H-PCFed: Hierarchical Privacy-Enhanced and Communication-Efficient using Federated Learning with Lightweight Deep Learning Model for Industrial IoTs

Ashwanthram A C, Dhivyadarshan G, Gautam R S, Ashwin S, Kirthiga S*
Department of Electronics and Communication Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India
*s_krithiga@cb.amrita.edu

***Abstract*—In IIoT ecosystems, privacy, scalability, real-time decision-making, and communication overhead are the major challenges. Traditional federated learning faces issues such as inefficient model aggregation and difficulty in handling non-IID data distributions. To address these, the proposed framework integrates federated learning with lightweight deep learning models within a fog-cloud hierarchical structure. This combination does improve the efficiency of the aggregation of models, reduces time taken for communication, and ensures privacy using differential privacy techniques. Lightweight deep learning is a critical role in these models, enhancing scalability with robustness to data variety and classification accuracy, fitting the approach particularly well for real-time IIoT applications. This framework highlights privacy preservation, efficiency, and improved model performance to show the potential of FL and lightweight DL models in creating privacy-preserving, real-time IIoT solutions by addressing communication overhead challenges.**

***Index Terms*—Industrial Internet of Things (IIoT), Federated Learning, Lightweight Deep Learning Models, Differential Privacy, Fog-Cloud Architecture, Privacy Preservation, Model Aggregation Efficiency, Non-IID Data Distributions, Communication Overhead.**

I. Introduction

The demand for real-time decision-making and secure communication in modern IIoT ecosystems has been crucial. Centralized machine learning, although accurate, brings tremendous risks such as breaches in privacy, high latency, and inefficiency in a distributed environment. FL tries to decentralize the process of learning but still carries communication overhead, suboptimal aggregation, and is difficult to handle non-IID data distributions across the devices. Therefore, it is less applicable in IIoT applications, especially latency-sensitive and resource-constrained ones.

Federated learning (FL) has drawn attention due to its ability to allow collaborative training with data privacy over distributed environments. One of the studies presents a systematic review of FL in federated cloud systems, discussing challenges such as data heterogeneity and communication efficiency, while emphasizing the need for scalable frameworks in IIoT applications [1]. Another paper presents a verifiable secure aggregation mechanism for FL, with an emphasis on protecting client data during model training and enhancing trust in cross-device FL systems, especially in IIoT environments that contain sensitive data [2]. Furthermore, a hierarchical federated learning approach that includes differential privacy introduces a new privacy budget allocation strategy that ensures high model accuracy while protecting sensitive data in IIoT applications [3].

Lightweight deep learning models have also been explored for the deployment in the IIoT environments, where resources are usually limited. In one research paper, an overview is presented of the lightweight techniques of model compression and pruning to point out the importance of the optimization of deep learning models for real-time applications with minimal energy consumption [4]. Another approach is a modified version of FedAVG which shares a globally shared model with improved convergence rates, while handling model drift. In fact, this approach makes the algorithm more suitable for scenarios in IIoT in which data variability is significant [5]. Additionally, communication efficiency is the issue in FL. It further presents strategies like gradient quantization and model compression with the aim of decreasing the communication overhead in network-constrained and hardware-limited IIoT systems [6].For privacy preservation, a method incorporating LDP with edge FL ensures that client data remains private by adding noise to model updates. This combination provides a promising solution for IIoT applications where data sensitivity is paramount [7].

Lightweight deep learning models, optimized through techniques like pruning and quantization, for the application of weed detection are designed for efficient computation and reduced model size. These models are ideal for IIoT applications, where edge devices require real-time processing with minimal energy and computational resources [8].The Fed Knowledge Distillation (FedKD) model reduces communication overhead and maintains high performance in remote sensing image recognition, while FedKD-HE incorporates homomorphic encryption to preserve privacy in IIoT scenar-

979-8-3315-2422-7/25/$31.00 ©2025 IEEE

ios [9]. A federated knowledge distillation strategy (FKD) addresses model heterogeneity, improving performance and convergence for aggregating local models across edge devices [10]. Additionally, the DaFKD model optimizes FL by considering local model diversity, boosting effectiveness in diverse IIoT settings [11]. The FKD method aggregates knowledge efficiently from local models, enhancing convergence [12]. Lastly, SHFL is introduced to optimize model performance in non-IID FL scenarios, improving communication efficiency and personalization for IIoT models with non-IID data [13].

