Scoring Aided Federated Learning on Long-tailed Data for Wireless IoMT based Healthcare System

Lianhong Zhang, Yuxin Wu, Lunyuan Chen, Lisheng Fan, and Arumugam Nallanathan, *Fellow, IEEE*

***Abstract*— In this paper, we propose a novel federated learning (FL) framework for wireless Internet of Medical Things (IoMT) based healthcare systems, where multiple mobile clients and one edge server (ES) collaboratively train a shared model on long-tail data through wireless channels. However, the presence of long-tailed data in this system may introduce a biased global model which fails to handle the tail classes. Additionally, the occurrence of severe fading in wireless channels may prevent mobile clients from successfully uploading local models to the ES, thereby excluding them from participating in the model aggregation. These situations adversely affect the performance of FL. To overcome these challenges, we propose a novel scoring aided FL framework that uses a scoring-based sampling strategy to select mobile clients with more tailed data and better transmission conditions to upload their local models. Specifically, we leverage the logits to explore the data distribution among local clients and propose a logits based scoring client selection method to alleviate the impact of long-tailed data. Moreover, we address the impact of severe fading by incorporating the channel state information (CSI) and data rate of clients into the logits based scoring and proposing a novel logits and model upload rate based client selection method. Experimental results demonstrate the effectiveness of our proposed framework. In particular, compared to the conventional FedAvg, the proposed framework can achieve accuracy gains ranging from 4.44% to 28.36% on the CIFAR-10-LT dataset with an imbalance factor (IF) of 50.**

***Index Terms*— FL, long-tailed data, IoMT, healthcare system, wireless transmission, severe fading.**

I. Introduction

In recent years, artificial intelligence (AI) has been widely applied in many Internet of Things (IoT) networks, where some intelligent algorithms should be employed to train a deep model to help make decisions for the system's operation. A typical application of IoT networks is the Internet of Medical Things (IoMT) based healthcare systems, where deep learning models are trained for automatic diagnosis [1], especially during COVID-19. In the IoMT networks, the medical clients have to transmit a large amount of data to the server for training a global model. As most of the medical content is private, it should be strictly restricted to access and use these data directly. In this case, machine learning may not reach its full potential and fail to learn a high-performance model. To tackle this issue, federated learning (FL), a new distributed learning paradigm, has been proposed, where the clients and server can collaboratively train a shared model through exchanging local models without sharing the raw data, which effectively avoids privacy disclosure [2]–[4]. However, although FL has advantages in privacy protection, it still has some challenges, detailed as follows.

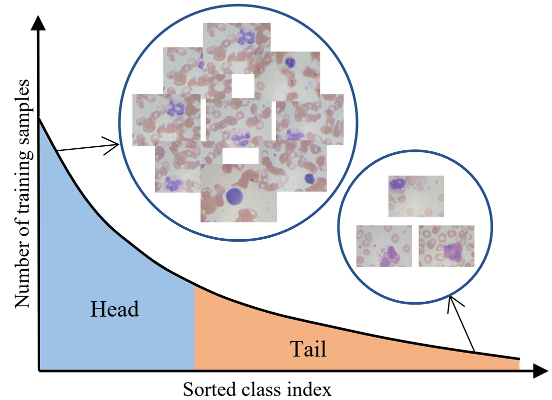


Fig. 1. A typical example of the long-tail distributed dataset.

One critical challenge in the FL training is the inhomogeneous data distribution of clients, which affects the FL training performance significantly. In practical scenarios, the occurrence of long-tailed data distribution is commonly observed. For instance, the home video industry's revenues across products and the sales of "infinite-inventory" retailers like Amazon follow long-tail distributions [5], [6]. As shown in Fig. 1, in the long-tailed dataset, the number of data samples on the head class is larger than that on the tail class. Although the tail class contains fewer number of data, it has a significant impact on the system training performance and test accuracy. Failure to exploit such extremely imbalanced long-tailed data characteristic will cause the training model perform poorly on the tail class [7]–[12]. The impact of long-tailed data on the deep learning models was initially studied in [13], where an

L. Zhang, Y. Wu, L. Chen and L. Fan are all with the School of Computer Science, Guangzhou University, Guangzhou 510006, China (e-mail: 2112106071@e.gzhu.edu.cn, 2112106230@e.gzhu.edu.cn, 2112019037@e.gzhu.edu.cn, lsfan@gzhu.edu.cn).

A. Nallanathan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London, U.K (e-mail: a.nallanathan@qmul.ac.uk).

This work was supported in part by the NSFC (Nos. 62271158/62101145), and supported in part by the Natural Science Foundation of Guangdong Province (No. 2021A1515011392).

The corresponding author of this paper is Lisheng Fan.

unbiased extension was proposed to the softmax function and meta learning was combined to estimate the optimal sampling rate of each class on the long-tailed dataset to improve the class re-balancing. In addition, the authors in [14] used the mean classification score to evaluate the training performance on each class, and obtained a performance improvement on the long-tailed data through adjusting the decision boundary and oversampling. Due to privacy protection, the user datas in FL are often invisible, which makes the aforementioned methods inapplicable. So far, the impact of the long-tailed data on FL has been seldom studied. In this direction, the work in [12] gave a comprehensive overview on the federated long-tailed learning, and discussed the experimental setting for the long-tailed data on FL. In addition, the authors in [15] used the knowledge distillation method and introduced extra neural networks to balance the knowledge in models and logits to improve the performance of FL on the long-tailed data, at the cost of increased resource consumption.

Besides the above challenge from the long-tailed data, wireless transmission is another challenge on the FL training in the IoMT based healthcare systems. Specifically, the clients and server in IoMT communicate with each other through wireless transmission which may experience severe fading, significantly affecting the transmission of the local models. When the clients fail to upload the local models within a latency threshold due to severe fading, they cannot participate into the FL aggregation, causing a poor FL training performance. To solve this problem, researchers have investigated the impact of limited communication resources of clients on the FL training, and proposed several resource allocation strategies to accelerate the model upload [16], [17]. In addition, relaying can be used to enhance the transmission quality of local models, which is helpful for the FL training performance [18]. Although the aforementioned works have investigated the impact of severe fading on FL performance, there has been little study on the impact of long-tailed data on FL in practical severe fadings, which motivates the work in this paper.

