Coordinated Federated Learning for Radio Resource Management in Future Wireless Networks

Ishtiaq Ahmad, Zdenek Becvar, Pavel Mach

***Abstract*—In this letter, we propose a coordinated federated learning framework for joint optimization of multiple radio resource management parameters in wireless networks. Like in a traditional federated learning, each user equipment (UE) trains a local machine learning model and shares parameters of the local models. However, the proposed approach also involves the communication-related aspects in the process of the federated learning and considers mutual dependencies among optimized variables. To this end, the UEs also share values of a new designed local loss functions specifically tailored to capture performance of the communication network to enable a local coordination for capturing mutual relation between optimized parameters. Then, base stations (BSs) leverage the received values of the new loss functions to produce a weighted global machine learning model. The weighting is based on the loss values so that the models of the UEs with low loss values are prioritized, since the low loss values indicate a proper setting of radio resource management parameters. The simulation results demonstrate a significant improvement from 10.6% to 18.9% and from 3.1% to 11.2% in sum capacity and ratio of UEs satisfied with the capacity, respectively, compared to the best performing state-of-the-art federated learning solution.**

***Index Terms*—Federated learning, radio resource management, transmission power, bandwidth, coordinated learning, 6G.**

I. Introduction

The combination of widely heterogeneous data and diverse quality of service (QoS) requirements in future wireless networks motivate to apply an intelligent optimization of radio resource management (RRM) [1]. Conventional centralized machine learning-based RRM approaches rely on collecting raw data from user equipments (UEs) at a base station (BS) for training of machine learning models [2]. While effective in small-scale systems, the centralized learning raises following challenges: (*i*) excessive communication overhead due to the continuous transfer of huge UEs' data, (*ii*) limited scalability as the number of UEs grows, and (*iii*) privacy risks associated with sharing sensitive data [3].

A promising extension of the centralized learning is seen in a federated learning (FL) enabling the UEs to collaboratively learn a global machine learning model without exchanging sensitive raw data [4]. In the FL, each UE trains a local machine learning model using its own gathered data and transmits only parameters of the local model to a serving BS. The serving BS, then, aggregates the local models from the UEs and produces the global machine learning model, which is subsequently shared back with the UEs for a local inference [5]. Thus, the FL reduces communication costs and preserves privacy [6]. However, a fundamental challenge is the inherent heterogeneity among the UEs (e.g., different channel conditions, computational capabilities, QoS requirements) [7]. Under such heterogeneity, the conventional aggregation of local models can degrade performance, as even few inaccurate or outdated local models can mislead the global model [8].

To overcome the limitations of the conventional aggregation for FL, a weighted aggregation of the local models is introduced. In [9], the authors propose a weighting of the local models according to a local dataset size to prioritize the UEs with more collected data. In [10], the weighting is designed so that the UEs with a higher historical contribution impact the global model more significantly. In [11], the authors employ reinforcement learning at the BS to optimize the weighting. In [12], the UEs whose updates reduce the global loss, i.e., the weighted average of local mean square error (MSE) losses across all UEs, are prioritized for the aggregation.

The main *limitation of existing FL* for wireless networks is consideration of only machine learning–level metrics (e.g., model accuracy, loss convergence) to prioritize local models of the UEs for a global aggregation at the BS. Such metrics do not reflect the actual communication performance, making the global aggregation process blind to wireless conditions. Moreover, the existing FL treats individual RRM parameters independently ignoring mutual dependencies among parameters. Neglecting the dependencies leads to suboptimal decisions and to a degraded network performance. To address the limitations of existing FL, we propose a coordinated federated learning (CFL) for RRM. The CFL incorporates communication-aware information into global model aggregation and accounts for the dependencies among RRM parameters.

The main contributions are summarized as follows:

- We introduce communication-driven loss functions directly reflecting wireless performance of the UEs. The loss functions steer construction of the global model by the BS based on the actual communication effectiveness.
- We capture natural dependencies among RRM variables to allow coordinated optimization of multiple parameters via the introduced loss functions exchanged mutually among individual machine learning models.
- We demonstrate benefits of the proposed CFL on joint optimization of transmission power and bandwidth allocation by two lightweight deep neural networks (DNNs) at each UE. The coordination between DNNs is enforced

I. Ahmad, Z. Becvar, and P. Mach are with the Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic, Email: {ahmadish, zdenek.becvar, machp2}@fel.cvut.cz

This work was supported by the project no. 23-05646S funded by the Czech Science Foundation.

