WiFederated: Scalable WiFi Sensing using Edge Based Federated Learning

Steven M. Hernandez *Student Member, IEEE,* and Eyuphan Bulut, *Senior Member, IEEE,*

***Abstract*—WiFi sensing using Channel State Information (CSI) offers a device-free and non-intrusive method for human activity monitoring. However, the data-hungry and location-specific training process hinders its scalable deployment at large sizes. In this work, we propose *WiFederated*, a federated learning (FL) approach to train machine learning models for WiFi sensing tasks. Using WiFederated, client devices can not only perform training in parallel at the edge instead of sequentially at a central server but can also collaboratively learn and share generalizable location-independent traits about physical actions being monitored. We demonstrate that an FL model trained on as few as 2-3 locations can provide high prediction accuracy in new locations even without any data available from them. We also demonstrate how new locations can achieve higher prediction accuracy even with a small number of available samples when using the pretrained FL model rather than training from scratch. The results show that the FL model can save local training epochs and reduce the need for large data collection at each new location. Thus, the proposed WiFederated system scales as more locations are added. We show that WiFederated provides a more accurate and time efficient solution compared to existing transfer learning and adversarial learning solutions thanks to the parallel training ability at multiple clients. By introducing new client selection methods during the FL process, we also show that accuracy can further increase. Finally, we evaluate the feasibility of training models at the edge and introduce continuous annotation to allow for continuous learning over time.**

***Index Terms*—WiFi sensing, federated learning, channel state information (CSI), device-free sensing.**

I. Introduction

WIFI sensing [1]–[3] aims to capture and analyze ambient WiFi signals to understand physical characteristics of a given environment. New applications of WiFi sensing have quickly gained popularity as it allows for low-cost (by leveraging existing WiFi infrastructure), device-free, and non-intrusive sensing of human presence and physical activities compared to sensor-based systems which typically require dedicated devices to be placed on the body. By leveraging existing infrastructure, it also adds more value for the owners of WiFi devices in several ways such as by providing physical occupancy analytics in retail shopping areas or other public buildings [4]–[6], reinforcing security through intrusion detection [7] and monitoring patients [8], [9] to protect from falls or other serious injuries.

Leveraging existing WiFi access points (AP) for sensing purposes provides a remarkable advantage compared to sensor-based methods by reducing costs associated with both hardware and labor for deploying the new sensors. However, typical WiFi sensing systems require large amounts of data to train their deep learning models [10]–[13], thus it is very cumbersome to build such systems. As the number of actions to be recognized increases for the model, the required amount of training data also increases. Moreover, the models produced in most existing works are specialized to a single given physical location. Thus, the physical actions must be performed over many repetitions for each individual location to train a model for it. This is time-consuming and impractical especially in cases where the locations are already occupied such as in the case of patient monitoring in hospital settings.

Consider the three physical environments illustrated in Fig. 1a-c for a typical office building. Each location has unique environmental properties (e.g., size, environmental clutter such as furniture), thus the characteristics of the signals received in one location will vary from that of each of the other locations. For example, in a less cluttered hallway environment such as in Fig. 1a, very few physical obstructions block the transmission path between WiFi devices. In some environments, the WiFi devices can only receive signals through the wall when performing sensing tasks such as in the private office rooms shown in Fig. 1b. Additionally, some other environments may have very dynamic physical features such as meeting rooms as shown in Fig. 1c where furniture like chairs and tables are constantly shifted around. As such, the core issue with deploying WiFi sensing systems is to develop generalizable models that can be used in new locations without requiring complete and extensive data recording and annotation steps for each new environment. Fig. 1d shows a typical WiFi sensing system diagram where for each environment, a technician must repeat many trials of each physical action before sending the manually labelled data to a server. Once at the server, the data must be run through data cleaning and preprocessing steps followed by extensive hyperparameter selection steps to determine the best model parameters for the given environment. After fully training, the location-specific model can finally be shared back to the edge for model inference.

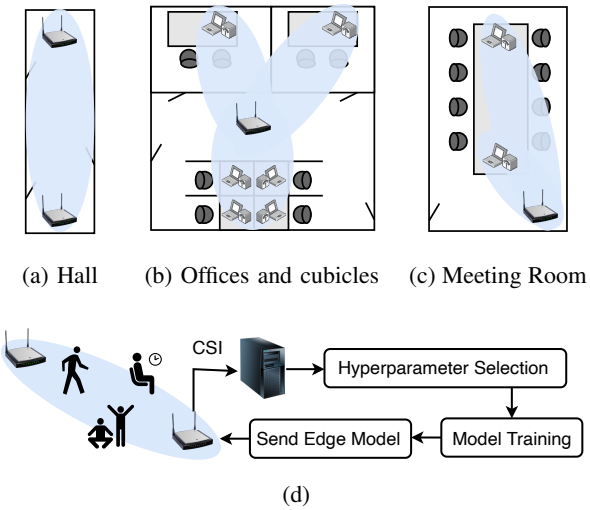
There are some recent studies [14]–[18] looking into the creation of generalizable models that can also be used in new environments, however these studies mostly focus on creating

S. M. Hernandez and E. Bulut are with the the Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA (e-mail: hernandezsm, ebulut@vcu.edu).

This work is supported in part by National Science Foundation (NSF) Graduate Research Fellowship Program (GRFP) under Grant No. 1744624, Virginia Commonwealth University Presidential Research Quest Fund (PeRQ) and Commonwealth Cyber Initiative (CCI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Manuscript received XXX XX, XXXX; revised XXX XX, XXXX.

Copyright (c) 2021 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.



(a) Hall (b) Offices and cubicles (c) Meeting Room

(d)

Fig. 1: (a)-(c) WiFi sensing environments in an office building. (a) Less cluttered environment. (b) Highly cluttered environment with through-wall sensing. (c) Highly cluttered dynamic environment. (d) Typical WiFi sensing system diagram.

a single model upfront rather than in designing a system which allows collaboration between new environments. Additionally, these studies still require a long duration of data collection and training from a large number of different locations. Instead, *our goal here is to develop a collaborative training framework for WiFi sensing that shares knowledge learned from one set of locations to improve the prediction capability of models at other new locations.* To account for this goal, we propose *WiFederated*, a system for collaboratively training machine learning models for WiFi sensing tasks at multiple unique environments using federated learning (FL) [19]. Through WiFederated, WiFi sensing tasks can scale to multi-location settings by (i) reducing the overall amount of annotated training data required per location; (ii) reducing the duration of training required for each new location; (iii) allowing for parallel edge training across multiple locations; and (iv) reducing the amount of data to be transmitted to a central server. Our contributions through this work are the following:

1) We perform human activity recognition experiments in ten locations and demonstrate that a model trained on data from one set of locations will not necessarily be generalized to new locations.
2) We propose *WiFederated*, a framework which introduces federated learning for the first time with WiFi sensing allowing for distributed model training across client devices within the network.
3) We evaluate WiFederated through extensive experiments and show that it outperforms the existing solutions.
4) We propose and evaluate client selection methods which select a subset of candidate locations based on local training-loss, resulting in a further increase in accuracy.
5) Lastly, we evaluate the feasibility of running WiFederated on real edge devices for inference and training and also introduce *continuous annotation* which allows clients to continue to capture representative CSI data and annotation labels for further model training over time.

The rest of the paper is organized as follows. Section II describes related work in WiFi sensing and reviews state of the art methods proposed for scalable multi-location WiFi sensing. In Section III, we provide the preliminaries of the proposed work including an introduction to Channel State Information (CSI), preprocessing steps used on the collected CSI data and the description of the underlying machine learning models used in this work. In Section IV, we demonstrate through experimental results the issues related to WiFi sensing across multiple locations to provide motivation for this study. Then, in Section V, we detail the design of our FL framework for WiFi sensing and provide its evaluation and comparison with existing works in Section VI. Section VII details some additional feasibility considerations for deploying our proposed system in real world scenarios. Finally, we provide our concluding remarks in Section VIII.

II. Related Works

Due to the non-intrusive and low-cost nature of WiFi sensing, it has been used in several applications including human activity recognition [14], [20], human body pose prediction [21], [22] and patient monitoring [9]. For example in patient monitoring, hospitals and elderly care facilities use WiFi sensing to detect falls without having the burden of attaching sensors on the body of patients. WiFi sensing is also used to detect fine-grained patient vital signs such as breathing rate and heart rate variability [23]. While a wide spectrum of studies using WiFi sensing can be found in recent surveys [1], typically these research studies consider only the case where a single model is pretrained for only one specific location.

In most of the current WiFi sensing research, experiments are designed such that a single WiFi transmitter (TX) and a single WiFi receiver (RX) are placed statically in a single physical location. Once placed, the set of actions to be detected are then performed over many repetitions to gather a significant distribution of signal variations per action. There are two core issues in this commonly adopted approach in WiFi sensing. *The first issue is*, because the CSI collecting devices are typically laptops with an attached Intel 5300 Network Interface Card (NIC) [24], the receivers can only be placed at limited locations which could be different than real world deployments of WiFi APs (e.g., ceiling in a hallway). Thus, some experiments might have been designed with the placement of TX and RX devices in unrealistic positions which give the best results, providing an unfair advantage when making predictions. Moreover, such configurations may result in models that may not work well when existing WiFi infrastructure and real world deployments of WiFi APs are used. *The second issue* is that the model which is trained with TX and RX at a single physical location only has knowledge of that one location. Thus, its predictions may not be of any use in new locations and we may need to repeat all of the data collection, annotation and training steps for each new location from scratch as it is done in [13], [25]. However, this is ultimately not scalable because each location then requires a time-consuming and potentially error-prone annotation step.

