Federated Learning with Heterogeneous Labels and Models for Mobile Activity Monitoring

Gautham Krishna Gudur
Global AI Accelerator, Ericsson
gautham.krishna.gudur@ericsson.com

Satheesh Kumar Perepu
Ericsson Research
perepu.satheesh.kumar@ericsson.com

Abstract

Various health-care applications such as assisted living, fall detection etc., require modeling of user behavior through Human Activity Recognition (HAR). Such applications demand characterization of insights from multiple resource-constrained user devices using machine learning techniques for effective personalized activity monitoring. On-device Federated Learning proves to be an effective approach for distributed and collaborative machine learning. 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 for federated label-based aggregation, which leverages overlapping information gain across activities using *Model Distillation Update*. We also propose that federated transfer of model scores is sufficient rather than model weight transfer from device to server. Empirical evaluation with the Heterogeneity Human Activity Recognition (HHAR) dataset (with four activities for effective elucidation of results) on Raspberry Pi 2 indicates an average deterministic accuracy increase of at least $\sim$11.01%, thus demonstrating the on-device capabilities of our proposed framework.

1 Introduction

Human Activity Recognition (HAR) is a technique of significant importance for modeling user behavior in applications like pervasive health monitoring, fitness tracking, fall detection, etc. With the ubiquitous proliferation of personalized sensor-based IoT devices, collaborative and distributed learning is now more possible than ever to help best utilize the behavioral information learnt from multiple users. With the advent of Federated Learning (FL) [1], we can effectively train a centralized model with a federation of users without compromising on sensitive data of users by combining local Stochastic Gradient Descent (SGD) of each local (client) model and aggregating weights on a server, instead of conventionally transferring sensitive data to cloud (*Federated Averaging* algorithm) [18].

On-device federated learning deals with various forms of heterogeneities like device, system, statistical heterogeneities, etc. [15]. Particularly, statistical heterogeneities have gained much research visibility predominantly due to the *non-IID* (non-independent and identically distributed) nature of data from distinct distributions, leading to challenges in personalized federation of devices. An important step in this direction is the ability of end-users to have the choice of architecting their own models, rather than using the pre-defined architectures mandated by the global model. FedMD [14] leverages knowledge distillation [9] to address this, wherein the local models distill their respective knowledge into a *student model* which has a common model architecture. 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.*
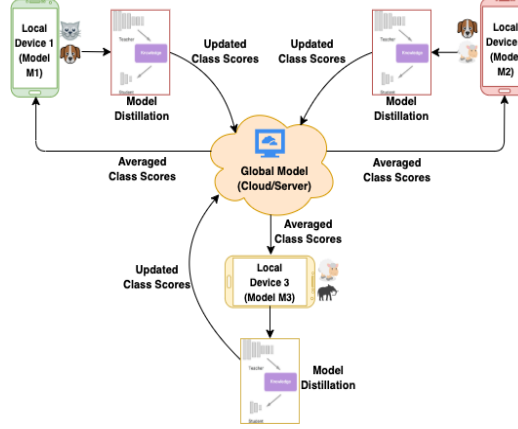
Figure 1: Overall Architecture with *Local Model Distillation*. Each mobile device can consist of disparate set of local labels and models, and they interact with the global model (cloud/server). Each local model is first distilled to a student model, the respective class scores are then aggregated in the global model, and the updated consensus is again distributed across local models.

Many such scenarios with heterogeneous labels and models typically exist in federated IoT settings, such as behaviour/health monitoring, activity tracking, 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. More related work on FL, HAR, etc. can be found in Appendix A.

Scientific contributions: **(1)** Enabling end-users to 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. **(2)** A framework to allow flexible heterogeneous selection of activities by showcasing scenarios with and without label overlap across different user devices, thereby leveraging the information learnt across devices pertaining to those overlapped activities. **(3)** Empirical demonstration of the framework's ability to handle real-world disparate data/label distributions (non-IID) on simple mobile and wearable devices with a public HAR dataset.

2   Our Approach

Problem Formulation: We assume the following scenario in federated learning. There are multiple local devices which can characterize different model architectures based on end users. We hypothesize that the incoming data to different devices also consist of heterogeneities in activities, with either unique or overlapping activities. We also have a public dataset with the label set consisting of all activities – 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 to the local models during learning (training).

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 here:

(a) **Build**: In this step, we build the model on the incoming data we have in each local user, i.e., local private data for a specific iteration. The users can choose their own model architecture which suits best for the data present in that iteration.