- via the developed loss functions and the aggregation of the UEs' local coordinated models at the BS is guided and weighted by real wireless performance.
- We show that the CFL increases the sum capacity by 10.6%–18.9% and the ratio of UEs satisfied with experienced capacity by 3.1%–11.2% compared to the best performing state-of-the-art FL baseline.

II. System Model and Problem Formulation

This section introduces network and communication models, models of DNNs for transmission power setting and bandwidth allocation, and formulates the targeted problem.

A. Network and Communication Models

We consider a wireless network comprising N UEs and K BSs. Without loss of generality, each UE is served by the BS with the highest channel gain. The achievable communication capacity of the n-th UE served by the k-th BS is:

$$c_{n,k} = B_{n,k}\log_2\left(1+\frac{p_n g_{n,k}}{\sigma^2+\sum_{n'\in\mathcal{I}_n} p_{n'} g_{n',k}}\right) \quad (1)$$

where $B_{n,k}$ is the bandwidth of the n-th UE, p_n and $p_{n'}$ are the transmission powers of the n-th UE and of the n'-th UE causing interference to the n-th UE's transmission, respectively, $g_{n,k}$ ($g_{n',k}$) is the channel gain between the n-th UE (n'-th UE) and the k-th BS, σ^2 is the noise power, and $\mathcal{I}_n = \{n' \neq n \mid B_{n'} \cap B_n \neq \emptyset\}$ denotes the set of all interfering UEs sharing the bandwidth with the n-th UE.

B. Machine Learning and Federated Learning Models

To demonstrate principle of the proposed CFL, we assume two DNNs deployed at each UE, one for the transmission power setting (denoted as P-DNN) and another for bandwidth allocation (B-DNN). Note that we do not present any novelty in this part and arbitrary DNNs solving any RRM problem can be adopted in the same way. The following subsections describe both DNNs and conventional FL models.

1) P-DNN: The DNN for transmission power setting consists of an input layer, H_c^{P} fully connected hidden layers, and an output layer, as in [13]. Let $\mathcal{F}_n^{\text{P}}(\cdot;\boldsymbol{\theta}_{n,k}^{\text{P}})$ denote the P-DNN model with trainable parameter $\boldsymbol{\theta}_{n,k}^{\text{P}}$ used by the n-th UE served by the k-th BS. The input layer receives the feature vector $\boldsymbol{I}_n^{\text{P}} \in \mathbb{R}$ for the n-th UE containing: (i) the channel gain $g_{n,k}$ and (ii) the vector $\mathbf{g}_{\mathcal{I}_n,k} = \{g_{n',k} \mid n' \in \mathcal{I}_n\}$ of channel gains of interfering UEs. For the input layer (layer 0), the activations are set as $\mathbf{a}^{(0,\text{P})} = \boldsymbol{I}_n^{\text{P}}$. For each hidden layer $h^{\text{P}} = [h_1^{\text{P}}, h_2^{\text{P}}, \ldots, H_c^{\text{P}}]$ with the neurons $i^{\text{P}} = [x_1^{\text{P}}, x_2^{\text{P}}, \ldots, X^{\text{P}}]$, the pre-activation value is labeled as $z_i^{(h,\text{P})}$ and we adopt a sigmoid pre-activation function, i.e., $1/(1+e^{-z_i^{(h,\text{P})}})$, which introduces non-linearity enabling the DNN to model complex functions and learn non-linear decision boundaries. Then, the output layer provides the transmission power $\hat{p}_n$ to be set by the n-th UE. The learning of P-DNN minimizes the MSE between the predicted and target values via a common internal loss function $\mathcal{L}_{n,k}^{\text{P}} = \frac{1}{M}\sum_{i=1}^{M}(p_n - \hat{p}_n)^2$, where M is the number of training samples, and $\bar{p}_n$ is the target value (see P-DNN in the left part of Fig. 1 with "Conventional FL").