In this paper, by taking into account the joint impact of long-tailed data and severe fading on the FL training, we propose a novel FL framework based on scoring, where the ES selects the clients with more tailed data and better transmission conditions to upload their local models. Specifically, the proposed method employs the output logits and transmission rate to evaluate the scores for clients in the client sampling process. Experiments are performed to show the advantage of our proposed methods in improving the FL training performance with long-tailed data. The contributions of this paper are summarized as follows:

- We propose a novel FL framework for IoMT based healthcare systems, where the joint impact of long-tailed data and severe fading is taken into account.
- We devise a novel scoring method based on the output logits, which can help select the clients with more tailed data samples to participate in the aggregation to alleviate the impact of long-tailed data in the FL training, and meanwhile protect the data privacy.
- In the FL, by taking the upload latency into account in practical severe fading, we further jointly exploit the output logits and transmission rate to evaluate the scores for clients.
- We conduct experiments on the long-tailed dataset under a given latency threshold. The experiment results show the advantage of our proposed methods in addressing the impact of long-tailed data and severe fading on FL.

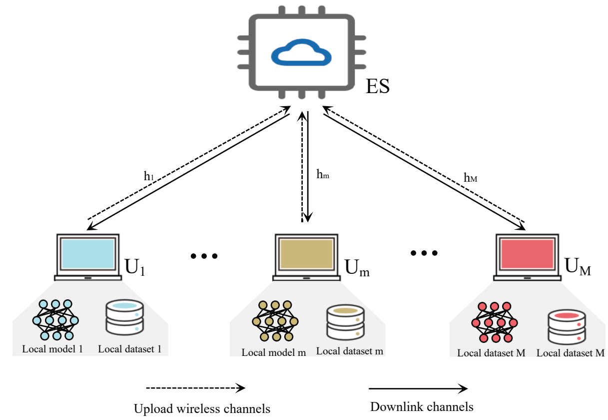


Fig. 2. System model of the federated learning network with M clients.

II. System model

Fig. 2 shows a federated learning network deployed in a wireless network, where an edge server (ES) and M mobile clients denoted by $\mathcal{U} = \{U_m | 1 \leq m \leq M\}$ cooperatively train a shared model on a long-tailed dataset $\mathcal{D}$, and the clients upload local models to ES through wireless fading channels h_m. Specifically, the global dataset $\mathcal{D}$ with C classes denoted by $\mathcal{C} = \{c | 1 \leq c \leq C\}$ follows a long-tail distribution, and we use n_c to denote the number of data samples on class c and sort the class index c in descending order of the number of data samples, i.e., we have $n_c > n_{c'}$, if $c < c'$. The long-tailed dataset $\mathcal{D}$ is non-independent and identically distributed (non-IID) among M clients, where $\mathcal{D} = \bigcup_{m=1}^{M} \mathcal{D}_m$ and $\mathcal{D}_m$ denotes the local dataset of client U_m.

In the considered federated learning network, at communication round r, ES firstly broadcasts the global model $\mathbf{w}_g^{r-1}$ to all clients through downlink channels. Then, from the global model $\mathbf{w}_g^{r-1}$, client U_m performs local training on $\mathcal{D}_m$ and updates its local model as

$$\mathbf{w}_m^r = \mathbf{w}_g^{r-1} - \eta \nabla F_{loss}(\mathbf{w}_g^{r-1}; D_m), \tag{1}$$

where $\mathbf{w}_m^r$ denotes the updated model parameters of client U_m, η and F_{loss} are the learning rate and loss function of local training, and ∇ represents the gradient operation.

After the local update, clients need to upload their models to ES. However, due to limited transmission resources in practice, only K clients can be selected from $\mathcal{U}$ to upload their models. The transmission rate of the selected client U_m is [19]–[21]

$$R_m = W \log_2 \left(1 + \frac{P|h_m|^2}{\sigma^2}\right), \tag{2}$$

where W is the wireless bandwidth, P is the transmit power, $|h_m|^2$ is the channel gain of the wireless channel from client

U_m to ES, and σ^2 is the variance of additive white Gaussian noise (AWGN). Note that to obtain the value of R_m, the client U_m sends some pilot signals to the server before uploading models. Then, the server estimates the channel parameters h_m based on the received signal and pilot signals and calculates the value of R_m by the Shannon formula. The details of this procedure of estimating channels and data rate R_m can be found in the literature, such as the works in [22]–[25].

From (2), the transmission latency of client U_m uploading its local model is

$$T_m = \frac{\zeta}{R_m}, \tag{3}$$

where ζ is the size of the local model parameter. Note that the randomness of h_m directly impacts the transmission rate R_m, leading to the variation in T_m, and hence affects whether the clients can upload models to ES in time and participate in the aggregation or not.

For synchronized federated learning, a latency threshold should be set for model upload. Specifically, the latency threshold γ_t is set for the transmission latency of clients. Therefore, ES aggregates the received local models and updates the global model as

$$\mathbf{w}_g^r = \frac{1}{\sum_{i=1}^{K} \mathbf{I}(T_m \leq \gamma_t)} \sum_{i=1}^{K} \mathbf{w}_i^r \cdot \mathbf{I}(T_m \leq \gamma_t), \tag{4}$$

where $\mathbf{I}(\cdot)$ is a indicating function which returns 1 if the condition holds or 0 otherwise. Note that the indicating function returns 1 when the transmission latency T_m of the client U_m does not exceed the latency threshold γ_t. Therefore, the aggregation function in (4) shows that, under the impact of severe fading, only the clients who upload local models to ES on time can successfully participate into the aggregation.

Note that there are two critical challenges in the system design of the considered federated learning networks. One critical challenge is the data imbalance caused by long-tail and non-IID data distribution among mobile clients, which will lead to a biased global model and low prediction accuracy on the tail classes. The other challenge is the impact of severe fading, which may cause the failure of model upload and accordingly deteriorate the FL training performance. More importantly, these two challenges have a joint impact on the FL training, where the failure upload caused by the severe fading may increase the bias caused by the long-tailed data. Thus, for the purpose of improving prediction accuracy for all classes, we will propose a scoring aided FL framework and corresponding scoring methods in the next section to address these two challenges.

