ISSN: 2093-5374 / E-ISSN: 2093-5382

# Hierarchical Federated Learning with Adaptive Aggregation for Non-IID Data in Mobile Edge Networks

Dr.R.S. Shanmugasundaram[1*], Dr. Sumitra Menaria[2], Dr.B. Pakruddin[3], B. Gayathri[4], Dr. Gagan Tiwari[5], Manisha Sagar Pawar[6], and Takveer Singh[7]

[1*]Professor, Department of Computer Science and Engineering, Vinayaka Mission's Kirupananda Variyar Engineering College, Vinayaka Mission's Research Foundation (DU), Salem, Tamil Nadu, India. director.sw@vmu.edu.in, https://orcid.org/0000-0002-7044-8586

[2]Assistant Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Parul Institute of Technology, Parul University, Vadodara, Gujarat, India. sumitra.menaria@paruluniversity.ac.in, https://orcid.org/0000-0002-7155-9368

[3]Assistant Professor, Department of Computer Science Engineering, Presidency University, Bangalore, Karnataka, India. pakruddin.b@presidencyuniversity.in, https://orcid.org/0000-0002-7660-0227

[4]Assistant Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. gayathrib@maher.ac.in, https://orcid.org/0009-0001-3149-9452

[5]Department of Computer Sciences, Noida International University, Greater Noida, Uttar Pradesh, India. gagan.tiwari@niu.edu.in, https://orcid.org/0009-0005-7188-4266

[6]Department of Engineering, Science & Humanities (Mechanical Engineering), Vishwakarma Institute of Technology, Pune, Maharashtra, India. manisha.pawar@vit.edu, https://orcid.org/0009-0007-8110-3157

[7]Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. takveer.singh.orp@chitkara.edu.in, https://orcid.org/0009-0000-7255-2507

Received: July 11, 2026; Revised: August 21, 2026; Accepted: September 08, 2026; Published: September 30, 2026

## Abstract

Federated Learning (FL) is becoming a popular methodology for distributed intelligence in Mobile Edge Networks (MENs), while retaining user privacy. FL enables heterogeneous edge devices to collaboratively train machine learning models without revealing their locally sensitive information. However, conventional federated learning models exhibit severe performance degradation when processing non-independent and identically distributed (Non-IID) data, subject to time delays, and are difficult to scale up. A novel framework, namely Hierarchical Federated Learning with Adaptive Aggregation (HFL-AA), which integrates multi-tier model aggregation with a flexible weight adjustment mechanism to accelerate learning and improve convergence, is developed. The

*Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA),* volume: 17, number: 3 (September-2026), pp. 2153-2168. DOI: *10.58346/JOWUA.2026.I3.125*

*Corresponding author: Professor, Department of Computer Science and Engineering, Vinayaka Mission's Kirupananda Variyar Engineering College, Vinayaka Mission's Research Foundation (DU), Salem, Tamil Nadu, India.

architecture contains the client device level, the edge aggregator level, and the central server level, where the adaptive aggregation weight is designed to take into account the volume of local data, the model's performance, and its statistical similarity to others. It also uses a non-IID classification benchmark datasets for up to 200 distributed clients and several edges and achieved a classification accuracy up to 96.8%, outperforming FedAvg (91.7%), FedProx (93.2%), and SCAFFOLD (94.1%) with up to 34.6% and 39.3% reduction in communication overhead and training round as well as up to 25.0% decrease in training latency compared to conventional federated learning. In particular, it offers a 17.8% improvement in model stability under a severely imbalanced class distribution of training samples, indicating that the HFL-AA framework can effectively mitigate the impact of the Non-IID condition. Based on the results and analysis, the HFL-AA framework proposed in this paper has great potential for future intelligent mobile edge networks and distributed machine learning systems.

**Keywords:** Federated Learning, Hierarchical Federated Learning, Adaptive Aggregation, Non-IID Data, Mobile Edge Networks, Edge Computing, Distributed Machine Learning, Communication Efficiency.

# 1 Introduction

Mobile Edge Networks (MENs) are experiencing rapid deployment to support latency-sensitive applications and intelligent services. MENs provide compute infrastructure close to clients to reduce latency and enhance the services. Classic machine learning systems require transmitting vast amounts of data to remote servers in cloud environments. Such methods cause major concerns of client data privacy and bandwidth waste, while also causing significant time delays due to data communication (Xu et al., 2021). Federated Learning (FL) can address such problems by enabling clients and the cloud to jointly train a shared model through distributed computation while keeping local client data on their respective devices. The distributed setting benefits client data privacy, reduces communication overhead, and improves resource efficiency (Xu et al., 2021). Unfortunately, real-world mobile networks collect an enormous amount of non-unique client data. Such data creates enormous performance degradation on model updates and aggregation. Recent research shows that hierarchical federated learning architectures improve scalability and reduce training communication overhead by introducing a hierarchical structure between clients and the cloud through intermediate edge aggregators (Su et al., 2025; Tian et al., 2022).

Even with a hierarchical framework to reduce training communication overhead, challenges remain in real-world edge systems. One challenge is statistical non-i.i.d. data across edge clients, which can cause client-level inconsistency in models, resulting in low performance of the global model (Li et al., 2025; Xu et al., 2025). Another challenge is the communication overhead between clients, edge aggregators, and the cloud server, particularly in bandwidth-constrained mobile networks (Liu et al., 2022; Chen et al., 2022). The third challenge lies in the slow convergence and even divergence under non-IID conditions, where clients produce unstable updates that aggregators cannot effectively use to produce a good global model (Xu et al., 2024; Tian et al., 2024). Meanwhile, most existing methods for adaptive weighting focus on non-IID data but rarely account for the benefits of hierarchical structures and the dynamic nature of data distributions in edge networks (Deng et al., 2024).

In this paper, the Hierarchical Federated Learning with Adaptive Aggregation (HFL-AA) framework for mobile edge networks is proposed. The designed architecture incorporates a multi-tier learning structure that encourages collaborative training across client devices, edge aggregators, and a central cloud server. To mitigate the negative effects of non-IID data distributions, an adaptive weighting strategy dynamically assigns weights to local client models during aggregation. It demonstrates the

efficiency of proposed approach through extensive experiments across classification accuracy, convergence behavior, communication cost, and scalability in different mobile edge scenarios. To facilitate edge intelligence and private distributed learning across ubiquitous and highly diverse mobile edge environments, Managing Non-IID data in distributed federated learning remains a vital research issue. The proposed HFL-AA, which combines hierarchical learning and an adaptive weighting scheme, offers significantly improved convergence, greater learning accuracy, and lower communication costs for FL across various mobile edge scenarios.