2) B-DNN: The DNN for bandwidth allocation consists of an input layer, H_c^{B} fully connected hidden layers, and an output layer, as in [14]. Let $\mathcal{F}_n^{\text{B}}(\cdot;\boldsymbol{\theta}_{n,k}^{\text{B}})$ represents the B-DNN model with trainable parameters $\boldsymbol{\theta}_{n,k}^{\text{B}}$ used by the n-th UE served by the k-th BS. The input to the B-DNN consists of: (i) the channel gain $g_{n,k}$ and (ii) the vector $\mathbf{g}_{\mathcal{I}_n,k} = \{g_{n',k} \mid n' \in \mathcal{I}_n\}$ of the channel gains of all interfering UEs. The input vector is denoted by $\boldsymbol{I}_n^{\text{B}}$, and the input layer activations are initialized as $a_j^{(0,\text{B})} = I_n^{\text{B}}$. For each hidden layer $h^{\text{B}} = [h_1^{\text{P}}, h_2^{\text{P}}, \ldots, H_c^{\text{B}}]$ and each neuron $i^{\text{B}} = [x_1^{\text{P}}, x_2^{\text{P}}, \ldots, X_h^{\text{B}}]$, the sigmoid activation function $1/(1+e^{-z_i^{(h,\text{B})}})$ is applied to the pre-activation value $z_i^{(h,\text{B})}$. The B-DNN outputs the bandwidth $\hat{B}_{n,k}$ allocated to the n-th UE by the k-th BS. The learning of B-DNN minimizes the MSE via a common internal loss function $\mathcal{L}_{n,k}^{\text{B}} = \frac{1}{M}\sum_{i=1}^{M}(B_{n,k} - \hat{B}_{n,k})^2$, where $\bar{B}_{n,k}$ is the target value (see B-DNN in the left part of Fig. 1).

3) Conventional Federated Learning Model: Each n-th UE served the k-th BS maintains and learns one local machine learning model for P-DNN and one for B-DNN with the parameters $\boldsymbol{\theta}_{n,k}^{\text{P}}$ and $\boldsymbol{\theta}_{n,k}^{\text{B}}$, respectively, see Fig. 1. After the local learning, each UE transmits models' parameters to the serving BS. The serving BS performs an aggregation of the local models from all served UEs to obtain the independent global machine learning models $\boldsymbol{\theta}_k^{\text{P,agg}}$ and $\boldsymbol{\theta}_k^{\text{B,agg}}$ for P-DNN and B-DNN, respectively. The aggregated global machine learning models are then returned to the UEs for local inference.

C. Problem Formulation

To demonstrate the novel CFL for RRM, we define the problem of the sum capacity maximization via optimization of two key RRM parameters: transmission power and bandwidth allocation for all UEs. The problem is formulated as:

$$\begin{aligned}
\boldsymbol{p}^*, \boldsymbol{B}^* = \quad & \underset{p,B}{\arg\max} \sum\nolimits_k \sum\nolimits_n c_{n,k} \\
\text{s.t.} \quad & a) \quad p_n \leq p_n^{\max}, \quad \forall_n, \\
& b) \quad \sum\nolimits_n B_{n,k} \leq B, \quad \forall_k, \\
& c) \quad B_{n,k} \geq 0, \quad \forall_n, \ \forall_k, \\
& d) \quad c_{n,k} \geq c_{\min}, \quad \forall_n,
\end{aligned} \quad (2)$$

where $\boldsymbol{p}^*$ and $\boldsymbol{B}^*$ represent the optimal transmission power and bandwidth matrices, respectively. Constraint (2a) limits

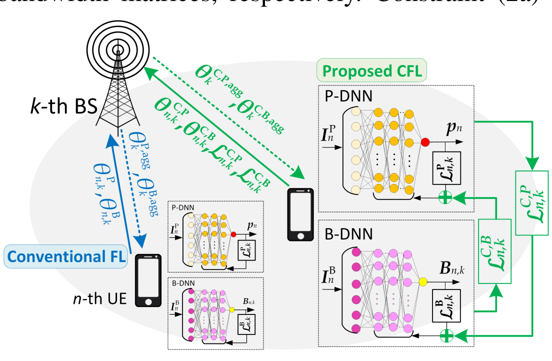

Fig. 1. Proposed CFL and conventional FL for joint optimization of transmission power and bandwidth (differences of the proposed CFL with respect to the conventional CL are highlighted by green color).

the transmission power to the maximum allowed $p_n^{\max}$, constraint (2b) ensures the total bandwidth assigned to the UEs at each BS fits the total available bandwidth B, constraint (2c) guarantees non-negativity of the assigned bandwidth, and constraint (2d) enforces fulfillment of the minimum required capacity c_{min} for the UEs. The joint resource management problem (2) is inherently non-convex and NP-hard due to the coupled decision variables and interference effects. Hence, in the next section, we propose the CFL to solve the problem.