III. Proposed method

In this section, we propose a scoring aided FL framework and corresponding scoring method to address the joint impact of long-tailed and severe fading. Specifically, to alleviate the deteriorated training performance caused by long-tailed data, we introduce a logits based scoring method, which can use the model output logits to score mobile clients, yielding a higher score and sampling importance to the clients with more tailed data. On this basis, we further jointly consider the impact of long-tailed data and severe fading on the FL training performance and design a scoring method by taking into account the channel state information (CSI) and the logits.

A. Scoring aided FL framework

Fig. 3 shows the proposed scoring aided framework, where all M clients receive the broadcast model from ES and conduct the local model training on their local datasets at each communication round. Subsequently, ES selects K clients among M clients to upload the updated models. To tackle the joint impact of long-tailed data and severe fading, we employ a scoring based sampling approach to select clients for global aggregation in order to enhance the performance of FL. This approach involves scoring all clients based on their data distribution and transmission rate, and then sampling clients according to their scores. In the following, we will detail the procedure of the proposed scoring method.

B. Logits based scoring method

To deal with the impact of long-tailed data, re-sampling can be used to balance and calibrate the imbalanced distribution by over-sampling the tail classes and under-sampling the head classes in centralized machine learning [26]–[28]. However, due to the privacy concern in FL, ES can not collect the raw data or data distribution of mobile clients, making it difficult to perform re-sampling on the training data. Thus, we turn to use a logits based scoring method to sample mobile clients instead of re-sampling the training data, where the logits are the model's raw output that can retain information on the local data distribution. In the following, we will detail the logits based scoring method.

To perform the logits based scoring at ES, we firstly need to transmit the logits $L_m = \{l_m^1, l_m^2, ..., l_m^c, ..., l_m^C\}$ to ES after the local training, where L_m is the logits of client m's updated local model on dataset $\mathcal{D}_m$ and l_m^c is the logit of dataset $\mathcal{D}_m$ on class c. After collecting the logits from all M clients, the global logit of class c on dataset $\mathcal{D}$ can be given by

$$l_g^c = \sum_{m=1}^{M} l_m^c. \tag{5}$$

In order to capture the data imbalance of each class, we further normalize each class's logit as

$$l^c = \frac{l_g^c}{\sum_{c=1}^{C} l_g^c}. \tag{6}$$

In the long-tailed learning, the data imbalance may cause a performance gap between the head classes and tail classes. For the purpose of balancing and calibrating the imbalanced distribution, we evaluate the tail classes with higher scores and assign lower scores to the head classes. Thus, the score of class c can be given by

$$S_c = (l^c)^{-a}, \tag{7}$$

where $a \geq 0$ is a hyperparameter to characterize the imbalance in the score of classes. Specifically, a large a indicates a severe

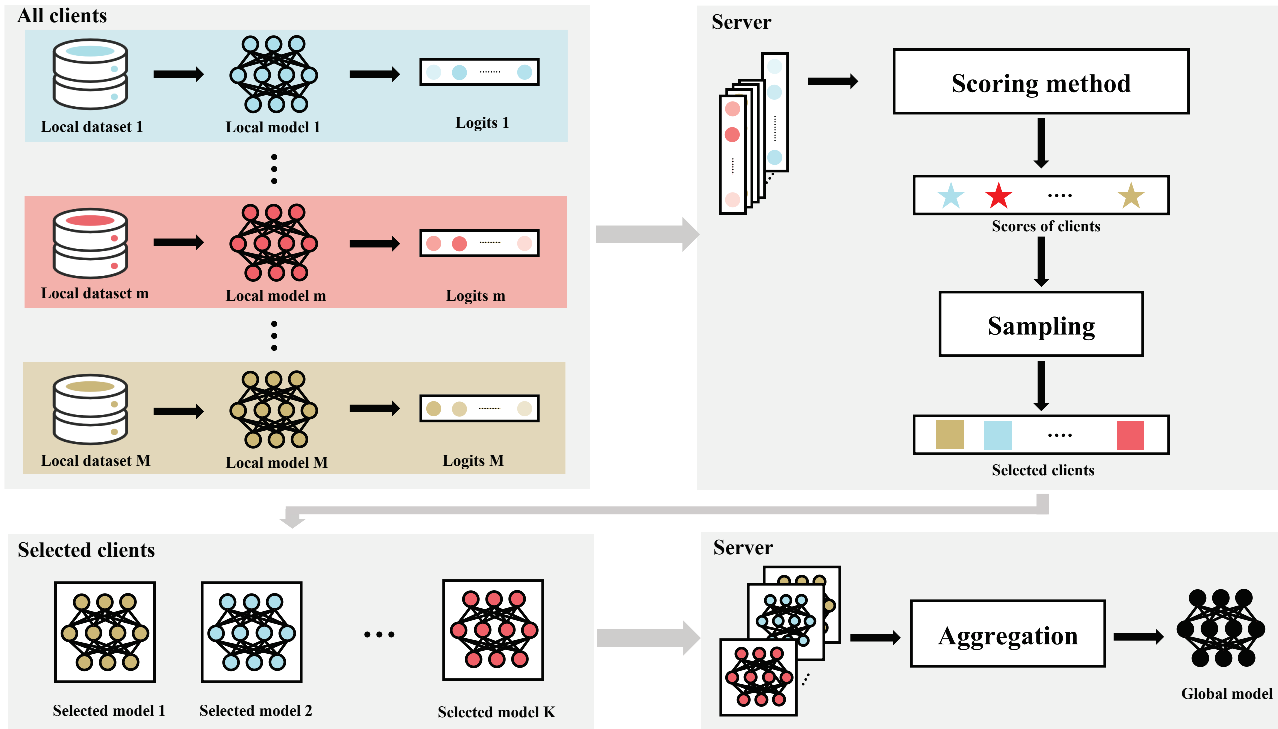


Fig. 3. Scoring aided FL framework with long-tailed dataset.

imbalance in the score of classes, while a smaller a leads to a flatter score of classes. In particular, $a = 0$ yields an equal score among all classes.

From (7), we can finally define the total score of client U_m by summing up the product of each class's logit and the associated score, given by

$$S_m^{\mathbf{I}} = \sum_{c=1}^{C} (S_c \cdot l_m^c). \tag{8}$$

From (5)-(8), we can find that, from the logits of one client, it is not possible to explore the global data distribution, and then it is challenging to distinguish which client has more tailed data. Therefore, based on the logits of all clients, we first explore the global data distribution and distinguish the tail classes through (5), (6) and (7). And then, the clients are scored based on the scores of classes and their local data through (8), which effectively scores the clients with more tailed data higher.

