Federated Transfer–Ordered–Personalized Learning for Driver Monitoring Application

Liangqi Yuan, *Student Member, IEEE*, Lu Su, *Member, IEEE*, and Ziran Wang, *Member, IEEE*

***Abstract*—Federated learning (FL) shines through in the Internet of Things (IoT) with its ability to realize collaborative learning and improve learning efficiency by sharing client model parameters trained on local data. Although FL has been successfully applied to various domains, including driver monitoring applications (DMAs) on the Internet of Vehicles (IoV), its usages still face some open issues, such as data and system heterogeneity, large-scale parallelism communication resources, malicious attacks, and data poisoning. This article proposes a federated transfer–ordered–personalized learning (FedTOP) framework to address the above problems and test on two real-world data sets with and without system heterogeneity. The performance of the three extensions, transfer, ordered, and personalized, is compared by an ablation study and achieves 92.32% and 95.96% accuracy on the test clients of two data sets, respectively. Compared to the baseline, there is a 462% improvement in accuracy and a 37.46% reduction in communication resource consumption. The results demonstrate that the proposed FedTOP can be used as a highly accurate, streamlined, privacy-preserving, cybersecurity-oriented, and personalized framework for DMA.**

***Index Terms*—Driver monitoring, federated learning (FL), Internet of Things (IoT), personalization, privacy protection.**

I. Introduction

WITH the rapid development of sensing, computing, and communication technologies, the Internet of Things (IoT) is a popular solution to solve problems in industry, agriculture, energy, transportation, etc. However, privacy issues in IoT are often a significant concern have been raised due to the intrusive behavior of sensors [1]. Specifically for the Internet of Vehicles (IoV), it massively parallels each vehicle and various sensors it carries, including global positioning system (GPS), radar, camera, light detection and ranging (LiDAR), etc., enabling pedestrian detection [2], automated driving [3], mobility digital twins [4], and other transportation applications. Federated learning (FL) has received extensive attention for protecting user privacy by sharing only model weights and not including users' raw data. FL is widely known for its successful business case in Google mobile keyboard prediction [5]. Nowadays, it has also become one of the mainstream and thriving solutions for privacy protection and efficient learning.

Manuscript received 11 March 2023; accepted 20 May 2023. Date of publication 23 May 2023; date of current version 9 October 2023. *(Corresponding author: Ziran Wang.)*

The authors are with the College of Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: liangqiy@purdue.edu; lusu@purdue.edu; ryanwang11@hotmail.com).

Digital Object Identifier 10.1109/JIOT.2023.3279273

A. Federated Learning and Related Work

FL is a potentially feasible solution to the privacy problem in IoT, which is able to avoid the proliferation, distribution, and exchange of local client data by sharing model parameters after training the model on local client data. FL frameworks are widely used in mobile service [6], healthcare [7], [8], industrial [9], [10], IoV [11], [12], etc., due to their usages of large scale and personalized data in an efficient and privacy-preserving way. Although FL has significant contributions to massively parallel devices and computations, it still has a notable drawback in that it cannot efficiently handle non-independent and identically distributed (non-i.i.d.) data. It is required to customize the applicable FL framework according to the features, resources, and constraints possessed by users, data, clients, and servers.

Non-i.i.d. data and heterogeneity have always been a challenge, and a key to research in FL [13], [14], [15]. Non-i.i.d. data is a common phenomenon for real-world clients that are scattered and not interoperable: taking IoV as an example, each driver is heterogeneous as a client. FedAvg [16], as one of the first proposed feasibility methods, has been the subject and center of research. FedAvg averages all local models to get the global model so that the local model may deviate far from the global optimum in the parameter space leading to some limitations in FedAvg. It is necessary to ensure that the local model does not deviate from the global model (prevent overfitting) and, simultaneously, that the local model can effectively learn the local client data set (prevent underfitting). Based on FedAvg, FedProx [17] is proposed to limit the deviation of the local model from the global model by adding a proximal term.

Besides considering accuracy, the FL framework in IoT should not underestimate communication and training resource constraints, cybersecurity, and ubiquity. Some of the recent surveys summarized challenges, threats, and solutions of the FL decentralization paradigm for IoT, including limited computing power, unreliable and limited availability, local training, accuracy, communication overhead, etc. [18], [19], [20], [21], [22], [23].

Transfer and edge learning are popular solutions to reduce communication resource consumption in FL frameworks. Zhang et al. [24] performed a federated transfer learning framework to detect driver drowsiness, where transfer learning was employed to save the communication cost in the FL framework. Su et al. [25] introduced edge servers as a collaborative mechanism, where local models were aggregated in the edge server and then sent to the global server to aggregate the

2327-4662 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

global model. The benefit of the additional edge server was that the communication between massively parallel clients and the edge server was consumed because the edge server was geographically close to the clients. High latency and intermittent connections could be mitigated. In addition, the edge server could also provide personalized aggregated local models due to the similarity of geographically adjacent clients.

Cyber attack is a problem that cannot be ignored for FL frameworks. Sun et al. [26] developed an attack method for FL framework in IoT, in which a bi-level optimization framework was proposed to compute optimal poisoning attacked FL framework, including direct, indirect, and hybrid attacks. Meanwhile, Zhang et al. [27] utilized a generative adversarial network (GAN)-based approach to attack the FL framework, especially since the attacker did not need any prior knowledge to carry out the attack.

Personalization is a common approach for FL frameworks to improve applicability for diverse users [28]. Fallah et al. [29] proposed a personalized variant of the FL, which allowed clients to perform several gradient descent iterations on an initial global model using local data to obtain a personalized local model. Wu et al. [30] explored a cloud edge-based personalized FL framework for in-home health monitoring, which addressed the problem that a single global model performed poorly on a specific client. Since the global model could only capture the common features of all clients, it lacked the ability to analyze fine-grained information of specific clients. Ma et al. [31] proposed a personalized FL framework of layer-wised weighted aggregation, which determines mutual contribution factors through a hypernetwork on the server to identify similarity between users at the layer granularity.

B. Federated Learning in Driver Monitoring Applications

Driver monitoring application (DMA) in IoV is adopted as the research direction in this article due to its real and visual image data, valuable application scenarios, and relatively blank research area. DMA also has challenges in terms of driver privacy issues, communication, and diverse and personalized driver state and behavior. Related DMA literature covers a wide variety of devices with algorithms to achieve different purposes, such as dangerous state detection [32], driver emotion recognition [33], driver lane change inference [34], etc. Compared to other methods [35], [36], [37], FL not only highlights efficient learning but also effectively protects the privacy of driver, passenger, and pedestrian biometric information, driving routes, and confidential driving areas such as military installations.