This paper presents a hierarchical federated learning framework that combines federated transfer learning with lightweight deep learning models to overcome these challenges. The hierarchical approach has incorporated a fog-cloud structure that allows the fog nodes to perform intermediate aggregation in order to eliminate redundant model updates, thereby reducing the overhead of communication. Differential privacy techniques ensure that local model updates remain secure. The use of FL further enhances the performance of the model on diverse and non-IID data, while lightweight deep learning models ensure computational efficiency. The framework is evaluated on real-time IIoT applications, demonstrating improved accuracy, reduced latency, and effective privacy preservation.

II. Methodology

This paper presents a hierarchical FL framework that integrates lightweight DL models within a fog-cloud architecture with the support of AWS services for the entire infrastructure. The proposed methodology addresses some major challenges in IIoT environments, such as communication overhead, non-IID data distributions, and privacy preservation, to enable efficient and secure defect classification in industrial applications.

A. Dataset Preprocessing

The Industrial Images Dataset from Kaggle, comprising 6890 labeled images (3542 defect, 3348 non-defect), simulates non-IID conditions across IIoT devices. Images are resized to 224×224 pixels and pixel values scaled to [0, 1] to align with MobileNetV2's input requirements. Data augmentation includes random rotations, horizontal flips, brightness adjustments, and Gaussian noise, enhancing the diversity of training and improving model robustness and reducing overfitting.

B. Lightweight Model Architecture

The proposed defect classification model is based on MobileNetV2, a lightweight convolutional neural network especially designed for resource-constrained environments, such as devices of IIoT. It applies depthwise separable convolutions and linear bottleneck layers to significantly reduce the computation and memory requirements while maintaining a high level of accuracy. Pre-trained weights from the ImageNet dataset are used to initialize the model so that the baseline for feature extraction can be provided strongly. The final layers are further fine-tuned for binary classification, which includes adding a global average pooling layer followed by a fully connected layer with softmax activation that gives probabilities over the classes of defect and non-defect. The proposed architecture is chosen based on a good balance between computational efficiency and high classification accuracy, which is highly suitable for real-time applications in IIoT.

C. Hierarchical Federated Learning Framework

This proposed framework relies on a hierarchical federated learning method implemented on top of a fog-cloud architecture. Specifically, the fog nodes employ AWS IoT Greengrass and the aggregation would happen at the cloud, implemented through AWS SageMaker. Here, every device performs a MobileNetV2 model's local training on its particular dataset. All data updates are kept secret while updating the trained models sent to the fog layer through AWS IoT Greengrass. Fog nodes collect and aggregate model updates from several devices in their proximity, eliminating redundant updates and reducing the volume of communication before sending the aggregated updates to the central cloud.The central cloud server, hosted on AWS SageMaker, performs global aggregation using Federated Averaging (FedAvg) to aggregate model updates from all the fog nodes. The Federated Averaging algorithm can be defined as follows:

$$\theta_{t+1} = \sum_{k=1}^{K} \frac{N_k}{N} \theta_k^t \tag{1}$$

$$\hat{w} = w + \mathcal{N}(0, \sigma^2) \tag{2}$$

Here Equation (1) denotes Federated Averaging Equation and Equation (2) denotes Differential Privacy Equation. Let be the number of data points across all the devices. This hierarchical aggregation resolves heterogeneity in data and helps reduce communication latency while further improving the overall performance of the model.

D. Privacy Preservation and Efficiency

Differential privacy techniques are applied to the model updates so that individual data points cannot be reconstructed from the aggregated updates. Adding noise to the model updates achieves differential privacy, as shown in the equation above. σ This ensures that the model updates cannot leak sensitive information about individual data points. AWS services such as S3 are used for storage and retrieval of datasets and model updates. Data is stored in S3 and retrieved by devices at the edge, fog, and cloud layers so that all components of the system can retrieve and contribute to model training. AWS S3 is a scalable and reliable solution for handling large volumes of data in IIoT environments. The framework offloads intermediate aggregation to AWS IoT Greengrass, reducing the communication burden on the cloud server and minimizing latency, thereby providing a highly scalable and efficient solution for large-scale IIoT networks with diverse data distributions. The lightweight design of MobileNetV2 reduces computational overhead, and the hierarchical aggregation approach makes the framework efficient and scalable.