III. CFL for Radio Resource Management

In this section, we present the proposed CFL to optimize RRM across multiple UEs, as illustrated in Fig. 1. We first define new communication-aware loss functions for the UEs to enable local coordination between P-DNN and B-DNN to not only maximize the sum capacity but also to handle potential constraint violation. Then, we describe a coordinated federated aggregation of the UEs' models at the BS.

A. Communication-aware Loss Functions and Coordination

The design of the UE's loss functions aims to update the parameters of the models for P-DNN and B-DNN. Each DNN is trained independently to allow easy light-weight training and scalability (easy addition of other parameters in the future). At the same time, both DNNs operate on shared observations and are coupled through a coordinated feedback mechanism to ensure efficient RRM. The objective of P-DNN is not only to stay within the allowed power budget but also to ensure that the resulting transmission power is sufficient to reach the minimum required c_{min}. To achieve this, P-DNN uses a new communication-aware loss function that balances following three aspects resulting from (2) and constraints: (*i*) maximizing capacity, (*ii*) meeting requirements on c_{min}, and (*iii*) ensuring the maximum transmission power limit is met. These aspects are integrated to the loss function as follows:

$$\begin{aligned}\mathcal{L}_{n,k}^{\text{C,P}} = &-\log\left(\frac{\hat{c}_{n,k}}{c_n^{\min}}\right) + \max\left(0, \frac{c_{\text{min}} - \hat{c}_{n,k}}{c_{\text{min}}}\right) \\ &+ \max\left(0, \frac{\hat{p}_n - p_n^{\max}}{p_n^{\max}}\right),\end{aligned} \tag{3}$$

where the first term encourages maximizing the capacity $\hat{c}_{n,k}$ of the n-th UE served by the k-th BS based on the determined power $\hat{p_n}$ by P-DNN, the second term penalizes the capacity below the minimum required capacity c_{min}, and the third term penalizes the transmission power exceeding the maximum allowed power $p_n^{\max}$.

The B-DNN predicts the optimal bandwidth $\hat{B}_{n,k}$ for the n-th UE served by the k-th BS by leveraging the communication-aware loss function reflecting (2) and constraints: (*i*) capacity maximization, (*ii*) fulfillment of c_{min}, and (*iii*) ensuring the total available system bandwidth is not exceeded. Hence, the loss function for B-DNN is defined as:

$$\begin{aligned}\mathcal{L}_{n,k}^{\text{C,B}} = &-\log\left(\frac{\hat{c}_{n,k}}{c_{\text{min}}}\right) + \max\left(0, \frac{c_{\text{min}} - \hat{c}_{n,k}}{c_{\text{min}}}\right) \\ &+ \max\left(0, \frac{\sum_{n=1}^{N} \hat{B}_{n,k} - B}{B}\right),\end{aligned} \tag{4}$$

The first and second terms are similar to those in (3) for P-DNN. The third term imposes a penalty if the total allocated bandwidth at the k-th BS exceeds the available bandwidth B.

Both introduced loss functions $\mathcal{L}_{n,k}^{\text{C,P}}$ and $\mathcal{L}_{n,k}^{\text{C,B}}$ generated by P-DNN and B-DNN are locally delivered to B-DNN and P-DNN, respectively, to keep both DNNs mutually aware of the performance to reflect interdependence between the power and the bandwidth. In the P-DNN and B-DNN, the loss functions $\mathcal{L}_{n,k}^{\text{C,P}}$ and $\mathcal{L}_{n,k}^{\text{C,B}}$ are added to the conventional internal loss functions $\mathcal{L}_{n,k}^{\text{P}}$ and $\mathcal{L}_{n,k}^{\text{B}}$, respectively, to effectively capture the network performance in both DNNs. The constraints in (2) are integrated into the learning process through the penalty terms ensuring the power and bandwidth limits as well as the minimum capacity requirements are respected.

B. Principle of Coordinated Federated Learning

This subsection describes the CFL framework for RRM with P-DNN and B-DNN. The inference proceeds in three iterative stages: (*i*) local inference at the UEs, (*ii*) global aggregation of the UEs models at the BS, and (*iii*) retraining/inference at the UE with the global model.