In this article, we introduce and adapt FL to DMA. Although some FL frameworks exist for DMA, they all suffer from some critical problems. Doshi and Yilmaz [38] proposed an FL edge-device framework to obtain a global model by aggregation feature representations and obtained considerable accuracy in recognizing driver activities. For the i.i.d. setting, the data set was partitioned for each edge node in a random way, while for the non-i.i.d. setting, the data set was assigned selectively. Zhao et al. [39] proposed an FL framework to monitor fatigue driving, where the non-i.i.d. setting was simulated by controlling the number of images per client. The above FL frameworks for DMA did not really take into account the actual situation of the application but artificially created a simulation scenario. Therefore, there is an urgent need for realistic analysis and research for real-world DMA, considering that the user (driver) should exist independently and be noninteroperable with different clients (vehicles). Moreover, in addition to the necessity of test data sets, the test client is also a critical evaluation criterion, which can reflect the universality of the FL framework. We summarize the existing neglects and challenges in the current FL for DMA frameworks as follows.

1) Clients in FL for DMA frameworks are often defined in unreasonable and incomprehensible forms. A real and natural definition of a client should be a driver or a vehicle.
2) There is no paper proposing to test on a testing client (not involved in the training process), which lacks universal testing for the FL framework.
3) For the DMA scenario, there is a great diversity and individuality of driver behaviors, postures, and facial expressions, which call for more personalized studies than other general IoV scenarios.
4) Similarly, DMA also has diverse scenarios, including diverse vehicle models, interior colors, seat positions, etc., which will greatly increase the learning difficulty.

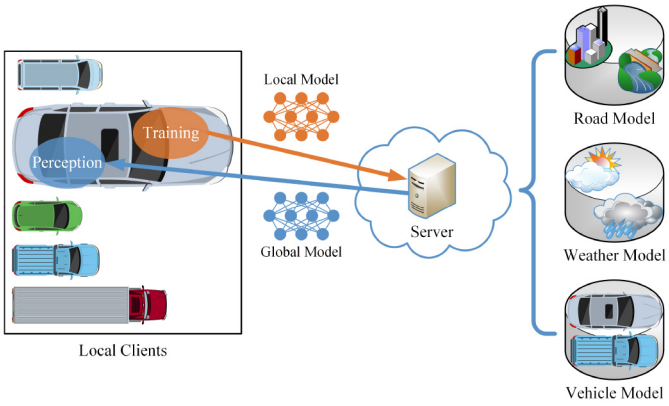

Fig. 1. Structure illustration of an FL framework for IoV. The server interacts with the local client and saves different scenarios as different models. Transparent neurons are nontrainable parameters, and nontransparent neurons are trainable parameters.

C. Proposed Solution and Contribution

In this article, we aim to propose an FL framework applicable and specific to practical applications in IoV, especially DMA, where an imaginary FL framework for IoV is illustrated in Fig. 1. Each local client, i.e., vehicle, includes a training module and a perception module. The training module uploads the model parameters to the server after learning and training the local data. After aggregation and optimizing the parameters of the local client models, the server downloads the global model parameters to the perception module in the local client. Moreover, transfer learning can be used to reduce the number of trainable parameters, resulting in reduced communication

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

consumption. The server can save different global models for different scenarios, such as road types, weather types, and vehicle types, so that the model can have better applicability.

Therefore, a federated transfer–ordered–personalized learning (FedTOP) framework is proposed to address the problems of accuracy, cybersecurity, communication resources, and diversified scenarios. In addition to the transfer-extension shown in Fig. 1, the FedTOP framework also enhances robustness and cybersecurity by orderly dropout clients due to their possible overfitting and poisoning of the data. Furthermore, the FedTOP framework is able to remarkably improve accuracy by adapting all clients through personalized-extension. The contributions of this article are as follows.

1) For realistic problems and usage scenarios in DMA, we propose a feasible FL framework FedTOP, realizing privacy protection, high accuracy, low communication requirements, cybersecurity, and pervasiveness. To the best of our knowledge, this is one of the first papers to establish a feasible FL framework for DMA.
2) The proposed FedTOP framework is tested on two real-world driver monitoring data sets with and without system heterogeneity, systematically characterizing system heterogeneity in real-world data sets and achieving considerable accuracies with 92.32% and 95.96%, respectively.
3) The experiments highlight a realistic and natural client setup, i.e., drivers and vehicles are naturally formed as clients. Moreover, we innovatively propose evaluation criteria for training and testing clients to test the generalization ability of the proposed FedTOP on different clients.
4) Through an ablation study, we demonstrate the performance and utility of the transfer, ordered, and personalized extensions. These detachable extensions can be selectively installed according to the task description, and the FL framework combined with different extensions can effectively adapt to different IoT application scenarios.

The presentation of this article is as follows. The problem statement and proposed solution are described in Section II. The experimental setup, heterogeneity, and results have been demonstrated in Section III. Section IV discusses the performances of three extensions of the proposed framework, followed by Section V summarizing this article and expounding on future work.

II. Methodologies

A. Problem Statement

FL frameworks are able to protect privacy, increase training efficiency, and save communication resources by sharing only model parameters in IoT. In this article, the FL framework is used to solve a driver activity classification task in DMA. Clients in real-world IoT are independent and heterogeneous due to the presence of only a minimal number of users per client. Considering the more general application scenarios, the global model ω for training clients C aggregation needs to be compatible with nontraining clients C' in addition to C.

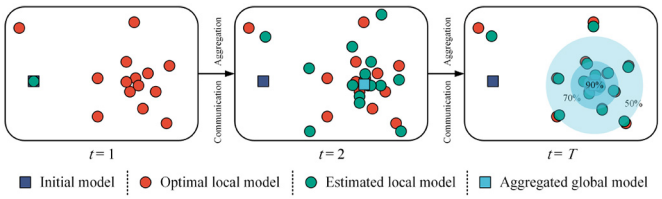

Fig. 2. Illustration of the FL algorithm finds the optimal global model solution in the parameter space. The shaded areas are accuracy contour areas. The farther the optimal local model dissociates from the global model, the lower the client accuracy. Local models enclosed by shaded areas have similar accuracies.

The data of each client D_c is non-i.i.d. when the data is not interoperable. We can consider a nested model

$$L_c = \omega_c(D_c) \tag{1}$$

where ω_c is the classifier model corresponding to client $c \in C$. $D_c \in \mathbb{R}^{n_c \times i \times j \times d}$ is the image set with n_c samples, i rows, j columns, and d channels. $L_c \in \mathbb{Z}^{n_c}$ is the corresponding label set. The global model ω is obtained by aggregating, e.g., averaging the weights of the local models

$$\omega = \sum_{c \in C} p_c \omega_c = \mathbb{E}[\omega_c | c \in C] \tag{2}$$

where $p_c \in [0, 1]$ is a weight density function of clients, for which $\sum p_c = 1$, p_c will be assigned according to the number of samples. Therefore, the optimization problem of the FL algorithm can be formulated as minimizing the global loss, which is equivalent to minimizing the sum of the local losses

$$\min_{\omega} \mathcal{L}(\omega) = \sum_{c \in C} p_c \mathcal{L}(\omega_c) = \mathbb{E}[\mathcal{L}(\omega_c) | c \in C] \tag{3}$$

where $\mathcal{L}$ is the loss function that will be assigned.

