Communication Efficient Privacy-Preserving Federated Learning via Knowledge Distillation for Human Activity Recognition Systems

Gad Gad [1], zubair fadlullah [2], Khaled Rabie [2], and Mostafa M. Fouda [2]

[1]Lakehead university
[2]Affiliation not available

June 16, 2023

Abstract

Emerging Internet of Things (IoT) applications, such as sensor-based Human Activity Recognition (HAR) systems, require efficient machine learning solutions due to their resource- constrained nature which raises the need to design heterogeneous model architectures. Federated Learning (FL) has been used to train distributed deep learning models. However, standard federated learning (fedAvg) does not allow the training of het-erogeneous models. Our work addresses the model and statistical heterogeneities of distributed HAR systems. We propose a Federated Learning via Augmented Knowledge Distillation (FedAKD) algorithm for heterogeneous HAR systems and evaluate it on a self-collected sensor-based HAR dataset. Then, Kullback-Leibler (KL) divergence loss is compared with Mean Squared Error (MSE) loss for the Knowledge Distillation (KD) mechanism. Our experiments demonstrate that MSE contributes to a better KD loss than KL. Experiments show that FedAKD is communication-efficient compared with model-dependent FL algorithms and outperforms other KD-based FL methods under the i.i.d. and non-i.i.d. scenarios.

Communication-Efficient Privacy-Preserving Federated Learning via Knowledge Distillation for Human Activity Recognition Systems

Gad Gad*[1], Zubair Md Fadlullah‡[2], Khaled Rabie§[3], and Mostafa M. Fouda¶[4].
*Department of Computer Science, Lakehead University, Thunder Bay, Ontario, Canada.
‡Department of Computer Science, Western University, London, ON, Canada.
§Department of Engineering, Manchester Metropolitan University, Manchester, UK.
¶Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA.
Emails: [1]ggad@lakeheadu.ca, [2]zfadlullah@ieee.org, [3]k.rabie@mmu.ac.uk, [4]mfouda@ieee.org

***Abstract*—Emerging Internet of Things (IoT) applications, such as sensor-based Human Activity Recognition (HAR) systems, require efficient machine learning solutions due to their resource-constrained nature which raises the need to design heterogeneous model architectures. Federated Learning (FL) has been used to train distributed deep learning models. However, standard federated learning (fedAvg) does not allow the training of heterogeneous models.Our work addresses the model and statistical heterogeneities of distributed HAR systems. We propose a Federated Learning via Augmented Knowledge Distillation (FedAKD) algorithm for heterogeneous HAR systems and evaluate it on a self-collected sensor-based HAR dataset. Then, Kullback-Leibler (KL) divergence loss is compared with Mean Squared Error (MSE) loss for the Knowledge Distillation (KD) mechanism. Our experiments demonstrate that MSE contributes to a better KD loss than KL. Experiments show that FedAKD is communication-efficient compared with model-dependent FL algorithms and outperforms other KD-based FL methods under the i.i.d. and non-i.i.d. scenarios.**

***Index Terms*—Deep Learning, Federated Learning, Knowledge Distillation, Human Activity Recognition (HAR), Kullback-Leibler divergence, privacy-preserving AI.**

I. Introduction

Human Activity Recognition (HAR) is one of the promising applications in fields like healthcare [1], [2] and surveillance [3] with the Internet of Things (IoT) [4]. HAR systems can be divided into vision-based or sensor-based data depending on the modality of the acquired data. While the former relies on a person's images or video streams to recognize the activity a person is performing, the latter provides a cheaper and more practical solution by incorporating low-power low-frequency cheap sensors. On one hand, these sensors provide less information than their vision-based counterparts. Still, on the other hand, wearable devices' sensors are 1) placed in close proximity to the skin to maximize the accuracy of the sensors' measurements, 2) since wearable devices such as smartwatches and fitness bands are worn most of the time, these low-power sensors get the chance to collect data over longer periods of time compared with sensing devices in vision-based HAR systems. The suite of sensors that wearable devices offer is getting larger in number, smaller in size, and less in cost and power consumption. Some of these sensors detect capture motion and gestures (e.g. accelerometer, gyroscope, and magnetometer), and some sensors detect physiological signals like heart rate which is usually detected using a Photoplethysmography sensor (an optical sensor) [5], and oxygen saturation in the blood (SpO2) which is detected using a blood oxygen saturation sensor.