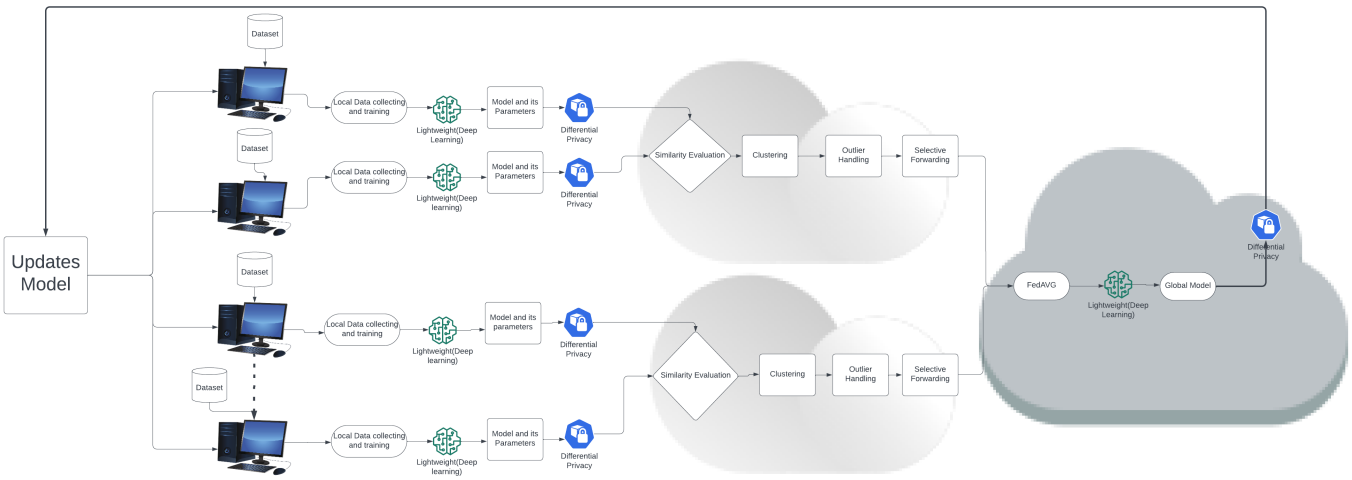

Fig. 1. Architecture of the proposed system.

Thus, it is capable of handling large-scale IIoT networks with minimal latency and is suitable for real-time industrial applications where performance and privacy are paramount.

E. Framework

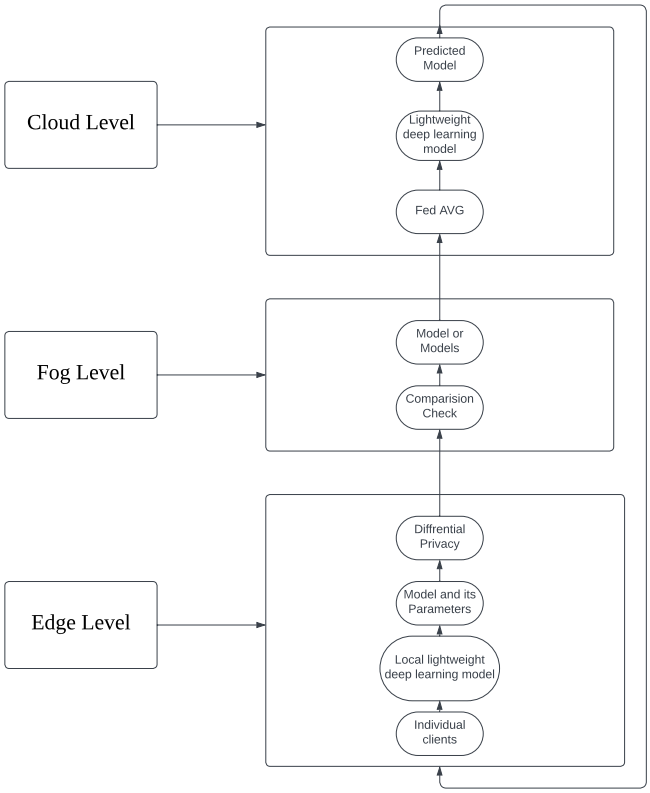

Fig. 2. Hierarchical federated learning framework.