For real-world classification tasks, we assume that the distribution of the local model in the parameter space presents a multivariate Normal distribution $\omega_c \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2)$, where μ_ω is the mean of all local models, and σ_ω^2 is the variance of all local models. Fig. 2 shows the process of the FL algorithm finding the optimal solution of the global model in the parameter space. After the initial model is trained locally, communicated, and aggregated globally, the final global model will be obtained by averaging and can be estimated as $\hat{\omega} = \mu_\omega$. Especially in the large-scale parallel application scenarios of IoT, according to the law of large numbers, $\hat{\omega} = \mu_\omega = \omega^*$ is an unbiased estimation.

However, there are still some defects in the method of obtaining the global model through average aggregation. First, we can confirm that there is enormous system heterogeneity in IoT, and the global model cannot ensure high accuracy for all clients. Second, we inevitably need a measure to prevent system heterogeneity and potential attacks and poisoning. As shown in Fig. 2, the farther the optimal local model is from the global model, the lower the accuracy, and vice versa. Therefore, it is conceivable that in the FL problem with heterogeneity, the clients' accuracy will also obey a Normal distribution.

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

B. Proposed Solution

According to the problem statement, we propose a FedTOP algorithm to address all of the following issues. First, the aggregation of global models needs to be more stable, which can be achieved by preventing the overfitting of local models. Second, considering the actual communication situation in IoT, we propose a transfer learning method to reduce the trainable parameters and hence reduce communication requirements. Third, the global model should have the ability to resist interference, attacks, and data poisoning, which can be achieved by orderly dropping out local models with a large loss. Fourth, a global model cannot take into account the situation of all clients, especially in the presence of data and system heterogeneity. Therefore, we recommend personalizing the global model to suit all the training and testing clients.

We refer to FedProx [17] using a proximal term to prevent local models ω_c from deviating from the global model ω. In which the proximal item $\mathcal{L}_p$ that computes the distance between the local and global model is added to the loss function

$$\mathcal{L}_p = \frac{\mu}{2}\|\omega_c - \omega\|^2 \tag{4}$$

where μ is the deviation coefficient, ω_c is local client model parameters, and ω is global model parameters. The overall loss function can be updated as

$$\mathcal{L} = \mathcal{L}_l + \mathcal{L}_p \tag{5}$$

where $\mathcal{L}_l$ is the loss between the true labels and the predicted labels, such as the negative log-likelihood loss used in our experiments.

Transfer-extension is a common and popular solution in many learning frameworks. In particular, FL frameworks are favored because they can effectively reduce local client training resources and communication resources. In our experiments, the base model is ResNet34 [40] pretrained on ImageNet, where only the last residual block and fully connected layer are trainable parameters. Although ImageNet is a large object classification data set far from DMA images, the lower layers are similar to convolutional neural networks (CNNs) and are used to extract image features. Therefore, the upper layers that are used to obtain high-level features and representations are given more attention. The ratio of reduced communication resource requirement in the network is approximately equal to the ratio of nontrainable parameters to total parameters

$$\text{Commun}_{\downarrow} \approx \frac{|\omega_{\text{non-trainable}}|}{|\omega|} = 37.46\% \tag{6}$$

where $\text{Commun}_{\downarrow}$ is the reduced communication resource requirement, $|\omega_{\text{non-trainable}}|$ is the number of nontrainable model parameters, and $|\omega|$ is the total number of model parameters. Therefore, the transfer-extension reduces the communication requirement by 37.46% by decreasing the trainable parameters.

Ordered-extension is for orderly dropout clients with enormous variance, which may be subject to malicious attacks and poisoning, extensive data and system heterogeneity, and model underfitting. These local clients with large losses should be discarded to enhance the generalizability of the global model. Ordered-extension not only enhances accuracy and robustness but also secures the global model. After all of the clients upload the local model parameters and the final training loss to the server, the server only aggregates the $q \in \mathbb{N} \leq |C|$ local models with the lowest loss as the global model. The set of q local models can be expressed as

$$C_q \in q - \arg\min_{c \in C} \mathcal{L}(\omega_c). \tag{7}$$

Therefore, only the models of clients in C_q with lower loss will be used to aggregate the global model.

Personalized-extension is to promote, popularize, and adapt the global model to the heterogeneity of all clients. As shown in Fig. 2, the global model cannot be applied to all clients due to the ubiquitous heterogeneity. The region of interest (ROI) of the model may vary depending on system heterogeneity, such as different camera angles, seat positions, and vehicle structures, resulting in differences in the relative position of the driver in the image. However, personalized-extension proposes to train the global model several times in each client to obtain a more personalized local model to improve accuracy. On the one hand, compared with the traditional FL algorithm, the personalized-extension can significantly and effectively improve accuracy and confidence. On the other hand, compared to the method that only trains locally, the personalized FL algorithm improves the training efficiency and avoids the overfitting of the local model. In particular, the personalized FL algorithm can help and generalize to other nontraining clients C', which may have minimal training resources. After receiving the global model, the nontraining clients C' can obtain a highly accurate and reliable local model with minimal training. The system diagram of the proposed FedTOP is shown in Fig. 3.

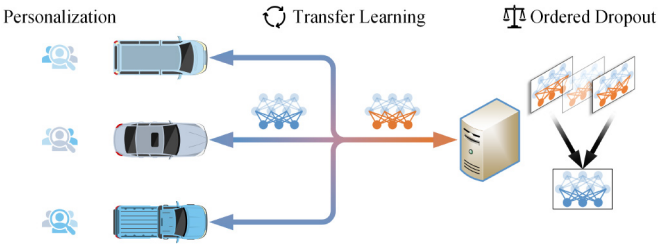

Fig. 3. Schematic of the proposed FedTOP system, where transfer, ordered, and personalized extensions are deployed on the communication, server, and client sides, respectively. They operate independently and do not interfere with each other.