Machine learning methods were applied to HAR systems [6] by training methods such as Support Vector Machine (SVM) [7], K-Nearest Neighbor (KNN) [8], [9], etc. on supervised datasets to classify activities. These methods produce lightweight models with acceptable performance. On the other hand, Deep Learning (DL) presents itself as a powerful feature extraction tool that learns directly from raw data and is able to differentiate the unique patterns of different activities [10]. Relative to ML, DL requires more computation and memory resources for training and running the model at inference time. While privacy preservation [11], [12] has been a key focus area of various researchers dealing with resource-constrained IoT systems, distributed learning with local deep learning models has not been systematically investigated in HAR applications.

Federated Learning (FL) [13], [14] provides a collaborative distributed learning framework to train deep learning models. Most FL methods rely on gradients/weights sharing and thus assume that participating models have the same neural network architecture. However, in some applications, clients/devices who participate in a federated learning process need to design their local model independently. This is referred to as model heterogeneity which arises in IoT scenarios due to the different resource limitations of each device and in the health care and supply chain due to privacy or intellectual property concerns.

In this paper, we propose Federated Learning with Augmented Knowledge Distillation (FedAKD) to train deep learning-based Human Activity Recognition systems with heterogeneous architectures. To train the proposed FL system, sensor-based data is collected from a commercial fitness band (Mi band 4) and annotated with human activities.

Our proposed system is more flexible than FedAvg [14]

since it allows the training of independently designed models. Furthermore, FedAKD can control its communication overhead by setting the size of the public dataset which is used for Knowledge Distillation. On the other hand, FedAvg's communication bandwidth depends on the participating models' size. Compared with other KD-based federated learning methods, FedAKD archives comparatively and outperforms existing methods in the non-i.i.d case, where a model is tested on activities that do not exist in its local dataset but are found in other clients' datasets. Finally, the Knowledge Distillation mechanism is further investigated by comparing two losses: KUllback-Leibler (KL) divergence loss, and Mean Square Error (MSE) loss. MSE loss is found to achieve better logit matching between teacher and student scores.

The remainder of this paper is organized as follows. The relevant research work are surveyed in section II. Then, our proposed methodology is presented in section III. The performance evaluation of our proposal is reported in section IV. Next, section V discusses this work's limitations and open research issues. Finally, the paper is concluded in section VI.

II. Related work

In this section, we overview the recent research work. The integration of Machine Learning (ML) and the Internet of Things (IoT) offers predictive capability at the edge, therefore enabling data collected at the edge to be utilized without sacrificing privacy or the overall cost of ML-IoT solutions. In the context of Human Activity Recognition (HAR), Machine learning (ML) models found their way to wearable devices to recognize movements due to their lightweight and good performance.

Researchers in [9] proposed a lightweight KNN-based HAR system with a reduced kernel to reduce the heavy computation of on large dataset. Since traditional methods suffer from a high positive rate, the research work in [15] employs a single-class SVM to distinguish abnormal activities from normal ones, the authors then use Kernel NonLinear Regression (KNLR) to reduce the false positive rate.

Deep Learning (DL) methods were also applied to HAR due to their superior expressive power relative to ML. In particular, DL models are able to extract complex features without manual feature engineering, which is often required in ML.

The work in [10] presented an indoor HAR system based on a hybrid deep learning model that integrates CNN, to extract spatial features, and LSTM to learn temporal patterns. On the other hand, an ensemble of Deep Stacked MultiLayered Perceptron (DS_MLP) was employed in [16] to classify human activity with high accuracy. In this approach, the output of five base models was leveraged to train a meta-learner model.

HAR systems are deployed on multiple devices, each of which collects data that might have different distributions or even activities. For example, one fitness band user A may use his band to record sensory data while walking, this data is used to train a local model to detect when user A is walking. When user A walks, the model is able to recognize this activity. Moreover, users can use online learning to automate new data labeling resulting in better performance. Now consider another user B who did the same as user A but for the activity $Sleeping$. That is, he collected, annotated, and trained his local model on sensor data mapped to the activity $Sleeping$. Each user has a model trained with a limited local dataset.

Federated Learning enabled distributed deep-learning-based HAR systems to train collaboratively without sharing local data. This is usually done by sharing gradients. However, if users A and B independently design their models, a model-agnostic communication medium is needed to transfer knowledge (about an activity) from one model to the other models. Knowledge Distillation (KD) [17] provides a communication protocol in which a shared dataset is employed to transfer knowledge.

ClusterFL was proposed by researchers in [18] to reduce the communication overhead of FL for HAR systems by clustering nodes based on the intrinsic relationships among local datasets. Next, the work in [19] presented a resource-constrained federated learning HAR system that addresses label and model heterogeneities that leverages transfer learning.

Our approach is similar to FedMD [20] in that both use KD. We introduce the following modifications, we propose $FedAKD$ which uses augmentation to improve accuracy. We also use weighted aggregation proportional to clients' performance, unlike the work in [20] which employs equal aggregation. Finally, we present two variants of $FedAKD$. $FedAKD_{KL}$ and $FedAKD_{MSE}$ to compare different losses for the KD mechanism.