Study	Recognition Task	Learning Type	Multiple Rooms	Distributed Training	Knowledge Sharing
CrossSense [14]	Activity	Multiple Global Expert Models	Yes	Possible	No
Ma et al. [15]	Activity	RNN, LSTM, State Machine	No	No	No
EI Framework [16]	Activity	Adversarial Network Model	Yes	No	No
WiHand [17]	Gesture	LRSD Decomposition	No	No	No
WiTransfer [18]	Activity	Global Transfer Learning	No	No	Yes
WiFederated (Ours)	Activity	Distributed Federated Learning	Yes	Yes	Yes

TABLE I: Comparison of related multi-location WiFi sensing literature.

In some recent works, the problem of developing WiFi sensing solutions that can be generalized to multiple environments has been studied. For example, in CrossSense [14], first, a set of expert models are pretrained at some initial locations. For each new location, the best fitting model out of all of them is used. However, the number of expert models required could be large depending on the environments, requiring once again a large collection of annotated samples and a long training process. Moreover, as the expert models do not collaborate, they cannot gain any information from each other. The work in [15] similarly looks at location-independent activity recognition through the use of deep learning techniques such as recurrent neural networks (RNN) and long short-term memory (LSTM) networks. However, using such complex models requires a long training duration (e.g., weeks) and results in relatively low prediction rates (e.g., 50Hz) even with powerful GPUs. In this study, we aim towards making predictions at the edge with devices having limited computation and power capabilities, thus GPUs cannot be used and much lower prediction rates are expected. The *EI Framework* [16] uses an adversarial network approach which trains three individual sub-networks. The feature extractor network transforms the raw data which is then fed into both an activity recognizer network and a separate domain discriminator network which predicts the location that a given sample comes from. The goal is to optimize the weights in the networks such that the features extracted by the feature extractor increase the accuracy for the activity recognizer while also decreasing the accuracy of the domain discriminator. The EI framework is expected to perform better as more locations are added because more samples will be available for the feature extractor to learn from, however because the activity recognizer and domain discriminator are working against one another, as more locations are added more training epochs will also be required. The *WiHand* [17] hand gesture recognition system suggests using low rank and sparse decomposition (LRSD) to split raw CSI data into two distinct parts, namely, a low rank part containing background information and a sparse part containing noise due to hand gesture movements. While the work considers multiple positions, each evaluated position is located in the same room which does not demonstrate how the model is applicable in physically unique environments. Finally, *WiTransfer* [18] suggests the use of transfer learning to pretrain a global transfer model which can then be used at a new location with minor personalization training performed using data from that location. However, because the model is trained globally at a central location, increasing the number of pretraining locations increases the time to fully train the model as the training must occur sequentially on the aggregated data.

In this study, we address the issues of these centrally trained methods by training in parallel through an FL based approach. FL has become a popular topic in the research community due to the ability to train machine learning models in parallel at the edge while also preserving data privacy [19]. Details of the FL process are discussed in Section V. In the past, FL has most commonly been evaluated on standard machine learning tasks such as handwriting recognition with the MNIST dataset and sentiment analysis with the Sentiment140 dataset [26] but has also been used in real-world applications such as improving next-word suggestions on smartphone keyboards [27] as well as leveraging electronic medical records from multiple hospitals to better predict medical conditions without compromising data privacy [28]. By using FL for WiFi sensing, we aim for a more scalable method of training models across disparate devices in unique physical locations while reducing the number of annotated data samples required per location and still achieving a higher prediction accuracy. Table I provides a comparison of the discussed works with our proposed WiFederated approach.

III. Preliminaries

For the experiments in this work, we use our ESP32-CSI Toolkit[1] [3] to collect CSI from the WiFi-enabled ESP32 microcontroller, which provides a small-size, lightweight and standalone solution compared to other existing methods (i.e., a host laptop with an updated NIC) and can be deployed anywhere. We transmit frames from an ESP32 transmitter and then collect WiFi CSI from a separate ESP32 receiver. CSI is a metric coming from orthogonal frequency-division multiplexing (OFDM) systems which allow multiple streams of data to be transmitted in parallel to increase the speed of data transfer by splitting the bitstreams into multiple subcarrier frequencies at transmission and then reconstructing the bitstreams into one cohesive message at the receiver. OFDM is used with 802.11 based WiFi systems to achieve high data throughput rate as well as to reduce multipath fading.

A. CSI Amplitude and Phase

When a TX transmits a signal x, it is received by the RX as $y = Hx + \eta$, where η is a vector representing environmental noise and H is the CSI matrix. Each element in H is composed of both real and imaginary components. Given the real component ($h_r^{(i)}$) and imaginary component ($h_{im}^{(i)}$) for each subcarrier i within H, we can calculate the amplitude $A^{(i)} = \sqrt{\left(h_r^{(i)}\right)^2 + \left(h_{im}^{(i)}\right)^2}$ as well as the phase $\phi^{(i)} = atan2\left(h_{im}^{(i)}, h_r^{(i)}\right)$. For our evaluations, we only consider

[1] https://github.com/StevenMHernandez/ESP32-CSI-Tool

amplitude because of the noise inherited in the received phase values as well as the requirement for a synchronized clock between the two devices.

B. Rolling Mean Filter

CSI amplitude measurements can have spurious noise which is unrelated to any physical movements in the environment as a result of variations caused by the physical radio hardware as well as other environmental noise. Denoising can be used to filter out such noise in CSI measurements. We apply a rolling mean denoising approach, which takes a window of size W samples, and applies the mean function over the previously collected $W-1$ samples along with the currently collected sample across a single subcarrier. More formally, we calculate

$$\hat{A}^{(i)}_{(t)} = \frac{1}{W}\sum_{w=0}^{W-1} A^{(i)}_{(t-w)}. \tag{1}$$

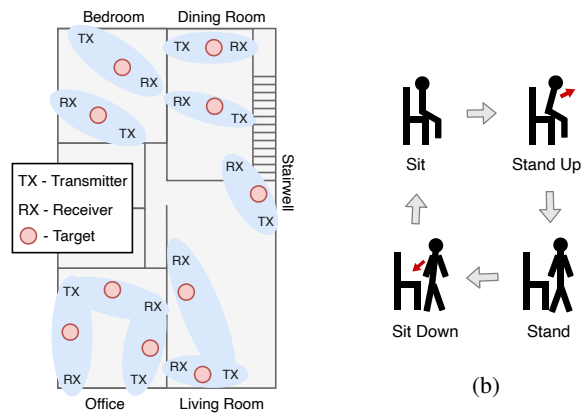
Assuming that we have 64 subcarriers, we then must keep a record of the previous $W-1$ values for each of these 64 subcarriers independently in memory on the edge device.

C. Machine Learning Training and Prediction

Using the preprocessed CSI data, we then train a machine learning model using a Dense Neural Network (DNN) architecture with three dense layers. Each layer has 100 hidden units and each unit uses a Rectified Linear Unit (ReLU) activation function. Each layer uses L_2 regularization for the kernel weights as well as an L_1 regularization penalty term for the activation output of each layer. The size of the input matrix for the model is 100×64, where 64 is the number of subcarriers and 100 is the number of CSI frames which is approximately one second worth of data since our sampling rate is set to 100Hz. The network terminates with a softmax multi-class output layer so that we can make predictions on multiple class types with a single network architecture. During training, we use a dropout value $d \in (0,1)$ between each layer to define the probability that a given weight is ignored during training to help prevent the model from overfitting to the training data. Dropout is ignored during model evaluation. We use Stochastic Gradient Descent (SGD) with learning rate $\eta = 10^{-5}$ to minimize the loss function

$$\mathcal{L}(\theta, x, y) = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{F}_\theta(x_i) - y_i)^2, \tag{2}$$

where θ is the set of parameters for the model ($\mathcal{F}$), $\mathcal{F}_\theta(x)$ is the prediction output of the model given input x, and y is the expected output based on the given action class which is a one-hot encoded vector of size N, where N is the number of possible predicted classes. The overall architecture is designed with edge devices in mind. This is why we do not consider more complex models such as the popular RNN or LSTM network architectures which would require more training data and longer training periods.



(a) (b)

Fig. 2: (a) Illustration of apartment environment where experiments are performed. TX, RX and human target are shown for each room location. (b) Four distinct actions (i.e., sit, stand up, stand and sit down) are recorded and annotated in each location in round-robin order.

IV. Motivation

To demonstrate our motivation for this work, we begin by explaining the experiments that are performed during our data collection and annotation process. We then review initial performance results on the collected data when using the standard training approach common to many other WiFi sensing research as illustrated in Fig. 1d.