For the proposed FedTOP framework, the client communicates with the server T rounds, and all clients C train E epochs in parallel between each communication. For our preliminary experiments, we set $T = 10$ and $E = 5$. For transfer-extension, the local model is the transfer learning model of ResNet34 pretrained on ImageNet. Only the last residual block and fully connected layer are set as trainable parameters. In addition, we add an additional fully connected layer to match the number of our classification categories. Based on FedProx, the activation function of the last layer is LogSoftmax, and the setting of the loss function $\mathcal{L}_l$ is a negative log-likelihood loss. ω^1 is the

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

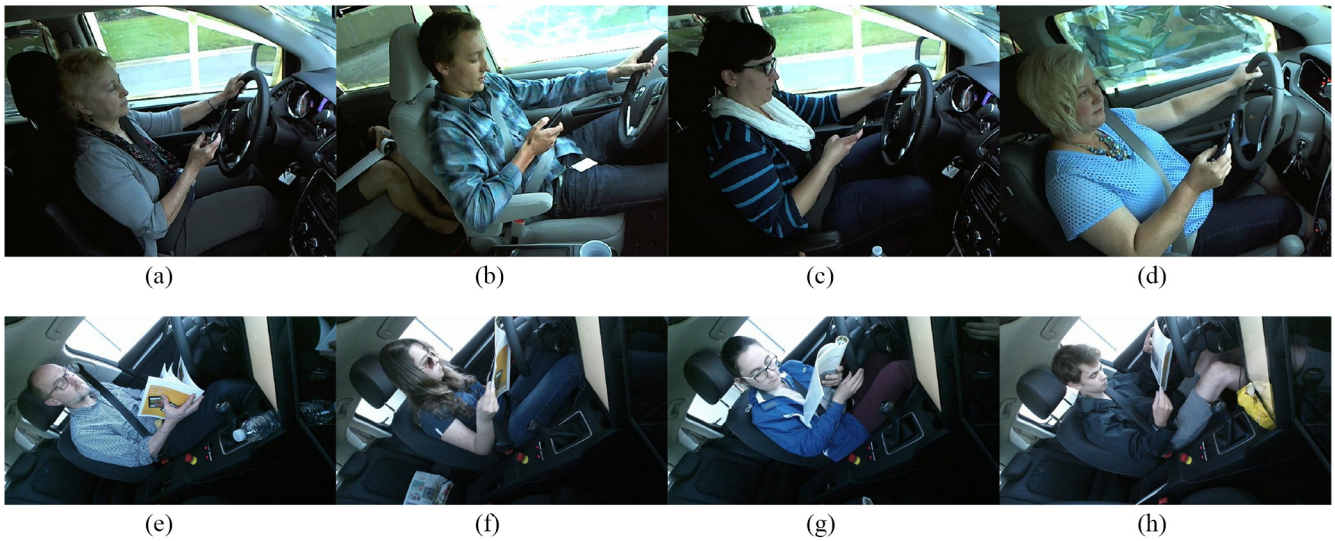

Fig. 4. Exampled activities of four drivers in each of SFDDD and DriveAct data sets. (a) SFDDD texting—right 1. (b) SFDDD texting—right 2. (c) SFDDD texting—right 3. (d) SFDDD texting—right 4. (e) DriveAct magazine 1. (f) DriveAct magazine 2. (g) DriveAct magazine 3. (h) DriveAct magazine 4.

Algorithm 1 FedTOP

Input: Communication rounds (T), training client set (C), training epoch (E), initial global model (ω^1), loss function ($\mathcal{L}_l$), deviation coefficient (μ), number of ordered clients (q)
Output: Trained global model (ω^T)
for $t = 1$ **to** $T - 1$ **do**
  for $c \in C$ **in parallel do**
    for $e = 1$ **to** $E - 1$ **do**
      Backpropagate the loss function and update the local model $\omega_c^{t^{e+1}} \leftarrow \arg\min_{\omega_c^{t^e}} \mathcal{L}_l(\omega_c^{t^e}) + \frac{\mu}{2}\|\omega_c^{t^e} - \omega^t\|^2$.
    end for
    Update the local model $\omega_c^t \leftarrow \omega_c^{t^E}$.
    Client sends ω_c^t to the server.
  end for
  Find a set C_q^t of top-q clients in C^t in term of loss values: $C^t \in q - \arg\min_{c \in C^t} \mathcal{L}(\omega_c^t)$.
  Server aggregates the ω as $\omega^{t+1} \leftarrow \frac{1}{q}\sum_{q \in C_q^t} \omega_q^t$.
end for
Send ω^T to clients $c \in \{C, C'\}$ do personalization.

Algorithm 2 Personalized-Extension

Input: Training client set (C), testing client set (C'), personalization epoch (E), Trained global model (ω^T), loss function ($\mathcal{L}_l$)
Output: Personalized local model (ω_c)
for $c \in \{C, C'\}$ **do**
  for $e = 1$ **to** $E - 1$ **do**
    Backpropagate the loss function and update the local model $\omega_c^{T^{e+1}} \leftarrow \arg\min_{\omega_c^{T^e}} \mathcal{L}_l(\omega_c^{T^e})$.
  end for
  Update the personalized local model $\omega_c \leftarrow \omega_c^{T^E}$.
end for

initial model parameter. The proposed FedTOP is described in Algorithm 1, and the personalization process is described in Algorithm 2.

III. Experiment and Results

Considering the data and system heterogeneity, experiments are conducted on two open real-world driver monitoring data sets, including State Farm Distracted Driver Detection (SFDDD) [41] and DriveAct [42]. In addition to comparing with FedProx as a baseline, this article also compares the performance of the transfer, ordered, and personalized extensions through an ablation study.

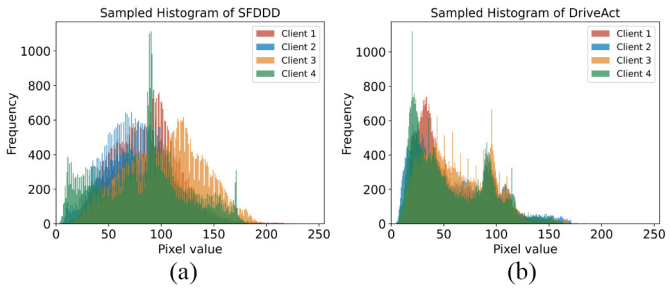

Fig. 5. Sampled client image histograms of (a) SFDDD and (b) DriveAct data sets.

A. Experiment Setup

To compare the impact of system heterogeneity on FL frameworks, the proposed FedTOP is tested on driver monitoring data sets with and without system heterogeneity. The SFDDD data set includes 26 drivers and 10 activities, and the DriveAct data set includes 15 drivers and 12 activities. The SFDDD data set considers system heterogeneity, that is, different drivers have different vehicles, different seat positions, different camera angles, etc., as shown in Fig. 4(a)–(d). The DriveAct data set does not take into account system heterogeneity, i.e., all subjects had their data collected in the same system. Recorded from the same camera angle, different drivers read the same magazine in the same vehicle, as shown in Fig. 4(e)–(h).