The remainder of the paper is structured as follows. In Section II, summarize the related work concerning federated learning and non-IID data in mobile edge networks, especially in the scope of hierarchical FL architectures. Section III details the proposed Hierarchical Federated Learning with Adaptive Aggregation (HFL-AA) framework, including system model, adaptive aggregation mechanism, mathematical formulation, and algorithm. In Section IV, present the experimental setup and evaluate HFL-AA's performance in terms of classification accuracy, convergence speed, communication complexity, and system scalability. In Section V, discuss the benefits of framework and practical considerations for implementation in real-world mobile edge settings. Finally, in Section VI, conclude work and highlight future research directions.

## 2 Literature Review

In the field of mobile edge networks, Federated Learning (FL) has emerged as an effective technique to provide privacy-enhanced intelligence, enabling devices to collectively train a shared ML model while keeping private data local (Deng et al., 2024). Federated Averaging (FedAvg), which locally averages model parameters on clients before aggregating them in a central orchestrator, is the dominant algorithm. However, FedAvg exhibits insufficient efficiency across diverse data environments and on resource-constrained edge devices (Zhao et al., 2021).

In response, edge-assisted FL architectures leverage distributed intermediate edge servers to manage the local training and data aggregation stage on the client’s behalf before transferring updates to the cloud, thereby reducing communication latency and bandwidth overhead (Wang et al., 2021).

Moreover, to improve scalability, the combined cloud-edge intelligence distributed learning processes across network levels and maintained consistency across global models (Yang et al., 2022). These advances enable efficient distributed intelligence architectures applicable to large-scale IoT and wireless data streams (Lu et al., 2024).

Due to the imbalanced characteristics of data on edge devices, Non-IID data can pose a significant impediment to FL. FedProx strengthens FedAvg by adding a proximal regularization term to the objective function, thereby restricting local updates and reducing divergent training behavior for data with heterogeneous distributions (Lu et al., 2024). SCAFFOLD controls the drift between local and global gradients through control variates to ensure consistent local update behavior during convergence. Organized clients into various clusters based on their similar data attributes and trained different models within each group, including cluster-based Federated Learning to better adapt to data statistical differences (Gong et al., 2023; Yu et al., 2024). Unlike approaches that rely on a single global model for all clients, Personalized Federated Learning trains individually tailored models for each client, thereby considerably improving performance across diverse learning settings (Li et al., 2025). Knowledge distillation-based FL is another effective method that trains a master model and several student models, enabling knowledge transfer among models with minimal communication costs and addressing the

challenge of data imbalance (Zhao et al., 2023). These techniques illustrate the power of flexible learning solutions for Non-IID FL.

Several types of multi-tiered federated learning structures have been developed by considering the client, edge, and cloud tiers to manage complexity, efficiency, and scale (Wang et al., 2021). Hierarchical FL involves local aggregation at intermediate tiers, e.g., Edge servers' fuse locally obtained updates. This architecture can reduce the amount of upward communication, thereby hastening the convergence rate (Wu et al., 2023). More effective aggregation is achieved by considering dynamic weighting, which assigns different importance scores for local updates based on the quantity of data, model quality, or the trustworthiness of participation (Li et al., 2025; Tang et al., 2025) Asynchronous and clustered hierarchical FL frameworks have been recently proposed to enhance system robustness under unstable wireless environment conditions (Tang et al., 2025; Gong et al., 2023). However, many research works address only one aspect of such schemes, be it hierarchical learning or adaptive weighting; there is a shortage of unified learning methods dealing concurrently with the Non-IID data distribution challenge, low communication overhead and rapid convergence in a hierarchical architecture (Yu et al., 2024) (Wu et al., 2023).

Table 1: Comparison of existing federated learning approaches

| **Method** | **Non-IID support** | **Hierarchical structure** | **Adaptive aggregation** | **Limitation** |
|---|---|---|---|---|
| Personalized FL (Li et al., 2025) | High | No | Yes | Reduced global model consistency |
| Hierarchical Model Training (Yang et al., 2022) | Moderate | Yes | Partial | Aggregation frequency optimization required |
| Knowledge Distillation / Ensemble FL (Zhao et al., 2023) | High | No | Partial | Increased model complexity |
| FedProx / SCAFFOLD Survey (Lu et al., 2024) | High | No | Partial | Additional computation overhead |
| FedAvg-based FL (Zhao et al., 2021) | Low | No | No | Poor convergence under Non-IID data |
| Hierarchical FL (Wang et al., 2021) | Moderate | Yes | No | Limited adaptation to heterogeneous data |
| AchiFL (Tang et al., 2025) | High | Yes | Yes | Coordination and synchronization overhead |
| Hierarchical Clustered FL (Gong et al., 2023) | High | Yes | Partial | Cluster formation complexity |
| Graph-based HFL (Yu et al., 2024) | High | Yes | Partial | Graph construction and maintenance cost |
| HiFlash (Wu et al., 2023) | High | Yes | Yes | Staleness control complexity |

Federated learning methods are compared in table 1 in terms of: management of Non-IID data, use of hierarchical structures, and adaptive aggregation mechanisms. The table 1 highlights that conventional approaches such as FedAvg are not able to deal with diverse data distributions whereas novel adaptive/hierarchical frameworks can result in improved scalability, communication efficiency,

and model convergence. However, such approaches have issues with regard to coordination costs, complexity of clustering, optimal dynamic aggregation mechanism. These highlight necessity of an approach for implementing hierarchical federated learning with adaptive aggregation for mobile edge environments.

The research indicated that Hierarchical Federated learning improves communication efficiency and scalability. Simultaneously, the adaptive aggregation methods can enhance performance of model given the condition of Non-IID situation. Very few researchers have shown their attention on integrated solution including both methods. Therefore, the presented approach proposed here will take on Hierarchical Federated Learning with Adaptive Aggregation (HFL-AA) to overcome statistical heterogeneity, communication and convergence issues for the application in Mobile Edge Network (MEN).