A. Experimental Setting

In this study, we consider a collaborative network of WiFi sensing devices in different and disconnected physical locations. The illustration in Fig. 2a shows the ten distinct areas that we record experiments at as well as the placement of the TX, RX and the human target. Each location has different positions for the TX and RX corresponding to the preexisting power outlets which were built into the building which provides a natural selection of locations for TX/RX rather than selecting the most optimal locations for performing the sensing tasks. The goal here is to demonstrate that the system can leverage existing infrastructure. The target performs four actions (sitting, standing up, standing, sitting down) as illustrated in Fig. 2b. In each location, we perform 50 individual repetitions of each of the four actions in round-robin order resulting in a total of $2,000$ annotated action-segments with corresponding CSI samples. For each location, we designate the first 25 repetitions for training our models and the final 25 for evaluating. This allows us to check if the traits learned by the model are generalized by evaluating the accuracy on these final 25 testing repetitions which are never seen when training the model.

B. Initial Results

The goal of this work is to demonstrate how collaborative WiFi sensing devices can achieve better prediction results compared to devices that work alone. This can be especially

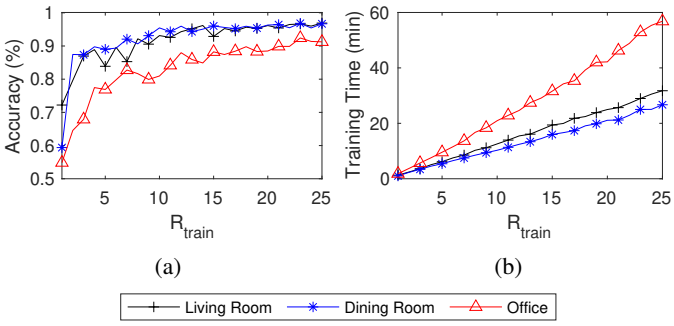


Fig. 3: (a) Accuracy for each locally trained model when different numbers of training repetitions of an action are performed in the location. (b) Training time to perform 100 epochs of training on a Raspberry Pi Edge device.

important when we aim to use preexisting WiFi infrastructure to build a WiFi sensing system [29]. In a non-collaborative system of WiFi sensing devices, we will need to record and annotate multiple new repetitions of each action before a local model can perform well in the environment. For example, Fig. 3a shows the accuracy of a non-collaborative local model with different numbers of training repetitions (R_{train}). We can see that training this local model on a small number of training repetitions produces a model which is overfit onto the training data and is unable to generalize to achieve high validation accuracy. Given our goal of large-scale deployments, performing a large number of repetitions at every location is not viable because it would increase the time and labor spent at each location. Similarly, this also increases the time to train the model as shown in Fig. 3b where we observe mostly a linear relation between the training time and the number of repetitions of each action when trained on a Raspberry Pi 4 model B edge device.

Training independent local models at each location without some collaboration mechanism means that we will not gain additional knowledge by adding new devices into the network. To account for this, an initial naïve approach is to collect and annotate CSI data at a few selected locations (we will denote this set of training locations as $\hat{L}$ throughout the paper) which we use to globally train a single machine learning model that will then be shared with all locations (L) including those which have no collected or annotated CSI data.

In Fig. 4 we show the prediction accuracy of our model ($\mathcal{F}_\theta$) when evaluated on each location $L_i \in L$, where $L = \{\text{Living Room}, \text{Dining Room}, \text{Office}\}$, after being trained on CSI data from different sets of $\hat{L} \subseteq L$. We select these three locations because they represent different multipath characteristics from one another. For example, the living room location provides a large open area with a low amount of environmental clutter; the dining room location offers a similar large environment but with higher amounts of environmental clutter; and finally the office location represents a smaller enclosed room environment with low clutter. We evaluate with the number of training repetitions $R_{train} \in \{10, 25\}$ to identify how increasing R_{train} affects the prediction capability of locations in $\hat{L}$ as well as locations not in $\hat{L}$ (i.e., $L - \hat{L}$).

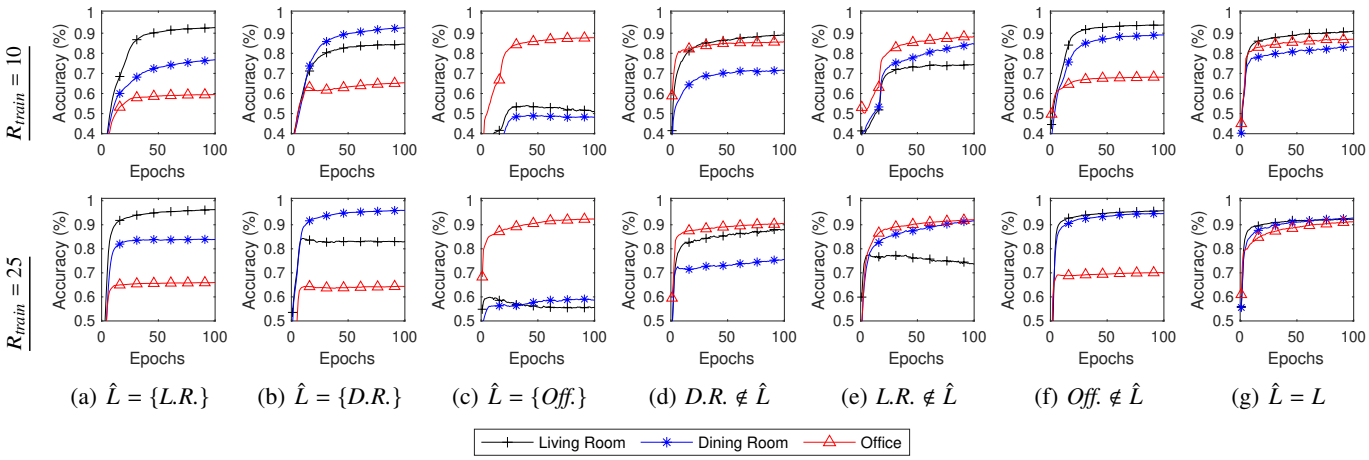
In Fig. 4a, we consider the case when $\mathcal{F}_\theta$ is trained on the Living Room location only. As we would expect, because the model is trained directly on data from this location (i.e., only first 25 repetitions), we can see that for both $R_{train} = 10$ and $R_{train} = 25$, the accuracy (on the last 25 repetitions) for the Living Room is high at 92.82% and 96.30%, respectively. Interestingly, we can see that increasing R_{train} from 10 up to 25 also allows for an increase in prediction accuracy for two unseen locations, namely Dining Room going up from 76.76% up to 83.90% and Office going from 59.49% up to 65.98%. However, we can still see a noticeable gap between the accuracy for each of these locations demonstrating that the model is still better fit to the data at the Living Room location.

Following this, in Fig. 4b we train only on the Dining Room location. When $R_{train} = 10$ both the Living Room location (seen during training) and Dining Room (unseen during training) achieve very close accuracy after 100 epochs of training. This shows that the model trained on the Dining Room location can also be used in a completely unseen new location. However, we can see that the accuracy for the Office location is still low at 65.38% demonstrating that even though the model is applicable to some locations, it is not necessarily a generalizable model that can be applied in all new locations.

On this note, we look at the case in Fig. 4c where we train only on the Office location. While the model is able to achieve a high evaluation accuracy for the location that it is trained on (i.e., Office), the model cannot be used in new and unseen locations. Comparing to Fig. 4a where a slight improvement is observed in unseen locations, the model trained exclusively on the Office location dataset is only able to achieve between 48.12% and 58.76% accuracy in the unseen locations with a gap of approximately 44% difference between the accuracies in the seen and unseen locations.

Training our model on data from two locations rather than one produces Fig. 4d-f. The model accuracy in the unseen locations mostly converge to a static value except in Fig. 4e when $R_{train} = 25$, where the gap in accuracy between seen and unseen locations grows as training continues. In Fig. 4e, when $R_{train} = 10$, the gap is only +10.04%, but when $R_{train} = 25$, the gap increases significantly to +18.05%. Out of each of these three cases where $|\hat{L}| = 2$, Fig. 4f shows the largest overall gap of +21.04% when $R_{train} = 10$ and +24.53% when $R_{train} = 25$. This again demonstrates that a model with high accuracy on trained locations may not be generalizable to unseen locations.

As such, we can only be sure that a high accuracy is achievable across each location when all locations are involved in training (i.e., $\hat{L} = L$) as we can also see from the results in Fig. 4g. However, for global models, annotated data needs to be shared with a central server for all devices. For example, if CSI data and annotations are collected continuously over time such as in [21], [30], [31], the amount of data transmitted to the server can be large. As the network of locations also increases in size, the server resources used to handle all of the incoming data may be too much. Thus, it would be more preferable to do training on an edge device at the physical location. However, to accomplish this, we will need a new method for sharing knowledge between multiple locations which reduces the amount of data transmitted over the network.