1) Local Training and Inference: Each n-th UE served by the k-th BS infers its local P-DNN and B-DNN models on local data with the following input features $\{g_{n,k}, \mathbf{g}_{\mathcal{I}_n,k}, \bar{p}_n, \bar{B}_{n,k}\}$, where $\bar{p}_n$, $\bar{B}_{n,k}$ are the targets of transmission power and bandwidth. The local models are updated independently by minimizing the new designed communication-aware loss functions $\mathcal{L}_{n,k}^{\text{C,P}}$ and $\mathcal{L}_{n,k}^{\text{C,B}}$ using gradient descent, i.e., $\boldsymbol{\theta}_{n,k}^{\text{C,P},(t+1)} \leftarrow \boldsymbol{\theta}_{n,k}^{\text{C,P},(t)} - \eta_1^{\text{C,P}} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{n,k}^{\text{C,P}}$, and $\boldsymbol{\theta}_{n,k}^{\text{C,B},(t+1)} \leftarrow \boldsymbol{\theta}_{n,k}^{\text{C,B},(t)} - \eta_1^{\text{C,B}} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{n,k}^{\text{C,B}}$, where $\eta_1^{\text{C,P}}$ and $\eta_1^{\text{C,B}}$ are the learning rates of P-DNN and B-DNN, respectively, and t is the index of the local training iteration. Like in the conventional FL, upon completion of the local inference, each UE transmits the updated model parameters $\boldsymbol{\theta}_{n,k}^{\text{C,P},(t+1)}$ and $\boldsymbol{\theta}_{n,k}^{\text{C,B},(t+1)}$ to the k-th BS. In the proposed CFL, however, the model parameters are determined considering also the new communication-aware loss functions allowing to directly reflect communication performance in the next step, i.e., in the global aggregation of local models. Besides, in the CFL, the UEs send also the new loss values $\mathcal{L}_{n,k}^{\text{C,P}}$ and $\mathcal{L}_{n,k}^{\text{C,B}}$ to allow the BS to adjust the global model according to the UEs communication-related performance.

2) Global Aggregation and Model Distribution: At the k-th BS, the parameters $\boldsymbol{\theta}_{n,k}^{\text{C,P},(t)}$ and $\boldsymbol{\theta}_{n,k}^{\text{C,B},(t)}$ and the loss values $\mathcal{L}_{n,k}^{\text{C,P}}$ and $\mathcal{L}_{n,k}^{\text{C,B}}$ are received from the UEs and are aggregated to generate updated BS-level global models $\boldsymbol{\theta}_{k}^{\text{C,P,agg},(t)}$ and $\boldsymbol{\theta}_{k}^{\text{C,B,agg},(t)}$ for P-DNN and B-DNN, respectively. Unlike conventional aggregation schemes that prioritize UEs based on indirect measures related to only machine learning [9]–[12], the proposed aggregation scheme leverages the loss values to implicitly prioritize contributions from the UEs whose local models demonstrate superior communication performance. Thus, in the CFL, the UEs with a lower loss values, i.e., lower $\mathcal{L}_{n,k}^{\text{C,P}}$ and $\mathcal{L}_{n,k}^{\text{C,B}}$, get larger aggregation weights, giving their models a higher impact on the global update. This integration

Algorithm 1 Proposed CFL for radio resource management.

1: **for** each UE **do**
2: Initialize model parameters $\boldsymbol{\theta}^{\mathrm{P}}_{n,k}$ and $\boldsymbol{\theta}^{\mathrm{B}}_{n,k}$
3: Compute $\mathcal{L}^{\mathrm{C,P}}_{n,k}$ and $\mathcal{L}^{\mathrm{C,B}}_{n,k}$ using (3) and (4)
4: Perform local learning and obtain $\boldsymbol{\theta}^{\mathrm{C,P}}_{n,k}$ and $\boldsymbol{\theta}^{\mathrm{C,B}}_{n,k}$
5: **end for**
6: **for** each aggregation round **do**
7: Transmit $\boldsymbol{\theta}^{\mathrm{C,P}}_{n,k}$ and $\boldsymbol{\theta}^{\mathrm{C,B}}_{n,k}$ to BS
8: BS determine $\boldsymbol{\theta}^{\mathrm{C,P,agg}}_{k}$ and $\boldsymbol{\theta}^{\mathrm{C,B,agg}}_{k}$ using (5) and (6)
9: **for** each UE **do**
10: $\boldsymbol{\theta}^{\mathrm{C,P}}_{n,k} \leftarrow \boldsymbol{\theta}^{\mathrm{C,P,agg}}_{k}$ and $\boldsymbol{\theta}^{\mathrm{C,B}}_{n,k} \leftarrow \boldsymbol{\theta}^{\mathrm{C,B,agg}}_{k}$
11: **end for**
12: **end for**

helps to mitigate the degradation of global performance caused by unreliable updates due to ignoring a real impact of the local models on communication. The aggregation for P-DNN at the k-th BS is computed as:

