

# Resource-Constrained Federated Learning with Heterogeneous Labels and Models for Human Activity Recognition

Gautham Krishna Gudur<sup>1</sup> and Satheesh Kumar Perepu<sup>2</sup>

<sup>1</sup> Global AI Accelerator, Ericsson  
gautham.krishna.gudur@ericsson.com

<sup>2</sup> Ericsson Research  
perepu.satheesh.kumar@ericsson.com

**Abstract.** One of the most significant applications in pervasive computing for modeling user behavior is Human Activity Recognition (HAR). Such applications necessitate us to characterize insights from multiple resource-constrained user devices using machine learning techniques for effective personalized activity monitoring. On-device Federated Learning proves to be an extremely viable option for distributed and collaborative machine learning in such scenarios, and is an active area of research. However, there are a variety of challenges in addressing statistical (non-IID data) and model heterogeneities across users. In addition, in this paper, we explore a new challenge of interest – to handle *heterogeneities in labels (activities)* across users during federated learning. To this end, we propose a framework with two different versions for federated label-based aggregation, which leverage overlapping information gain across activities – one using *Model Distillation Update*, and the other using *Weighted  $\alpha$ -update*. Empirical evaluation on the Heterogeneity Human Activity Recognition (HHAR) dataset (with four activities for effective elucidation of results) indicates an average deterministic accuracy increase of at least  $\sim 11.01\%$  with the model distillation update strategy and  $\sim 9.16\%$  with the weighted  $\alpha$ -update strategy. We demonstrate the on-device capabilities of our proposed framework by using Raspberry Pi 2, a single-board computing platform.

**Keywords:** Human Activity Recognition · On-Device Deep Learning · Federated Learning · Heterogeneous Labels · Heterogeneous Models · Knowledge Distillation.

## 1 Introduction

Contemporary machine learning, particularly deep learning has led to major breakthroughs in various domains, such as vision, speech, Internet of Things (IoT), etc. Particularly, on-device deep learning has spiked up a huge interest in the research community owing to their automatic feature extraction mechanisms and the compute capabilities vested in resource-constrained mobile and wearable

devices. Sensors embedded in such IoT devices have a vast amount of incoming data which have massive potential to leverage such on-device machine learning techniques on-the-fly to transform them into meaningful information coupled with supervised, unsupervised and/or other learning mechanisms. Human Activity Recognition (HAR) in such personalized IoT devices is a technique of significant importance for our community as it plays a key role in modeling user behavior across a variety of applications like pervasive health monitoring, fitness tracking, fall detection, etc. With the ubiquitous proliferation of such personalized IoT devices, collaborative and distributed learning is now more possible than ever to help best utilize the behavioral information learnt from multiple devices.

However, such collaborative data sharing across devices might always not be feasible owing to privacy concerns from multiple participants. Users might not have any interest in sending their private data to a remote server/cloud, particularly in areas like healthcare. With the advent of *Federated Learning (FL)* [17, 1], it is now possible to effectively train a global/centralized model without compromising on sensitive data of various users by enabling the transfer of model weights and updates from local devices to the cloud, instead of conventionally transferring the sensitive data to the cloud. A server has the role of coordinating between models, however most of the work is not performed by a central entity anymore, but by a federation of clients/devices. The *Federated Averaging (FedAvg)* algorithm was first proposed by McMahan et al. in [17] which combines local Stochastic Gradient Descent (SGD) of each client (local device) with a server that aggregates the model weights. Federated learning has been an active and challenging area of research in solving problems pertaining to secure communication protocols, optimization, privacy preserving networks, etc. [14].

Federated Learning deals with various forms of heterogeneities like device, system, statistical heterogeneities, etc. [14]. Particularly in Federated Learning with IoT scenarios, statistical heterogeneities have gained much visibility as a research problem predominantly owing to the non-IID (non-independent and identically distributed) nature of the vast amounts of streaming real-world data incoming from distinct distributions across devices. This leads to challenges in personalized federation of devices, and necessitates us to address various heterogeneities in data and learning processes for effective model aggregation.