(a) $\hat{L} = \{L.R.\}$ (b) $\hat{L} = \{D.R.\}$ (c) $\hat{L} = \{Off.\}$ (d) $D.R. \notin \hat{L}$ (e) $L.R. \notin \hat{L}$ (f) $Off. \notin \hat{L}$ (g) $\hat{L} = L$

Fig. 4: Prediction accuracy in three locations (Living Room (*L.R.*), Dining Room (*D.R.*), Office (*Off.*)) when trained with data from only locations in $\hat{L}$. Columns (a-c) train on a single location, (d-f) train on two locations and (g) trains on all locations.

V. Federated Learning Framework (*WiFederated*)

In order to develop a system which can collaboratively train a machine learning model in parallel across many edge devices (or clients) in disparate physical locations, we propose a collaborative WiFi sensing framework using federated learning [19], which we call *WiFederated*. FL is useful in our case because we expect that each client will have a different data distribution corresponding to their unique physical environment, and we want to benefit from all data while also reducing the amount of data shared to any central server so as to reduce network usage. This also has a secondary purpose in allowing for massively parallel machine learning by enabling each client to perform machine learning locally rather than at the central server.

Our overall goal is to find model weights (θ) minimizing:

$$\min_{\theta \in \mathbb{R}^d} \sum_{k=1}^{|L|} \frac{1}{n_k} \sum_{x_i, y_i \in P_k} \mathcal{L}(\theta, x_i, y_i), \tag{3}$$

where L is the set of the clients at diverse physical locations, $P_k = \{\mathbf{x}_k, \mathbf{y}_k\}$ is the set of annotated data points for a client k where $\mathbf{x}_k$ is the set of CSI input and $\mathbf{y}_k$ is the corresponding set of annotated labels recorded at the location of client k such that $n_k = |P_k|$, and $\mathcal{L}(\cdot)$ is a loss function as described in (2).

The issue with this optimization is that $|L|$, the total number of clients, is expected to be large which will result in a high computation and communication cost and thus slower model training. Furthermore, some locations may have either small amount of data or low quality of data which will only poison the prediction quality for other clients. To combat this, our proposed FL system selects a subset of clients ($\hat{L} \subseteq L$) to iteratively update the model weights. For each client ($k \in \hat{L}$), we learn unique model parameters (θ_k) by optimizing:

$$\min_{\theta_k \in \mathbb{R}^d} \frac{1}{n_k} \sum_{i \in P_k} \mathcal{L}(\theta_k, x_i, y_i). \tag{4}$$

After N_{epochs} of training per client, each client sends $\Delta\theta_k$ to a central server to perform *Federated Averaging (FedAvg)* [19] to determine new model parameters for the next round $(r+1)$:

$$\theta^{(r+1)} = \theta^{(r)} + \sum_{k=1}^{K} \frac{n_k}{n} \Delta\theta_k, \tag{5}$$

where $\Delta\theta_k$ is gradient change over top of $\theta^{(r)}$ learned from the data available to client k. Whenever new clients join the network, they can use the pretrained federated model parameters ($\theta^{(r)}$) for the given round as-is if no additional labelled data points are available for the locations or alternatively can perform $N_{post\text{-}epochs}$ additional training epochs (i.e., model personalization) on top of the federated model to better fit to their own unique data distribution for the client.

Note that the FL approach is different from transfer learning [32] in which a single parent model is trained over a large number of training epochs on a large amount of annotated data, typically on a powerful computation system. The goal of transfer learning is to create a model which can be reused in a new somewhat related task by using a small amount of additional data from the new task. While FL gains from this same idea, the goal of our FL framework is not to just reuse a pretrained model for some new task, but instead to allow many disparate clients to massively train a model in parallel from scratch so that new clients can gain directly from this pretrained model upon joining the network. Furthermore, transfer learning requires all data to be located at a single central location for training, thus removing the parallel training capabilities found with FL systems. While some WiFi sensing studies utilize transfer learning as mentioned in Section II, to the best of our knowledge, FL based training has not been proposed for WiFi sensing tasks.

Algorithm 1 shows the steps of the proposed WiFederated learning model. L is the set of locations where we aim to perform WiFi sensing and G is our global federated model where $G.W$ is the set of weights for all layers of the model. L can potentially change over time as new locations are added or as locations are removed due to battery power loss or disconnection from the network. Note that in the algorithm, we only perform a predetermined number of rounds (N_{rounds}) before terminating, however federated training can be per-

Algorithm 1: WiFederated Learning

Input: Set of all WiFi sensing locations L.
Global model G where $G.W$ is the set of weights for the model.
Local model M_l $\forall l \in L$.

$G.W = RandomWeights()$
forall $r \in \{1, \ldots, N_{rounds}\}$ **do**
  forall $l \in L$ **do** `// Share weights.`
    $M_l.W = G.W$
  $\hat{L} = SelectClients_{Train}(L)$
  forall $l \in \hat{L}$ *s.t.* $\hat{L} \subseteq L$ **do** `// In parallel.`
    $M_l.train(N_{epochs}, R_{train})$
  $\hat{L}' = SelectClients_{FedAvg}(\hat{L})$
  $G.W = \frac{1}{|\hat{L}'|} \sum_{l \in \hat{L}'} M_l.W$ `// FedAvg.`
forall $l \in L$ **do** `// Post-Train.`
  $M_l.train(N_{post\text{-}epochs}, R_{post})$

formed continuously to allow the network to recognize newly annotated data over time. At the beginning of each round, the weights of the global model are shared with each location so that they can each begin from a similar starting point. Within each round, we select our set of clients ($\hat{L}$) to train locally for some fixed number of training epochs (N_{epochs}). This step can be thought of as a model-personalization step because at the beginning of the round, the model is initially fit on the global distribution of CSI data and at the end of the round, the models at each selected client are slightly more personalized to their own distribution of data. It is important that we do not perform too many training epochs locally during each round because then the local models will become more overfit onto location-specific data. Additionally, high N_{epochs} will result in over-utilization of individual devices which will result in a higher power consumption for these edge devices. Moreover, as mentioned in [33], [34], if we perform federated averaging on models which are fit to very different distributions, the resulting federated averaged weights will not be fit to either of the distributions of location-specific data but will also not be fit to the global distribution of CSI data. Thus, as long as N_{epochs} is not too large, at the end of each round, the FedAvg step will help prevent the global federated model from diverging too far away from fitting onto the global CSI data distribution. To understand how federated learning behaves with varying amounts of training samples, we also limit the number of training-repetitions to R_{train} for each location. Furthermore, after the personalization step, we may decide to further refine our selected clients down to $\hat{L}' \subseteq \hat{L} \subseteq L$ before applying FedAvg. This can be important because we may find that the weights proposed by the initially selected clients may cause the federated model to converge in a negative way. Finally, after training our model over N_{rounds}, each location in L can perform $N_{post\text{-}epochs}$ of post-training using R_{post} repetitions. While this step will not produce completely new models per location, it will be able to take the generalizable features discovered through the federated learning process and alter the model weights slightly to give better location-specific results.

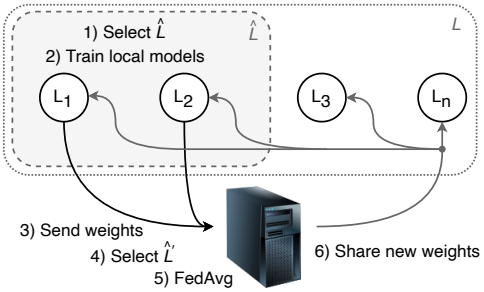


Fig. 5: Illustration of one round of the WiFederated system.

The illustration in Fig. 5 demonstrates the six key steps taken for performing FL in our *WiFederated* framework on multiple locations and a central server. In the illustration, L_1 and L_2 are selected to train locally and then share their weights to the central server. Each location only transmits the new model weights rather than sharing the actual CSI data and annotation labels through the network. This means that we can keep a constant bound on the amount of data transmitted over the network by selecting an appropriately designed and sized machine learning model architecture. If we were to instead share annotated CSI data, then the size of the data could be unbounded especially in cases where our devices are able to self-annotate. Once the server receives the weights from each client location $L_i \in \hat{L}$, the server can perform another round of client selection to select $\hat{L}' \subseteq \hat{L}$ before performing the federated averaging step. This second client selection phase allows the server to filter out the weights received from locations that are deemed to be lower quality based on some metric such as loss. After the federated averaging step is performed, we end the loop by resharing the central model to each of the locations. After performing these steps over subsequent rounds, any location can perform a final set of personalization training epochs on their locally available data. This step takes the shared federated model and performs a final number of personalization training epochs which can be useful for totally new locations or locations with few training locations. The key here is that this can be performed without requiring the location to share the results back to the central server. The goal accomplished by this framework is that new locations can achieve higher prediction accuracy by using the federated model as a base rather than starting from scratch.

VI. Evaluation

In this section, we evaluate the proposed *WiFederated* learning framework for use in the setting described in Section IV-A.