(b) **Local Update (Model Distillation Update)**: In this step, the local models trained on local private data, and distilled to a student model (with public data) with corresponding labels. Distillation acts like a summarization of information captured from previous FL iterations. The global averaged consensus are distributed to the local models before a local update.

(c) **Global update**: In this step, *only the scores of each local model are sent to the server* rather than traditional transfer of weights. The model scores are then averaged with parameter β, where β governs the weightage given to overlapping labels across users using test accuracies of corresponding labels on public data. This module gives the global update scores.

Algorithm 1 Our Proposed Framework

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 (Model Distillation):
        Build a distilled model only on respective local model labels 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.
    end for
    Global Update: Update label wise
$$f_G^{i+1} = \sum_{m=1}^{M} \beta_m f_{\mathcal{D}_m^i}(x_0), \text{ where}$$
$$\beta = \begin{cases} 1 & \text{If labels are unique} \\ \text{acc}(f_{\mathcal{D}_m^{i+1}}(x_0)) & \text{if labels are not unique} \end{cases}$$
    where $\text{acc}(f_{\mathcal{D}_m^{i+1}}(x_0))$ 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

---

3 Experiments and Results

Table 1: Model Architectures (filters in each layer), Labels and 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
Architecture	2-Layer CNN (16, 32) Softmax Activation	3-Layer CNN (16, 16, 32) ReLU Activation	3-Layer ANN (16, 16, 32) ReLU Activation	–
Activities	{Sit, Walk}	{Walk, Stand}	{Stand, StairsUp}	{Sit, Walk, Stand, StairsUp}
Activity Windows per iteration	{2000, 2000} = 4000	{2000, 2000} = 4000	{2000, 2000} = 4000	{2000, 2000, 2000, 2000} = 8000

We simulate a Federated Learning scenario with multiple iterations of incremental data across users' devices. We use the ***Heterogeneity Human Activity Recognition dataset*** [22] for our experiment. More details regarding the dataset and data preprocessing are discussed in Appendix B.

Label Heterogeneities: In our experiment, we consider four activities – {*Sit, Walk, Stand, StairsUp*} across three user devices from the dataset (Table 1). 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 each user, for convenience of showcasing the advantage of overlapping activities. We also create a non-IID environment across different federated learning iterations wherein, the activity data are split with disparities in both the aforementioned labels and distributions in data (*Statistical Heterogeneities*).

Model Heterogeneities: We choose three different model architectures for the three different local users (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 independence, we induce a change in the model architectures across various FL iterations as shown in Appendix C Table 3.

We create a Public Dataset (D_0) with 8000 activity windows, and 2000 activity windows corresponding to each activity. Also, we consider 2000 activity windows per label per iteration in a user (Table 1). In total, we run 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 classification, and use the Adam optimizer [11] to optimize the classification loss. We simulate all our experiments – both federated learning and inference on a *Raspberry Pi 2*.

3.1 Discussion on Results

In Table 2, *Local Update* and *Global Update* signify the accuracies of each local updated model and the corresponding global updated model (after i^{th} iteration) on Public Dataset respectively. We
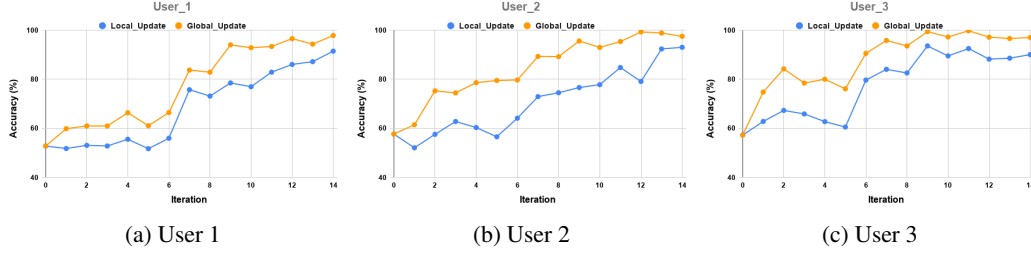
Figure 2: Iterations vs *Local Update* and *Global Update* Accuracies across all three users.

can clearly observe that the global updates (governed by β), are higher for all three users than the accuracies of their respective local updates. The average accuracy increase across all iterations from local to global updates is deterministically $\sim$11.01%.

Table 2: Average Accuracies (%) of Local and Global Updates, and their respective Accuracy increase.