C. Logits and model upload rate based scoring method

Although the above logits based scoring method can balance and calibrate the imbalanced training dataset by scoring clients based on the logits, it ignores the impact of the severe fading on the FL training performance. In practice, due to the severe fading, the clients may not be able to upload models to the server within the latency threshold γ_t and fail to participate in the aggregation. The reduction in the number of clients participating in the aggregation deteriorates the training performance of FL [16]. In particular, in the FL with long-tailed data, the clients with high scores but poor transmission conditions may fail to take part in the model aggregation, resulting in a mismatch of the logits based scoring method. Therefore, we should further design a scoring method that takes into account the joint impact of the failure of model upload caused by the severe fading and long-tailed data to enhance the FL training performance.

From (2)-(4), we can find that the model upload rate directly affects the model upload latency which affects whether the clients can participate in the aggregation or not. Thus, a feasible way to overcome the joint impact of the severe fading and long-tailed data is to further design the scoring method by multiplying the exponential function of the transmission rate, which can be written as

$$S_m^{\mathbf{II}} = \exp\{R_m\} \cdot S_m^{\mathbf{I}}, \tag{9}$$

where the first part $\exp\{R_m\}$ is an exponential function of R_m, and the second part $S_m^{\mathbf{I}}$ is a score obtained by the logits based method in (8). Note that $S_m^{\mathbf{I}}$ inherits the solution of the impact of long-tailed data from the logits based method, and the explosive function $\exp\{R_m\}$ explosively increases the score of client U_m with the transmission rate. Therefore, the selected client has better transmission conditions to increase the likelihood of the successful upload of local models. By taking the logarithmic operation, we can further write the score as

$$\ln S_m^{\mathbf{II}} = R_m + \ln\left(\sum_{c=1}^{C} (S_c \cdot l_m^c)\right), \tag{10}$$

where R_m is the transmission-related term showing the transmission conditions of client U_m, and $\ln\left(\sum_{c=1}^{C} (S_c \cdot l_m^c)\right)$ is the data-related term which is decided by the data of client

U_m. From (10), we can discuss the design of scoring method II more essentially. From the viewpoint of information theory, the term $\ln\left(\sum_{c=1}^{C}(S_c \cdot l_m^c)\right)$ in (10) can be presented as a kind of data rate related to the score of client U_m. If client U_m has a higher score on the classes of local dataset, the overall data rate of client U_m will increase, which yields a better opportunity to participate in the FL training. On the contrary, if client U_m has a lower score on the classes of local dataset, the overall data rate of client U_m will decrease, yielding a worse opportunity to participate in the FL training. With the latency constraint, the update of the global model is affected by the number of uploaded local models arriving at the server within the latency threshold γ_t. Therefore, the communication state of clients is the first factor for scoring clients. After ensuring the update of the global model, the scores of the long-tailed data in (8) can be used to improve the update quality, in order to enhance the FL training performance.

To summarize, we provide the procedure of the proposed scoring aided FL in Algorithm 1. Specifically, after the local training, each client firstly transmits its local logits to ES, and then ES computes the corresponding scores according to (8). Then, ES assigns sampling probability to all mobile clients by normalizing their scores and samples K clients to upload their models according to the sampling probability.

Algorithm 1: Proposed scoring aided FL

Input: $\mathcal{U}$, K
Output: global model $\mathbf{w}_g$
ES chooses a scoring method $\mathbf{F}_S$.
for *Round r = 0, ..., R-1* **do**
  Each client obtains the global model $\mathbf{w}_g^r$ from ES.
  Each client updates the local model as $\mathbf{w}_m^r = \mathbf{w}_g^r - \eta \nabla F_{loss}(\mathbf{w}_g^r; D_m)$.
  Each client uploads the local logits L_m to ES.
  ES computes client scores S_m based on $\mathbf{F}_S$ and assigns the sampling probability of clients as $P_m = \frac{S_m}{\sum_{u=1}^{|\mathcal{U}|} S_u}$.
  ES samples K clients from $\mathcal{U}$ according to P_m.
  The sampled K clients upload their local models $\mathbf{w}_m^r$.
  ES updates the global model as $\mathbf{w}_g^{r+1} = \frac{1}{\sum_{i=1}^{K} \mathbf{I}(T_i \le \gamma_t)} \sum_{i=1}^{K} \mathbf{w}_i^r \cdot \mathbf{I}(T_i \le \gamma_t)$.
end
ES obtains the output global model as $\mathbf{w}_g = \mathbf{w}_g^R$.

IV. Experiments

In this part, we perform some experiments to verify the proposed studies on the FL framework. Specifically, following the basic FL settings in [15], the DL model is ResNet-8, and the dataset $\mathcal{D}$ is CIFAR-10-LT and CIFAR-100-LT. By default, the federated learning runs 200 communication rounds with 20 clients, among which 40% clients are selected in each round. The clients perform the local training by using the mini-batch SGD as the optimizer, where the batch size is set to 128, the local training epoch is 10, and the learning rate is set to 0.1. For the non-IID data setting, the dirichlet distribution is used to generate the non-IID data with the concentration parameter $\alpha = 0.1$. Moreover, for the experimental setting of the communication environment, the transmit power of each client is 3W, the noise variance γ^2 is set to 0.01, the bandwidth of each client is 5MHz, and the local model upload latency threshold is 0.20s. In particular, in the network, the upload wireless channels used by the clients to upload local models and the downlink channels used by the server to broadcast the global model all experience Rayleigh flat fading. Without loss of generality, the average channel gains of the upload wireless channels and downlink channels are set to unity. For the sake of brevity, we use proposed method I to denote logits based scoring method, and proposed method II to denote logits and model upload rate based scoring method.

