A Robust Aggregation Approach for Heterogeneous Federated Learning

Dost Muhammad Saqib Bhatti
Department of Electrical and Electronic Engineering
Hanyang University
Ansan, South Korea
saqib@hanyang.ac.kr

Haewoon Nam
Department of Electrical and Electronic Engineering
Hanyang University
Ansan, South Korea
hnam@hanyang.ac.kr

***Abstract*—Federated learning is a cutting-edge method of model training, which leverages the end users to train the global model on the server. The end users are responsible for training locally on their datasets and update the shared global model. Once the local training is executed, the local trained models are forwarded back to the server to further upgrade the global model by performing aggregation. This process of global training is carried out for certain number of rounds. Practically, the datasets of clients are distributed heterogeneously. Thus, the updated local models by clients emanate broad variation among local models due to heterogeneity. In other words, the aggregation of local models plays a vital role in federated learning. Specifically, aggregating the diversified local models may deliver unsatisfactory output if not performed efficiently. This article presents a performance efficient and robust aggregation approach for heterogeneous federated learning called FedLbl. Our approach takes the diversity of data among clients into consideration before conducting the aggregation of local models.Our study compares the proposed method with conventional federated learning techniques, resulting in a 28% increase in accuracy and a 19% reduction in loss.**

***Index Terms*—Heterogeneous networks, federated learning, deep learning.**

I. Introduction

Federated learning has emerged as a promising approach for global model training by utilizing the end-users' devices in a network. This approach is transforming the field of artificial intelligence by preserving the privacy of users. [1]. By incorporating intelligence at the edge of the network, It has become an efficient model for privacy-preserving learning [2]. The end-users, referred to as clients of the network, can deploy any neural network based on the network's needs. Typically, deep neural networks (DNNs), convolutional neural networks (CNNs) or recurrent neural networks (RNNs) are deployed [3], [4]. It is important to note that since the datasets of clients are non-independent and identically distributed (Non-IID), the local models trained by clients exhibit significant variations as a result of their varying characteristics [5]. Therefore, the process of aggregating highly diverse local models in a heterogeneous environment affects the global training performance [6], [7].

This paper proposes a performance efficient aggregation method for heterogeneous federated learning called federated labels (FedLbl), which takes the data heterogeneity into the consideration before performing aggregation of local models.

A. Related Work

The schema of distributed learning mechanism that utilizes end-users was first proposed by Konecny et al. [8]. Since federated learning involves diverse kind of clients, it is not a realistic approach to assume even distribution among clients, known as IID, as discussed in [9] and [10]. Studies by Zhao et al. [11] have shown that when clients with heterogeneous data distribution are involved in global training, the prediction accuracy of the global model is affected. In order to address the issue of significant diversity in the weights of local and global models due to data heterogeneity, a data-sharing strategy was proposed by [11]. This strategy aims to mitigate the impact of varying data distributions across different clients by sharing data across clients with similar data distributions. A subset of data samples representing each class of label is created and distributed among all clients connected to the network. The authors on [12] and [13] also developed the method of creating a smaller dataset for distribution among clients, as described above. However, creating a smaller dataset using data from all clients may not protect their privacy. The authors on [14] suggested a technique that involves adjusting the global model prior to aggregating the local models. However, this increases the overhead for the server. A method termed FedProx was put forth by [15], which incorporates a proximity function to limit the local updates and keep them closer to the initial weights of the model at server. The authors of [16] explored a personalized version of federated learning that aims to find an initial shared model that can be easily adapted by new clients. A clustered federated learning approach was proposed by [17], which utilizes the geometric characteristics of the federated learning loss to improve performance. According to [18], the performance of federated learning is primarily determined by how the server aggregates the weights of local models. A well-known aggregation method called FedAvg was proposed by [18], which calculates the weighted average of local models to update the global model. In [19], an aggregation method called temporally weighted aggregation was introduced, which utilizes past local models to improve performance.