An important step in this direction is the ability of end-users to have the choice of architecting their own models, rather than being constrained by the pre-defined architectures mandated by the global model. One effective way to circumvent this problem is by leveraging the concept of knowledge distillation [8], wherein the disparate local models distill their respective knowledge into various *student models* which have a common model architecture, thereby effectively incorporating model independence and heterogeneity. This was proposed by Li et al. in FedMD [13]. However, as much independence and heterogeneity in architecting the users' own models is ensured in their work, *they do not guarantee heterogeneity and independence in labels across users.*

Many such scenarios with heterogeneous labels and models exist in federated IoT settings, such as behaviour/health monitoring, activity tracking, keyword spotting, next-word prediction, etc. Few works address handling new labels in typical machine learning scenarios, however, to the best of our knowledge, there is no work which addresses this important problem of *label and model heterogeneities* in non-IID federated learning scenarios.

The main scientific contributions in this work are as follows:

- Enabling end-users to build and characterize their own preferred local architectures in a federated learning scenario for HAR, so that effective transfer learning and federated aggregation happens between global and local models.
- A framework with two different versions to allow flexible heterogeneous selection of activity labels by showcasing scenarios with and without overlap across different user devices, thereby leveraging the information learnt across devices pertaining to those overlapped activities.
- Empirical demonstration of the framework’s ability to handle real-world disparate data/label distributions (non-IID) on-device independent of users on a public HAR dataset, capable of running on simple mobile and wearable devices.

## 2 Related Work

Deep learning for HAR, particularly inertial/sensor-based HAR measured from devices like accelerometer, gyroscope, etc. for improving pervasive healthcare has been an active area of research [7, 18]. Particularly, mobile- and wearable-based deep learning techniques for HAR have proven to be an extremely fruitful area of research with neural network models being able to efficiently run on such resource-constrained devices [23, 12, 22]. Few other challenges with deep learning for HAR have been explored like handling unlabeled data using semi-supervised and active learning mechanisms [24, 6], domain adaptation [2], few-shot learning [4], and many more.

Federated Learning has contributed vividly in enabling distributed and collective machine learning across various such devices. Federated learning and differentially private machine learning have, or soon will emerge to become the de facto mechanisms for dealing with sensitive data, data protected by Intellectual Property rights, GDPR, etc. [1]. Federated Learning was first introduced in [17], and new challenges and open problems to be solved [14] and multiple advancements [9] have been proposed and addressed in many interesting recent works.

Multiple device and system heterogeneities making them optimization problems are addressed in [15]. Personalized federated learning closely deals with optimizing the degree of personalization and contribution from various clients, thereby enabling effective aggregation as discussed in [3]. Federated learning on the edge with disparate data distributions – non-IID data, and creating a small subset of data globally shared between all devices is discussed in [25].

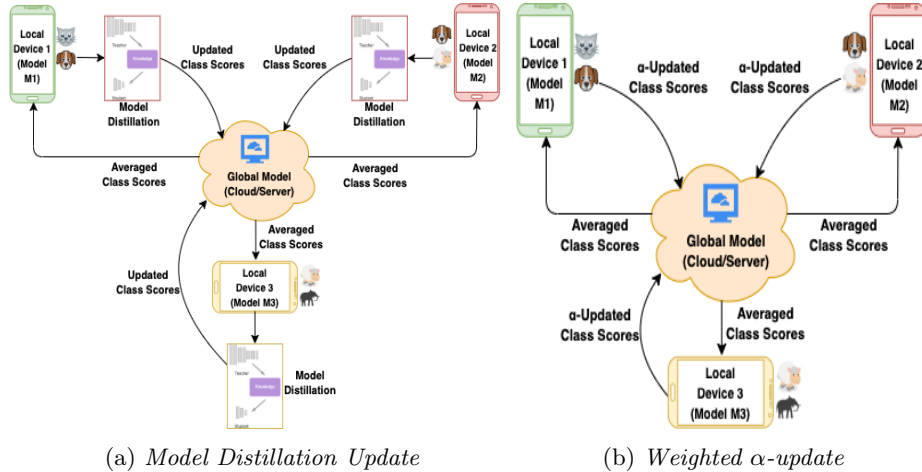


Fig. 1: Overall Architecture with both proposed versions. Each device consists of disparate sets of local labels and models, and they interact with the global model (cloud/server). The models in each local device are first updated using one of the two strategies, the respective class scores are then aggregated in the global model, and the updated consensus is again distributed across local models.