$$\boldsymbol{\theta}^{\mathrm{C,P,agg},(t)}_{k} = \frac{1}{\mathcal{N}_k} \sum_{n \in \mathcal{N}_k} \frac{1}{\mathcal{L}^{\mathrm{C,P}}_{n,k}} \boldsymbol{\theta}^{\mathrm{C,P},(t)}_{n,k}, \tag{5}$$

where $\mathcal{N}_k$ is the number of UEs associated with the k-th BS. Similarly, the aggregation for B-DNN is done as:

$$\boldsymbol{\theta}^{\mathrm{C,B,agg},(t)}_{k} = \frac{1}{\mathcal{N}_k} \sum_{n \in \mathcal{N}_k} \frac{1}{\mathcal{L}^{\mathrm{C,B}}_{n,k}} \boldsymbol{\theta}^{\mathrm{C,B},(t)}_{n,k}. \tag{6}$$

The communication-aware weighting enables the global model to prioritize local models with lower loss indicating that the UE's local RRM decisions are more effective in achieving the overall network objectives. Upon completing the aggregation at the BS, the updated global model parameters $\boldsymbol{\theta}^{\mathrm{C,P,agg},(t)}_{k}$ and $\boldsymbol{\theta}^{\mathrm{C,B,agg},(t)}_{k}$ are transmitted back to the UEs.

3) Retraining/Inference at the UEs: During this stage, the UEs perform local retraining using the updated global models from the serving BSs so that $\boldsymbol{\theta}^{\mathrm{C,P},(t)}_{n,k} = \boldsymbol{\theta}^{\mathrm{C,P,agg},(t)}_{k}$ and $\boldsymbol{\theta}^{\mathrm{C,B},(t)}_{n,k} = \boldsymbol{\theta}^{\mathrm{C,B,agg},(t)}_{k}$. The iterative process of the local learning at UE, global aggregation at BSs, and retraining at UEs continues for a predefined number of rounds or until convergence criteria are met. Via the new communication-aware local losses and consideration of the communication-related aspects in the global aggregation at BSs, the CFL enables an effective coordination of training among spatially distributed UEs while preserving privacy by keeping training data on the UE and sharing only model parameters with communication-driven loss values during the aggregation.

C. Summary of Proposed CFL and Implementation Aspects

We summarize the proposed CFL in Algorithm 1. The process begins with each UE initializing DNN models parameters for transmission power and bandwidth allocation, i.e., $\boldsymbol{\theta}^{\mathrm{P}}_{n,k}$ and $\boldsymbol{\theta}^{\mathrm{B}}_{n,k}$, respectively (line 2). The communication-aware loss functions $\mathcal{L}^{\mathrm{C,P}}_{n,k}$ and $\mathcal{L}^{\mathrm{C,B}}_{n,k}$ are computed using (3) and (4) (line 3). Each UE then performs local learning to obtain the updated parameters of the DNN models $\boldsymbol{\theta}^{\mathrm{C,P}}_{n,k}$ and $\boldsymbol{\theta}^{\mathrm{C,B}}_{n,k}$ (line 4). For each aggregation round, the UEs send model parameters to the BS (line 7), which computes the aggregated $\boldsymbol{\theta}^{\mathrm{C,P,agg}}_{k}$ and $\boldsymbol{\theta}^{\mathrm{C,B,agg}}_{k}$ using (5) and (6) (line 8). The aggregated model is transmitted back to the UEs, and each UE replaces its local models with the aggregated one (lines 9–11).

The proposed CFL framework aligns with emerging 6G architectures and objectives by enabling AI integration and distributed intelligence via lightweight machine learning models running locally at the UEs while the BS performs communication-aware aggregation. The CFL design requires no architectural changes and scales efficiently across dense networks thanks to marginal signaling overhead. Relevant 6G use-cases include dense urban and cell-free massive deployments, mission-critical industrial networks requiring low-latency decisions, vehicular scenarios with high mobility, or large-scale IoT systems demanding energy-efficiency.

Complexity of DNNs in our CFL is relatively low corresponding to approximately $5,770$ operations per training round and $1,785$ trainable parameters per DNN (3,570 parameters for two DNNs in our case). At the BS, the aggregation of models from N UEs requires $N \times 3,570$ operations per aggregation round, demonstrating low computational cost and practical deployability in 6G-oriented environments.

IV. Performance Evaluation

In this section, we outline the simulation scenario, machine learning models, and competitive algorithms, followed by discussion of simulation results.