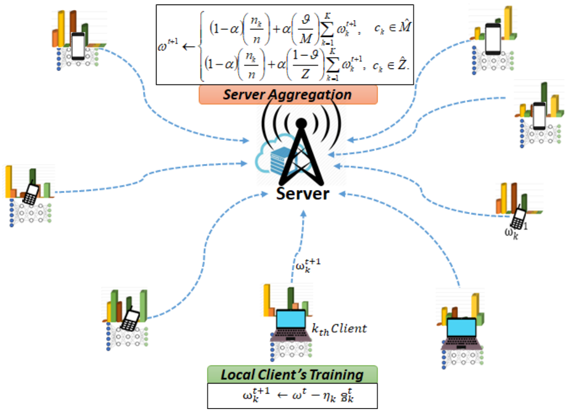


Fig. 1: System model of FedLbl

B. Contributions

Our primary goal is to obtain a global model with improved accuracy and minimal loss while maintaining the clients'privacy. The key contributions of proposed FedLbl method are summarized as follows.

- First of all, we have proposed a novel method called FedLbl, that considers the parameter of diversity in client data distributions prior to the aggregation of the local models to strive towards the minimum global loss point.
- Secondly, we have studied the impact of varying data distribution among clients on global training, and have shown that heterogeneity can negatively affect the performance of global training.
- Finally, our proposed FedLbl method attains the highest classification accuracy and the lowest loss in comparison to traditional methods, as it considers the variance and volume of data possessed by a client during aggregation.

To the best of the author's knowledge, while most of the methods proposed in the literature assume non-IID data distribution among clients in federated learning, only a few of them focus specifically on addressing the issue of data heterogeneity.

II. System Model

Assuming a network, which consists of K number of clients that collaborate with a single server for global training. Initially, the server shares the global model at round t as $\omega^{\mathbf{t}}$. The k-th client can be denoted by c_k, and it has D_k data, consisting of n_k samples and L_k labels. After receiving the global model, the client initializes local training and updates the model accordingly. The locally trained model of the k-th client is forwarded to the server as $\omega_{\mathbf{k}}^{\mathbf{t+1}}$. Afterwards, the server aggregates the received local model weights of all K number of clients and updates the global model at round $t+1$ as $\omega^{\mathbf{t+1}}$.

The proposed model divides the local models into two groups based on heterogeneity level as shown in Fig. 1.

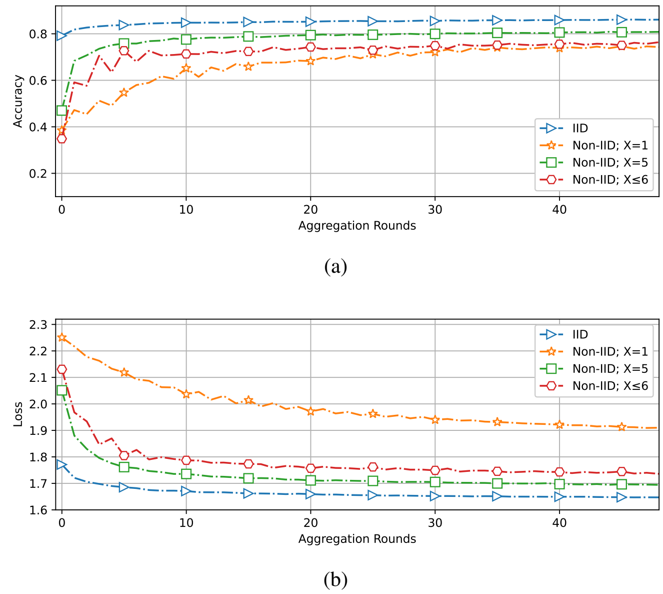


Fig. 2: Comparison Global Accuracy and Loss with various distribution of data among clients: (a) Accuracy (b) Loss

III. Problem Statement

The idea of conventional distributed learning is to improve the global model by simple averaging. On the contrary, the FedAvg trains the global model by evaluating the weighted average of local models [4], [18]. When server forwards the global model, the k-th client computes gradients with a fixed learning rate of η_k as $\eth_{\mathbf{k}} = \nabla F_k(\omega_{\mathbf{k}}^{\mathbf{t}})$. The local model is updated as

$$\omega_{\mathbf{k}}^{\mathbf{t+1}} \leftarrow \omega^{\mathbf{t}} - \eta_k \eth_{\mathbf{k}}. \tag{1}$$

The received updated local models from K number of clients at round $t+1$ are aggregated to update the global model as

$$\omega^{\mathbf{t+1}} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \omega_{\mathbf{k}}^{\mathbf{t+1}}. \tag{2}$$

The weighted average of local models is calculated based on the number of samples of each client's data. However, this method can lead to the global model being influenced too heavily by a client with a large number of samples, potentially causing the global model to converge on a local minimum loss point instead of the global minimum.