A. Results of the proposed method I for long-tailed data

Table I shows the test accuracy of the proposed method I on CIFAR-10-LT and CIFAR-100-LT, where IF is set to 50 and 100. For comparison, we list the test accuracy of several typical methods under the same experiment setting [15]. As observed from this table, we can find that the proposed method I achieves the highest test accuracy on both long-tailed datasets in different IF cases. Specifically, for the FL methods on CIFAR-10-LT, the test accuracy of the proposed method I is 28.36% and 21.23% higher than that of the baseline FedAvg when IF = 50 and IF = 100, respectively. Such accuracy improvement is achieved due to the fact that the proposed method I samples clients based on the score of clients' logits exploring the data distribution of clients, which is helpful in improving the performance of the global model on the long-tailed dataset. Moreover, the proposed method I performs better than the imbalance-oriented FL methods. This is because the proposed method I employs the logits to score clients, making the local models trained on the data of tail classes successfully uploaded to contribute to the global model. In further, the slightly inferior performance of the distillation-based FL methods in comparison to our proposed method could be attributed to the fact that the knowledge used for distillation is intrinsically imbalanced, which exacerbates the test error of the global model.

Table II displays the test accuracy of different class cases of the proposed method I on CIFAR-10-LT with IF= 50 and IF = 100, where the CIFAR-10-LT is divided into three class cases as Many case (class 1 $\sim$ class 3), Medium case (class 4 $\sim$ class 6), and Few case (class 7 $\sim$ class 10). Note that it is a general way to use an auxiliary dataset to help train the global model in FL. Hence, we compare the proposed method I with or without an auxiliary dataset to the corresponding FL methods. As observed from Table II, we can find that, in Few case, the proposed method I has a higher test accuracy than that of other methods. Specifically, without the auxiliary dataset, the accuracy of the proposed method I is about 7.73% higher than that of the baseline FedAvg on CIFAR-10-LT with IF= 50. Such performance superiority is due to that the proposed method I assigns higher scores and sampling probabilities to the clients who have more tail class data, so that the clients

TABLE I
TEST ACCURACY (%) FOR PROPOSED METHOD I ON CIFAR-10-LT AND CIFAR-100-LT WITH DIFFERENT IFS.

Family	Method	CIFAR-10-LT		CIFAR-100-LT	
		IF=100	IF=50	IF=100	IF=50
FL methods	FedAvg [4]	52.12	52.43	25.81	28.19
	FedAvgM [29]	53.64	54.42	25.11	28.82
	FedProx [30]	52.75	55.07	25.43	27.77
	FedNova [31]	52.93	56.53	26.81	28.91
Imbalance-oriented FL Methods	Fed-Focal Loss [32]	49.66	52.02	24.66	26.04
	Ratio Loss [33]	54.15	57.77	26.72	28.83
	FedAvg+cRt [34]	51.74	55.87	30.73	31.47
	FedAvg+τ-norm [34]	44.38	45.59	19.59	22.07
	FedAvg+LWS [34]	44.48	46.20	20.70	23.24
Distillation-based FL Methods	FedDF [35]	50.33	52.58	25.60	28.79
	FedBE [36]	44.05	50.66	22.46	23.77
	FEDIC [15]	63.11	63.82	33.67	34.74
Proposed method	Proposed method I	**63.19**	**67.30**	**35.66**	**38.62**

TABLE II
TEST ACCURACY (%) OF DIFFERENT CLASS CASES FOR DIFFERENT METHODS ON CIFAR-10-LT WITH DIFFERENT IFS.

	Auxiliary dataset	Method	Many	Medium	Few
IF=50	✘	FedAvg	84.37	65.98	55.67
		FedProx	85.90	**66.65**	49.33
		Proposed method I	**88.33**	61.43	**63.40**
	✔	FEDIC	65.40	60.01	71.03
		Proposed method I	**75.53**	**66.33**	**78.33**
IF=100	✘	FedAvg	**88.93**	59.65	47.63
		FedProx	87.27	**64.43**	42.50
		Proposed method I	88.30	61.65	**53.17**
	✔	FEDIC	65.63	58.00	68.23
		Proposed method I	**73.93**	**63.50**	**74.97**

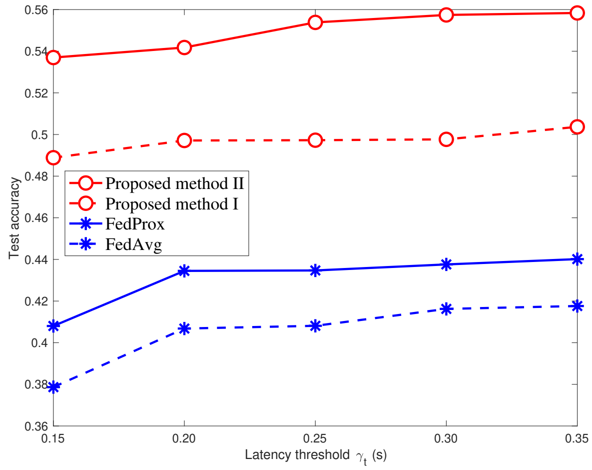
can be selected to participate in aggregation. In contrast, the proposed method I is slightly inferior to other methods in Many case and Medium case. Such performance deterioration can be tolerable, as it is worth improving the test accuracy on the tail classes significantly at a slight expense of head classes in the long-tailed data.

B. Results of proposed methods for long-tailed data under the given latency threshold

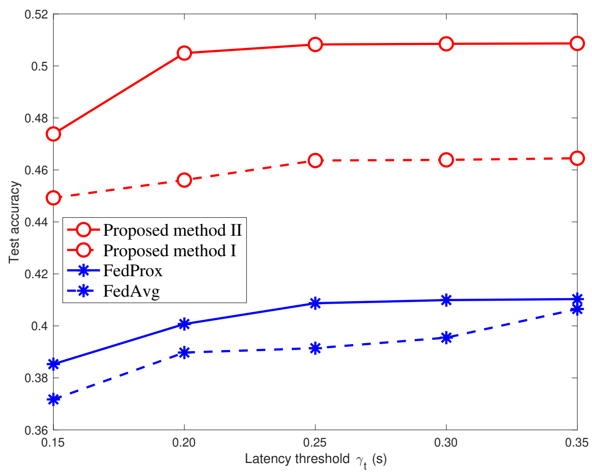
In this part, we plot the test accuracy of the proposed method II, the proposed method I, FedProx [30] and FedAvg [4] on long-tailed data to show the performance of the proposed method II under the given latency threshold. To perform a comprehensive comparison, we also provide the results of transmission-oriented methods, namely CSI-based Sampling and CSI-based greedy client selection (CSI-based GCS), where only CSI is used for client selection and the impact of data distribution is ignored. Specifically, the CSI-based Sampling method employs the proposed scoring-based sampling method with (9) to select clients, but the score of data distribution $S_m^{\mathbf{I}}$ is set to a constant value, which allows us to isolate the influence of data distribution. Moreover, the CSI-based GCS method performs client selection in a greedy approach, where the clients with better CSI will be prioritized to select until the desired number of clients is reached [16].

