neubauer Memorial Award (Best System Paper Award, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY) from the IEEE Vehicular Technology Society, in 2008, and the LG Research and Development Award: Industry-Academic Cooperation Prize, LG Electronics, in 2013. From 2006 to 2012, he served as an Editor for the *Journal of the Korean Institute of Communications and Information Sciences* (KICS). From 2013 to 2016, he served as the Editor-in-Chief for the *Journal of KICS*. Since 2008, he has been serving as an Editor for the *Journal of Communications and Networks* (JCN). From 2009 to 2014, he served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

QAZWAN ABDULLAH (Senior Member, IEEE) was born in Taiz, Yemen. He received the bachelor's and Master of Science degrees in electrical and electronic engineering from the Universiti Tun Hussein Onn Malaysia (UTHM), in 2013 and 2015, respectively. Currently, he is a Research Assistant. He has more than 40 scientific publications. His research interests include control theory, adaptive fuzzy logic controller, mobile communications (5G/6G), fuzzy logic control and its applications, motor drive, electric vehicles, and antenna filter design.

YAHYA M. AL-MOLIKI (Member, IEEE) worked as an Assistant Lecturer for some undergraduate electrical engineering courses. He was an Assistant Lecturer for undergraduate courses with Taiz University, Yemen, from 2008 to 2010. He is currently a part-time Assistant Professor with King Saud University (KSU), Muzahimiyah Campus.

KHALED A. ALJALOUD (Member, IEEE) received the B.S. degree from King Saud University, Riyadh, Saudi Arabia, in 2009, the M.S. degree from the University of Southern California, Los Angles, CA, USA, in 2013, and the Ph.D. degree from University College London, London, U.K., in 2018. He is currently an Assistant Professor with the College of Applied Engineering, King Saud University. His current research interests include RF circuits, embedded systems and deep learning-based embedded systems, and their uses in robotics and the IoT technology.

HAIRUL NIZAM TALIB (Member, IEEE) was born in Malaysia, in 1976. He received the B.S. degree in electrical engineering from the Universiti Teknologi Malaysia (UTM), Johor, Malaysia, in 1999, the M.S. degree in electrical engineering from the University of Nottingham, Nottingham, U.K., in 2005, and the Ph.D. degree from the Universiti Teknikal Malaysia Melaka (UTeM), Malaysia, in 2016. He is currently a Senior Lecturer with UTeM. His main research interests include power electronics, fuzzy logic control, and motor drives.