By sharing soft labels instead of gradients, clients not only have the freedom to build unique model architectures that meet changing resource requirements on heterogeneous edge devices, but it also enables clients to control their communication bandwidth, which is itself a precious resource, especially for IoT devices that are not internet-enabled.

III. Proposed Methodology

In this section, we present our proposed methodology. First, we describe the tools and procedure involved in experimental data collection. Next, we delineate the heterogeneous model architectures used for centralized learning for the HAR system. Then, we develop the idea behind our proposed distributed solution, i.e., FL via Augmented Knowledge Distillation (FedAKD) algorithm.

A. Data Collection: Tools and procedure

We performed data collection using a commercial fitness band: Mi band 4. We used an open-source library [1] to communicate with the band over Bluetooth Low Energy (BLE) using python. A number of student volunteers were invited to participate in data collection where each volunteer is given a Mi band 4, a raspberry pi 3B (RPI), and a 5000mAh battery. To start collecting data for an activity, the volunteer would power on the RPI while wearing the band and perform the activity. Once the RPI is powered on, a custom Linux service executes on-boot, and the RPI searches for nearby Mi band

[1] https://github.com/satcar77/miband4

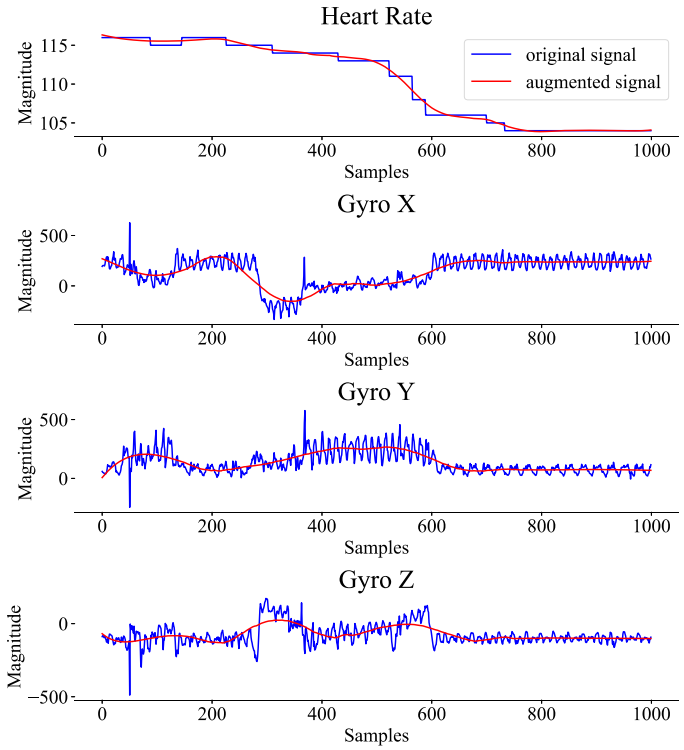

Fig. 1: Plots show the four collected input signals to predict human activities. From top to bottom, the signals shown are Heart rate, Gyroscope X, Gyroscope Y, and Gyroscope Z. The blue signal shows the raw signal. The red signal represents the Sav-Gol filter applied to the raw signal as a low-pass filter.

4 devices that have a given MAC address and authentication key. After the RPI connects to a Mi band 4, both devices use BLE to communicate the IMU and PPG sensor readings. It is worth mentioning that the IMU data retrieved is only the Gyroscope readings. Also, PPG sensor data is processed by the band, and sends the calculated heart rate to the RPI. The data collection process is called a data collection session. It starts by powering on the RPI which is responsible for connecting to the wearable device, requesting sensor data using BLE, and storing the received data in a format that is easy to read and process. A data collection session is terminated when the volunteer shuts down the RPI or when it terminates automatically after a defined amount of time.

During a data collection session, four data streams are collected by the RPI: Heart rate, Gyroscope X, Gyroscope Y, and Gyroscope Z. Figure 1 shows a 2000 data points sample of each of the four signals. The figure also shows a low-frequency version (red) produced by applying a Sav-GOl filter to the raw signals (blue) of the signal. We can observe that Gyroscope signals fluctuate more aggressively (have a relatively bigger variance) and therefore are expected to work as strong indicators to extract features relevant to recognizing motion and activities. In addition to data collection, volunteers are asked to manually label each data collection session by renaming the file that was generated (initially named with the timestamp) to the name of the activity. These activities are $Walking$, $Studying$, and $Sleeping$. We also add another class: $Idle$ which essentially represents an unworn fitness band. That is, The data for the fourth class $Idle$ was collected while the band was placed on a table.