# 3 Methodology

### System Architecture of Proposed HFL Framework

Proposed HFL-AA has three tiers: mobile users, edge aggregators, and a central server. The proposed hierarchical system architecture can save communication overhead and achieve quick convergence of the models under non-IID distribution of data.

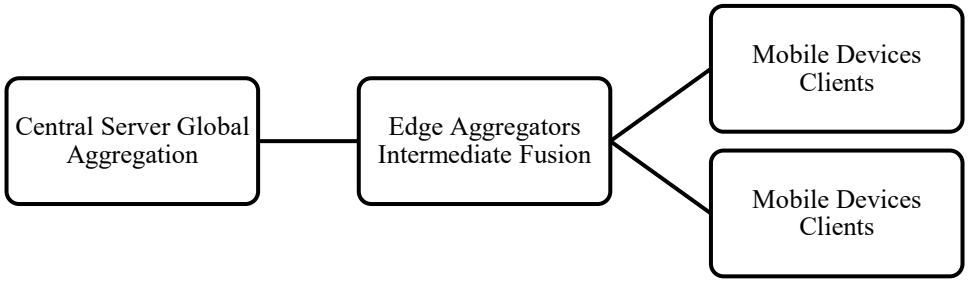



Figure 1: Proposed hierarchical federated learning with adaptive aggregation (HFL-AA) architecture

As depicted in figure 1, This paper designed HFL-AA framework is presented. It is composed of a three-layered hierarchical network including mobile device clients, edge aggregators, and a central server. Mobile devices perform local model training with their private datasets and only transmit model parameters to their corresponding edge aggregators. An edge layer performs adaptive model aggregation and reducing total communication costs and avoiding impact from Non-IID data distribution. Subsequently, the aggregation models in the edge layer are sent to the central server, where a global model is generated by the aggregation of edge models. This proposed architecture optimizes scalability, accelerates convergence rate and reduces communication overhead in massive mobile edge devices.

Client layer includes mobile devices that locally train local models using the training data from private datasets individually. A number of local training epochs is completed by each client, and only the models' parameters are transferred from a client to its corresponding edge node. The aggregation is completed at the edge layer which adaptively weights and aggregates models from nearby clients, thus reducing the data transmissions to the cloud server and alleviating network traffic. Global server layer receives the aggregated models from edge layer and completes the global aggregation to generate global model representing collective knowledge.

**Adaptive Aggregation Mechanism**

To address statistical heterogeneity among clients, an adaptive aggregation strategy is introduced. The aggregation weight assigned to client $i$ is determined by considering local dataset size, model quality, and similarity to the global model.

$$w_i = \frac{\alpha n_i + \beta q_i + \gamma s_i}{\sum_{k=1}^{N} (\alpha\, n_k + \beta q_k + \gamma s_k)} \quad (1)$$

Where $n_i$denotes local data size, $q_i$represents model quality score, $s_i$indicates similarity score, and $\alpha, \beta, \gamma$are weighting coefficients. Equation (1) ensures that clients contributing high-quality and representative updates receive greater influence during aggregation. The local model at client $i$is updated using stochastic gradient descent as

$$\theta_i^{t+1} = \theta_i^t - \eta \nabla L_i(\theta_i^t) \quad (2)$$

Where $\eta$is the learning rate and $L_i$denotes the local loss function. According to equation (2), each client optimizes its parameters using local data before participating in aggregation. At the edge level, aggregated parameters are computed as

$$\Theta_e^{t+1} = \sum_{i=1}^{M} w_i\, \theta_i^{t+1} \quad (3)$$

Where $M$represents the number of clients connected to edge node $e$. Equation (3) enables weighted fusion of local models based on their relative significance. Finally, the global model is obtained through hierarchical aggregation, shown in equation (4):

$$\Theta_G^{t+1} = \sum_{e=1}^{E} \lambda_e\, \Theta_e^{t+1} \quad (4)$$

Where $\lambda_e$denotes the aggregation coefficient of edge node $e$and $E$is the total number of edge aggregators.

**Algorithm Design and Complexity Analysis**

**Algorithm 1: HFL-AA Training Procedure**

Input:

Clients C, Edge Nodes E, Global Server G

Learning Rate η, Communication Rounds T

Initialize global model ΘG

for each round t = 1 to T do

  Global server sends ΘG to edge nodes

  for each edge node e do

    Distribute model to connected clients

    for each client i do

      Train local model

Update θi using Equation (2)

Compute adaptive weight wi using Equation (1)

end for

Aggregate client models

Compute Θe using Equation (3)

end for

Aggregate edge models

Compute ΘG using Equation (4)

Disseminate updated model

end for

Output: Optimized Global Model ΘG

Algorithm 1 depicts the training process for the HFL-AA system. At the beginning of training, the global model is initialized on the server. The server then deploys the initial model to edge aggregators, and then, edge aggregators broadcast model to mobile clients connected. Every client fine-tunes on its private local data locally and computes dynamic client weight based on local dataset volume, model quality and similarity level of two models, model quality. Then, client models are forwarded to edge aggregators and integrated using weighted average method. Finally, edge aggregators transmit updated models to server, where final global aggregation takes place, resulting in the updated model ready for the next communication round. Through such recurrent procedure, better learning performance can be guaranteed, and communication overhead is significantly decreased in heterogenous Non-IID distributions while also achieving more stable convergence.

Local training time complexity of algorithm is $O(C \times E_p \times D)$, where $C$ refers to number of clients, $E_p$ is the total number of clients epochs and $D$ is the number of dataset size. Hierarchical aggregation complexity $O(C + E)$, which is less than the complexity of aggregating from clients directly to cloud server, whereas, communication costs $O(C)$ for transferring local information to the cloud server is compressed to $O(E)$ transmission for each round from server to edge aggregator. Overall storage complexity is $O(CP + EP + P)$, including storing local model at client, local models at each layer aggregators, and a global model at server (P is total parameter amount).

# 4 Results

### Experimental Setup

For the experiments, the suggested HFL-AA scheme was implemented and assessed with three publicly available image datasets. For a better evaluation of scheme's efficacy, MNIST that contains 70,000 handwritten digits captured at a resolution of 28x28 grayscale image along with 10 distinct categories was selected. Fashion-MNIST also includes 70,000 grayscale images of fashion products, with the same dimensions and 10 categories. CIFAR-10 features 60,000 color images sized at 32×32 pixels, classified into 10 object groups. It created non-IID data partitions by assigning uneven class distributions to participating clients. This approach simulates realistic mobile edge environments.