To show more clearly and visually the heterogeneity between different clients in the two data sets, Fig. 5 shows histograms of the sample images of the two data sets. It can be seen that the SFDDD data set with system heterogeneity has a

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

TABLE I
PERFORMANCE OF FEDTOP AND ABLATION STUDY ON SFDDD AND DRIVEACT DATA SETS

Dataset	Method [1]	$\|C\|$	q	μ	Transfer	Accuracy (%) [2] Training	Accuracy (%) [2] Testing	Time$_\downarrow$ (%) [3]	Commun$_\downarrow$ (%) [4]	Cybersecurity
SFDDD	FedProx (baseline)	20	20	1	No	54.63	16.44	$\sim$	$\sim$	$\sim$
	FedOP	20	15	1	No	97.69	96.37	1.45 $\downarrow$	$\sim$	$\uparrow$
	FedTP	20	20	1	Yes	94.76	92.8	17.3 $\downarrow$	37.46 $\downarrow$	$\sim$
	FedTO	20	15	1	Yes	46.16	16.43	18.91 $\downarrow$	37.46 $\downarrow$	$\uparrow$
	FedTOP	**20**	**15**	**1**	**Yes**	**94.65**	**92.32**	**18.91** $\downarrow$	**37.46** $\downarrow$	$\uparrow$
DriveAct	FedProx (baseline)	12	12	1	No	73.18	23.96	$\sim$	$\sim$	$\sim$
	FedOP	12	10	1	No	98.07	97.97	0.44 $\downarrow$	$\sim$	$\uparrow$
	FedTP	12	12	1	Yes	97.00	95.71	16.83 $\downarrow$	37.46 $\downarrow$	$\sim$
	FedTO	12	10	1	Yes	62.30	22.89	19.18 $\downarrow$	37.46 $\downarrow$	$\uparrow$
	FedTOP	**12**	**10**	**1**	**Yes**	**97.04**	**95.96**	**19.18** $\downarrow$	**37.46** $\downarrow$	$\uparrow$

[1] FedOP, FedTP, and FedTO refer to ablating the transfer, ordered, and personalized extensions of the FL framework, respectively. FedOP refers to training and transmitting the entire model without using the transfer learning paradigm. FedTP refers to the aggregation of all models without ordered dropout of larger loss models. FedTO refers to the use of the global model without local personalization. [2] Accuracy refers to the testing sets of training clients and testing clients, which is described in Section III-A. [3] Time$_\downarrow$ refers to the ratio of reduced training time per client to the baseline. [4] Commun$_\downarrow$ refers to the ratio of reduced communication consumption to the baseline, which is described in (6).

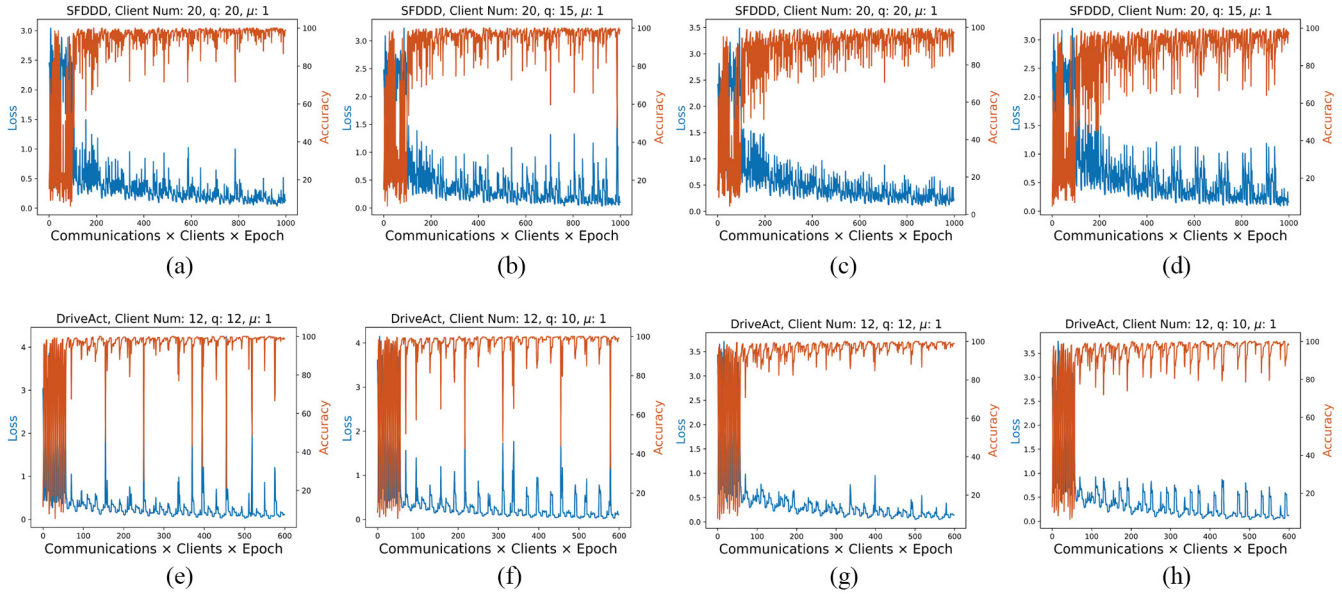

Fig. 6. Accuracy and loss curves of the FL framework and its extensions on the SFDDD and DriveAct data sets, which is the training process of Algorithm 1. Note that personalization does not affect the convergence of the global model in FL frameworks, i.e., FedTO converges the same as FedTOP. FedO and FedT refer to ablating the transfer and ordered extensions of the FedTOP framework, respectively. (a) and (e) FedProx. (b) and (f) FedO. (c) and (g) FedT. (d) and (h) FedTO.

more considerable difference in the distribution of histograms than the DriveAct data set without system heterogeneity, and the mean value of the SFDDD images is larger. The possible reason is that the vehicle interiors of the DriveAct data set view are darker, resulting in most of the pixel values being lower. Therefore, FL frameworks may be more challenged by the scene information when training on the SFDDD data set, such as different vehicle interiors.

Clients are naturally divided based on the drivers. In order to better demonstrate the role of personalized-extension, the data sets are first divided into training clients and testing clients at a ratio of about 0.8, 0.2, with $|C_{\text{SFDDD}}| = 20$, $|C'_{\text{SFDDD}}| = 6$, $|C_{\text{DriveAct}}| = 12$, and $|C'_{\text{DriveAct}}| = 3$. And then, the data sets for each client are divided into a training set, verification set, and testing set at a ratio of 0.7, 0.15, and 0.15, respectively. After the global model is trained by the training data set of training clients, the final trained global model is shared with all clients for personalization. The personalization of the global model will only be processed on the training sets, while the personalized local model will be tested on the unseen testing sets. FL architectures are established on Pytorch and trained on an Intel Core i9-10850K CPU @ 3.60 GHz, and a Nvidia GeForce RTX 3080 GPU.