Fig. 4 shows the impact of latency threshold on the test accuracy of the proposed methods, where the training dataset is CIFAR-10-LT and the latency threshold γ_t varies from 0.15s to 0.35s. Specifically, Fig. 4(a) and Fig. 4(b) depict the test accuracy with IF = 50 and IF = 100, respectively. According to Fig. 4, for either IF = 50 or IF = 100, the test accuracy of all methods improves as γ_t increases, since more clients can participate in aggregation under a larger latency threshold. Moreover, all methods perform worse on IF = 100 compared to that on IF=50, among which the proposed method II still performs the best, indicating the robustness of the proposed method II with various IFs. In further, the proposed method II performs the best when the latency threshold γ_t varies, as it jointly considers the logits and the CSI to select clients to upload local models. Furthermore, under the same latency threshold, the proposed method I achieves higher test accuracy than Fedprox and FedAvg, which proves the superiority of the proposed method I for the FL network on the long-tailed data.

Fig. 5 demonstrates the test accuracy of the proposed method II on CIFAR-10-LT versus the wireless bandwidth W, where Fig. 5(a) and Fig. 5(b) correspond to the test accuracy with IF = 50 and IF = 100, respectively. From Fig. 5, we can observe that the performances of all methods improve when the wireless bandwidth W increases from 2MHz to 6MHz, as a larger bandwidth leads to a smaller model upload latency, which allows more selected clients to upload local models successfully under the given latency threshold. Moreover, the proposed methods outperform other methods both on IF = 50 and IF = 100, which proves that our proposed methods are robust on different IFs. In further, except for the case of $W = 2$ MHz, where the clients are hard to upload local model successfully causing the inability to aggregate the global model, the proposed method II outperforms the baseline methods including FedProx and FedAvg as well as the CSI-based methods such as CSI-based Sampling and CSI-based GCS, for various values of W. This is because the proposed method II assigns more sampling probability for the clients with better transmission conditions and more tailed data in client sampling.

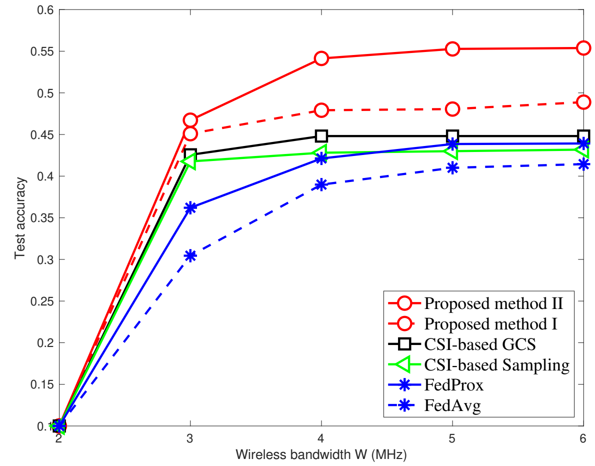


(a) IF = 50

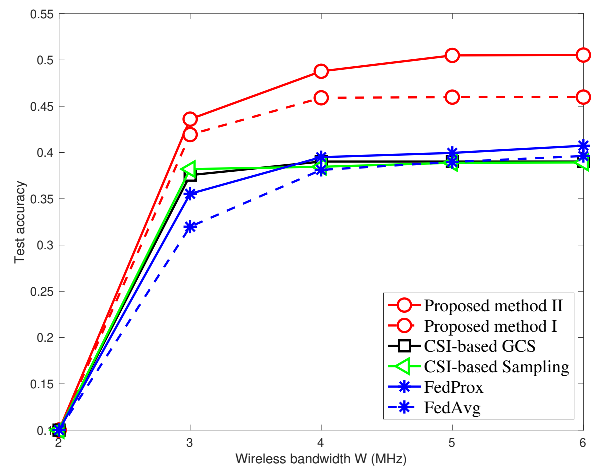


(b) IF = 100

Fig. 4. Test accuracy of the proposed methods on CIFAR-10-LT versus the latency threshold γ_t.



(a) IF = 50



(b) IF = 100

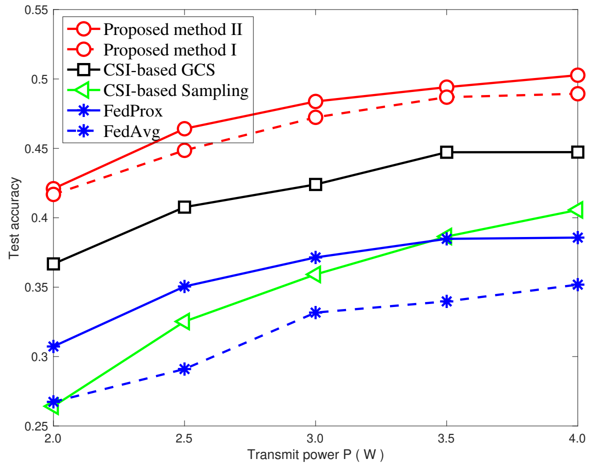
Fig. 5. Test accuracy of the proposed methods on CIFAR-10-LT versus the wireless bandwidth.

Fig. 6 describes the impact of transmit power on the test accuracy of the proposed methods, where the wireless bandwidth of each client is 3MHz, the training dataset is CIFAR-10-LT and the transmit power P varies from 2W to 4W. Specifically, Fig. 6(a) and Fig. 6(b) show the test accuracy with IF = 50 and IF = 100, respectively. From Fig. 6, we can see that all methods achieve a higher test accuracy as the transmit power P becomes larger, since a larger transmit power results in a smaller model upload latency and the clients are easier to upload their local modes successfully. Then, the number of clients participating in the aggregation increases, which improves the test accuracy of the global model [16]. Moreover, though the performance of all methods with IF = 100 is worse than that with IF =50, the proposed methods still outperform other methods, as the output logits used in the proposed methods are obtained according to the dataset. In further, compared with the baseline methods and the CSI-based methods, the proposed method II achieves a better performance, as it combines the logits and the CSI to overcome the joint impact of long-tailed data and severe fading.