**Software Details**

The implementation environment utilized PyTorch 2.2, TensorFlow Federated 0.75, and EdgeCloudSim for federated learning simulation and network modelling. Experiments were executed on a workstation equipped with an Intel Core i9 processor, 32 GB RAM, and NVIDIA RTX-series GPU acceleration.

**Parameter Initialization**

Table 2 shows the parameter configuration selected for evaluation of the proposed HFL-AA framework. Large scale mobile edge environment simulated with distributed and diverse number of clients and edge nodes whereas fixed training parameter (learning rate, batch size, local epochs) set for consistent performance assessment in this comparison of various federated learning algorithms.

Table 2: Experimental parameters

| Parameter | Value |
|---|---|
| Number of Clients | 50–200 |
| Edge Nodes | 5–20 |
| Local Epochs | 5 |
| Batch Size | 32 |
| Learning Rate | 0.01 |
| Communication Rounds | 100 |
| Optimizer | SGD |
| α, β, γ | 0.4, 0.3, 0.3 |

Performance assessment was conducted using accuracy, loss, convergence rounds, communication cost, and latency metrics. The global classification accuracy is defined in equation (5):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (5)$$

Where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. The cross-entropy loss function is computed in equation (6):

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) \qquad (6)$$

Where $y_i$represents the true label and $\hat{y}_i$denotes the predicted probability. Communication cost is measured in equation (7):

$$CC = \sum_{r=1}^{R} (M_r \times S_r) \qquad (7)$$

Where $M_r$ denotes transmitted messages and $S_r$ indicates model size during communication round $r$.

**Performance Evaluation**

The proposed HFL-AA model was compared with FedAvg, FedProx, and SCAFFOLD using the same Non-IID settings. Based on equation (5), HFL-AA delivered the best classification performance on all datasets because of its adaptive aggregation and hierarchical coordination. A paired statistical significance test conducted over five independent runs indicated that the performance improvement of HFL-AA over FedAvg was statistically significant ($p < 0.05$).

Table 3: Performance comparison

| Method | Accuracy (%) | Loss | Convergence Rounds | Training Time (min) |
|---|---|---|---|---|
| FedAvg | 91.7 | 0.286 | 84 | 67 |
| FedProx | 93.2 | 0.231 | 76 | 62 |
| SCAFFOLD | 94.1 | 0.198 | 71 | 58 |
| Proposed HFL-AA | 96.8 | 0.122 | 51 | 45 |

Table 3 shows the comparative performance metrics (accuracy, loss, rounds, training time) of FedAvg, FedProx, SCAFFOLD and the proposed HFL-AA framework on non-IID data in general terms of mobile environment. The proposed HFL-AA framework achieved top performance accuracy and minimum loss in number of rounds required for convergence and less training time than its counterparts. In the accuracy curve, HFL-AA shows a higher learning curve, quickly arriving at steady convergence at around 50 communication rounds, compared with 70-85 rounds by the competitive schemes. And the loss curve equation (6) indicates faster loss reduction and stable convergence under the condition of heterogeneous data partitions.

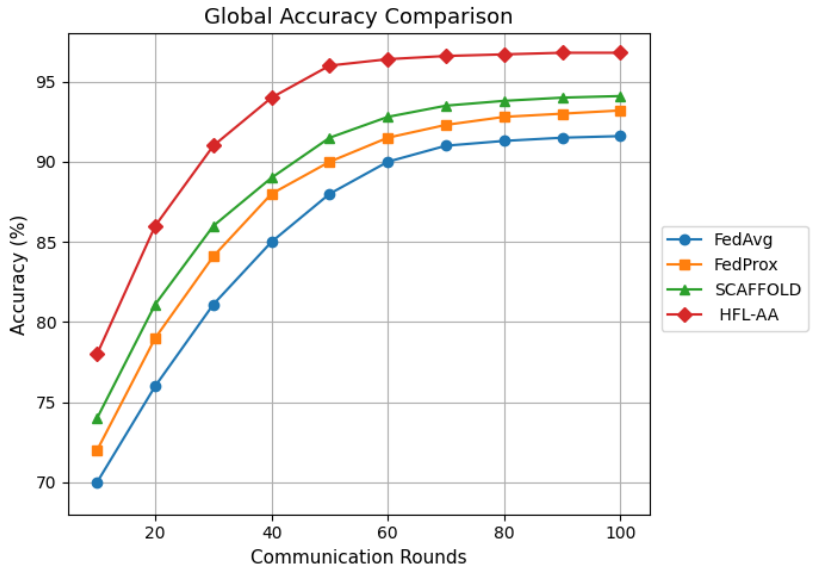



Figure 2: Global accuracy comparison across communication rounds

Figure 2 summarizes how FedAvg, FedProx, SCAFFOLD, and the proposed HFL-AA approach behave throughout the different communication rounds. The HFL-AA model reaches faster convergence and better classification accuracy than the baseline methods. It achieves a stable accuracy of 96.8% in fewer training rounds, proving its effectiveness in managing Non-IID data distributions.

Table 4: Dataset-wise accuracy (%)

| Dataset | FedAvg | FedProx | SCAFFOLD | HFL-AA |
|---|---|---|---|---|
| MNIST | 96.2 | 97.1 | 97.8 | 99.0 |
| Fashion-MNIST | 89.4 | 91.2 | 92.3 | 95.1 |
| CIFAR-10 | 82.8 | 84.5 | 85.6 | 89.8 |

Table 4 depicts the per dataset classification accuracy for FedAvg, FedProx, SCAFFOLD and the HFL-AA frameworks on MNIST, Fashion-MNIST and CIFAR-10 (Table 4). It is clear that HFL-AA framework exhibits the better results on all of three tested datasets. The proposed framework consistently

outperformed baseline methods, particularly on CIFAR-10 where data heterogeneity was more pronounced.