B. Ablation Study and Results

We explore the role of each FedTOP extension on two real-world data sets through an ablation study. FedProx is used as a baseline for comparison. According to the experimental setup described in the previous section, the experimental results are shown in Table I.

The results and comparisons for two data sets and three extensions are shown in Fig. 6, which is equivalent to demonstrating Algorithm 1. By observing the accuracy and loss curves on the two data sets, it can be concluded that the SFDDD data set with system heterogeneity is fundamentally different from the DriveAct data set without system heterogeneity. It can be clearly seen that the SFDDD data set

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

with system heterogeneity requires more communication to converge, while the DriveAct data set without system heterogeneity has a fast convergence speed, especially at the first communication. Therefore, for real-world data sets, system heterogeneity can be mitigated by more communication times.

By observing Fig. 6(b), (d), (f), and (h), it can be found that the ordered-extension diminishes the stability of the system. Although the anomalous large-loss local model is discarded to reduce the bias of the global model, it also increases the variance of the global model resulting in reduced generalizability. By observing Fig. 6(c), (d), (g), and (h), we can see that the effect of transfer-extension is different for data sets with and without system heterogeneity. On the one hand, transfer-extension increases the variance of the model on the SFDDD data set and leads to a reduced and unstable model convergence. On the other hand, transfer-extension improves the speed of model convergence on DriveAct, and the convergence effect is more stable. The possible reason is that the transfer-extension retains only a small number of trainable parameters, resulting in the neural network model not being able to learn human behavioral features effectively in the SFDDD data set with system heterogeneity. However, for the DriveAct data set without system heterogeneity, the factors are constant except for the driver, and the local model does not need to focus on these exact same pixels, but only on the changing pixels, including objects, such as drivers, computers, and magazines. Therefore, for the DriveAct data set, transfer-extension can effectively increase convergence and stability. The proposed FedTOP framework is able to obtain 92.32% and 95.96% accuracy on the SFDDD and DriveAct data sets, respectively, when considering five times of personalization training. Compared to FedProx as a baseline, FedTOP can effectively improve the accuracy by 462% in addition to considering a 37.46% reduction in communication resources. The results demonstrate the feasibility of the proposed FedTOP in terms of communication resource-saving, accuracy improvement, robustness, and cybersecurity.

C. Performance of Personalized-Extension

Personalized-extension needs to be further discussed and analyzed as the most effective approach to improve accuracy. Based on the division of training and testing clients in Section III-A, in this section, we further discuss how the trained and aggregated global model is adapted to both training and testing clients. The results of the personalized-extension on the two data sets are shown in Fig. 7 with different personalization epochs, which is equivalent to demonstrating Algorithm 2. It can be seen that the personalization process differs significantly on the data sets with and without system heterogeneity, which is similar to the results in Fig. 6. The clients in the DriveAct data set have faster convergence, minor accuracy variance, and higher final accuracy. On the contrary, the clients in the SFDDD data set not only converge slower but also have an anomalous client with relatively low accuracy. The possible reason is that the anomalous client has a huge data and system heterogeneity, causing the optimal model to deviate significantly far from the aggregated global model.

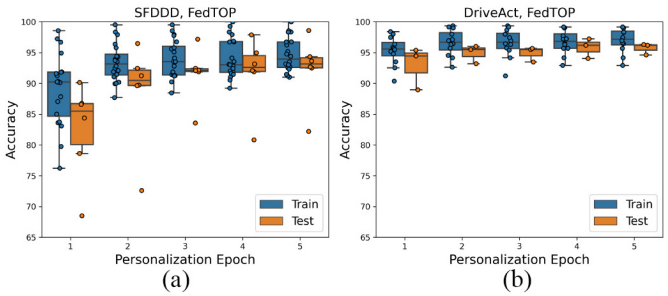

Fig. 7. Error bars and accuracy distributions for training and testing clients on (a) SFDDD and (b) DriveAct data sets. Each scatter represents a client. Testing accuracy of the training and testing clients varies with personalized epoch, which is the result of Algorithm 2. (a) SFDDD, FedTOP. (b) DriveAct, FedTOP.

Fig. 8 further demonstrates that the trained global model repositions the ROI during the personalized training process via class activation map (CAM) [43]. The test client of the SFDDD data set can be seen struggling with the personalization process. The trained global model focuses the ROI on the seat backrest, driver's chest, hand, and knee, and vehicle door. Due to the system heterogeneity present in the SFDDD data set, the positions of the driver, seat, and steering wheel, as shown in Fig. 8(a), are different from other clients, as shown in Fig. 4(b)–(d). Therefore, the initial ROI is likely to be a driver's position among other clients. During the five personalization training processes, the local model is able to effectively reposition the ROI to the driver, which is what the personalized-extension is intended to show. Moreover, the personalization process also reduces the number of ROI while targeting more attention to a specific area.

On the contrary, for the test clients in the DriveAct data set, the adjustment of the ROI is negligible. Note that the ROI does not necessarily have to cover the driver's body or an object such as the magazine. The ROI should cover those pixels that can distinguish between different activities, such as static activities like reading the magazine, and dynamic activities like wearing a seatbelt in the DriveAct data set activity setting. The ROI focuses on areas where large differences are likely to occur. The fact that the ROI in the DriveAct data set covers almost the same pixels during the personalization process can also prove the negative impact of system heterogeneity on the FL framework.

IV. Discussion

The proposed FedTOP is tested on two data sets: 1) SFDDD and 2) DriveAct. Although both data sets involve driver activity classification tasks, they have different objectives. The SFDDD data set aims to detect driver distraction under normal driving conditions, while the DriveAct data set is focused on activity recognition in autonomous driving scenarios. The DMA highlighted in this article is interested in monitoring driver distractions or other risky behaviors, as different activities carry different risk factors. For instance, reaching behind is more hazardous than talking to a passenger. Therefore, in situations where there is a high risk of distracted state or dangerous behavior, the autopilot system can