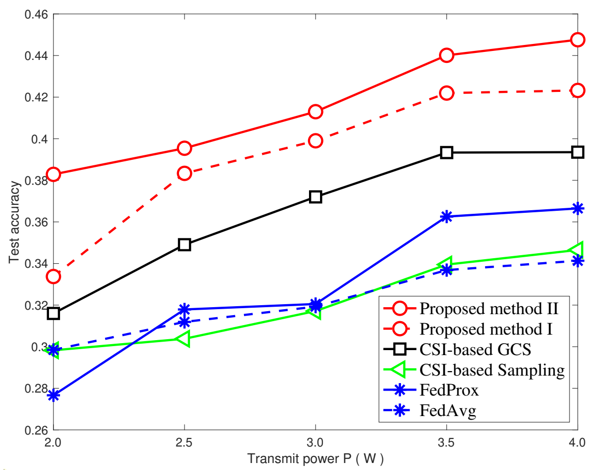
Overall, the performance comparison in Figs. 4-6 illustrate the effectiveness of the proposed methods for the considered FL system. Specifically, with the development of 6G, it is a major trend to deploy FL in wireless networks, where severe fading may deteriorate the training of FL. Fig. 4 shows this impact by plotting the changes in test accuracy of the global model under different latency thresholds, and it is very meaningful to guide the deployment of FL on wireless networks. Moreover, Figs. 5-6 present the performance of several methods for FL with various transmission conditions, where our methods have the best performance under severe fading. This guides that, it is a good way to enhance the performance of the FL deployed in wireless networks on long-tailed data through jointly exploiting logits and CSI.

V. Conclusion

In this paper, we investigated the FL deployed in wireless networks with long-tailed data, where the long-tailed data led to the global model bias and the severe fading affected the upload of local models. To enhance the performance of the FL, we proposed the scoring aided FL framework, where two client scoring methods were designed. Specifically, the proposed method I used the logits to explore the local data



(a) IF = 50



(b) IF = 100

Fig. 6. Test accuracy of the proposed methods on CIFAR-10-LT versus the transmit power.

distribution, making the clients with more tailed data have higher scores to be selected to participate in the aggregation, and thus enhancing the performance of the FL with long-tailed data. Moreover, the proposed method II integrated the logits and the CSI to overcome the joint impact of the long-tailed data and model upload failure caused by the severe fading. Experiments were finally given to show the effectiveness of the proposed methods. In particular, the test accuracy of the proposed method I is about 28.36% higher than that of the baseline FedAvg on the dataset CIFAR-10-LT with IF = 50. In the present work, the clients with more tailed data and poor transmission conditions may fail to participate in the aggregation, resulting in a negative impact on the performance of the FL on long-tailed data. Therefore, in future work, we will focus on exploring communication resource allocation schemes that enable clients with more tailed data and poor transmission conditions to successfully upload their local models to participate in the aggregation.

[1] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. H. Maier-Hein, S. Ourselin, M. J. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *Digit. Medicine*, vol. 3, 2020.

[2] L. Chen and X. Lei, "Relay-assisted federated edge learning:Performance analysis and system optimization," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3387–3401, 2023.

[3] S. Zheng, C. Shen, and X. Chen, "Design and analysis of uplink and downlink communications for federated learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2150–2167, 2021.

[4] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*, vol. 54, 2017, pp. 1273–1282.

[5] A. Elberse and F. Oberholzer-Gee, *Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales*, 2006, vol. 7.

[6] S. Goel, A. Z. Broder, E. Gabrilovich, and B. Pang, "Anatomy of the long tail: ordinary people with extraordinary tastes," in *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010*, 2010, pp. 201–210.

[7] Y. Zhang, B. Kang, B. Hooi, S. Yan, and J. Feng, "Deep long-tailed learning: A survey," *CoRR*, vol. abs/2110.04596, 2021.

[8] T. Wang, Y. Li, B. Kang, J. Li, J. H. Liew, S. Tang, S. C. H. Hoi, and J. Feng, "The devil is in classification: A simple framework for long-tail instance segmentation," in *16th European Conference on Computer Vision, ECCV 2020*, vol. 12359, 2020, pp. 728–744.

[9] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, "Decoupling representation and classifier for long-tailed recognition," in *8th International Conference on Learning Representations, ICLR 2020*, 2020.

[10] S. Li, K. Gong, C. H. Liu, Y. Wang, F. Qiao, and X. Cheng, "Metasaug: Meta semantic augmentation for long-tailed visual recognition," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021*, 2021, pp. 5212–5221.

[11] B. Zhou, Q. Cui, X. Wei, and Z. Chen, "BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, 2020, pp. 9716–9725.

[12] Z. Chen, S. Liu, H. Wang, H. H. Yang, T. Q. S. Quek, and Z. Liu, "Towards federated long-tailed learning," *CoRR*, vol. abs/2206.14988, 2022.

[13] J. Ren, C. Yu, S. Sheng, X. Ma, H. Zhao, S. Yi, and H. Li, "Balanced meta-softmax for long-tailed visual recognition," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.

[14] C. Feng, Y. Zhong, and W. Huang, "Exploring classification equilibrium in long-tailed object detection," in *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021*, 2021, pp. 3397–3406.

[15] X. Shang, Y. Lu, Y. Cheung, and H. Wang, "FEDIC: Federated learning on non-iid and long-tailed data via calibrated distillation," in *IEEE International Conference on Multimedia and Expo, ICME 2022, Taipei, Taiwan, July 18-22, 2022*, 2022, pp. 1–6.

[16] Z. Zhao, J. Xia, L. Fan, X. Lei, G. K. Karagiannidis, and A. Nallanathan, "System optimization of federated learning networks with a constrained latency," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 1095–1100, 2022.

[17] J. Xu and H. Wang, "Client selection and bandwidth allocation in wireless federated learning networks: A long-term perspective," *IEEE Trans. Wirel. Commun.*, vol. 20, no. 2, pp. 1188–1200, 2021.

[18] Z. Qu, S. Guo, H. Wang, B. Ye, Y. Wang, A. Y. Zomaya, and B. Tang, "Partial synchronization to accelerate federated learning over relay-assisted edge networks," *IEEE Trans. Mob. Comput.*, vol. 21, no. 12, pp. 4502–4516, 2022.