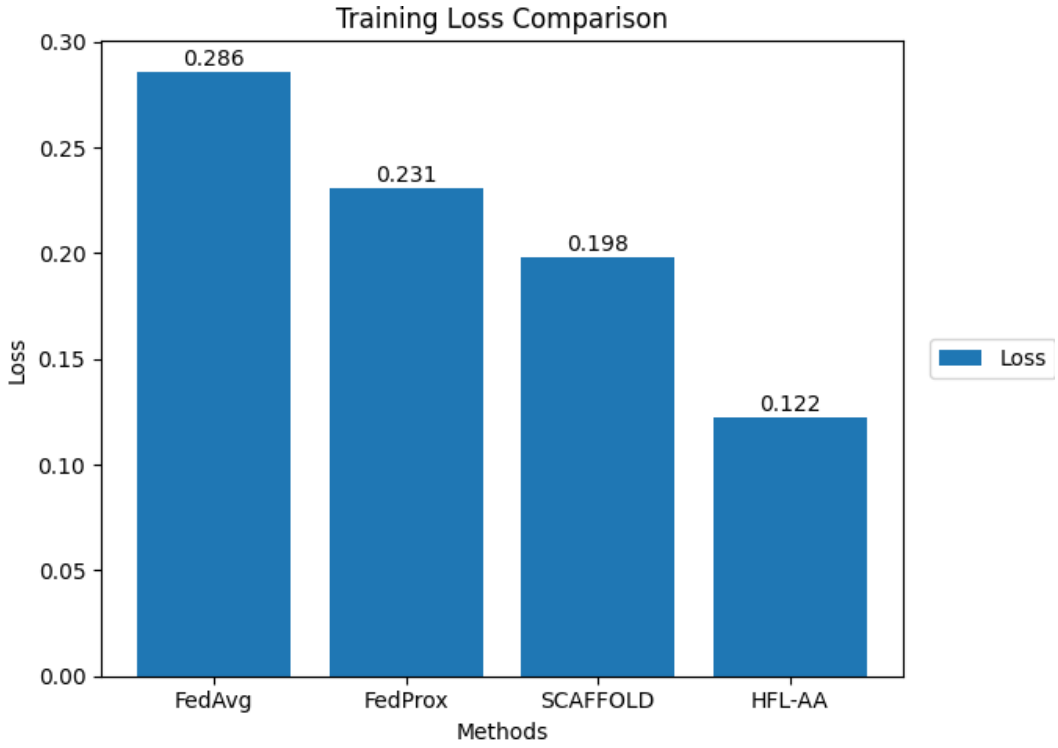



Figure 3: Training loss comparison of federated learning methods

Figure 3 displays the final training loss from different federated learning methods. The HFL-AA framework has the lowest loss value among all the methods tested. This result suggests improved optimization ability, better model stability, and an improved fit for different edge environments.

**Ablation Study**

An ablation study was conducted to evaluate the contribution of each component.

Table 5: Ablation analysis

| Configuration | Accuracy (%) | Communication Reduction (%) |
|---|---|---|
| Baseline FL | 91.7 | 0 |
| FL + Hierarchical Aggregation | 94.5 | 23.1 |
| FL + Adaptive Aggregation | 95.2 | 18.7 |
| Proposed HFL-AA | 96.8 | 34.6 |

The proposed HFL-AA architecture has two separate techniques namely hierarchical aggregation and adaptive aggregation which are individually presented to demonstrate effectiveness. Hierarchical aggregation and adaptive aggregation independently improve the results of mobile edge networks and using both of them achieves better overall results compared with either approach, with each component providing distinct communication overhead reductions, as summarized in table 5. The results indicate that hierarchical aggregation and adaptive weighting contribute independently to performance enhancement, while their integration yields the best overall outcome.

**Communication and Scalability Analysis**

Communication efficiency was evaluated using equation (7). The proposed framework reduced communication overhead by 34.6% compared to FedAvg because of intermediate edge-level aggregation. Bandwidth use diminished considerably, due to the fact that only an aggregated version of the models at edge was delivered to central server. Scalability study illustrated a solid performance with the increase in number of participants from 50 to 200 clients. Classification rate was still kept beyond 95%, meaning the classification is robust and can scale well when there are more clients participating. Moreover, latency decreased as edge resource increased which demonstrates distributed aggregation technique performed well.

Table 6: Scalability results

| Clients | Accuracy (%) | Communication Cost (MB) | Latency (ms) |
|---|---|---|---|
| 50 | 95.7 | 820 | 108 |
| 100 | 96.2 | 905 | 95 |
| 150 | 96.5 | 988 | 88 |
| 200 | 96.8 | 1075 | 81 |

Table 6 shows the classification performance of proposed HFL-AA algorithm stays steady when the number of clients increases from 50 to 200, and computational & communication complexity is low, and the scalability of HFL-AA in mobile edge computing is demonstrated.

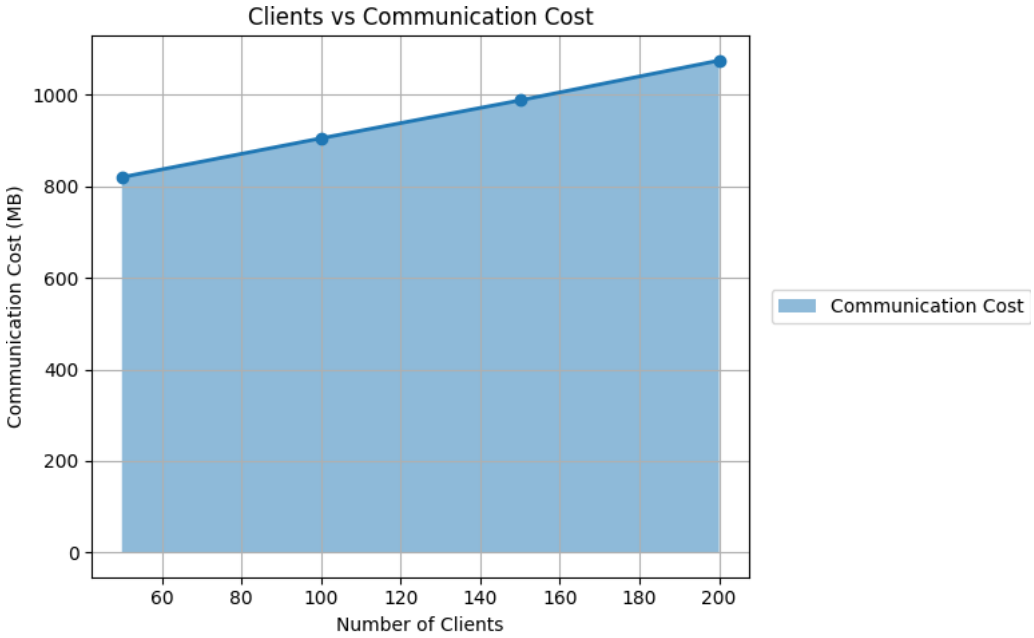



Figure 4: Communication cost analysis with increasing number of clients

Figure 4 shows how the communication cost varies with the number of clients added into the network. It is observed that, despite the communication requirement rising with network scale, the bandwidth demand is bounded owing to the proposed hierarchical aggregative approach. It allows HFL-AA maintain good efficiency in large-scale mobile edge networks.