We studied the effect of client data heterogeneity by dividing it into different label classes. The figure shown below illustrates how the uneven distribution of data across clients affects the global training process. The simulation parameter X controls the number of label classes assigned to each client. The figure clearly shows the degradation of performance when heterogeneity is considered.

IV. Proposed Model

The proposed method incorporates the key parameters of the system, which are essential to develop a realistic environment,

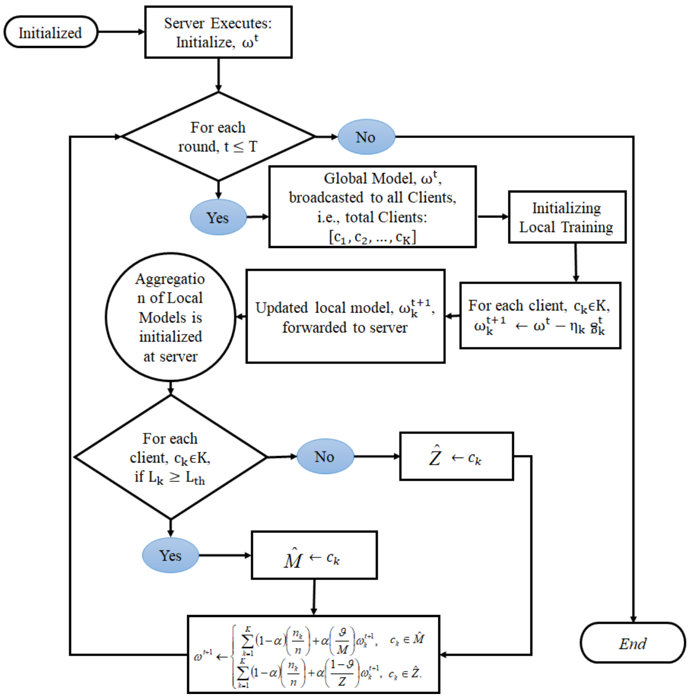
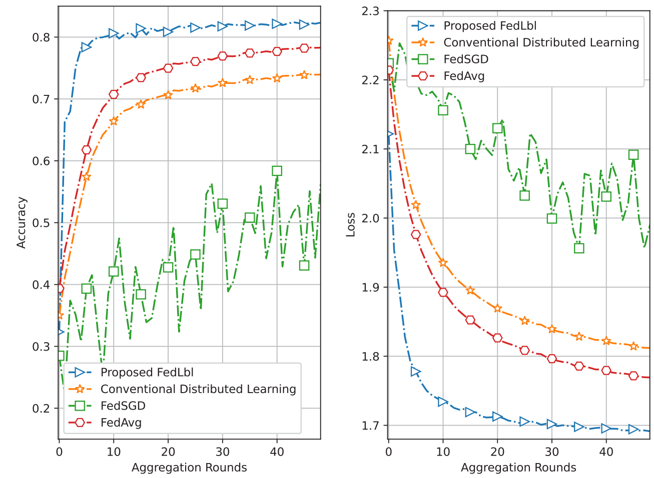


Fig. 3: Proposed FedLbl Method



(a) (b)