	Local Update	Global Update	Accuracy Increase
User_1	68.38	77.61	9.23
User_2	70.82	84.4	**13.58**
User_3	77.68	87.9	10.22
Average	**72.293**	**83.303**	**11.01**

The significant contribution of overlap in activities towards increase in global accuracies (β-weighted global update information gain) is vividly visible in User 2 (Figure 2b – {*Walk, Stand*}), where, in spite of an accuracy dip in local update at iterations 5 and 12, the global updates at those iterations do not spike down. This can be primarily attributed to the robustness of overlapping label information gain from User 1 (*Walk*) and User 3 (*Stand*) on User 2. On the contrary, for User 3 (Figure 2c), when a dip in local update accuracies are observed at iterations 5 and 8, the global update accuracies in those respective iterations also spike down in a similar fashion. Similar trends in local and global accuracies can also be observed in User 1 (Figure 2a). This also accounts for the highest accuracy increase from local to global updates in User 2 over the rest (Table 2). The final overall global model accuracies averaged across all users after each iteration are also elucidated in Appendix E Figure 3.

3.2 On-Device Performance

Raspberry Pi 2 is used for evaluating our proposed framework as it has similar hardware and software (HW/SW) specifications to predominant contemporary IoT/mobile devices. The computation times taken for execution of our proposed framework are reported in Appendix D Table 4. The distillation mechanism accounts for higher computation time overheads on edge/mobile devices, which depends on the temperature parameter (default set at 1) and the distilled student model architecture chosen.

4 Conclusion

This paper presents a framework for flexibly handling heterogeneous labels and model architectures in federated learning for Human Activity Recognition. We propose a framework with model distillation in local models, and leverage the effectiveness of global model updates with label based averaging (weighted β-update) to obtain higher efficiencies. Moreover, overlapping activities across user devices are found to make our framework robust, and also aid 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 across different iterations. We expect a good amount of research focus hereon in handling statistical, model and label based heterogeneities for HAR and other pervasive mobile health tasks.

References

[1] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konecny, Stefano Mazzocchi, H Brendan McMahan, Timon Van Overveldt, David Petrou, Daniel Ramage, and Jason Roselander. Towards federated learning at scale: System design. In *SysML 2019*, 2019.

[2] Youngjae Chang, Akhil Mathur, Anton Isopoussu, Junehwa Song, and Fahim Kawsar. 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] Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. Adaptive personalized federated learning. *arXiv preprint arXiv:2003.13461*, 2020.

[4] Siwei Feng and Marco F Duarte. Few-shot learning-based human activity recognition. *Expert Systems with Applications*, 138:112782, 2019.

[5] Gautham Krishna Gudur, Bala Shyamala Balaji, and Satheesh K Perepu. Resource-constrained federated learning with heterogeneous labels and models. *arXiv preprint arXiv:2011.03206*, 2020.

[6] Gautham Krishna Gudur and Satheesh K Perepu. Resource-constrained federated learning with heterogeneous labels and models for human activity recognition. In *Deep Learning for Human Activity Recognition Workshop, IJCAI*, 2020 [Accepted].

[7] Gautham Krishna Gudur, Prahalathan Sundaramoorthy, and Venkatesh Umaashankar. Active-harnet: Towards on-device deep bayesian active learning for human activity recognition. In *The 3rd International Workshop on Deep Learning for Mobile Systems and Applications*, pages 7–12, 2019.

[8] Nils Y. Hammerla, Shane Halloran, and Thomas Plötz. Deep, convolutional, and recurrent models for human activity recognition using wearables. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 1533–1540. AAAI Press, 2016.

[9] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. In *NIPS Deep Learning and Representation Learning Workshop*, 2015.

[10] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.

[11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[12] Jakub Konečný, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*, 2016.

[13] Nicholas D. Lane, Sourav Bhattacharya, Petko Georgiev, Claudio Forlivesi, and Fahim Kawsar. 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*, IoT-App '15, pages 7–12, 2015.

[14] Daliang Li and Junpu Wang. Fedmd: Heterogenous federated learning via model distillation. *arXiv preprint arXiv:1910.03581*, 2019.

[15] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37:50–60, 2020.

[16] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems 2020*, pages 429–450, 2020.

[17] Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, Yutao Jiao, Ying-Chang Liang, Qiang Yang, Dusit Niyato, and Chunyan Miao. Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 2020.

[18] H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54, pages 1273–1282, 2017.

[19] Venet Osmani, Sasitharan Balasubramaniam, and Dmitri Botvich. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of Network and Computer Applications*, 31:628–655, 2008.