[19] W. Zhou and X. Lei, "Priority-aware resource scheduling for uav-mounted mobile edge computing networks," *IEEE Trans. Vehic. Tech.*, vol. 72, no. 7, pp. 9682–9687, 2023.

[20] L. He and X. Tang, "Learning-based MIMO detection with dynamic spatial modulation," *IEEE Trans. Cog. Commun. and Netw.*, vol. PP, no. 99, pp. 1–12, 2023.

[21] W. Zhou and F. Zhou, "Profit maximization for cache-enabled vehicular mobile edge computing networks," *IEEE Trans. Vehic. Tech.*, vol. PP, no. 99, pp. 1–6, 2023.

[22] Y. Deng, Z. Chen, X. Chen, and Y. Fang, "Task offloading in multi-hop relay-aided multi-access edge computing," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 1372–1376, 2023.

[23] W. Fan, Z. Chen, Z. Hao, F. Wu, and Y. Liu, "Joint task offloading and resource allocation for quality-aware edge-assisted machine learning task inference," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6739–6752, 2023.

[24] F. Ke, Y. Lin, Y. Liu, H. Zhou, M. Wen, and Q. Zhang, “Task offloading, caching and matching in ultra-dense relay networks,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 4010–4025, 2023.
[25] H. Li, S. Wu, J. Jiao, X. Lin, N. Zhang, and Q. Zhang, “Energy-efficient task offloading of edge-aided maritime UAV systems,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 1116–1126, 2023.
[26] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, “Decoupling representation and classifier for long-tailed recognition,” in *8th International Conference on Learning Representations, ICLR 2020*, 2020.
[27] A. Estabrooks, T. Jo, and N. Japkowicz, “A multiple resampling method for learning from imbalanced data sets,” *Comput. Intell.*, vol. 20, no. 1, pp. 18–36, 2004.
[28] X. Liu, J. Wu, and Z. Zhou, “Exploratory undersampling for class-imbalance learning,” *IEEE Trans. Syst. Man Cybern. Part B*, vol. 39, no. 2, pp. 539–550, 2009.
[29] T. H. Hsu, H. Qi, and M. Brown, “Measuring the effects of non-identical data distribution for federated visual classification,” *CoRR*, vol. abs/1909.06335, 2019.
[30] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” in *Proceedings of Machine Learning and Systems 2020, MLSys 2020*, 2020.
[31] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, “Tackling the objective inconsistency problem in heterogeneous federated optimization,” in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.
[32] D. Sarkar, A. Narang, and S. Rai, “Fed-focal loss for imbalanced data classification in federated learning,” *CoRR*, vol. abs/2011.06283, 2020.
[33] L. Wang, S. Xu, X. Wang, and Q. Zhu, “Addressing class imbalance in federated learning,” in *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, 2021, pp. 10 165–10 173.
[34] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis, “Decoupling representation and classifier for long-tailed recognition,” in *8th International Conference on Learning Representations, ICLR 2020*, 2020.
[35] T. Lin, L. Kong, S. U. Stich, and M. Jaggi, “Ensemble distillation for robust model fusion in federated learning,” in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.
[36] H. Chen and W. Chao, “Fedbe: Making bayesian model ensemble applicable to federated learning,” in *9th International Conference on Learning Representations, ICLR 2021*, 2021.

Lianhong Zhang received the bachelor degree in computer science and technology from Huaqiao University in June, 2019. She is currently pursuing the master’s degree with the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, China. Her current research interests include federated learning and mobile edge computing.

Yuxin Wu received the bachelor degree in Software Engineering from Northeastern university in 2018. He is currently pursuing the master degree with the school of Electronic and Information Engineering, Guangzhou University. His current research interests focus on mobile edge computing and deep learning.

Lunyuan Chen received the bachelor’s degree in Communication Engineering from Xidian University in 2019. He is currently pursuing the master’s degree with the school of Electronics and Communication Engineering, Guangzhou University. His current interests focus on mobile edge computing and deep learning.

Lisheng Fan received the bachelor and master degrees from Fudan University and Tsinghua University, China, in 2002 and 2005, respectively, both from the Department of Electronic Engineering. He received the Ph.D degree from the Department of Communications and Integrated Systems of Tokyo Institute of Technology, Japan, in 2008. He is now a Professor with GuangZhou University. His research interests span in the areas of wireless cooperative communications, physical-layer secure communications, interference modeling, and system performance evaluation. Lisheng Fan has published many papers in international journals such as IEEE Transactions on Wireless Communications, IEEE Transactions on Communications, IEEE Transactions on Information Theory, as well as papers in conferences such as IEEE ICC, IEEE Globecom, and IEEE WCNC. He is a guest editor of EURASIP Journal on Wireless Communications and Networking, and served as the chair of Wireless Communications and Networking Symposium for Chinacom 2014. He has also served as a member of Technical Program Committees for IEEE conferences such as Globecom, ICC, WCNC, and VTC.

Arumugam Nallanathan (Fellow, IEEE) was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, from August 2000 to December 2007. He was with the Department of Informatics, King’s College London, from December 2007 to August 2017, where he was a Professor of wireless communications from April 2013 to August 2017, and has been a Visiting Professor since September 2017. He has been a Professor of wireless communications and the Head of the Communication Systems Research (CSR) Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, since September 2017. He has published nearly 500 technical papers in scientific journals and international conferences. His research interests include artificial intelligence for wireless systems, beyond 5G wireless networks, the Internet of Things (IoT), and molecular communications. He is an IEEE Distinguished Lecturer. He was a co-recipient of the Best Paper Awards presented at the IEEE International Conference on Communications 2016 (ICC’2016), the IEEE Global Communications Conference 2017 (GLOBECOM’2017), and the IEEE Vehicular Technology Conference 2018 (VTC’2018). He received the IEEE Communications Society SPCE Outstanding Service Award in 2012 and the IEEE Communications Society RCC Outstanding Service Award in 2014. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society and the Technical Program Chair and a member of technical program committees for numerous IEEE conferences. He was an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2006–2011), IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2006–2017), and IEEE SIGNAL PROCESSING LETTERS. He is the Editor-atLarge of IEEE TRANSACTIONS ON COMMUNICATIONS and a Senior Editor of IEEE WIRELESS COMMUNICATIONS LETTERS. He has been selected as a Web of Science Highly Cited Researcher in 2016 and AI 2000 Internet of Things Most Influential Scholar in 2020.