We consider four BSs and 15 to 120 UEs randomly distributed within an area of 1000×1000 m. The system operates at carrier frequency of 2 GHz with bandwidth of 20 MHz corresponding to thermal noise of –110 dBm. The scenario is mix of Line-of-Sight (LoS) and Non-LoS (NLoS) conditions. For both, the path loss is modeled according to the 3GPP outdoor-to-outdoor environment [15]. For NLoS, an additional attenuation of 10 dB per wall is applied if the signal is obstructed by building(s). Both P-DNN and B-DNN consist of five hidden layers with 25, 18, 15, 12 and 5 neurons. The batch size is set to 32, with a learning rate of 0.01. All settings are determined through a trial-and-error process. The code of the implemented proposal with all datasets used for evaluation are publicly available[1].

We compare performance of our *Proposed CFL* with: *i)* state-of-the-art *adaptive federated learning*, where aggregation is dynamically adjusted based on each UE's MSE [12], and *ii) basic federated learning*, which is conventional FL, where UE models aggregation is performed equally without any weighting [2]. We evaluate the performance of the proposed and baseline approaches using two metrics: *i) Sum Capacity:*, i.e., the sum of communication capacities of all UEs, and *ii) Ratio of Satisfied UEs*, i.e., the ratio of UEs that meet the minimum capacity requirement.

In Fig. 2, we show sum capacity for varying number of UEs and different values of $c_{\min}$. The sum capacity generally increases with the number of UEs, as each additional UE

[1] https://gitlab.fel.cvut.cz/mobile-and-wireless/codes/publications/coordinated-federated-learning

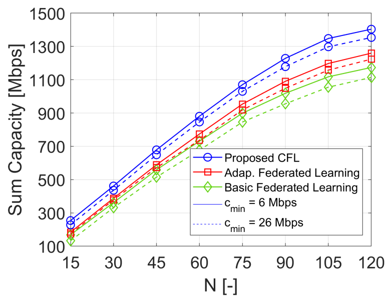

Fig. 2. Impact of the number of UEs on the sum capacity of all investigated algorithms.

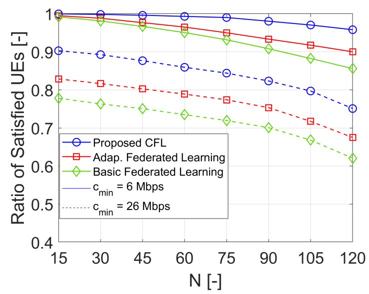

Fig. 3. Impact of number of UEs on the ratio of UEs satisfied with the capacity ($N = 60$).

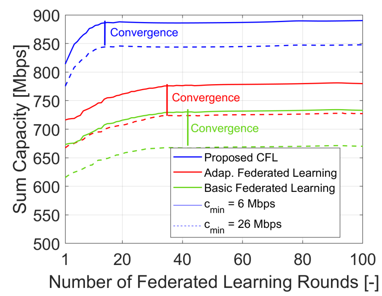

Fig. 4. Convergence of all algorithms ($N = 60$).

directly contributes to sum capacity. However, due to limited amount of radio resources, sum capacity for a higher number of UEs starts saturating. The gain of proposed CFL is from 11.4% to 18.9% compared to adaptive federated aggregation and from 19.1% to 31.4% compared to basic federated learning if $c_{\min} = 6$ Mbps. The performance gain is attributed to the aggregation by CFL, which prioritizes reliable updates with awareness of actual local learning losses leading to more efficient RRM. The gain in sum capacity is slightly smaller (i.e., 10.6%–15.3% compared to adaptive and 17.5%–27.6% compared to basic) if $c_{\min} = 26$ Mbps. The smaller gain for a higher $c_{\min}$ is due to stricter capacity requirements, which reduce resource allocation flexibility.

Fig. 3 depicts impact of number of UEs on ratio of satisfied UEs with the communication capacity. As number of UEs in the system increases, satisfaction ratio decreases for all algorithms due to limited amount of available resources. The proposed CFL approach surpasses adaptive federated learning and basic federated learning by 3.1% to 7.4%, and 5.4% to 11.9%, respectively, for $c_{\min} = 6$ Mbps due to same reasons as indicated in Fig. 2. The gain of the proposal notably grows if $c_{\min}$ is increased to 26 Mbps, i.e., proposed CFL increases satisfaction of UEs by 8.9% to 11.2% and 13.9% to 19.1% over adaptive and basic federated learning, respectively.