This hierarchical federated learning framework, as shown in Figures 1 and 2, adopts a hierarchical architecture that brings efficiency, privacy, and scalability in distributed machine learning. At the edge level, individual devices—IoT sensors—collect and process local data. Lightweight deep learning models are trained on this data, and their parameters are updated locally. These parameters undergo differential privacy techniques, such as noise addition, before being shared with the fog layer to preserve the privacy of the data collected at the edge.

The fog layer aggregates model parameters from multiple edge devices and only sends unique or aggregated parameters to the cloud, thereby significantly reducing the communication Overhead. This approach not only minimizes bandwidth consumption but also ensures that updates are optimized for efficiency, addressing network constraints common in IIoT environments.

At the cloud layer, global model aggregation is performed using techniques such as Federated Averaging (FedAvg). Lightweight deep learning models are also deployed at the cloud level to decrease overhead computational work and enhance scalability. The updated global model is then disseminated to the fog layer, which subsequently forwards it to the edge devices for further training. This cyclic process ensures continuous refinement of the models.

This iterative process allows the models to learn collaboratively, thereby improving overall accuracy and performance across devices. The introduction of the fog layer enhances communication efficiency, reduces latency, and ensures scalability for large-scale IIoT networks. Moreover, the framework is designed to balance privacy, computational efficiency, and real-time processing needs, making it well-suited for diverse industrial applications.

III. Results

Simulation evaluates H-PCFed, a hierarchical federated learning framework for IIoT defect classification with MobileNetV2 and the Industrial Images Dataset within a fog-cloud framework. The framework has Edge, Fog, and Cloud

TABLE I
DATA TRANSMISSION

Time(s)	H-PCFed(Data transmission in MB)	PCFed(Data transmission in MB)	DPFed(Data transmission in MB)
10	7.2	8.5	9.8
20	11.8	13.7	19.4
30	14.3	20.5	28.2
40	19.1	29.2	37.5
50	23.7	38.4	47.1

TABLE II
BANDWIDTH USAGE

Time(s)	H-PCFed(bandwidth usage in Mbps)	PCFed(bandwidth usage in Mbps)	DPFed(bandwidth usage in Mbps)
10	4.2	6.2	8.8
20	4.0	6.1	8.7
30	3.9	5.9	8.5
40	3.8	5.8	8.4
50	3.7	5.7	8.3

layers, where IIoT edge devices train local MobileNetV2 models. Model updates are federated at the fog layer (AWS IoT Greengrass) and uploaded to the cloud layer (AWS SageMaker) for global aggregation using FedAvg. Differential privacy is implemented by adding Gaussian noise to model updates prior to transmission. Learning rate, batch size, epochs, and differential privacy noise (σ) tuned in relation to the privacy budget (ϵ) are some of the important hyperparameters. The simulation is executed on AWS, with AWS IoT Greengrass being used for fog-level aggregation, AWS SageMaker for cloud-level processing, and AWS S3 for global model update storage, which are fetched by IIoT edge devices for the subsequent round of training.

Table I illustrates the data transmission (in megabytes) for each method across various time intervals. Hierarchical-PCFed has much less data transmission overhead than PCFed and DPFed, especially at higher time intervals. For example, at 50 seconds, Hierarchical-PCFed transmits 23.7 MB, while PCFed and DPFed transmit 38.4 MB and 47.1 MB, respectively. This is indicative of how Hierarchical-PCFed can save redundant communication due to its hierarchical structure.

Table II reports bandwidth utilization in Mbps by the same algorithms and within the same intervals. Hierarchical-PCFed consistently features lower bandwidth usage than the PCFed and DPFed techniques, thus validating its communications optimization capabilities. At 50 s, the Hierarchical-PCFed procedure consumes a mere 3.7 Mbps, contrasted by PCFed: 5.7 Mbps; DPFed: 8.3 Mbps. This should make the Hierarchical-PCFed scalable for deployments in less resource-intensive infrastructures.