A. Impact of Averaging Interval

Two types of clients exist in our system, clients *participating* in federated averaging (i.e., the clients in $\hat{L}$) and clients with data that is *unseen* during the training phases (i.e., the clients not in $\hat{L}$).[2] We first look at how the federated averaging process affects the prediction capability of the locations

[2]For the initial evaluations in this section, we say $\hat{L}' = \hat{L}$. Client selection methods for selecting $\hat{L}'$ are further evaluated in Section VI-F.

participating in training the federated model. To evaluate this, we set $\hat{L} = L = \{\text{Living Room}, \text{Dining Room}, \text{Office}\}$ and evaluate on each of these clients individually. Fig. 6a shows the prediction accuracy when the number of epochs per round $N_{epochs} = 50$. Initially, for all three locations, the accuracy remains exactly the same between the local model and the federated model. This is because, up until epoch 50, the models are the same; no federated averaging has occurred. However, after 50 epochs, a sudden dip in accuracy is found for the federated models. This dip is expected because at the end of this epoch, the first round of local training has concluded and the three local models are aggregated together through the federated averaging step. The key observation is that the federated averaging step aims to create model parameters which are generalizable to many locations rather than specialized to any one location. On the other hand, when we train a local model solely on the data available at a single location, the model will be better fit to the distribution of data at the given location, but it will be too specialized to aid other new locations when they join the network. *The goal for all of the training clients in $\hat{L}$ is to sacrifice some amount of predictive capability so as to benefit other new locations.* Even so, we can see in Fig. 6a that after the sudden dip caused by the FedAvg step, the accuracy quickly returns to a similar accuracy as we would see if we had simply trained the local model. This shows that the federated averaging step does not cause the accuracy to deteriorate too much for the participating locations. Fig. 6b shows the accuracy comparison between locally trained models and federated models during the final FedAvg aggregation. We can see that, as N_{epochs} increases, the accuracy for the Office location decreases indicating that if we set N_{epochs} too large, the FedAvg step has a big impact on the prediction capability of the model at that location. On the other hand, because we assume a fixed total budget of $N_{budget} = N_{epochs} \times N_{rounds} = 100$ epochs, increasing N_{epochs} decreases the number of rounds and thus consequently decreases the network communication. This is because at the end of each round, $|\hat{L}|$ locations must first communicate $\Delta\theta_k$ back to the central server for FedAvg aggregation and then the server must communicate $\theta^{(r+1)}$ back to $|L|$ clients.

Thus, we must find a balance between communication overhead and prediction accuracy. To this end, for the following experiments, we set $N_{epochs} = 10$ which gives a minor decrease in prediction capability when compared to $N_{epochs} \in \{1, 2\}$ but greatly reduces the amount of communication required.

B. Impact on Unseen Locations

It is useful to see how FL performs from the perspective of clients within $\hat{L}$. However, our primary goal is to see how such an FL framework can be helpful for new locations which are added to the network with zero or only a small number of available annotated training repetitions. To show this, in the following results when we evaluate a given client L_i, we set $\hat{L} = L - \{L_i\}$ when training our federated model so that we can show that the parameters learned by the model are learned generally from all other locations and the model does not have any beforehand knowledge of the data or distribution of data at

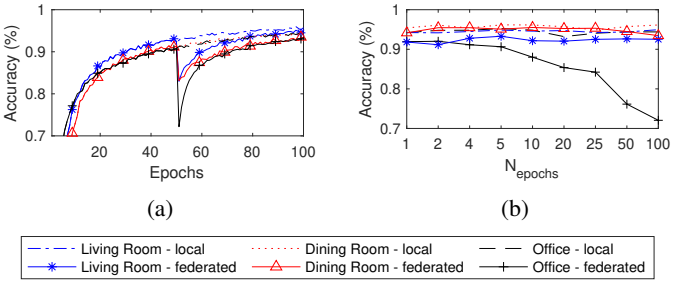


Fig. 6: (a) Accuracy of federated learning over 100 epochs when $N_{epochs} = 50$ and $\hat{L} = L$ versus local machine learning. (b) Accuracy after applying final round of FedAvg for different values of N_{epochs}.

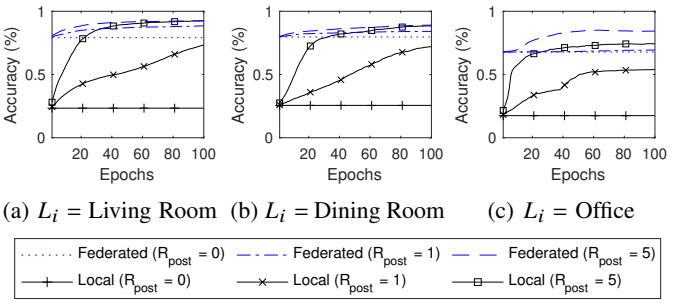
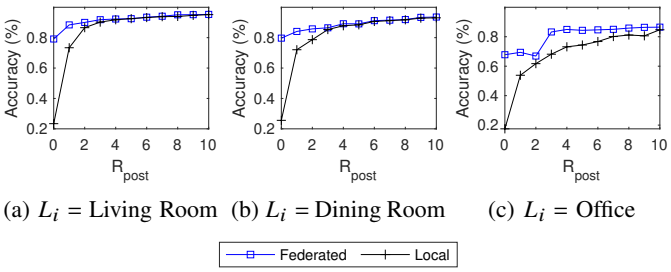


Fig. 7: Accuracy during post-training (personalization) over 100 epochs starting with a randomly initialized local model versus starting with a federated model trained on $\hat{L} = L - L_i$.

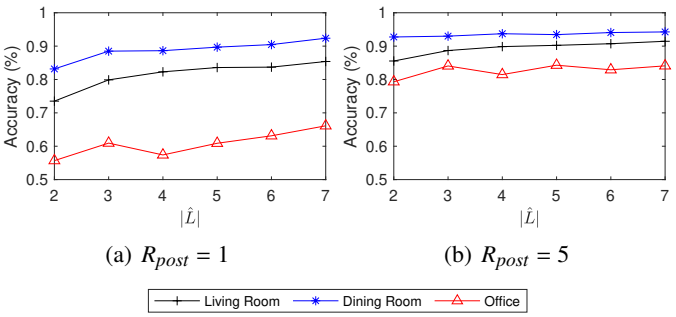
client L_i. This is important to evaluate because when deployed into a real world system, we may not have annotated data for all clients. In those cases, we can select $\hat{L}$ based on attributes such as the availability of annotated data or even in cases of battery powered units, we can select only those which have surplus power to complete a given training round. As such, $\hat{L}$ can also change for every subsequent round to prevent wasting networking resources or power at any single location and also to prevent overfitting on data from the selected locations.

Consider the three subfigures in Fig. 7 where L_i is set to a different client for each. The figures show the accuracy over all 100 training epochs for the two training methods. We begin by comparing the *local* training method. For this method, the models are randomly initialized and then trained on some number of training repetitions (R_{post}) from location L_i. For the following evaluations note that a single post-training repetition ($R_{post} = 1$) includes a sample for each class type, as we discussed in Section IV-A. Consider the case where a new client is added to the network without any post-training repetitions (i.e., $R_{post} = 0$). For this case, the model cannot be trained for these 100 epochs, which means that the accuracy of the model remains constant at whatever value it started at. Since the model was initialized to have random values for θ, the accuracy when $R_{post} = 0$ is approximately 25% for each value of L_i because there are 4 classes to predict from. We should next compare this to a model which is instead initialized on a federated model pretrained on $\hat{L}$. For the federated model, these 100 post-training epochs are represented in Algorithm 1 as $N_{post\text{-}epochs}$. We can see that, with



(a) L_i = Living Room (b) L_i = Dining Room (c) L_i = Office

Fig. 8: Accuracy for federated model versus randomly initialized local model after 100 epochs of post-training (personalization) with different post-training repetitions (R_{post}) at L_i.



(a) R_{post} = 1 (b) R_{post} = 5

Fig. 9: Accuracy for WiFederated as $|\hat{L}|$ increases.

the federated model, the accuracy for $R_{post} = 0$ is also constant over the epochs because we do not have any post-training repetitions to train the model further. However, because the model learns generalizable traits from $\hat{L}$, the initial accuracy is much higher than the randomly initialized local model. This is a very important feature for ensuring that WiFi sensing can scale without requiring all new client locations to pass through extensive CSI data collection and annotation steps.

Suppose now that a new client is added to the network, but we are able to perform very few repetitions of our actions in the environment (i.e., $R_{post} \in \{1, 2, 3\}$). We can see in Fig. 8 that the addition of these post-training repetitions allows both models to increase their predictive accuracy by the end of the 100 post-training epochs. However, with small values for R_{post}, the local model is still unable to surpass the initial accuracy achieved by the federated model. Thus, even when some number of training samples are available, using the pretrained federated model can achieve higher prediction accuracy compared to training from scratch at each location.