Fig. 4: Evaluation of the proposed approach with conventional methods using MNIST dataset with DNN (a) accuracy (b) loss

i.e. the volume and variance of the dataset on which the local client trained the model. When the model at server is shared with all clients at round t as ω^t, the client calculate the gradients as

$$\eth_{\mathbf{k}}^{\mathbf{t}} = \nabla_{\omega_{\mathbf{k}}^{\mathbf{t}}} \ell(\omega_{\mathbf{k}}^{\mathbf{t}}, D_k), \tag{3}$$

where, D_k is dataset of kth client with n_k samples $(x_{k_i}, y_{k_i}), 1 < i < n_k$ as defined earlier. The training objective of that client can written as

$$\min_{\omega_{\mathbf{k}}^{\mathbf{t}} \in \mathbb{R}} \ell(\omega_{\mathbf{k}}^{\mathbf{t}}, D_k), \tag{4}$$

where $\ell(\omega_{\mathbf{k}}^{\mathbf{t}}, D_k)$ is the loss of prediction, which is defined as $\ell(\omega_{\mathbf{k}}^{\mathbf{t}}, D_k) = \frac{1}{n_k} \sum_{(x_{k_i}, y_{k_i}) \in D_k} f_{k_i}(\omega_{\mathbf{k}}^{\mathbf{t}})$. The k-th client updates its local model using the calculated gradients and forwards it to the server as

$$\omega_{\mathbf{k}}^{\mathbf{t+1}} \leftarrow \omega_{\mathbf{k}}^{\mathbf{t}} - \eta_k \eth_{\mathbf{k}}^{\mathbf{t}}. \tag{5}$$

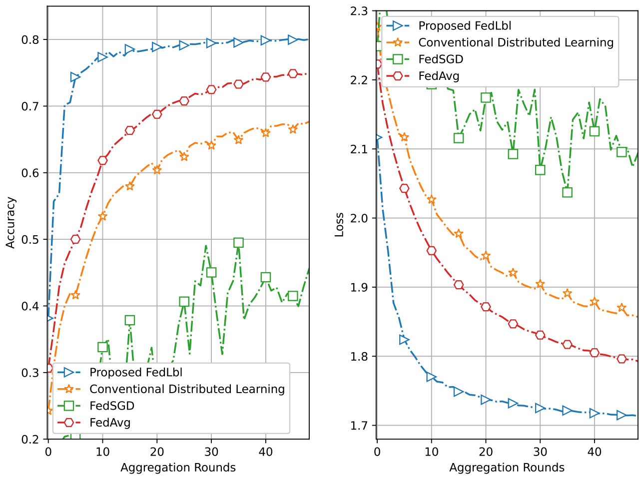
When the local models of all clients are received at the server, they are divided into two groups based on the number of classes of labels. If k-th client's number of classes of labels is less than predefined threshold, i.e., $L_k \leq L_{th}$, the local model of that client is added to the group $\hat{M}$, else it is added to $\hat{Z}$. If we defined M as added local models in $\hat{M}$ and Z is the local models added in $\hat{Z}$, the global model is optimized as

$$\omega^{\mathbf{t+1}} \leftarrow \begin{cases} \sum_{k=1}^{K} (1-\alpha) \left(\frac{n_k}{n}\right) & \\ +\alpha \left(\frac{\vartheta}{M}\right) \omega_{\mathbf{k}}^{\mathbf{t+1}}, & \text{if } c_k \in \hat{M}, \\ \sum_{k=1}^{K} (1-\alpha) \left(\frac{n_k}{n}\right) & \\ +\alpha \left(\frac{(1-\vartheta)}{Z}\right) \omega_{\mathbf{k}}^{\mathbf{t+1}}, & \text{if } c_k \in \hat{Z}, \end{cases} \tag{6}$$

The parameter α is a weightage given to the labels' classes, and it ranges from 0 to 1. During the aggregation process, α controls the weightage given to the clients' local models. If α is 0, the weighted average is entirely based on the volume of clients' training data. On the other hand, if α is 1, the weighted average depends entirely on the variety of training data. Furthermore, the parameter ϑ is the preference given to clients whose models were trained with more classes of labels. Specifically, it represents the weightage given to the group of clients with a more diverse range of label classes. When the server receives the clients' local models, it divides them into two groups: one with more label class variety and the other with less. If ϑ is 1, the weightage is given to the former group, and if it is 0, the whole weightage goes to the latter group. Furthermore, the Fig. 3 describes the proposed FedLbl method's flowchart.