Particularly for Federated Learning in IoT and pervasive (mobile or wearable) devices, important problems and research directions on mobile and edge networks are addressed in this survey [16], while federated optimization for on-device applications is discussed in [11]. Federated Learning for HAR is addressed in [20] which deals with activity sensing with a smart service adapter, while [19] compares between centralized and federated learning approaches.

FedMD [13], which we believe to be our most closest work, deals with heterogeneities in model architectures, and addresses this problem using transfer learning and knowledge distillation [8], and also uses an initial public dataset across all labels (which can be accessed by any device during federated learning). Current federated learning approaches predominantly handle same labels across all the users and do not provide the flexibility to handle unique labels. However, in many practical applications, having unique labels for each local client/model is a very viable scenario owing to their dependencies and constraints on specific regions, demographics, privacy constraints, etc. A version of the proposed work is discussed for vision tasks in [5]. However, to the best of our knowledge, none of the works take into account label and model heterogeneities in the context of HAR.

The rest of the paper is organized as follows. Section 3.1 discusses the problem formulation of handling heterogeneous labels and models in on-device federated learning scenarios, and Section 3.2 presents the overall proposed framework and the methods used to address these challenges. Systematic experimentation and evaluation of the framework across different users, devices, iterations, models,

activities in a federated learning setting is showcased in Section 4, while also proving feasibility of the same on resource-constrained devices (Section 4.2). Finally, Section 5 concludes the paper.

### 3 Our Approach

In this section, we discuss in detail about the problem formulation of heterogeneity in labels and models, and our proposed framework to handle the same (showcased in Figures 1a and 1b).

#### 3.1 Problem Formulation

We assume the following scenario in federated learning. There are multiple local devices which can characterize different model architectures based on the end users. We hypothesize that the incoming data to different devices also consist of heterogeneities in labels, with either unique or overlapping labels. We also have a public dataset with the label set consisting of all labels – this can be accessed by any device anytime, and acts as an initial template of the data and labels that can stream through, over different iterations. We re-purpose this public dataset as the test set also, so that consistency is maintained while testing. To make FL iterations independent from the public dataset, we do not expose the public dataset during learning (training) to the local models. The research problem here is to create a unified framework to handle heterogeneous labels, models and data distributions (non-IID nature) in a federated learning setting.

#### 3.2 Proposed Framework

Our proposed framework to handle heterogeneous labels and models in a federated learning setting is presented in Algorithm 1. There are three important steps in our proposed method.

1. **Build:** In this step, we build the model on the incoming data we have in each local user, i.e., local private data for the specific iteration. The users can choose their own model architecture which suits best for the data present in that iteration.
2. **Local Update:** In this step, we update the averaged global model scores (on public data) for the  $i^{th}$  iteration on the local private data. For the first iteration, we do not have any global scores and we initialize the scores to be zero in this case. For the rest of iterations, we have global averaged scores which we can use to update the local model scores according to Algorithm 1. We propose two versions in the local update.
  - (a) **Model Distillation Update**, where the local model is distilled based only on labels corresponding to the local user. Distillation acts a summarization of the information captured from the older models in different FL iterations.