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

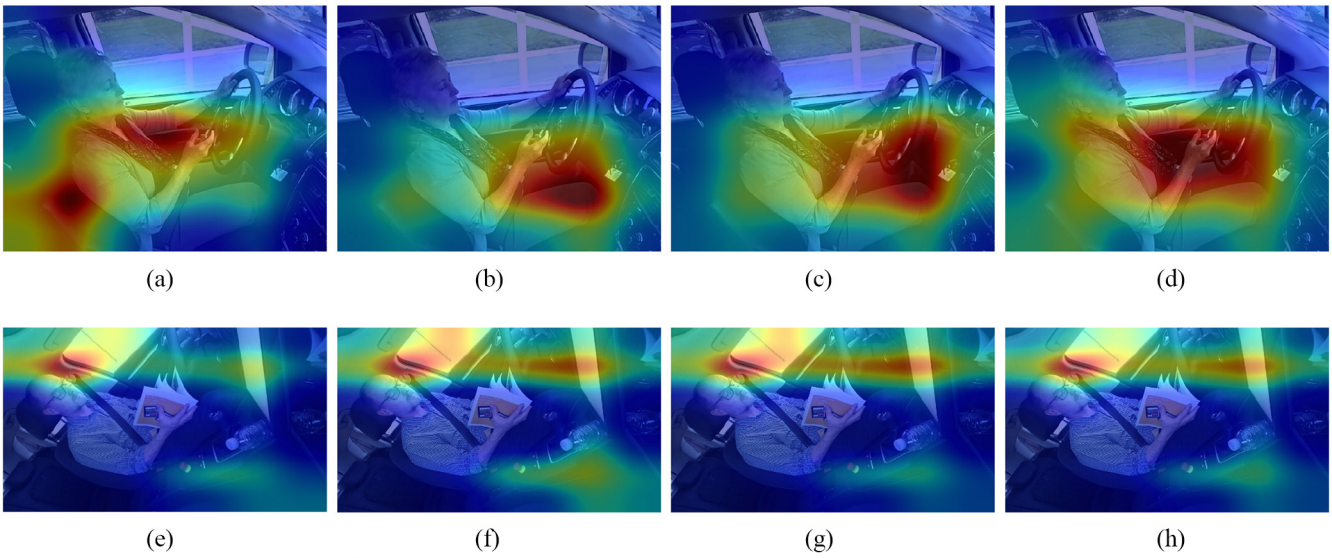

Fig. 8. CAMs of the test clients in SFDDD and DriveAct data sets during the personalization process. (a)–(d) are test clients in the SFDDD data set, which is the same as Fig. 4(a). (e)–(h) are test clients in the DriveAct data set, which is the same as Fig. 4(e). (a) and (e) Trained global model ω^T. (b) and (f) Personalization Epoch 1 ω^{T^1}. (c) and (g) Personalization Epoch 3 ω^{T^3}. (d) and (h) Personalization Epoch 5 ω^{T^5}.

be employed to take control of the vehicle and prevent traffic accidents.

The two data sets used, SFDDD and DriveAct, still have some flaws. First, although the SFDDD data set takes system heterogeneity into account, quite a few drivers collect data in the same vehicle. Therefore, there are still some differences between the data set and the real-world data, which leads to the fact that the proposed FedTOP may need more communication rounds to achieve similar accuracy on a real-world data set. Second, there is currently no driver monitoring data set with real poisoning data currently existing, resulting in the effect of ordered-extension not being reflected. The different modalities, positions, and angles of the camera or the method of generating fake data may be a hypothesis for poisoned data, but it cannot be highlighted as real. Moreover, due to road safety guidelines, the current data set is only driving on safe roads or simulated driving. Therefore, the driver's posture, demeanor, facial concentration, etc., are far from the real driving behavior. Therefore, there is an urgent need for a more realistic data set that can include camera images of different positions and angles, different vehicle scenes, and more drivers driving on real roads.

For an FL framework in IoT, in addition to accuracy being the evaluation criterion, factors, such as communication requirements, robustness, fairness, cybersecurity, etc., also need to be considered. Although it seems that transfer and ordered extensions may not improve accuracy but rather reduce it in the current experimental results, it can potentially improve the performance of the FL framework. Therefore, we keep two extensions as one of our future directions. Personalized-extension is an approach similar to transfer learning and incremental learning. On the one hand, the local client is incrementally learned based on the trained global model, but it does not intentionally retain the previously learned knowledge. On the other hand, the global model is transferred to the client data set as in transfer-extension, but the low-level nontrainable weights are still pretrained on ImageNet. Therefore, the proposed personalized-extension actually uses the trained global model weights to fit different client data, such as the reposition of ROI. Although the personalized-extension requires additional training locally for each client, there are many benefits, including high accuracy, applicability to nontraining clients, customization, etc. Conceivably, personalized-extension can effectively address the problem of system heterogeneity, e.g., it can be applied to different cameras, camera angles, vehicle interiors, etc.

V. Conclusion

In this article, we propose an FL framework FedTOP for DMA to address the issues of privacy preservation, efficient training, communication resource-saving, poisoned data, and diversified scenarios. Through the ablation study, the impact, role, and performance of three extensions, including transfer, ordered, and personalized on the model, are disclosed. Moreover, the experiments demonstrate dramatic differences between data sets with and without system heterogeneity. In addition to the proposed FedTOP being able to exhibit 92.32% and 95.96% accuracy in two data sets for testing clients, it is also appreciated that FedTOP reduces communication consumption by 37.46% and potentially improves cybersecurity. The experimental results show that the proposed FedTOP is a highly accurate, lightweight, privacy-preserving, robust, cybersecure, and universally applicable FL framework for potential DMAs.

Future work lies in the continued research of extensions. For the ordered-extension, a possible plan is to introduce some malicious local clients to attack and poison with the global model. For example, subjects may not place the camera on the side as instructed but place it on the front or behind instead. Such outliers may cause the global model to deviate significantly from the optimal solution, so in this

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

case, ordered-expansion can prevent the deviation of the global model by discarding the larger value of the losses. For the transfer-extension, there is currently a lack of a general driver monitoring model, so we used a model pretrained on ImageNet. Future work can pretrain a driver model ourselves as a base model, which will get better performance in DMA. Fig. 1 shows the FL framework for foresight in IoV, but the data set used does not contain scenario information, such as road, weather, vehicle models, etc. Therefore, we expect a well-developed real-world data set to include such scenario information, data and system heterogeneity, etc.

REFERENCES

[1] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1250–1258, Oct. 2017.

[2] J. Cao, Y. Pang, J. Xie, F. S. Khan, and L. Shao, "From handcrafted to deep features for pedestrian detection: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4913–4934, Sep. 2022.

[3] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 829–846, Apr. 2018.

[4] Z. Wang et al., "Mobility digital twin: Concept, architecture, case study, and future challenges," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17452–17467, Sep. 2022.

[5] A. Hard et al., "Federated learning for mobile keyboard prediction," 2018, *arXiv:1811.03604*.

[6] X. Zhou, W. Liang, J. Ma, Z. Yan, and K. I.-K. Wang, "2D federated learning for personalized human activity recognition in cyber-physical-social systems," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 6, pp. 3934–3944, Nov./Dec. 2022.

[7] I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," *Nat. Med.*, vol. 27, no. 10, pp. 1735–1743, Oct. 2021.

[8] N. Rieke et al., "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–7, Sep. 2020.