C. Impact of the Number of Training Locations

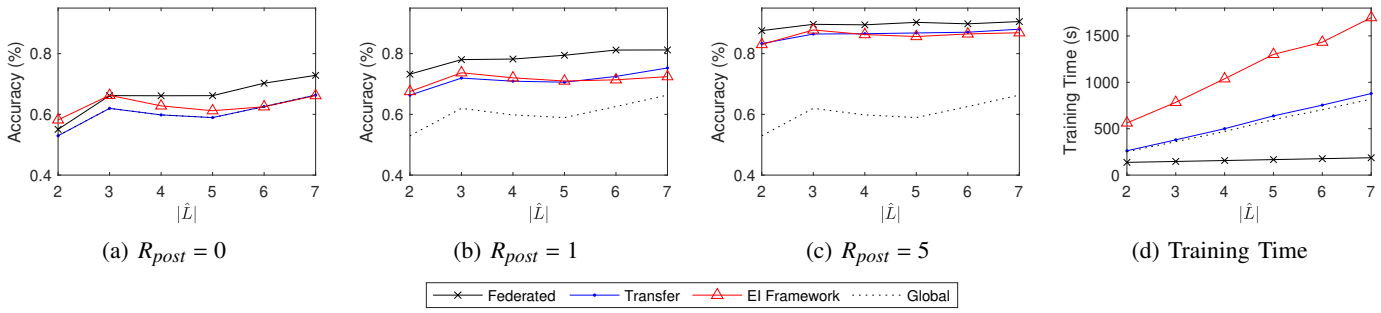
So far, we have evaluated the WiFederated system when the number of training clients $|\hat{L}| \in \{2, 3\}$, however FL is able to accommodate larger number of clients especially considering that the training is processed in parallel across each selected client. To evaluate larger number of training locations, we collected data at ten total locations. We continue to use the same three locations (i.e., {Living Room, Dining Room, Office}) to evaluate our system since they have unique multipath characteristics and also to keep results consistent with our evaluations in the previous sections. Thus, seven remaining candidate locations are available for pretraining such that $\hat{L} \subseteq L - \{\text{Living Room, Dining Room, Office}\}$.

Fig. 9 shows the accuracy for each of the evaluation locations as $|\hat{L}|$ increases up to 7 which corresponds to an increasing trend in the accuracy for the evaluation locations. When $|\hat{L}| = 7$ all seven candidate locations are selected for federated averaging. When $|\hat{L}| < 7$, different combinations of clients could be selected for $\hat{L}$. To account for this, we repeated each experiment 10 times with randomly selected value for $\hat{L}$ each time. This ensures that we do not accidently select values for $\hat{L}$ which are consequently overly good or overly bad. Furthermore, we use a static client-selection process where the clients in $\hat{L}$ are selected randomly at the first round of federated training and then reused for all subsequent rounds without further client selection. This process emulates the case where CSI data is collected at some random initial set of $|\hat{L}|$ client locations so that future client locations require a much shorter data collection and annotation period. When the technician selects these first $|\hat{L}|$ clients, they cannot be sure whether the locations they selected will be useful for building a generalizable federated model. Even so, we can see that increasing $|\hat{L}|$ from 2 up to 7 clients offers an accuracy increase of +11.88% for L_i = Dining Room, +9.21% for L_i = Living Room and +10.43% for L_i = Office when the number of post-training repetitions $R_{post} = 1$ and +5.91% for L_i = Dining Room, +1.52% for L_i = Living Room and +4.77% for L_i = Office when $R_{post} = 5$. This shows that increasing the number of locations involved in training the federated model improves the prediction accuracy of the models for new locations which are unseen during the FL steps.

D. Comparison with State of the Art Approaches

In some recent studies [18], [20], in order to address scalability issue of WiFi sensing, a transfer learning based approach is proposed. However, for transfer learning based model training, data from $\hat{L}$ must be aggregated to a central location to train the model. Another approach is the EI framework [16] which uses adversarial networks. EI models are composed of two separate network branches connected by a parent feature extraction network. The first branch acts as a single activity recognizer with loss value $\mathcal{L}_a$ while the second branch acts as a domain discriminator with loss value $\mathcal{L}_d$. The domain discriminator attempts to recognize the domain or physical location where the CSI data was collected. The goal is to minimize $\mathcal{L}_{EI} = \mathcal{L}_a - \mathcal{L}_d$ which can be read as minimizing the loss of the activity recognizer while maximizing the loss of the domain discriminator. By training in this method, it is expected that the model will be generalizable to unseen locations by learning features that are representative of the activities being performed but not specific to any domain.

For the remaining evaluations, we look at the average accuracy across all three evaluation locations rather than individual client accuracy values. Fig. 10a-c show a comparison of four methods: our federated approach, the commonly used transfer learning approach, the adversarial EI approach and the globally trained model approach. In Fig. 10a, both *Transfer* and *Global* achieve the same accuracy. This is because when $R_{post} = 0$,



(a) $R_{post} = 0$ (b) $R_{post} = 1$ (c) $R_{post} = 5$ (d) Training Time

Fig. 10: (a-c) Comparison of four methods when using different numbers of pretraining locations with different post-training repetitions. (d) Training times required for each method with federated learning being the fastest thanks to parallelism.

there are no training repetitions to personalize the transfer learning model and as such, the transfer learning method and the global training method are the exact same. Both methods train their model on the same $|\hat{L}|$ locations at a global server before sharing the model with the client. Fig. 10b demonstrates how *Transfer* is different from *Global*. Namely, because $R_{post} = 1$, transfer learning allows some additional post-training or personalization steps over top of the globally trained model. Even so, we can see that our federated model consistently achieves higher accuracy than the *Transfer* model by +7.69% when $R_{post} = 0$, +8.87% when $R_{post} = 1$ and +4.26% when $R_{post} = 5$. With EI, we find that when $R_{post} = 0$, and $|\hat{L}| \in \{2, 3, 4\}$, the accuracy is similar to our federated approach but when $|\hat{L}| \in \{5, 6, 7\}$ the accuracy is more similar to the transfer learning approach. We find that as $|\hat{L}|$ increases, the number of training epochs for the EI framework must also increase. Thus, because we limit all models to a total budget of 100 epochs, the accuracy for the EI framework decreases. However, we find that even increasing the total budget to 500 epochs for EI still achieves a lower accuracy than the FL model trained with a budget of 100 epochs. When $R_{post} \in \{1, 5\}$, EI exhibits similar accuracy to transfer learning suggesting that the extracted features are similar between both methods.

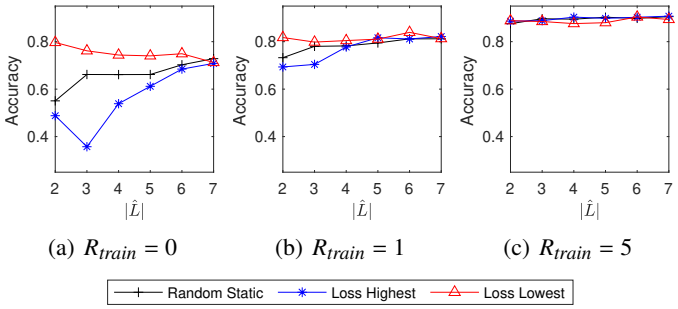
All methods have a general trend towards improvement as $|\hat{L}|$ increases. However, it is hugely important to consider the training time required for each of these methods. The key important insight is to see that our FL method is able to train in parallel across clients while the other methods must first aggregate the data to a central server which must then train the model sequentially on all of the same data. In Fig. 10d, we demonstrate how the time to train global and transfer learning models increases as more locations are used for training while the time required for FL remains relatively constant as a result of parallel training. EI method requires twice the amount of time required by the transfer learning method because it is essentially training two models, the activity recognizer and the domain discriminator. Note, unlike Fig. 3b results which are obtained on a Raspberry Pi, because of the memory overhead required for global, transfer learning and the EI methods, this evaluation is completed on a more powerful server.

E. Run Time Complexity Comparison

The training process for all of the discussed methods is composed of four stages: (i) data collection, (ii) preprocessing, (iii) training the model, and (iv) post-training, after which evaluation is followed. Out of these steps, the most time consuming step is the training step followed by the post-training step, thus they both define the overall duration for the model development. For the results in Fig. 10a-c, we allow all models to initially be trained on the datasets available in $\hat{L}$ for a total of $N_{budget} = 100$ epochs after which the federated models, transfer learning models and adversarial EI Framework models are trained for an additional $N_{post\text{-}epochs} = 100$ epochs. For simplicity in our notation, we say that $e = N_{budget} = N_{post\text{-}epochs}$. The time complexity of training is related to the product of the number of epochs (e) and the number of samples (S) used which we can denote as $O(e \times S)$. In the case of FL, the data from each location is trained independently and distributed in parallel, thus the time complexity would be $O(e \times max(S_l \ \forall \ l \in \hat{L}))$. On the other hand, transfer learning would have a dataset size $S = \sum_{l \in \hat{L}} S_l$, thus the time complexity is $O(e \times \sum_{l \in \hat{L}} S_l)$, which becomes $O(e \times S_l \times |\hat{L}|)$ if we assume $S_l = S_k \ \forall l, k \in L$ for simplicity. The difference between the transfer and EI methods is that EI essentially trains two model networks, an activity recognizer and a domain discriminator while transfer learning only trains a single activity recognizer model. However, this will still keep the time complexity for EI asymptotically similar to transfer learning at $O(e \times S_l \times |\hat{L}|)$. The post training step at a given location for FL, transfer learning and EI is performed on S_{post} number of post training samples, where typically $S_{post} << S_l$, for an additional e epochs which results in an added time complexity of $O(e \times S_{post})$. Thus, we can conclude that FL has a time complexity of $O\left(e \times \left(S_l + S_{post}\right)\right)$, global learning has a time complexity of $O(e \times S_l \times |\hat{L}|)$, and transfer learning and EI both have a time complexity of $O\left(e \times \left(S_l \times |\hat{L}| + S_{post}\right)\right)$, where $|\hat{L}|$ is the number of locations from which samples are aggregated for training at the central server. We can see that the complexity of global learning, transfer learning and EI methods all increase as $|\hat{L}|$ increases while FL is able to maintain a consistent time-complexity showing that FL will be a much better option for scalability as additional locations are added to the system.