The communication time varies for federated learning methods depicted in Figure 3 based on the sampling interval. The hierarchical structure ensures low communication time for Hierarchical-PCFed even for increasing sampling intervals, proving to be an effective way to reduce redundant communication. In case of PCFed, there is a drastic reduction in communication time with an increase in the sampling interval;

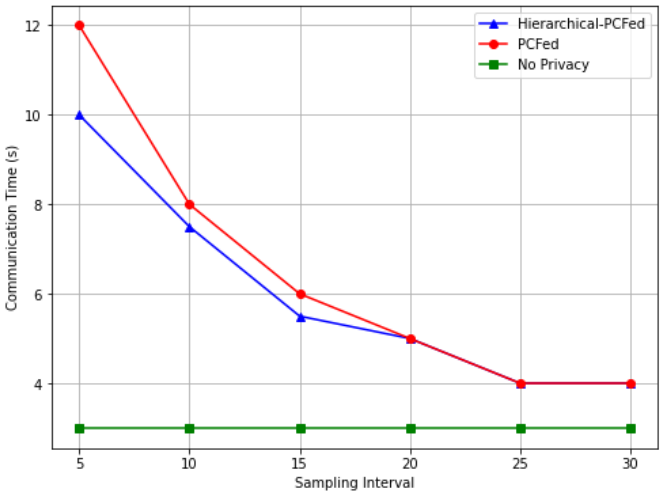

Fig. 3. Communication Consumption versus Sampling Interval.

however, it is greater than the overhead of Hierarchical-PCFed. This "No Privacy" approach has the lowest communication time but compromises data privacy. In general, Hierarchical-PCFed is a very promising approach that balances privacy and efficiency.

Figure 4 shows the accuracy of Hierarchical-PCFed and PCFed for different time slots. Both have a similar trend, but with an increase in the number of training iterations, the accuracy increases. However, Hierarchical-PCFed always performs better than PCFed, which indicates that the hierarchical structure in Hierarchical-PCFed helps in more efficient learning and better generalization. While both methods converge to high accuracy, Hierarchical-PCFed shows faster convergence and better performance, which makes it a more effective solution for federated learning scenarios.

Figure 5 depicts the trade-off between prediction accuracy and privacy budget among various federated learning strategies. In all strategies, with an increasing privacy budget, i.e.,

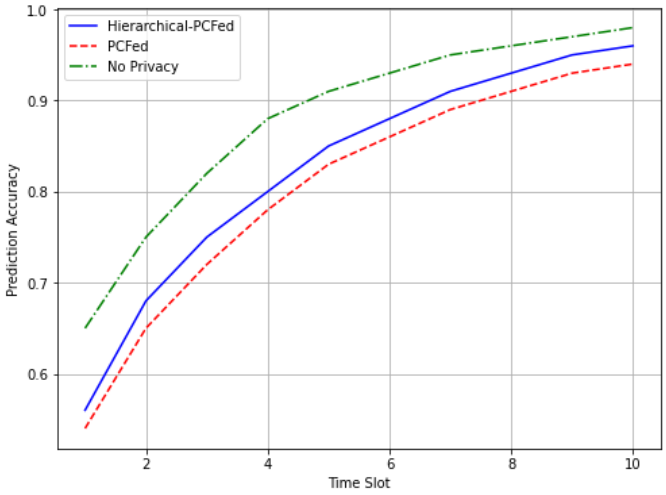

Fig. 4. Model Accuracy versus Time.

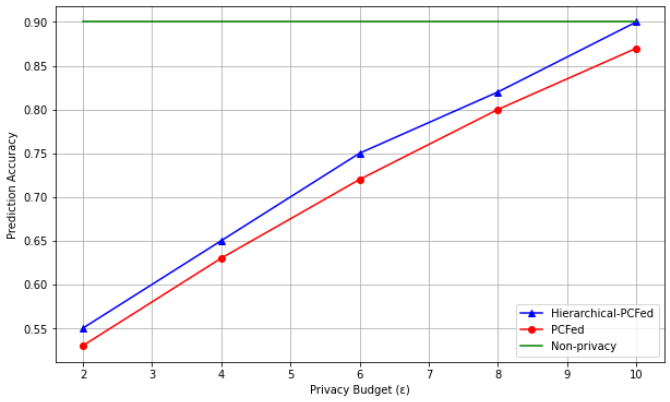

Fig. 5. Model Accuracy versus Privacy Budget.