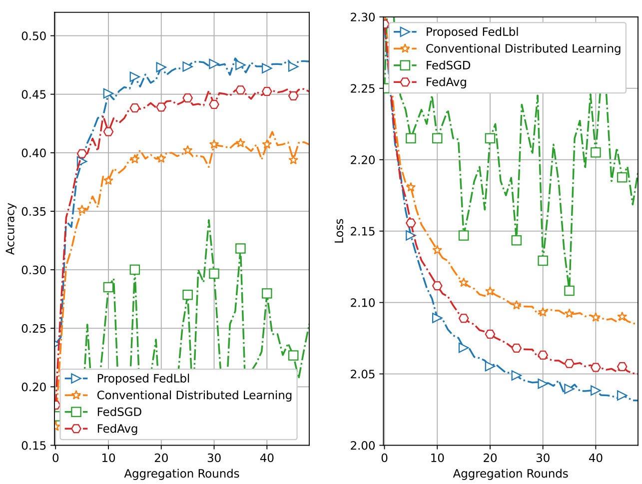
V. Simulation Results

We have validated the effectiveness of our proposed method by training local models on a variety of datasets; MNIST [20], Fashion-MNIST [21], CIFAR-10 [22] and spam/ham emails dataset. For Fashion-MNIST and MNIST datasets, we have employed the DNN architecture featuring one input layer, 4 hidden layers of 200 neurons with a 'relu' activation function, and an output layer with a softmax activation function. Moreover, for CIFAR-10 dataset, clients were deployed with the CNN model which consists of a 2D convolutional layer with 64 filters, where each filter is of size 3×3 with 'Relu' activation function, two hidden layers of 200 neurons with the same activation function and an output layer with 'Softmax' function. Furthermore, for text classification, RNN model with BERT encoder comprised of 12 layers (L-12), a hidden state of 768 (H-768) and A-12 is used.



(a) (b)

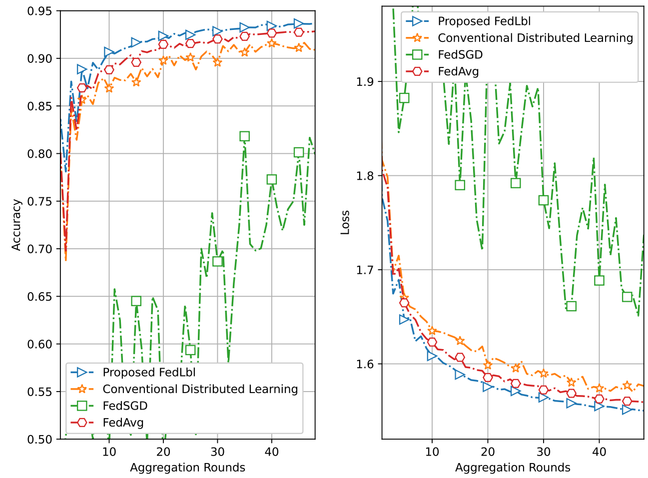
Fig. 5: Evaluation of the proposed approach with conventional methods using Fashion-MNIST dataset with DNN (a) accuracy (b) loss



(a) (b)

Fig. 6: Evaluation of the proposed approach with conventional methods using CIFAR dataset with CNN (a) accuracy (b) loss

Our proposed method is compared with FedAvg along with conventional distributed learning and FedSGD. Fig. 4 and 5 compares the accuracy and loss using MNIST and Fashion-MNIST datasets. It is clear from the figures that the proposed Fedlbl outperforms the conventional methods. Furthermore, we have verified our proposed method by deploying CNN model on end users having CIFAR-10 Data. The proposed method achieved the significant performance as shown in Fig. 6. We have also compared our proposed model by text classification using RNN model with BERT encoder. The proposed method has achieved an enhanced performance in that setting also as shown in Fig. 7.



(a) (b)

Fig. 7: Evaluation of the proposed approach with conventional methods using text dataset with RNN (a) accuracy (b) loss

VI. Conclusion

In this paper, an analysis was conducted to evaluate the performance of federated learning in a diverse and practical setting. The study demonstrated that the heterogeneity in the division of data among clients can have an impact on the performance of global model training. In addition to that, an aggregation algorithm is proposed to overcome this issue, which is well-suited to this environment and obtains better accuracy while minimizing prediction loss. Considering heterogeneous distribution, our proposed approach of segregating the local models for aggregation achieves a significant improvements over the conventional federated learning methods.