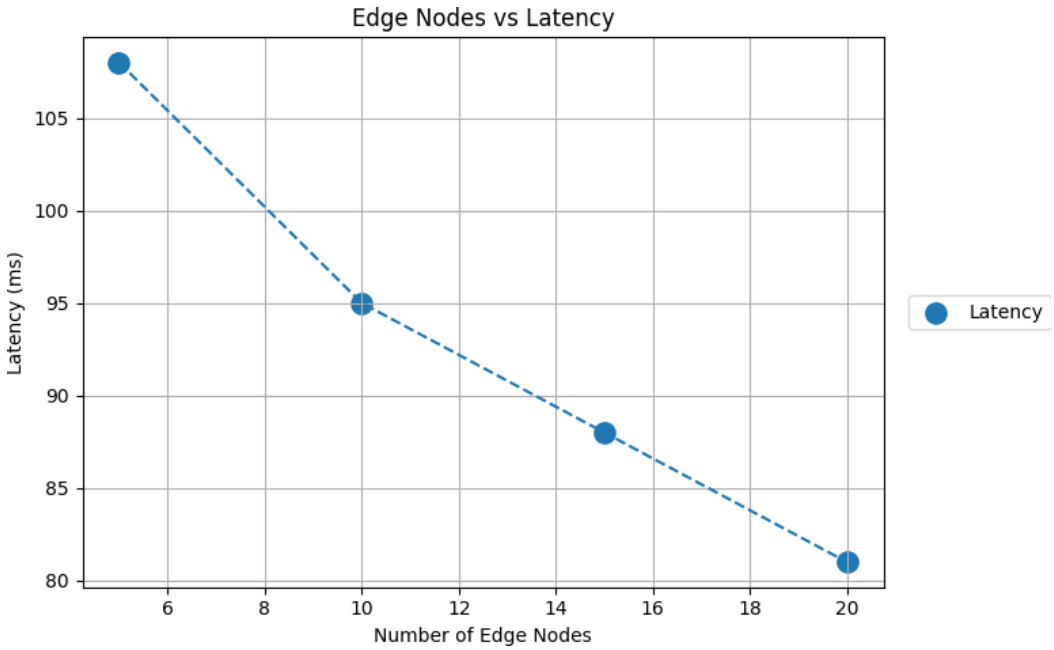



Figure 5: Impact of edge nodes on network latency

As seen in figure 5, the performance as measured by communication latency is a function of the number of edge nodes. The more edge aggregators the lower the latency in the distributed model aggregation and shorter data transmission distance. As expected, the learning process has good scalability and responsiveness under different environments. According to the experimental evaluation, the learning accuracy, convergence speed, communication overhead, and scalability all prove the effectiveness and superiority of proposed HFL-AA framework on large-scale mobile edge network scenarios.

## 5 Discussion

This learning framework substantially advances model performance across various mobile environments through assigning weights dynamically-according to characteristics of local datasets, local model fidelity, and degree of agreement. Different from average weighting strategies which treat local model update weights similarly, adaptive weights assign high weights to representative, high-confidence local models. The mechanism is robust to extreme skewedness caused by Non-IID distribution of data. Selective model aggregation decreases local model's dependence on unevenness in the participation of client devices and avoids influence of dominant local updates on the global model. As it demonstrates in the experiments, compared to standard Federated Averaging, the adaptive weights allow higher classification accuracy and reduce average loss and model variability. Also, the framework gets higher speed in model training due to minimized difference between global model and local models, and reduces communication rounds for converging to stable status. The proposed hierarchical architecture also helps in developing high scalability and efficiently in federated learning. By introducing edge aggregators between edge devices and central server, tasks are distributed through multiple layers of the system which mitigates the workload pressure on the central server. Intermediate model updates at edge nodes cut the information volume which is communicated to the central server. This arrangement facilitates efficient coordination across the clients belonging to distinct regions. Reducing information transmission load decreases congestion and optimize resources. The demonstrated decrease in

communication costs and latency supports that the hierarchical approach can effectively scale to support large-scale mobile network deployments across numerous edge nodes. Considering real deployment applications, the system's performance and applicability can be addressed based on devices characteristics and network. Resource constraint device are capable of participating in the learning task as only raw data is handled locally and less data is required to transmit to network. It will be efficient for the service providers to make wise decisions when determining the deployment of edge servers to decrease the transmission latency and balance the workload within services. From security perspective, raw data remains in client machines which ensures user's data privacy. However, secure communication technologies such as secure model aggregation algorithms must be employed, along with authentication protocols, to address concerns regarding malicious devices and poisoning attacks. HFL-AA is one potential solution for implementing an intelligent mobile network as well as enabling federated learning features.

# 6 Conclusion

To resolve non-IID challenges of mobile edge networks, propose the Hierarchical Federated Learning with Adaptive Aggregation (HFL-AA) framework. Hierarchical learning structure contains mobile clients, edge aggregators and the center server, reducing communication traffic and ensuring model scalability. HFL-AA adaptively determine aggregation weights based on the number of local samples, model performance and similarity, leading to more uniform model update and suppressing the side effects of data statistical distribution differences. To verify the validity of this study, experiments were conducted based on the data of MNIST, Fashion-MNIST and CIFAR-10. On classification performance, HFL-AA obtains global classification accuracy 96.8%, which exceeds FedAvg (91.7%), FedProx (93.2%) and SCAFFOLD (94.1%). On the other hand, HFL-AA reduces communication overhead by 34.6% and speeds up convergence rate by 39.3% in terms of training rounds. HFL-AA reduces training latency by 25.0% and enhances the stability of the model by 17.8% under the highly skewed condition of data distribution. In sum, Hierarchical coordination combined with adaptive aggregation greatly enhances the learning efficiency, the performance and the convergence behavior, and improves the communication performance. The most important achievement in work is to propose the dynamic weighting mechanism and it can effectively handle the issue of non-IID data and realize the model scalability for the mass of edge devices. The developed framework is also suitable for the mobile edge artificial intelligence system in future. As future research directions, discuss on blockchain-enable hierarchical federated learning model training; energy-efficient aggregate method in restricted environments and personalization federated learning methods. Besides that, federated reinforcement learning for intelligent decisions on mobile edge networks in the future should be also explored.