We found that the sampling rate of Mi band 4's IMU is between 10-15 Hz. Whereas, the sampling rate of the heart rate pulses is 2-3 Hz only. As mentioned earlier, the heart rate pulses we retrieve from the band are calculated by the band based on the PPG sensor, which justifies the lower frequency (as the band takes time to calculate the heart rate from PPG).

The structure of the data directory and the format in which data is stored is described next.

The dataset directory is split into folders, one for each volunteer, named $Person1, Person2, ...PersonN$ for N volunteers. The data collected by a volunteer are stored in his designated folder, a file for each data collection session. During a data collection session, the RPI labels files by the timestamp at which the data collection session started. The volunteer then has to keep track of which label corresponds to each file he collects, and manually renames that file with the activity. For example, a volunteer that is collecting $Walking$ data performs the data collection session as described, a file is created once data collection starts and data is appended to that file throughout the data collection session. After that, the volunteer manually renames that file to $WalkingXX.txt$ where XX is the number of the file.

For each text file in the dataset containing the data of a single activity, data is appended line by line. Due to frequency inconsistency (each of the two sensors has a different sampling rate), each line will either contain IMU or heart rate values.

B. Heterogeneous model architectures

The main focus of this work is to perform federated learning on heterogeneous Human Activity Recognition (HAR) systems. These HAR systems are practically deep learning models. In the previous subsection, we introduced the collection and annotation of a wrist-based sensor-based HAR dataset. In this subsection, we construct heterogeneous deep learning model architectures that are designed to take time-series data as inputs and produce a probability distribution of the target human activities. Finally, in the next subsection, we propose Federated Learning via Augmented Knowledge Distillation (FedAKD) to train these heterogeneous deep learning models on the collected dataset.

We build a custom deep learning model that consists of a stack of 1D-convolution layers, Long Short-Term Memory (LSTM) units, and non-linear activation functions of various types. Then, some hyperparameters of these layers are tuned to produce ten different versions of the initial template model we started with. One version, for example, would use $ReLU$ as an activation function, set the number of filters in the 1D-convolution layer to 10, and set the drop rate of the dropout layer to 0.2. Additionally, we also include the optimizer and the learning rate value among the hyperparameters to be tuned.

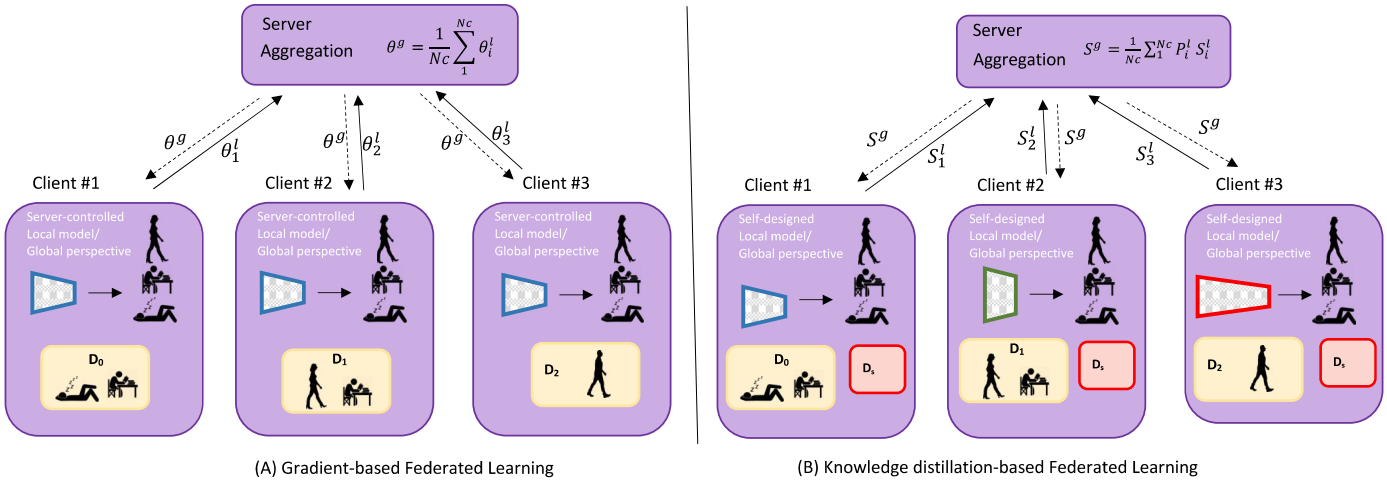

Fig. 2: Local models use federated learning to gain a global perspective of a distributed dataset, allowing a client's model to predict activities not found in his local data. However, this approach assumes all clients have the same architecture. Knowledge distillation is a technique that relaxes this assumption. Instead of having clients send their gradients to the server, KD sends soft scores calculated on a shared dataset. Because gradients are model-dependent while soft scores are not, KD-based FL is able to train models of heterogeneous architectures. Here, D_i is the local dataset of client i and D_s is the shared public set.