---

**Algorithm 1** Our Proposed Framework (with two version choices)

---

**Input:** Public Dataset  $\mathcal{D}_0\{x_0, y_0\}$ , Private Datasets  $\mathcal{D}_m^i$ , Total users  $M$ , Total iterations  $I$ , LabelSet  $l_m$  for each user  
**Output:** Trained Model scores  $f_G^I$   
Initialize  $f_G^0 = \mathbf{0}$  (Global Model Scores)  
**for**  $i = 1$  **to**  $I$  **do**  
  **for**  $m = 1$  **to**  $M$  **do**  
    **Build:** Model  $\mathcal{D}_m^i$  and predict  $f_{\mathcal{D}_m^i}(x_0)$   
    **Local Update:**  
      **Choice 1 – Model Distillation Update:**  
      Build a distilled model on only labels corresponding to local user’s model with global averaged probabilities on public dataset  $D_0$ . Now, update the model with the new data  $\mathcal{D}_m^i$  arriving in this iteration.  
      **Choice 2 – Weighted  $\alpha$ -update:**  
       $f_{\mathcal{D}_m^i}(x_0) = f_G^I(x_0^{l_m}) + \alpha f_{\mathcal{D}_m^i}(x_0)$ , where  $f_G^I(x_0^{l_m})$  are the global scores of only the set of labels  $l_m$  with the  $m^{th}$  user,  $\alpha = \frac{\text{len}(\mathcal{D}_m^i)}{\text{len}(\mathcal{D}_0)}$   
    **end for**  
  **Global Update:** Update label wise  
 $f_G^{i+1} = \sum_{m=1}^M \beta_m f_{\mathcal{D}_m^i}(x_0)$ , where  

$$\beta = \begin{cases} 1 & \text{If labels are unique} \\ \text{acc}(f_{\mathcal{D}_m^i}(x_0)) & \text{if labels are not unique} \end{cases}$$
where  $\text{acc}(f_{\mathcal{D}_m^i}(x_0))$  is the accuracy function of the given model, and is defined by the ratio of correctly classified samples to the total samples for the given local model  
  **end for**

---

- (b) **Weighted  $\alpha$ -update**, where  $\alpha$  is the ratio between the size of current private dataset and the size of public dataset. This parameter governs the contributions of the new and the old models across different FL iterations.
3. **Global update:** In this step, we first train the local model on the respective private datasets for that FL iteration. Further, we evaluate (test) this trained model on the public data, thereby obtaining the model scores on public data. We then perform such label-based averaging across all the users using the  $\beta$  parameter, where  $\beta$  governs the weightage given to unique and overlapping labels across users using test accuracies of the corresponding labels on public data (as given in Algorithm 1). This module gives the global averaged scores.

## 4 Experiments and Results

We simulate a federated learning scenario with multiple iterations of small chunks of incremental data incoming (details in Table 1), across three different users to test our approach, and assume that the activities arrive in real-time in

Table 1: Model Architectures (filters/units in each layer), Labels (Activities) and Number of Activity Windows per federated learning iteration across user devices. Note the disparate model architectures and labels across users.

	User_1	User_2	User_3	Global_User
<b>Architecture</b>	2-Layer CNN (16, 32) Softmax Activation	3-Layer CNN (16, 16, 32) ReLU Activation	3-Layer ANN (16, 16, 32) ReLU Activation	-
<b>Activities</b>	{Sit, Walk}	{Walk, Stand}	{Stand, StairsUp}	{Sit, Walk, Stand, StairsUp}
<b>Activity Windows per iteration</b>	{2000, 2000} = 4000	{2000, 2000} = 4000	{2000, 2000} = 4000	{2000, 2000, 2000, 2000} = 8000

the users’ devices. We use the *Heterogeneity Human Activity Recognition dataset* [21], which consists of inertial data from four different mobile phones across nine users performing six daily activities: Biking, Sitting, Standing, Walking, Stairs-Up, Stairs-Down in heterogeneous conditions.

**Data Preprocessing:** In this experiment, we perform similar preprocessing techniques as stated in [22]. As discussed, we use the mobile phone accelerometer data only and not gyroscope, due to the reduction in data size without substantial accuracy decrease. We initially segment the triaxial accelerometer data into two-second non-overlapping windows and then perform *Decimation* to downsample (normalize) all activity windows to the least sampling frequency (50 Hz). Following this, Discrete Wavelet Transform (DWT) is performed for obtaining temporal and frequency information and we use Approximation coefficients only, all together is stated to have a substantial decrease in data size.