F. Impact of Client Selection

Client selection can be performed at two different times during the entire FL process as denoted in Algorithm 1. First,



(a) $R_{train} = 0$ (b) $R_{train} = 1$ (c) $R_{train} = 5$

Fig. 11: Impact of client selection with different number of training repetitions (R_{train}).

we select $\hat{L}$ to be the clients that are used to train in parallel. After this, we can further reduce this selection to $\hat{L}' \subseteq \hat{L} \subseteq L$. Typically the first client selection step is used to distribute tasks fairly across the network. For example, it would not be fair to require any single client to take part in all FL rounds. This would consume more power and waste time for this individual client. As such, it can be useful to be selective when choosing $\hat{L}$ so that we do not overburden any single device. Furthermore for battery powered devices, it is important to select clients which have a battery level above some threshold to filter out any devices that may lose power during training.

The second client selection step can further guide the optimization of the federated model. For example, there may be clients with poor quality training repetitions or clients (k) whose $\Delta\theta_k$ will cause some negative impact onto the model parameters for the federated model as a whole. As such, we consider some other client selection methods. Specifically, we use the calculated loss $\mathcal{L}(\cdot)$ for all clients in $\hat{L}$ to determine which clients should be used for FedAvg. It was previously recognized [35] that selecting the client with the highest loss during client selection will increase the accuracy of the federated model overall. The idea is that a high loss is directly related to a high amount of error. If a client has a high error, then the distribution of the data at the client must have some amount of novelty which may be applicable to other clients as well. In Fig. 11 we can see a comparison between client selection methods. Comparing the *Random Selection* method we used for previous evaluations to the *Loss Highest* method, we can see that using this loss-based approach to guide the federated system actually results in lower accuracy overall. Our intuition here is that if we guide the training with high-loss clients, then we are selecting clients with the worst fit to the federated model from the previous round. This means that the clients are likely to have low-quality data available for training and thus the federated model as a whole will suffer. Alternatively, if we use the low-loss clients, then data available from these clients is similar to what is expected by the federated model and by extension, the data would also be similar to the data found in the locations selected for $\hat{L}'$ in previous rounds. In fact, if we take this opposite approach and guide the FL process through a *Loss Lowest* approach, we achieve a greater accuracy. When $|\hat{L}| = 2$ we achieve an accuracy of 79.89% with *Loss Lowest* versus 55.05% for *Random Static*. Similarly with $R_{post} = 1$, the *Loss Lowest* approach achieves 81.74% compared to 73.25%.

VII. Feasibility of WiFederated at the Client

We demonstrated that our proposed WiFederated system achieves increased predictive accuracy by training local models across different locations using a federated averaging and client selection process in comparison to starting from a randomly initialized model. Moreover, we demonstrated that our framework reduces the amount of training data that is required at each location when compared to training a local model at a single location independently without some form of collaboration. For the development and deployment of a full system leveraging this FL framework for WiFi sensing, we must also consider some additional issues.

A. Training and Inference at the Edge

Training deep learning models at a desktop computer or at a server benefits from the use of GPUs to speed up the training time especially as more data is added. However, as we discussed, sending all data to a central server for processing may not be feasible and with the use of FL is no longer a requirement. However, we must consider that local training at the edge requires special consideration. When designing the model used throughout this work, special care was placed to ensure that the complexity of the model will not require GPU-based training. Thus, we can use less powerful devices to train our model at each location.

Single-Board Edge Computers: Single board computers such as the Nvidia Jetson series of boards are designed specifically for machine learning at the edge with an on-board GPU and a full Ubuntu operating system. This means that any software written to run on a standard computer or server can directly be ported to the Jetson single-board computer, allowing for fast development time. Alternatively, lower cost single-board computers such as a Raspberry Pi can also be used in the same way, however without direct access to an onboard GPU. In Fig. 12, we show the time required for training a local model on different values of R_{train} for both a Raspberry Pi 4 B and an Nvidia Jetson Xavier NX. We can see that the GPU on the Jetson speeds up the computation considerably, but at a higher cost compared to the Raspberry Pi. Even so, the Raspberry Pi can still train the model in under 10 minutes when $R_{train} \leq 6$. This is useful considering that post-training for most edge devices in the WiFederated system can still achieve good prediction accuracy even with only a small number of repetitions.

Using standard ESP32s: To allow for the fewest additional hardware components in the system at each location, it would be most beneficial to train on the ESP32 microcontroller used to collect CSI. Recent research literature shows that training machine learning models on such low-powered devices is desirable [36], [37] and code libraries such as MicroMLP [38] have appeared for training machine learning models on low resource microcontrollers such as the ESP32. We attempted to train our federated model using this library, but we find that the model does not fit in memory due to some inefficient memory

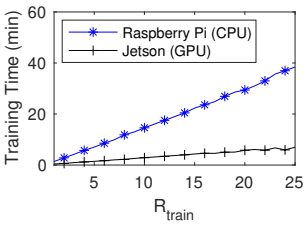


Fig. 12: Average training time for edge devices.

Device	Rate (Hz)
ESP32	5.726
Raspberry Pi	2252
Jetson	5359

TABLE II: Average prediction rate for edge devices.

allocation within the library itself. Instead we look at using a popular model inference library called Tensorflow Lite. Using this library, with model quantization, we are able to store our model directly in memory on the ESP32 and we are able to achieve a prediction rate of 5.726Hz. In Table II, we can see a comparison of prediction rates possible with each edge device. Even though both Raspberry Pi and Nvidia Jetson devices are able to achieve prediction rates of greater than $1,000$Hz, CSI collection rate is only 100Hz. This shows that the single-board computers can make predictions for every received CSI sample. Many applications may not need such high prediction rate because human activities should not change at such a high rate, so a standalone ESP32 is still useful for inference due to its low power consumption and small size.

B. Continuous Annotation

The process of annotating and recording CSI for a single location can be time consuming, repetitive and error prone. Luckily, as demonstrated in the previous section, leveraging a model obtained through FL in multiple different locations can greatly reduce the number of action repetitions which are needed to obtain a useful model.

Furthermore, we find that some sensing tasks have the ability to be continuously annotated over time, thus allowing for the model to continuously be trained on new data from the environment. Wearable respiratory monitoring belts have commonly been used in WiFi sensing research to label ground truth data for monitoring patient breathing patterns [39]. However, these works assume the belts are used only to train the initial model, but are never used again afterwards. An alternative approach is to use the sensor data to train the CSI based model continuously over time. Fig. 13a illustrates that CSI can be collected in the background at the same time a wearable sensor is being worn and Fig. 13b shows a time series view of the data being collected. While CSI is continuously collected in the background, there is a span of time where wearable sensor data is missing. This may happen because the user removed the wearable for example before going to bed at nighttime or due to discomfort. However, even when the sensor is not being worn, it can be important to continue to track the user. To do this, whenever sensor data is available (i.e., sensor is worn), the sensor data can be used to automatically annotate the collected CSI data to train WiFi sensing model $\mathcal{F}$. Then, whenever sensor data is unavailable, $\mathcal{F}$ can continue to monitor the user using the available CSI data. This ensures that the model can continue to learn over time and the person can

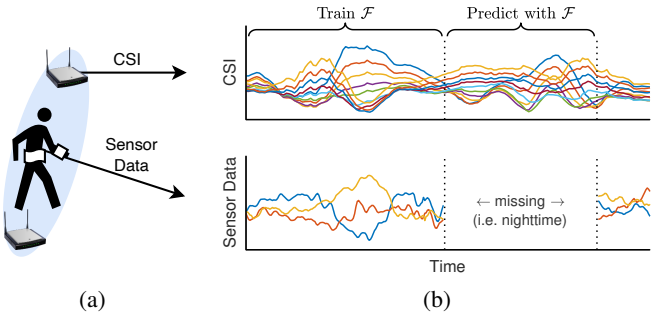


Fig. 13: Example scenario for continuous learning. (a) User with a wearable sensor while CSI is collected in the background. (b) When sensor data is available, both CSI and sensor data can be used to train $\mathcal{F}$. When sensor data is unavailable (i.e., at nighttime when wearable sensors are removed), CSI can be used with $\mathcal{F}$ to continue monitoring.

be safely monitored even when the sensor data is unavailable. Continuous data labelling and thus continuous model training means that the proposed WiFederated system can continue to learn without requiring time-consuming manual data collection steps and can also help reduce the effects of data drift [40] over time due to environmental changes.