Each of these custom Neural Networks (NN) has a different learning capacity, and our goal is to verify that a knowledge distillation-based federated learning scheme can push the performance of these NNs beyond local effort. For that, we first create a smaller version of the collected dataset that limits the number of samples for each activity. This is done to give each model a local dataset that smaller than the full dataset in order to force each model to learn from the other models in a federated learning setting.

C. Proposed Federated Learning via Augmented Knowledge Distillation (FedAKD)

An overview of the proposed Federated Learning via Augmented Knowledge Distillation (FedAKD) and how it compares to the standard federated learning algorithm (FedAvg) is shown in figure 2. In the standard FL algorithm [14], clients share model weights and receive aggregated weights. Clients in the proposed algorithm share soft labels on a shared dataset, and train on the aggregated consensus soft labels using knowledge distillation loss. In the FL setting, we have $\mathbf{K} = K_1, K_2, ..., K_{N_k}$ distributed devices collaborating to learn some task, where N_k is the number of clients, and K_i is the i-th client. K_i possess a local dataset D_i, and a server-designed model f_i. Let $\mathbf{D} = \{D_1, D_2, ..., D_{N_k}\}$ be a dataset that combines all local datasets. The task of FL is to collaboratively train f_i to approach the would-be-performance if f_i were to be trained on $\mathbf{D}$ without explicitly sharing the local dataset D_i for all participating clients. In FedSGD, client K_i would calculate the gradient g_i for a set of data points B_i as

$$g_i = \Delta \frac{1}{N_i} \sum_{j \in B_i} L_{CCE}(x_j, y_j, f^r), \tag{1}$$

where N_i is the number of data points on client K_i, and $L_{CCE}(x_j, y_j, f^r)$ is the cross-entropy loss applied to the data point pair (x_i, y_i) on the model f^r which is the global model of the r-th round. Gradients calculated by each client are sent to the server and the global model's weights are updated by

$$f^{r+1} = f^r - \eta \sum_{i=1}^{N_k} \frac{|D_i|}{|\mathbf{D}|} g_i. \tag{2}$$

The global model f^{r+1} is then broadcasted to clients. The problem with this approach is that the server assumes control over the design of f_i. Knowledge Distillation (KD) is a technique that employs a trained model to train a to-be-trained model. KD can be implemented in different ways. We use an approach similar to FedMD [20], where clients have a public/shared dataset D_s.

Clients first calculate soft labels on the shared dataset D_s as:

$$S_i^r = \mathbf{f}_i(D_{Aug}^r), \tag{3}$$

with the server, which aggregates them and broadcasts consensus soft labels back to clients.

$$S^r = \sum_{i=1}^{N_k} \frac{P_i^r S_i^r}{\sum_{k=0}^{N_k} P_k^r} \tag{4}$$

Aggregated soft labels are then used by clients to train on D_{Aug}^r. Here, S_i^r denotes the soft scores of the i-th client on round r, $\mathbf{f}_i$ produces logits (without applying softmax), and D_{Aug}^r is produced by applying augmentation using [21] on D_s. D_{Aug}^r has the round superscript because each new round uses augmentation. Our hypothesis is that this approach produces different variants of D_s which encourages the soft labels S_i^r to change and distill more knowledge. We weigh clients' contribution to the $teacherscores$ proportional to their performance

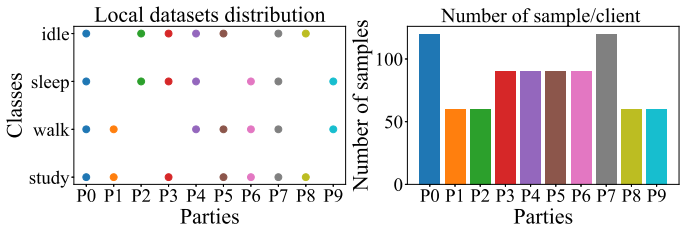

Fig. 3: Class-level local data distribution (left) and number of samples per client (right) both scenarios are from the non-i.i.d case.