ϵ, prediction accuracy improves. However, Hierarchical-PCFed consistently outperforms both PCFed and the non-privacy baseline, demonstrating its ability to balance privacy and accuracy effectively. At lower privacy budgets, Hierarchical-PCFed shows a significant advantage over PCFed, ensuring robust performance even when privacy constraints are stringent.

With a larger privacy budget, the accuracy difference between Hierarchical-PCFed and PCFed decreases, but Hierarchical-PCFed is stronger and more privacy-sustaining, particularly at lower privacy budgets. Its high accuracy without sacrificing privacy in IIoT applications and scalability to industrial environments make it suitable for many industrial settings. The approach is superior to conventional federated learning in communication efficiency, privacy, and accuracy. Incorporation of the fog layer and optimised communication save overhead while preserving privacy. Findings from Tables I and II, and Figures 3, 4, and 5 substantiate its excellence for large-scale, privacy-sensitive, and resource-limited deployments.

IV. Conclusion

This study proposed a hierarchical federated learning framework using lightweight MobileNetV2 models within a fog-cloud architecture for efficient and privacy-preserving defect classification in IIoT environments. The approach addresses communication overhead, non-IID data, and privacy issues while ensuring scalability and real-time performance. simulated results on the Industrial Images Dataset validate the framework's efficiency, privacy, and accuracy. In future work, we will extend our framework to HPCFed+ for managing infinite data streams while maintaining privacy protection, and further investigate personalized differential privacy with adaptable budget allocation to assess its effects on privacy preservation and communication efficiency.

References

[1] A. Ajao, O. Jonathan, and E. Adetiba, "The applications of federated learning algorithm in the federated cloud environment: A systematic review," in *2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG)*, pp. 1–15, IEEE, 2024.

[2] F. Luo, H. Wang, and X. Yan, "Comments on "versa: Verifiable secure aggregation for cross-device federated learning"," *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 1, pp. 499–500, 2023.

[3] W. Yuwen, G. Yu, and L. Xiangjun, "Differential privacy hierarchical federated learning method based on privacy budget allocation," in *2023 9th International Conference on Computer and Communications (ICCC)*, pp. 2177–2181, IEEE, 2023.

[4] C.-H. Wang, K.-Y. Huang, Y. Yao, J.-C. Chen, H.-H. Shuai, and W.-H. Cheng, "Lightweight deep learning: An overview," *IEEE consumer electronics magazine*, 2022.

[5] O. B. Tanmoy, M. Al Mamun, S. Hasan, and A. Anwar, "Enhancing federated learning with globally shared model: A modified fedavg approach (gsm-fedavg)," in *2023 6th International Conference on Electrical Information and Communication Technology (EICT)*, pp. 1–6, IEEE, 2023.

[6] O. R. A. Almanifi, C.-O. Chow, M.-L. Tham, J. H. Chuah, and J. Kanesan, "Communication and computation efficiency in federated learning: A survey," *Internet of Things*, vol. 22, p. 100742, 2023.

[7] A. Aminifar, M. Shokri, and A. Aminifar, "Privacy-preserving edge federated learning for intelligent mobile-health systems," *arXiv preprint arXiv:2405.05611*, 2024.

[8] U. Farooq, A. Rehman, T. Khanam, A. Amtullah, M. A. Bou-Rabee, and M. Tariq, "Lightweight deep learning model for weed detection for iot devices," in *2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*, pp. 1–5, IEEE, 2022.

[9] D. G. Nair, C. V. Aswartha Narayana, K. Jaideep Reddy, and J. J. Nair, "Exploring svm for federated machine learning applications," in *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2022*, pp. 295–305, Springer, 2022.

[10] N. S. Bisht and S. Duttagupta, "Deploying a federated learning based ai solution in a hierarchical edge architecture," in *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 247–252, IEEE, 2022.

[11] V. P. Pillai and R. K. Megalingam, "System partitioning with virtualization for federated and distributed machine learning on critical iot edge systems," in *Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2*, pp. 443–453, Springer, 2022.

[12] S. M. Rajagopal, M. Supriya, and R. Buyya, "Fedsdm: Federated learning based smart decision making module for ecg data in iot integrated edge–fog–cloud computing environments," *Internet of Things*, vol. 22, p. 100784, 2023.

[13] F.-H. Tseng and Y.-T. Lai, "Shfl: Selective hierarchical federated learning for non-iid data distribution," in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, pp. 1–6, IEEE, 2024.