Now, we discuss the settings for label and model heterogeneities in our experiment.

**Label Heterogeneities:** In our experiment, we consider only four activities – {*Sit, Walk, Stand, StairsUp*} from the dataset as shown in Table 1. Also, we include the number of activity windows considered per user per iteration (2000 activity windows per iteration). The activities in each local user can either be unique (present only in that single user) or overlapping across users (present in more one user). We split the four activities into three pairs of two activities each, for convenience of showcasing the advantage of overlapping activities in experimentation. We also create a non-IID environment across different federated learning iterations wherein, the activity data across different iterations are split with disparities in both the aforementioned labels and distributions in data (*Statistical Heterogeneities*).

**Model Heterogeneities:** We choose three different model architectures (CNNs and ANNs) for the three different local users. This is clearly elucidated in Table 1. We also use a simple two-layer ANN model with (8, 16) filters as the *distilled student architecture*. To truly showcase near-real-time heterogeneity and model

Table 2: Details of Model Architectures (filters/units in each layers) changed across federated learning iterations and users.

Iteration	New Model Architecture
User_1 Iteration_10	3-Layer ANN (16, 16, 32) ReLU Activation
User_1 Iteration_14	1-Layer CNN (16) Softmax Activation
User_2 Iteration_6	3-Layer CNN (16, 16, 32) Softmax activation
User_3 Iteration_5	4-Layer CNN (8, 16, 16, 32) Softmax activation

independence, we induce a change in the model architectures across and within various FL iterations as shown in Table 2.

Initially, we divide the activity windows across the three different users according to the four activity labels. We create a Public Dataset ( $D_0$ ) with 8000 activity windows, with 2000 activity windows corresponding to each activity. Next, we sample 2000 activity windows in every iteration for each label of a user (as shown in Table 1). In total, we ran 15 federated learning iterations in this whole experiment, with each iteration running with early stopping (with a maximum 5 epochs). We track the loss using categorical cross-entropy loss function for multi-class classification, and use the Adam optimizer [10] to optimize the classification loss. We simulate all our experiments – both federated learning and inference on a *Raspberry Pi 2*.

#### 4.1 Discussion on Results

Table 3: Average Accuracies (%) of Local and Global Updates, and their respective Accuracy increase with *Model Distillation Update* and *Weighted  $\alpha$ -update*.

	Model Distillation			Weighted $\alpha$ -update		
	Local_Update	Global_Update	Increase	Local_Update	Global_Update	Increase
User_1	68.38	77.61	9.23	66.98	74.29	7.31
User_2	70.82	84.4	13.58	68.88	81.9	13.02
User_3	77.68	87.9	10.22	76.57	83.7	7.13
Average	<b>72.293</b>	<b>83.303</b>	<b>11.01</b>	<b>70.81</b>	<b>79.963</b>	<b>9.153</b>

Figure 2 represents the results across all three users for both proposed versions of our framework on the HHAR dataset. Also, from Table 3, we can clearly