[9] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu, "Efficient and privacy-enhanced federated learning for industrial artificial intelligence," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6532–6542, Oct. 2020.

[10] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 4177–4186, Jun. 2020.

[11] Z. Du, C. Wu, T. Yoshinaga, K.-L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular Internet of Things: Recent advances and open issues," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 45–61, 2020.

[12] X. Kong et al., "A federated learning-based license plate recognition scheme for 5G-enabled Internet of Vehicles," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 8523–8530, Dec. 2021.

[13] F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-iid data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3400–3413, Sep. 2020.

[14] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5132–5143.

[15] S. Horvath, S. Laskaridis, M. Almeida, I. Leontiadis, S. Venieris, and N. Lane, "FjORD: Fair and accurate federated learning under heterogeneous targets with ordered dropout," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 12876–12889.

[16] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Stat.*, 2017, pp. 1273–1282.

[17] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. Mach. Learn. Syst.*, vol. 2, 2020, pp. 429–450.

[18] B. Ghimire and D. B. Rawat, "Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8229–8249, Jun. 2022.

[19] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, May 2020.

[20] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.

[21] L. Lyu, H. Yu, and Q. Yang, "Threats to federated learning: A survey," Mar. 2020, *arXiv:2003.02133*.

[22] P. Kairouz et al., "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, nos. 1–2, pp. 1–210, 2021.

[23] Q. Li et al., "A survey on federated learning systems: Vision, hype and reality for data privacy and protection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3347–3366, Apr. 2023.

[24] L. Zhang, H. Saito, L. Yang, and J. Wu, "Privacy-preserving federated transfer learning for driver drowsiness detection," *IEEE Access*, vol. 10, pp. 80565–80574, 2022.

[25] Z. Su et al., "Secure and efficient federated learning for smart grid with edge-cloud collaboration," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1333–1344, Feb. 2022.

[26] G. Sun, Y. Cong, J. Dong, Q. Wang, L. Lyu, and J. Liu, "Data poisoning attacks on federated machine learning," *IEEE Internet Things J.*, vol. 9, no. 13, pp. 11365–11375, Jul. 2022.

[27] J. Zhang, B. Chen, X. Cheng, H. T. T. Binh, and S. Yu, "PoisonGAN: Generative poisoning attacks against federated learning in edge computing systems," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3310–3322, Mar. 2020.

[28] A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 28, 2022, doi: 10.1109/TNNLS.2022.3160699.

[29] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning: A meta-learning approach," Feb. 2020, *arXiv:2002.07948*.

[30] Q. Wu, X. Chen, Z. Zhou, and J. Zhang, "FedHome: Cloud-edge based personalized federated learning for in-home health monitoring," *IEEE Trans. Mobile Comput.*, vol. 21, no. 8, pp. 2818–2832, Aug. 2022.

[31] X. Ma, J. Zhang, S. Guo, and W. Xu, "Layer-wised model aggregation for personalized federated learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10092–10101.

[32] A. Kashevnik, I. Lashkov, and A. Gurtov, "Methodology and mobile application for driver behavior analysis and accident prevention," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2427–2436, Jun. 2020.

[33] S. Zepf, J. Hernandez, A. Schmitt, W. Minker, and R. W. Picard, "Driver emotion recognition for intelligent vehicles: A survey," *ACM Comput. Surveys*, vol. 53, no. 3, pp. 1–30, Jun. 2020.

[34] Y. Xing et al., "Driver lane change intention inference for intelligent vehicles: Framework, survey, and challenges," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4377–4390, May 2019.

[35] A. Masood, D. S. Lakew, and S. Cho, "Security and privacy challenges in connected vehicular cloud computing," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2725–2764, 4th Quart., 2020.

[36] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. Fallah, "A survey of deep learning applications to autonomous vehicle control," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 712–733, Feb. 2021.

[37] M. Ramzan, H. U. Khan, S. M. Awan, A. Ismail, M. Ilyas, and A. Mahmood, "A survey on state-of-the-art drowsiness detection techniques," *IEEE Access*, vol. 7, pp. 61904–61919, 2019.

[38] K. Doshi and Y. Yilmaz, "Federated learning-based driver activity recognition for edge devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 3338–3346.

[39] C. Zhao, Z. Gao, Q. Wang, K. Xiao, Z. Mo, and M. J. Deen, "FedSup: A communication-efficient federated learning fatigue driving behaviors supervision approach," *Future Gener. Comput. Syst.*, vol. 138, pp. 52–60, Jan. 2023.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[41] "State farm distracted driver detection." State Farm. Apr. 2016. [Online]. Available: https://www.kaggle.com/competitions/state-farm-distracted-driver-detection/overview/description

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.

[42] M. Martin et al., "Drive&Act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2801–2810.
[43] D. Omeiza, S. Speakman, C. Cintas, and K. Weldermariam, "Smooth Grad-CAM++: An enhanced inference level visualization technique for deep convolutional neural network models," 2019, *arXiv:1908.01224*.

Liangqi Yuan (Student Member, IEEE) received the B.E. degree from Beijing Information Science and Technology University, Beijing, China, in 2020, and the M.S. degree from Oakland University, Rochester, MI, USA, in 2022. He is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA.

His research interests are in the areas of sensors, Internet of Things, signal processing, and machine learning.

Lu Su (Member, IEEE) received the M.S. degree in statistics and the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2012 and 2013, respectively.

He is an Associate Professor with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. He has also worked with IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, and the National Center for Supercomputing Applications, Urbana, IL, USA. He has published more than 100 papers in referred journals and conferences His research interests are in the general areas of Internet of Things and cyber–physical systems, with a current focus on wireless, mobile, and crowdsensing systems.

Dr. Su is the recipient of the NSF CAREER Award, the University at Buffalo Young Investigator Award, the ICCPS'17 Best Paper Award, and the ICDCS'17 Best Student Paper Award. He serves as an Associate Editor for *ACM Transactions on Sensor Networks*. He is a member of ACM.

Ziran Wang (Member, IEEE) received the Ph.D. degree from the University of California at Riverside, Riverside, CA, USA, in 2019.

He is an Assistant Professor with the College of Engineering, Purdue University, West Lafayette, IN, USA and was a Principal Researcher with Toyota Motor North America, Plano, TX, USA. His research focuses on automated driving, human–autonomy teaming, and digital twin.

Dr. Wang serves as the Founding Chair of the IEEE Technical Committee on Internet of Things in Intelligent Transportation Systems, and an Associate Editor of four academic journals, including IEEE INTERNET OF THINGS JOURNAL and IEEE TRANSACTIONS ON INTELLIGENT VEHICLES.

Authorized licensed use limited to: Purdue University. Downloaded on October 23,2024 at 03:29:19 UTC from IEEE Xplore. Restrictions apply.