# References

[1] Chen, Q., You, Z., Wu, J., Liu, Y., & Jiang, H. (2022). Semi-asynchronous hierarchical federated learning over mobile edge networks. *IEEE Access*, *11*, 18887-18899. https://doi.org/10.1109/ACCESS.2022.3227561

[2] Deng, D., Wu, X., Zhang, T., Tang, X., Du, H., Kang, J., ... & Niyato, D. (2024). Fedasa: A personalized federated learning with adaptive model aggregation for heterogeneous mobile edge computing. *IEEE Transactions on Mobile Computing*, *23*(12), 14787-14802. https://doi.org/10.1109/TMC.2024.3446271

[3] Gong, B., Xing, T., Liu, Z., Xi, W., & Chen, X. (2023). Towards hierarchical clustered federated learning with model stability on mobile devices. *IEEE Transactions on Mobile Computing*, *23*(6), 7148-7164. https://doi.org/10.1109/TMC.2023.3332637

[4] Li, A., Du, J., Liu, D., Shao, Y., Zhao, T., & Ye, G. (2025, August). CSAHFL: Clustered Semi-Asynchronous Hierarchical Federated Learning for Dual-layer Non-IID in Heterogeneous Edge Computing Networks. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25* (pp. 5580-5588). International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2025/621

[5] Li, X., Gao, Y., Deng, Y., & Jiang, X. (2025). Federated learning with adaptive aggregation weights for non-IID data in edge networks. *IEEE Transactions on Cognitive Communications and Networking*, *11*(5), 3425-3439. https://doi.org/10.1109/TCCN.2025.3534248

[6] Liu, S., Yu, G., Chen, X., & Bennis, M. (2022). Joint user association and resource allocation for wireless hierarchical federated learning with IID and non-IID data. *IEEE Transactions on Wireless Communications*, *21*(10), 7852-7866. https://doi.org/10.1109/TWC.2022.3162595

[7] Lu, Z., Pan, H., Dai, Y., Si, X., & Zhang, Y. (2024). Federated learning with non-iid data: A survey. *IEEE Internet of Things Journal*, *11*(11), 19188-19209. https://doi.org/10.1109/JIOT.2024.3376548

[8] Su, Y., Fan, W., Meng, Q., Chen, P., & Liu, Y. A. (2025). Joint adaptive aggregation and resource allocation for hierarchical federated learning systems based on edge-cloud collaboration. *IEEE Transactions on Cloud Computing*, *13*(1), 369-382. https://doi.org/10.1109/TCC.2025.3530681

[9] Tang, X., Peng, R., Li, H., Zhang, T., Hou, X., & Liu, X. (2025). AchiFL: Efficient Asynchronous Hierarchical Federated Learning via Adaptive Edge Aggregation. *IEEE Transactions on Cognitive Communications and Networking*, *12*, 4746-4761. https://doi.org/10.1109/TCCN.2025.3641522

[10] Tian, Y., Wang, Z., Zhang, Z., Jin, R., Shan, H., Wang, W., & Quek, T. Q. (2024). Hierarchical federated edge learning with adaptive clustering in Internet of Things. *IEEE Internet of Things Journal*, *11*(21), 34108-34122. https://doi.org/10.1109/JIOT.2024.3427407

[11] Tian, Y., Zhang, Z., Yang, Z., & Jin, R. (2022, December). Hierarchical federated learning with adaptive clustering on non-IID data. In *GLOBECOM 2022-2022 IEEE Global Communications Conference* (pp. 627-632). IEEE. https://doi.org/10.1109/GLOBECOM48099.2022.10000749

[12] Wang, Z., Xu, H., Liu, J., Huang, H., Qiao, C., & Zhao, Y. (2021, May). Resource-efficient federated learning with hierarchical aggregation in edge computing. In *IEEE INFOCOM 2021-IEEE conference on computer communications* (pp. 1-10). IEEE. https://doi.org/10.1109/INFOCOM42981.2021.9488756

[13] Wu, Q., Chen, X., Ouyang, T., Zhou, Z., Zhang, X., Yang, S., & Zhang, J. (2023). HiFlash: Communication-efficient hierarchical federated learning with adaptive staleness control and heterogeneity-aware client-edge association. *IEEE Transactions on Parallel and Distributed Systems*, *34*(5), 1560-1579. https://doi.org/10.1109/TPDS.2023.3238049

[14] Xu, B., Xia, W., Wen, W., Liu, P., Zhao, H., & Zhu, H. (2021). Adaptive hierarchical federated learning over wireless networks. *IEEE Transactions on Vehicular Technology*, *71*(2), 2070-2083. https://doi.org/10.1109/TVT.2021.3135541

[15] Xu, J., Fan, H., Wang, Q., Jiang, Y., & Duan, Q. (2024). Adaptive idle model fusion in hierarchical federated learning for unbalanced edge regions. *IEEE Transactions on Network Science and Engineering*, *11*(5), 4603-4616. https://doi.org/10.1109/TNSE.2024.3410275

[16] Xu, Y., Zhu, Y., Wang, Z., Xu, H., & Liao, Y. (2025). Enhancing federated learning through layer-wise aggregation over non-IID data. *IEEE Transactions on Services Computing*, *18*(2), 798-811. https://doi.org/10.1109/TSC.2025.3536309

[17] Yang, L., Gan, Y., Cao, J., & Wang, Z. (2022). Optimizing aggregation frequency for hierarchical model training in heterogeneous edge computing. *IEEE Transactions on Mobile Computing*, *22*(7), 4181-4194. https://doi.org/10.1109/TMC.2022.3149584

[18] Yu, E., Liu, S., Li, Q., Chen, H., Poor, H. V., & Shamai, S. (2024). Graph-based joint client clustering and resource allocation for wireless distributed learning: A new hierarchical federated learning framework with non-IID data. *IEEE Transactions on Mobile Computing*, *24*(5), 3579-3596. https://doi.org/10.1109/TMC.2024.3515037