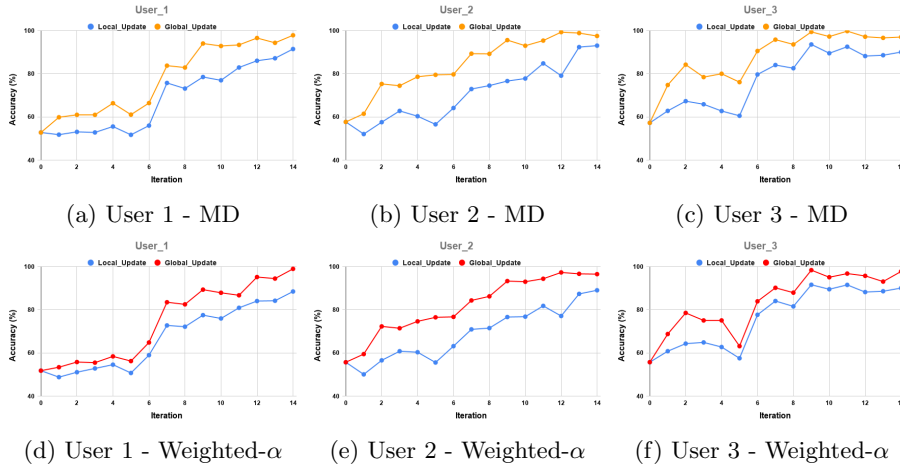


Fig. 2: Iterations vs Accuracy across all three users with *Model Distillation Update* (MD) and *Weighted  $\alpha$ -update*. *Local\_Update* signifies the accuracy of each local updated model (after  $i^{th}$  iteration) on Public Dataset. *Global\_Update* signifies the accuracy of the corresponding global updated model (averaged across all the users after  $i^{th}$  iteration) on Public Dataset.

observe that the global updates – which represent the accuracy of the global updated model (and averaged across all users’ labels in the  $i$ th iteration governed by  $\beta$ ), are higher for all three users than the accuracies of their respective local updates. For instance, from Figures 2a and 2d, we can infer that the corresponding accuracies of labels  $\{Sit, Walk\}$  (User 1 labels) after global updates in each iteration are deterministically higher than their respective local updates by an average of  $\sim 9.23\%$  and  $\sim 7.31\%$  across all iterations with model distillation and  $\alpha$ -update versions respectively. Similarly for User 2 labels consisting of  $\{Walk, Stand\}$ , we observe an average accuracy increase of  $\sim 13.58\%$  and  $\sim 13.02\%$  respectively from local updates to the global updates, while for User 3 labels consisting of  $\{Stand, StairsUp\}$ , we observe an average increase of  $\sim 10.22\%$  and  $\sim 7.13\%$  respectively from local updates to global updates in model distillation and  $\alpha$ -update versions.

We would like to particularly point out that the overlap in activities significantly contributes to highest increase in accuracies, since information gain (weighted global update) happens only for overlapping labels. This is vividly visible in User 2 (Figure 2b and Figure 2e), whose labels are  $\{Walk, Stand\}$ , where, in spite of an accuracy dip in local update at FL iterations 5 and 12, the global update at those iterations do not take a spike down which can be primarily attributed to the information gain from overlapping activity labels between User 1 and User 3 (in this case, *Walk* and *Stand* respectively), thereby showcasing the robustness of overlapping label information gain in User 2. On the contrary, when we observe User 3 (Figure 2c and Figure 2f), in spite of the accuracies of

global updates being inherently better than local updates, when a dip in accuracies of local updates are observed at iterations 5 and 8, the accuracies of global updates at those iterations also spike down in a similar fashion. Similar trends of local and global accuracy trends like those observed in User 3 can also be observed in User 1 (Figure 2a and Figure 2d). This clearly shows that when there are lesser overlapping activity labels (User 1 and User 3), the global model does not learn the activities’ characteristics much, while the global updates are more robust in spite of spikes and dips in local updates with such overlapping labels (User 2), thereby leading to higher average increase in accuracies (as observed in Table 3).