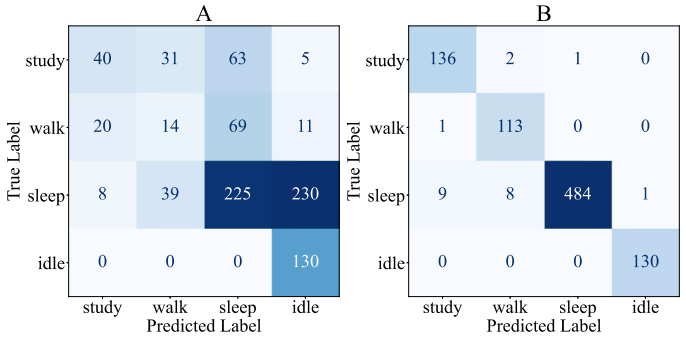

Fig. 4: The confusion matrix of locally trained (A) and FL-trained under non-i.i.d scenario (B) models.

on their respective local datasets P_i^r. Other weighting schemes include weighting contributions proportional to the size of the local datasets. We compare our method with [20] which uses uniform (equal) weighting.

A factor $\alpha^r \in (0, 1)$ is set by the server each round r to calculate D_{Aug}^r as

$$D_{Aug}^r = \alpha^r D_z^r + (1 - \alpha^r) D_s \tag{5}$$

where D_z^r is a permuted version of D_s by another factor β^r also set by the server.

IV. Results and Discussion

In this section, we evaluate the performance of the proposal. First, we compare Knowledge Distillation (KD) with local training. We also compare the communication overhead of FedMD [20], FedAvg [14], and the proposed algorithm, FedAKD. Finally, we compare the average accuracy gains of our proposed solution with benchmark approaches.

A. Knowledge Distillation vs Local Training

The data distribution of each party in the federated learning experiment statistical scenarios considered is shown in figure 3. The model which belongs to $P1$ gets, under the non-i.i.d scenario, 60 samples. These 60 samples are distributed as 30 samples over 2 classes, which means that the model trained on this local data has only two classes out of four. Figure 4 shows the confusion matrices of that particular model (the model which belongs to $P1$) when that model is trained on his local dataset (which lacks the data of two activities) vs when

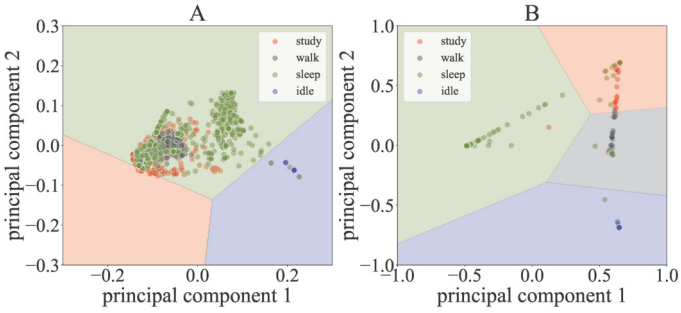

Fig. 5: The principle components of the output of the last dense layer of (A) a locally trained model, and (B) an FL-trained model. Colors refer to ground truth labels. Background refers to classes recognized by a Logistic Regression model on the PCA-reduced 2D samples.

TABLE I: Communication overhead between model-based and model-agnostic FL methods. The proposed algorithm (FedAKD) communicates the soft labels, which are denoted as Z or S for the soft labels calculated by FedMD and FedAKD, respectively. On the other hand, FedAvg communicates model weights θ.

Federated learning algorithm	Communicated entities	Size
FedMD	$\|Z\|$	10 KB
FedAKD	$\|S\|$	10 KB
FedAvg	$\|\theta\|$	250 KB

this model participates in a federated learning process, on the left and right sides, respectively. The confusion matrix of the FL-trained model shows an improvement in predicting data of both missing classes: $Walk$ and $Sleep$. This is attributed to distilling knowledge from other clients who have that class in their local dataset. Figure 6 also presents a detailed performance comparison between locally trained (right) and collaboratively trained models (left) with FedKD. The Area Under Curve (AUC) of classes like $Sleep$ and $Walk$ increased when the model is trained with Knowledge Distillation-based federated learning compared with local training.

B. Communication efficiency of model-based vs model-agnostic federated learning methods

Table I demonstrates the communication cost of each of the considered federated learning algorithms: FedMD [20], FedAvg [14], and the proposed algorithm, FedAKD. Both FedMD and FedAKD rely on soft labels to distill knowledge from clients to the server and then to the clients. On the other hand, FedAvg sends model weights, denoted by θ. The size of soft labels depends on the size of the public dataset on which the soft labels are calculated and the number of labels, which is usually much smaller than the size of the deep learning model. In our case, the average size of the used models is 250 Kilo Bytes (KB) which is 10 times the size of the soft labels.