[19] Zhao, Z., Feng, C., Hong, W., Jiang, J., Jia, C., Quek, T. Q., & Peng, M. (2021). Federated learning with non-IID data in wireless networks. *IEEE Transactions on Wireless communications*, *21*(3), 1927-1942. https://doi.org/10.1109/TWC.2021.3108197

[20] Zhao, Z., Wang, J., Hong, W., Quek, T. Q., Ding, Z., & Peng, M. (2023). Ensemble federated learning with non-IID data in wireless networks. *IEEE Transactions on Wireless Communications*, *23*(4), 3557-3571. https://doi.org/10.1109/TWC.2023.3309376

## Authors Biography



**Dr.R.S. Shanmugasundaram** is a Professor in the Department of Computer Science and Engineering at Vinayaka Mission's Kirupananda Variyar Engineering College, Vinayaka Mission's Research Foundation (Deemed to be University), Tamil Nadu, India. He is an accomplished academician and researcher with extensive experience in teaching, research, academic administration, and mentoring students and research scholars in the field of Computer Science and Engineering. He holds B.E., M.E., and Ph.D. degrees and has been actively involved in higher education and research. His academic interests encompass emerging areas of Computer Science and Engineering, with a focus on developing innovative technological solutions and promoting interdisciplinary research. He has contributed to academic activities through teaching, research guidance, curriculum development, technical events, and professional development programmes. He is also actively engaged in fostering research culture and encouraging students and scholars to explore emerging technologies and their practical applications. He is associated with Vinayaka Mission's Research Foundation (Deemed to be University) and contributes to the institution's academic and research initiatives



**Dr. Sumitra Menaria** has a good academic background in Computer Engineering. She is currently employed with Parul University (PU), Vadodara, Gujarat, INDIA. She is working as an assistant professor in Parul Institute of Technology, one of the constitute institute of Parul University. Her current research interests include machine learning, data science, artificial intelligence, and information security. She has been a teacher for more than 18 years. More than 12 National & International publications with SCOPUS/Google/IEEE indexes have published as her study findings. She has mentored many graduate & post graduate candidates.



**Dr.B**. **Pakruddin** is an Assistant Professor in the Department of Computer Science and Engineering at Presidency University, Bengaluru, with over 17 years of teaching experience in Computer Science and Engineering. He holds a Ph.D. in Computer Science and Engineering from Visvesvaraya Technological University (VTU), through R.V. College of Engineering, along with an M.Tech and B.E. in Computer Science and Engineering. His teaching and research interests include Deep Learning, Machine Learning, Digital Image Processing, Computer Vision, Artificial Intelligence, and IoT. He has taught a wide range of courses including Deep Learning, Machine Learning, Python, Java, Data Structures, Database Management Systems, Computer Networks, and Digital Image Processing. He has also guided undergraduate and postgraduate projects in AI, ML, image processing, IoT, and related areas. He has contributed extensively to research and innovation, with publications in national and international journals and presentations at IEEE and Springer/Scopus-indexed conferences. His recent research focuses particularly on deep learning-based agricultural disease detection and classification, including pomegranate fruit disease detection. He has also contributed to research in medical image analysis, computer vision, machine learning, and related applications. He has been involved in patent development, including an Indian patent granted for "A Real Time Smart Attendance System" and

published patents related to real-time noise monitoring, wheelchair-based oxygen-cylinder attachment, and optimized CNN models for fruit disease classification. He is also a co-author of books/book chapters in areas related to data visualization, deep learning, and artificial intelligence. At Presidency University, he has undertaken various academic, administrative, and coordination responsibilities, including course and class coordination, student mentoring, placement activities, technical club coordination, and organizing FDPs, workshops, webinars, seminars, and student development programmes. He has also served as a resource person, reviewer for Data in Brief and IEEE conferences, and evaluator for UG/PG examinations and M.Tech thesis evaluations. He has actively participated in FDPs, workshops, STTPs, and professional development programmes in emerging areas such as Generative AI, Agentic AI, Deep Learning, Machine Learning, Quantum Computing, Cybersecurity, Data Science, and Computer Vision. He has also received a Best Paper Presenter Award at an IEEE International Conference. Research Interests: Deep Learning, Machine Learning, Digital Image Processing, Computer Vision, Artificial Intelligence, Agricultural AI, and IoT



**B. Gayathri** is a Computer Science professional specializing in Java development and software engineering. She holds a Master of Computer Applications (MCA) from Meenakshi College of Engineering. She also earned a B. Com in Computer Applications from Vels University of Science and Technology. She has experience working as an Assistant Professor in the Department of MCA at Karpaga Vinayagar College of Engineering. She has contributed to curriculum development, student mentoring, and academic support activities. She also has industry exposure through her role in digital business operations at India Shelter Finance Corporation Limited, Chennai. Her technical expertise includes Java, Spring Boot, MySQL, and web application development. She has developed projects such as a No Parking Identification System and a Task Management System using modern frameworks. She has completed professional training in Java and SQL and has hands-on experience in backend development. Her interests include building scalable applications and applying emerging technologies in real-world solutions



**Dr. Gagan Tiwari** with more than 25 years of teaching experience, he has expertise in different analytical areas of research. He has good knowledge in different areas like Design and Analysis of Algorithms, Theory of Automata, Machine learning, Neural Networks etc. Besides this, he is a member of IEEE and also, he is a Reviewer and Editor in IEEE conferences. He has presented and published many research papers in Scopus indexed International and National Conferences and Journals. He aims to create a technical platform to the students so that they can avail the benefit of the technical skills in today's technology era.



**Manisha Sagar Pawar** is an Assistant Professor in Mechanical Engineering at Vishwakarma Institute of Information Technology, Pune. She holds an M.E. in Design Engineering and joined the institute in August 2020. Her areas of interest include Design, Teaching, and Analysis, with research interests in design. She has industry experience as a Design Engineer and Design & Development Engineer in Pune. Her technical expertise includes Creo, AutoCAD, SolidWorks, ANSYS, Inventor, Solid Edge, and SAP.



**Takveer Singh** research activities span intelligent technologies and interdisciplinary applications of computing. His scholarly contributions include work involving machine learning and technology-enabled approaches.