VIII. Conclusion

In this work, we introduce WiFederated, a federated learning framework designed for scalable deployment of multi-location CSI-based WiFi sensing systems. By training local models at each location in parallel and then performing federated averaging of the model weights at the central server over multiple FL rounds, we are able to train a federated model which is generalized to location-independent features. We show that we can leverage this federated model to reduce the number of training repetitions required per physical action when training at new locations. We conclude that this reduction of training repetitions allows for more rapid deployment of WiFi sensing devices into new locations without increasing the complexity or workload for the technician installing any new WiFi hardware for the system. Additionally, in cases where hardware such as WiFi access points are already deployed throughout a building, the technician does not need to enter each location to perform a large number of time-consuming training repetitions of the actions before seeing useful prediction accuracy. We also find that using WiFederated can reduce the number of local training epochs required compared to other methods. Finally, we evaluate training and evaluation at the edge and consider possible methods for continuous annotation to ensure that clients are able to continue to capture accurate annotation labels over time to further train the federated model.

References

[1] Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, and Y. Guo, "A survey on human behavior recognition using channel state information," *IEEE Access*, vol. 7, pp. 155 986–156 024, 2019.

[2] Y. He, Y. Chen, Y. Hu, and B. Zeng, "Wifi vision: Sensing, recognition, and detection with commodity mimo-ofdm wifi," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8296–8317, 2020.

[3] S. M. Hernandez and E. Bulut, "Lightweight and standalone IoT based WiFi sensing for active repositioning and mobility," in *21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, Cork, Ireland, Jun. 2020.
[4] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Analyzing shopper's behavior through wifi signals," in *Proceedings of the 2nd workshop on Workshop on Physical Analytics*, 2015, pp. 13–18.
[5] S. Depatla and Y. Mostofi, "Passive Crowd Speed Estimation in Adjacent Regions With Minimal WiFi Sensing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 10, pp. 2429–2444, 2019.
[6] C. Tang, W. Li, S. Vishwakarma, K. Chetty, S. Julier, and K. Woodbridge, "Occupancy Detection and People Counting Using WiFi Passive Radar," in *IEEE Radar Conference (RadarConf20)*, 2020, pp. 1–6.
[7] E. Ding, X. Li, T. Zhao, L. Zhang, and Y. Hu, "A robust passive intrusion detection system with commodity wifi devices," *J. Sensors*, pp. 1–12, 2018.
[8] B. Yu, Y. Wang, K. Niu, Y. Zeng, T. Gu, L. Wang, C. Guan, and D. Zhang, "Wifi-sleep: Sleep stage monitoring using commodity wi-fi devices," *IEEE Internet of Things Journal*, 2021.
[9] B. Tan, Q. Chen, K. Chetty, K. Woodbridge, W. Li, and R. Piechocki, "Exploiting wifi channel state information for residential healthcare informatics," *IEEE Communications Magazine*, vol. 56, no. 5, pp. 130–137, 2018.
[10] S. Liu, Y. Zhao, F. Xue, B. Chen, and X. Chen, "Deepcount: Crowd counting with wifi via deep learning," *arXiv preprint arXiv:1903.05316*, 2019.
[11] F. Wang, W. Gong, J. Liu, and K. Wu, "Channel selective activity recognition with wifi: A deep learning approach exploring wideband information," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 181–192, 2020.
[12] F. Wang, S. Panev, Z. Dai, J. Han, and D. Huang, "Can WiFi estimate person pose?" *arXiv preprint arXiv:1904.00277*, 2019.
[13] H. Zou, Y. Zhou, J. Yang, and C. J. Spanos, "Towards occupant activity driven smart buildings via WiFi-enabled IoT devices and deep learning," *Energy and Buildings*, vol. 177, pp. 12–22, 2018.
[14] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "CrossSense: Towards Cross-Site and Large-Scale WiFi Sensing," in *Proc. of the 24th Annual International Conf. on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: ACM, 2018, pp. 305–320.
[15] Y. Ma, S. Arshad, S. Muniraju, E. Torkildson, E. Rantala, K. Doppler, and G. Zhou, "Location-and person-independent activity recognition with wifi, deep neural networks, and reinforcement learning," *ACM Transactions on Internet of Things*, vol. 2, no. 1, pp. 1–25, 2021.
[16] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas *et al.*, "Towards environment independent device free human activity recognition," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 2018, pp. 289–304.
[17] Y. Lu, S. Lv, and X. Wang, "Towards Location Independent Gesture Recognition with Commodity WiFi Devices," *Electronics*, vol. 8, no. 10, p. 1069, 2019.
[18] Y. Fang, B. Sheng, H. Wang, and F. Xiao, "WiTransfer: A Cross-scene Transfer Activity Recognition System Using WiFi," in *Proceedings of the ACM Turing Celebration Conference-China*, 2020, pp. 59–63.
[19] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.
[20] S. Arshad, C. Feng, R. Yu, and Y. Liu, "Leveraging transfer learning in multiple human activity recognition using wifi signal," in *IEEE 20th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, 2019, pp. 1–10.
[21] L. Guo, Z. Lu, X. Wen, S. Zhou, and Z. Han, "From signal to image: Capturing fine-grained human poses with commodity wi-fi," *IEEE Communications Letters*, vol. 24, no. 4, pp. 802–806, 2019.
[22] L. Feng, Z. Li, C. Liu, X. Chen, X. Yin, and D. Fang, "SitR: Sitting Posture Recognition Using RF Signals," *IEEE Internet Things Journal*, vol. 7, no. 12, pp. 11 492–11 504, 2020.
[23] X. Wang, C. Yang, and S. Mao, "PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices," in *IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 1230–1239.
[24] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM CCR*, vol. 41, no. 1, p. 53, Jan. 2011.
[25] Z. Shi, J. A. Zhang, R. Y. Xu, and Q. Cheng, "WiFi-Based Activity Recognition using Activity Filter and Enhanced Correlation with Deep Learning," in *IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2020, pp. 1–6.
[26] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," 2020.
[27] A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, "Federated learning for mobile keyboard prediction," *arXiv preprint arXiv:1811.03604*, 2018.
[28] L. Huang, A. L. Shea, H. Qian, A. Masurkar, H. Deng, and D. Liu, "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records," *Journal of biomedical informatics*, vol. 99, p. 103291, 2019.
[29] H. Jiang, C. Cai, X. Ma, Y. Yang, and J. Liu, "Smart home based on WiFi sensing: A survey," *IEEE Access*, vol. 6, pp. 13 317–13 325, 2018.
[30] H. Zou, J. Yang, H. P. Das, H. Liu, Y. Zhou, and C. J. Spanos, "WiFi and Vision Multimodal Learning for Accurate and Robust Device-Free Human Activity Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Long Beach, CA, USA, June 16-20*, 2019, pp. 426–433.
[31] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 5452–5461.
[32] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
[33] M. Yurochkin, M. Agarwal, S. Ghosh, K. Greenewald, T. N. Hoang, and Y. Khazaeni, "Bayesian nonparametric federated learning of neural networks," 2019.
[34] H. Wang, M. Yurochkin, Y. Sun, D. Papailiopoulos, and Y. Khazaeni, "Federated learning with matched averaging," in *International Conference on Learning Representations*, 2020.
[35] Y. J. Cho, J. Wang, and G. Joshi, "Client selection in federated learning: Convergence analysis and power-of-choice selection strategies," *arXiv preprint arXiv:2010.01243*, 2020.
[36] S. Dhar, J. Guo, J. Liu, S. Tripathi, U. Kurup, and M. Shah, "On-device machine learning: An algorithms and learning theory perspective," *arXiv preprint arXiv:1911.00623*, 2019.
[37] F. Liang, W. Yu, X. Liu, D. Griffith, and N. Golmie, "Toward edge-based deep learning in industrial internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4329–4341, 2020.
[38] J.-C. Bos, "MicroMLP," 2018. [Online]. Available: https://github.com/jczic/MicroMLP.
[39] X. Wang, C. Yang, and S. Mao, "On csi-based vital sign monitoring using commodity wifi," *ACM Transactions on Computing for Healthcare*, vol. 1, no. 3, pp. 1–27, 2020.
[40] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM computing surveys (CSUR)*, vol. 46, no. 4, pp. 1–37, 2014.

Steven M. Hernandez received a B.S. degree in Computer Science from Virginia Commonwealth University in 2018. He is now completing his Ph.D. in the Computer Science Department of Virginia Commonwealth University with funding through the National Science Foundation Graduate Research Fellowship under the supervision of Dr. Eyuphan Bulut. His research interests include WiFi sensing, Device-to-Device (D2D) communications, machine learning for edge devices. He is a member of IEEE.

Eyuphan Bulut (SM'20) received the Ph.D. degree in the Computer Science department of Rensselaer Polytechnic Institute (RPI), Troy, NY, in 2011. He then worked as a senior engineer in Mobile Internet Technology Group (MITG) group of Cisco Systems in Richardson, TX for 4.5 years. He is now an Associate Professor with the Department of Computer Science, Virginia Commonwealth University (VCU), Richmond, VA. His research interests include mobile and wireless computing, network security and privacy, mobile social networks and crowd-sensing. Dr. Bulut is an Associate Editor in IEEE Access and has been serving in the organizing committee of several conferences. He is also a member of ACM.