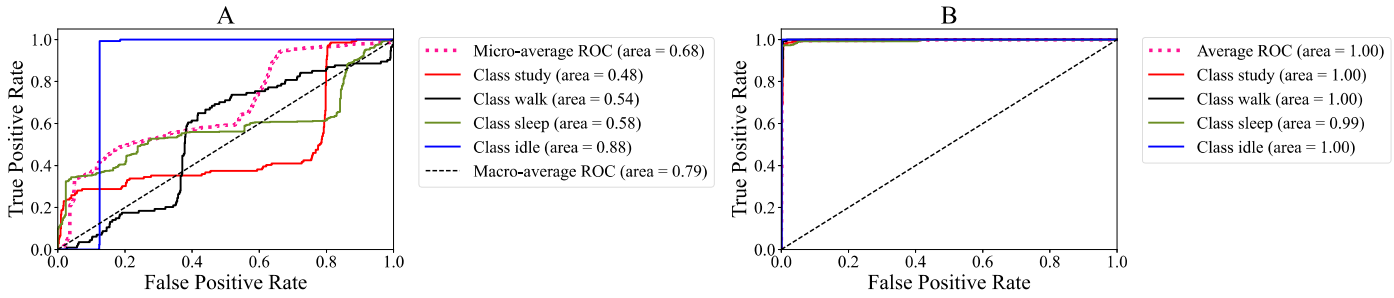

Fig. 6: A: ROC curve of a locally trained model. B: ROC curve of a model trained with knowledge-distillation-based federated learning. The locally trained model data for two labels: $Walk$ and $Sleep$, therefore the Area Under Curve (AUC) of both activities is relatively low. KD increased activity recognition accuracy for all classes.

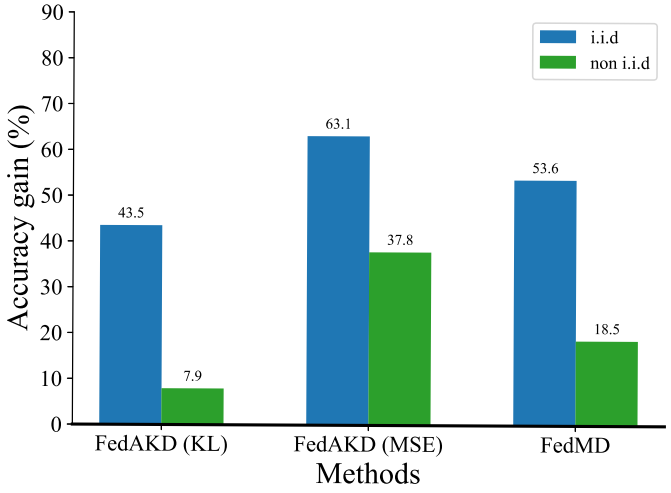

Fig. 7: Comparative accuracy of the various federated learning frameworks under i.i.d. and non-i.i.d. scenarios.

C. Comparing $FedAKD$ vs $FedMD$ using different loss functions

Figure 7 compares the average accuracy gains achieved by our proposed $FedAKD$ methods: $FedAKD_{KL}$ and $FedAKD_{MSE}$ vs Federated Learning via Model Distillation ($FedMD$) [20]. It can be observed that using MSE as the loss function for the Knowledge Distillation (KD) mechanism produced the highest average accuracy gains compared to using Kullback-Leibler divergence loss or $FedMD$. While $FedMD$ also uses MSE for KD, our method $FedAKD_{MSE}$ applies mixup augmentation [21] to the shared dataset D_s to increase the variance of the shared soft labels V resulting in higher accuracy than $FedMD$.

V. Limitations and Future Work

There are two limitations of our adopted approach which need to be highlighted regarding the knowledge distillation mechanism:

1) The distribution of the public dataset. In our experiments, we considered a public dataset D_p with a similar distribution to the local datasets. This is important for efficient knowledge distillation as noted by [20] who chose D_p as the CIFAR 100 dataset when training on the CIFAR 10 dataset. This is a clear limitation of the knowledge distillation technique because not all datasets have a companion dataset with a similar distribution. Also, more work can be done on the efficiency of knowledge distillation using proxy datasets (like D_p) with distribution different from that of the target dataset.
2) While FedAKD is a model-agnostic federated learning algorithm, more work can be done regarding the impact of changing the model architecture of clients on knowledge distillation efficiency and overall performance.

VI. Conclusion

In this work, we proposed a Knowledge Distillation (KD)-based Federated Learning (FL) framework called FedAKD. Unlike FedAvg, FedAKD allows participant parties to have control over the design of their neural network architectures. Different experiments were conducted to show the utility and communication efficiency of the FedAKD. Mean Square Error (MSE) and Kullback Leibler (KL) losses were also tested as the KD loss.