In Fig. 4, we analyze convergence of CFL and baselines. As number of FL rounds increases, accuracy of machine learning models improves leading to higher sum capacity. The proposed coordinated framework converges quickly after only 15 rounds, whereas the adaptive federated learning and basic federated learning converge after 39 and 52 rounds, respectively. The performance gain of the CFL is due to joint optimization of RMM parameters, coordination between local DNNs, and weighted aggregation that prioritizes high-quality updates, enabling fast convergence and high sum capacity. A similar trend is observed disregarding $c_{\min}$.

V. Conclusion

In this letter, we have proposed a novel CFL framework for joint optimization of multiple RRM parameters in the wireless networks. The CFL employs multiple local machine learning models at the UEs, each model handling one RRM parameter. The local models are coordinated through a communication-aware loss functions. The local models are aggregated at the BS prioritizing well-performing local models based on the loss functions, thereby accelerating convergence and enhancing overall network performance compared to the best performing baseline, with the sum capacity and user satisfaction improved by 10.6%–18.9% and 3.1%–11.2%, respectively.

References

[1] X. Zhou *et al.*, "Adaptive Segmentation Enhanced Asynchronous Federated Learning for Sustainable Intelligent Transportation Systems," *IEEE Trans. Int. Trans. Sys.*, vol. 25, no. 7, 2024.

[2] T. Sivalingam, B. Gour, S. Yadav, V. Bhatia and N. Rajatheva, "Federated Learning Based Base Station Selection Using LiDAR," *IEEE Trans. Veh. Tech.*, vol. 74, no. 7, pp. 11621-11625, 2025.

[3] J. Lee *et al.*, "Federated Learning-Empowered Mobile Network Management for 5G and Beyond Networks: From Access to Core," *IEEE Commun. Surveys Tut.*, vol. 26, no. 3, pp. 2176-2212, 2024.

[4] C. Ren et al., "Advances and Open Challenges in Federated Foundation Models," *IEEE Commun. Surveys Tut.*, 2025.

[5] O. Aouedi *et al.*, "Federated Learning for Human Activity Recognition: Overview, Advances, and Challenges," *IEEE Open J. Commun. Society*, vol. 5, pp. 7341-7367, 2024.

[6] M. M. Fouda, Z. M. Fadlullah, M. I. Ibrahem and N. Kato, "Privacy-Preserving Data-Driven Learning Models for Emerging Communication Networks: A Comprehensive Survey," *IEEE Commun. Surveys Tut.*, vol. 27, no. 4, pp. 2505-2542, Aug. 2025.

[7] S. Wang, *et al.*, "Adaptive Federated Learning in Resource Constrained Edge Computing Systems," *IEEE J. Selected Areas Commun.*, vol. 37, no. 6, pp. 1205-1221, 2019.

[8] L. U. Khan, W. Saad, Z. Han, E. Hossain and C. S. Hong, "Federated Learning for Internet of Things: Recent Advances, Taxonomy, and Open Challenges," *IEEE Commun. Surveys Tut.*, vol. 23, no. 3, 2021.

[9] 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, 2021.

[10] M. M. Rashid, *et al.*, "Trustworthy and Fair Federated Learning via Reputation-Based Consensus and Adaptive Incentives," *IEEE Trans. Inf. Forensics and Security*, vol. 20, pp. 2868-2882, 2025.

[11] R. Albelaihi, A. Alasandagutti, L. Yu, J. Yao and X. Sun, "Deep-Reinforcement-Learning-Assisted Client Selection in Nonorthogonal-Multiple-Access-Based Federated Learning," *IEEE Internet of Things J.*, vol. 10, no. 17, pp. 15515-15525, 2023.

[12] W. B. Kou, *et al.*, "Fast-Convergent and Communication-Alleviated Heterogeneous Hierarchical Federated Learning in Autonomous Driving," *IEEE Trans. Int. Trans. Sys.*, vol. 26, pp. 10496-10511, 2025.

[13] M. Najla *et al.*, "Machine Learning for Power Control in D2D Communication based on Cellular Channel Gains," in *IEEE Globecom Workshop*, 2019.

[14] R. Lihua *et al.*, "Deep neural network supervised bandwidth allocation decisions for low-latency heterogeneous e-health networks." *J. Lightwave Tech.*, pp. 4147-4154, 2019.

[15] 3GPP TS 36.814, "Further advancements for E-UTRA physical layer aspects," *3GPP Tech. Spec. #36.814*, vol. 9.2.0, 2017.