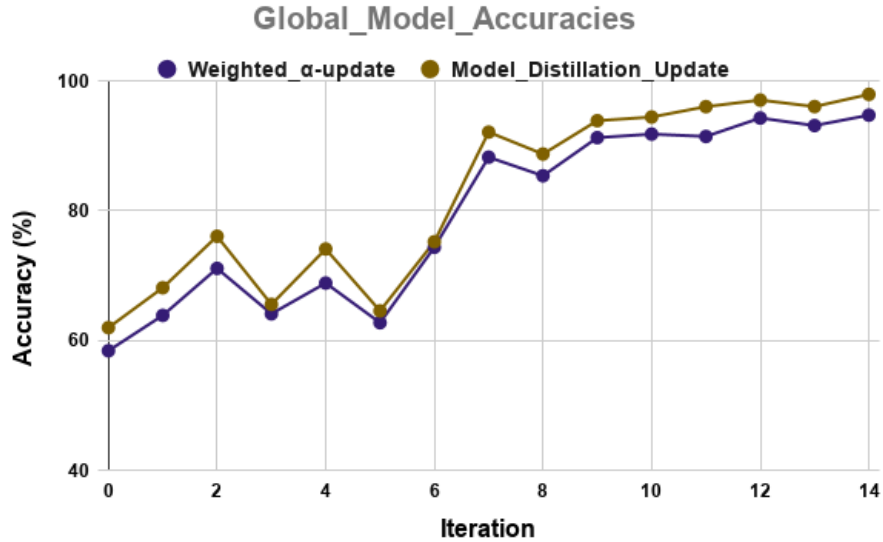


Fig. 3: Iterations vs Final Global Average Accuracies (%) with *Model Distillation Update* and *Weighted  $\alpha$ -update*

Overall average deterministic (not relative) increase in accuracies of  $\sim 11.01\%$  and  $\sim 9.153\%$  are observed respectively with the model distillation and  $\alpha$ -update versions on the HHAR dataset, which are calculated from the global model updates (Table 3). The overall global model accuracies averaged across all users after each iteration (which is different from global update accuracies after each iteration observed in Figure 2) are also elucidated in Figure 3. We can observe that the distillation version performs better than the  $\alpha$ -update version with a  $\sim 3.21\%$  deterministic accuracy increase. With our current framework, *communication (transfer) of just the model scores* of respective activity labels between clients (local devices) and the central cloud is performed, without necessitat-

ing transfer of the entire model weights, which significantly reduces latency and memory overheads.

## 4.2 On-Device Performance

We observe the on-device performance of our proposed framework by experimenting on a Raspberry Pi 2. We choose this single-board computing platform since it has similar hardware and software (HW/SW) specifications with that of predominant contemporary IoT/mobile devices. The computation times taken for execution of on-device federated learning and inference are reported in Table 4. This clearly shows the feasibility of our proposed system on embedded devices. Also, the distillation mechanism accounts for higher computation overheads in time on edge/mobile devices, and depend on the temperature parameters (default set at 1) and the distilled student model architecture chosen. The end-user can typically make the trade-off of choosing the local distillation version or the  $\alpha$ -update version depending on their compute capabilities and accuracy requirements.

Table 4: Time taken for Execution

Process	Computation Time
Training time per epoch in an FL iteration ( $i$ )	$\sim 1.8$ sec
Inference time	$\sim 16$ ms
Discrete Wavelet Transform	$\sim 0.45$ ms
Decimation	$\sim 4.6$ ms

## 5 Conclusion

This paper presents a unified framework for flexibly handling heterogeneous labels and model architectures in federated learning for Human Activity Recognition (HAR). By leveraging transfer learning along with simple scenario changes in the federated learning setting, we propose a framework with two versions – *Model Distillation Update* and *Weighted  $\alpha$ -update* aggregation in local models, and we are able to leverage the effectiveness of global model updates with activity label based averaging across all devices and obtain higher efficiencies. Moreover, overlapping activities are found to make our framework robust, and also helps in effective accuracy increase. We also experiment by sending only model scores rather than model weights from user device to server, which reduces latency and memory overheads multifold. We empirically showcase the successful feasibility of our framework on-device, for federated learning/training across different iterations on the widely used HHAR dataset. We expect a good amount of research focus hereon in handling statistical, model and label based heterogeneities for HAR and other pervasive sensing tasks.