Acknowledgment

This research was supported by Brain Pool program funded by the Ministry of Science and ICT through the National Research Foundation of Korea (grant number 2021H1D3A2A02039326) .

References

[1] J. Mills, J. Hu, and G. Min, "Multi-task federated learning for personalised deep neural networks in edge computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 3, pp. 630–641, 2022.

[2] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated learning for 6g communications: Challenges, methods, and future directions," *China Commun.*, vol. 17, no. 9, pp. 105–118, 2020.

[3] D. M. S. Bhatti, R. Khalil, N. Saeed, and H. Nam, "Detection and spatial correlation analysis of infectious diseases using wireless body area network under imperfect wireless channel," *Big Data*, vol. 10, no. 1, pp. 54–56, 2021.

[4] Y. Zhou, Q. Ye, and J. Lv, "Communication-efficient federated learning with compensated overlap-fedavg," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 1, pp. 192–205, 2022.

[5] D. M. S. Bhatti, Y. Rehman, P. S. Rajput, P. Kumar, and D. Kumar, "Machine learning based cluster formation in vehicular communication," *Telecommun Syst*, vol. 78, pp. 39–47, 2021.

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

[7] D. M. S. Bhatti, M. Haris, and H. Nam, "A communication efficient approach of global training in federated learning," in *Int. Conf. on ICTC Convergence (ICTC 2022)*, Jeju, Korea, 2022.
[8] J. Konečný, B. McMahan, and D. Ramage, "Federated optimization: Distributed optimization beyond the datacenter," *CoRR*, vol. abs/1511.03575, 2015. [Online]. Available: http://arxiv.org/abs/1511.03575
[9] D. M. S. Bhatti and H. Nam, "A performance efficient approach of global training in federated learning," in *Int. Conf. on Artificial Intell. in Info. and Commun. (ICAIIC)*, 2023, pp. 112–115.
[10] M. Ben Ghorbel, H. Nam, and M.-S. Alouini, "Soft cooperative spectrum sensing performance under imperfect and non identical reporting channels," *IEEE Commun. Lett.*, vol. 19, no. 2, pp. 227–230, 2015.
[11] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated learning with non-iid data," *CoRR*, vol. abs/1806.00582, 2018. [Online]. Available: http://arxiv.org/abs/1806.00582
[12] E. Jeong, S. Oh, H. Kim, J. Park, M. Bennis, and S. Kim, "Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data," *CoRR*, vol. abs/1811.11479, 2018. [Online]. Available: http://arxiv.org/abs/1811.11479
[13] X. Yao, T. Huang, R. Zhang, R. Li, and L. Sun, "Federated learning with unbiased gradient aggregation and controllable meta updating," *CoRR*, vol. abs/1910.08234, 2019. [Online]. Available: http://arxiv.org/abs/1910.08234
[14] L. Zhang, L. Shen, L. Ding, D. Tao, and L.-Y. Duan, "Fine-tuning global model via data-free knowledge distillation for non-iid federated learning," 2022. [Online]. Available: https://arxiv.org/abs/2203.09249
[15] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," 2018. [Online]. Available: https://arxiv.org/abs/1812.06127
[16] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning: A meta-learning approach," 2020. [Online]. Available: https://arxiv.org/abs/2002.07948
[17] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3710–3722, 2021.
[18] H. B. McMahan, E. Moore, D. Ramage, and B. A. Y. Arcas, "Federated learning of deep networks using model averaging," *CoRR*, vol. abs/1602.05629, 2016. [Online]. Available: http://arxiv.org/abs/1602.05629
[19] Y. Chen, X. Sun, and Y. Jin, "Communication-efficient federated deep learning with layerwise asynchronous model update and temporally weighted aggregation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4229–4238, 2020.
[20] L. Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, 2012.
[21] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," *CoRR*, vol. abs/1708.07747, 2017. [Online]. Available: http://arxiv.org/abs/1708.07747
[22] K. Alex, N. Vinod, and H. Geoffrey, "The cifar-10 dataset," http://www.cs.toronto.edu/ kriz/cifar.html, 25-01-2022.