[20] Arun Kishore Ramakrishnan, Nayyab Zia Naqvi, Davy Preuveneers, and Yolande Berbers. Federated mobile activity recognition using a smart service adapter for cloud offloading. In *Human Centric Technology and Service in Smart Space*, pages 173–180. Springer, 2012.

[21] Konstantin Sozinov, Vladimir Vlassov, and Sarunas Girdzijauskas. 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)*, pages 1103–1111. IEEE, 2018.

[22] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 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*, SenSys '15, pages 127–140, 2015.

[23] Prahalathan Sundaramoorthy, Gautham Krishna Gudur, Manav Rajiv Moorthy, R. Nidhi Bhandari, and Vineeth Vijayaraghavan. 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*, EMDL'18, pages 31–36, 2018.

[24] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 351–360, 2017.

[25] Shuochao Yao, Yiran Zhao, Huajie Shao, Chao Zhang, Aston Zhang, Shaohan Hu, Dongxin Liu, Shengzhong Liu, Lu Su, and Tarek Abdelzaher. 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.

[26] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. Federated learning with non-iid data. *arXiv preprint arXiv:1806.00582*, 2018.

A Appendix: Related Work

Deep learning for HAR, particularly inertial/sensor-based HAR for improving pervasive healthcare has been an active area of research [8, 19]. 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 [24, 13, 23]. Many other challenges in deep learning for HAR tasks have been explored like handling unlabeled data using semi-supervised mechanisms [25, 7], 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 [18], and new challenges and open problems to be solved [15] and multiple advancements [10] have been proposed and addressed in many interesting recent works.

Multiple device and system heterogeneities making them optimization problems are addressed [16]. 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 [26].

Particularly for Federated Learning in IoT and pervasive (mobile/wearable/edge) devices, important problems and research directions on mobile and edge networks are addressed in this survey [17], while federated optimization for on-device applications is discussed in [12]. Federated Learning for HAR is addressed in [21] which deals with activity sensing with a smart service adapter, while [20] compares between centralized and federated learning approaches.

FedMD [14], 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 [9], 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 users and do not provide the flexibility to handle unique labels. However, in many practical applications, having unique labels or overlapping labels for each local client/model is a very viable scenario owing to their dependencies and constraints on specific regions, demographics, privacy constrains, etc. A version of the proposed work is discussed for vision tasks in [5] and for HAR tasks in [6].

B Appendix: Dataset and Dataset Preprocessing

The Heterogeneity Human Activity Recognition dataset [22] 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 [23]. As discussed in [23], 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 with much loss in information.

C Appendix: Model Heterogeneities across Iterations

The model heterogeneities across and within different iterations are observed in Table 3.

Table 3: Details of Model Architectures (filters in each layer) changed across federated learning iterations and users for both datasets.

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

D Appendix: On-Device Performance

The computation times for on-device performance on Raspberry Pi 2 for our proposed framework are observed in Table 4.

Table 4: Computation Time taken for Execution for HHAR dataset

Process	Computation Time
Training time per epoch in an FL iteration (i)	~1.7 sec
Inference time	~15 ms
Discrete Wavelet Transform	~0.45 ms
Decimation	~4.6 ms

E Appendix: Overall Global Average Accuracies

The final overall global model accuracies averaged across all users in each iteration is observed in Figure 3.
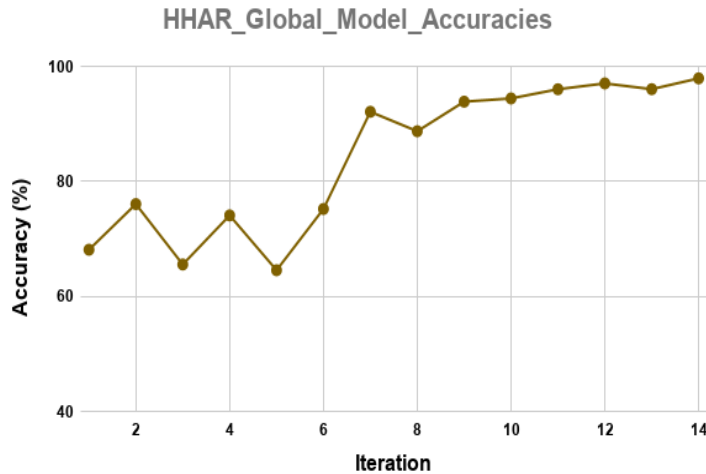

Figure 3: Iterations vs Final Global Average Accuracies (%) with *Local Model Distillation*.