## References

1. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konecny, J., Mazzocchi, S., McMahan, H.B., Overveldt, T.V., Petrou, D., Ramage, D., Roselander, J.: Towards federated learning at scale: System design. In: SysML 2019 (2019)
2. Chang, Y., Mathur, A., Isopoussu, A., Song, J., Kawsar, F.: A systematic study of unsupervised domain adaptation for robust human-activity recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **4**(1), 1–30 (2020)
3. Deng, Y., Kamani, M.M., Mahdavi, M.: Adaptive personalized federated learning. *arXiv preprint arXiv:2003.13461* (2020)
4. Feng, S., Duarte, M.F.: Few-shot learning-based human activity recognition. *Expert Systems with Applications* **138**, 112782 (2019)
5. Gudur, G.K., Balaji, B.S., Perepu, S.K.: Resource-constrained federated learning with heterogeneous labels and models. *arXiv preprint arXiv:2011.03206* (2020)
6. Gudur, G.K., Sundaramoorthy, P., Umaashankar, V.: Activeharnet: Towards on-device deep bayesian active learning for human activity recognition. In: *The 3rd International Workshop on Deep Learning for Mobile Systems and Applications*. pp. 7–12 (2019)
7. Hammerla, N.Y., Halloran, S., Plötz, T.: Deep, convolutional, and recurrent models for human activity recognition using wearables. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. pp. 1533–1540. IJCAI’16, AAAI Press (2016)
8. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. In: *NIPS Deep Learning and Representation Learning Workshop* (2015)
9. Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., et al.: Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977* (2019)
10. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
11. Konečný, J., McMahan, H.B., Ramage, D., Richtárik, P.: Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527* (2016)
12. Lane, N.D., Bhattacharya, S., Georgiev, P., Forlivesi, C., Kawsar, F.: An early resource characterization of deep learning on wearables, smartphones and internet-of-things devices. In: *Proceedings of the 2015 International Workshop on Internet of Things Towards Applications*. pp. 7–12. IoT-App ’15 (2015)
13. Li, D., Wang, J.: Fedmd: Heterogenous federated learning via model distillation. *arXiv preprint arXiv:1910.03581* (2019)
14. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine* **37**, 50–60 (2020)
15. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems 2020* pp. 429–450 (2020)
16. Lim, W.Y.B., Luong, N.C., Hoang, D.T., Jiao, Y., Liang, Y.C., Yang, Q., Niyato, D., Miao, C.: Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials* (2020)
17. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y.: Communication-efficient learning of deep networks from decentralized data. In:

- Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. vol. 54, pp. 1273–1282 (2017)
18. Osmani, V., Balasubramaniam, S., Botvich, D.: Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of Network and Computer Applications* **31**, 628–655 (2008)
  19. Ramakrishnan, A.K., Naqvi, N.Z., Preuveneers, D., Berbers, Y.: Federated mobile activity recognition using a smart service adapter for cloud offloading. In: *Human Centric Technology and Service in Smart Space*, pp. 173–180. Springer (2012)
  20. Sozinov, K., Vlassov, V., Girdzijauskas, S.: Human activity recognition using federated learning. In: *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCLOUD/SocialCom/SustainCom)*. pp. 1103–1111. IEEE (2018)
  21. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T.S., Kjærgaard, M.B., Dey, A., Sonne, T., Jensen, M.M.: Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In: *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. pp. 127–140. SenSys '15 (2015)
  22. Sundaramoorthy, P., Gudur, G.K., Moorthy, M.R., Bhandari, R.N., Vijayaraghavan, V.: Harnet: Towards on-device incremental learning using deep ensembles on constrained devices. In: *Proceedings of the 2nd International Workshop on Embedded and Mobile Deep Learning*. pp. 31–36. EMDL'18 (2018)
  23. Yao, S., Hu, S., Zhao, Y., Zhang, A., Abdelzaher, T.: Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In: *Proceedings of the 26th International Conference on World Wide Web*. pp. 351–360. WWW '17 (2017)
  24. Yao, S., Zhao, Y., Shao, H., Zhang, C., Zhang, A., Hu, S., Liu, D., Liu, S., Su, L., Abdelzaher, T.: Sensegan: Enabling deep learning for internet of things with a semi-supervised framework. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2**(3), 1–21 (2018)
  25. Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., Chandra, V.: Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582* (2018)