Our conducted experiments considered both statistical and model heterogeneity; ten custom deep Neural Networks (NNs) with heterogeneous model architectures were constructed and collaboratively trained under i.i.d and non-i.i.d conditions. The proposed algorithm performed comparably to other KD-based FL frameworks. Compared to standard federated learning (FedAvg), our proposed FL algorithm is more communication efficient and allows clients to have control over the design of their local model.

We demonstrated that a device participating in KD-based FL gains, on average 40 percentage points, over its local performance. A client is able to distill knowledge about an activity that does not exist in his dataset. MSE loss was found to be a better choice than KL loss for KD, which is also reported by [22]. The proposed solution (FedAKD) can be used to train heterogeneous distributed HAR systems to push their performance beyond their local effort.

ACKNOWLEDGMENT

This research was partly supported by a New Frontiers in Research Fund (NFRF) Explore grant: 11-50-16110115.

DATA AND CODE AVAILABILITY

Code for analysis in the work can be accessed at https://github.com/gadm21/Knowledge-distillation-based-federated-learning (accessed on Jan 22, 2023).

REFERENCES

[1] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Systems with Applications*, vol. 137, pp. 167–190, 2019.

[2] G. Ogbuabor and R. La, "Human activity recognition for healthcare using smartphones," in *Proceedings of the 2018 10th international conference on machine learning and computing*, 2018, pp. 41–46.

[3] W. Lin, M.-T. Sun, R. Poovandran, and Z. Zhang, "Human activity recognition for video surveillance," in *2008 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2008, pp. 2737–2740.

[4] R. Hussain, D. Kim, J. Son, J. Lee, C. A. Kerrache, A. Benslimane, and H. Oh, "Secure and privacy-aware incentives-based witness service in social internet of vehicles clouds," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2441–2448, 2018.

[5] S. Das, S. Pal, and M. Mitra, "Real time heart rate detection from PPG signal in noisy environment," in *2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI)*. IEEE, 2016, pp. 70–73.

[6] R. Bhattarai and V. Bohara, "A review on human activity recognition techniques and comparative performance analysis," *World Journal of Innovative Research (WJIR)*, vol. 12, no. 1, pp. 37–39, 2022.

[7] C. Tang, A. Tong, A. Zheng, H. Peng, and W. Li, "Using a selective ensemble support vector machine to fuse multimodal features for human action recognition," *Computational Intelligence and Neuroscience*, vol. 2022, article no. 1877464, 2022.

[8] A. Al-Taei, M. F. Ibrahim, and N. J. Habeeb, "Optimizing the performance of KNN classifier for human activity recognition," in *International Conference on Advances in Computing and Data Sciences*, 2021, pp. 373–385.

[9] Z. Liu, S. Li, J. Hao, J. Hu, and M. Pan, "An efficient and fast model reduced kernel KNN for human activity recognition," *Journal of Advanced Transportation*, vol. 2021, article no. 2026895, 2021.

[10] I. U. Khan, S. Afzal, and J. W. Lee, "Human activity recognition via hybrid deep learning based model," *Sensors*, vol. 22, no. 1, article no. 323, 2022.

[11] C. Boudagdigue, A. Benslimane, A. Kobbane, and J. Liu, "Trust management in industrial internet of things," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3667–3682, 2020.

[12] H. Benaddi, K. Ibrahimi, A. Benslimane, M. Jouhari, and J. Qadir, "Robust enhancement of intrusion detection systems using deep reinforcement learning and stochastic game," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 11 089–11 102, 2022.

[13] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, article no. 106854, 2020.

[14] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

[15] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[16] F. Rustam, A. A. Reshi, I. Ashraf, A. Mehmood, S. Ullah, D. M. Khan, and G. S. Choi, "Sensor-based human activity recognition using deep stacked multilayered perceptron model," *IEEE Access*, vol. 8, pp. 218 898–218 910, 2020.

[17] G. Hinton, O. Vinyals, J. Dean *et al.*, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, vol. 2, no. 7, 2015.

[18] X. Ouyang, Z. Xie, J. Zhou, J. Huang, and G. Xing, "ClusterFL: a similarity-aware federated learning system for human activity recognition," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021, pp. 54–66.

[19] G. K. Gudur and S. K. Perepu, "Resource-constrained federated learning with heterogeneous labels and models for human activity recognition," in *International Workshop on Deep Learning for Human Activity Recognition*. Springer, 2021, pp. 57–69.

[20] D. Li and J. Wang, "FedMD: Heterogenous federated learning via model distillation," *arXiv preprint arXiv:1910.03581*, 2019.

[21] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[22] T. Kim, J. Oh, N. Kim, S. Cho, and S.-Y. Yun, "Comparing kullback-leibler divergence and mean squared error loss in knowledge distillation," *arXiv preprint arXiv:2105.08919*, 2021.