**Clustered Federated Learning: A Study of Approaches for Optimizing Communication and Privacy while Reducing Central Server Burden**

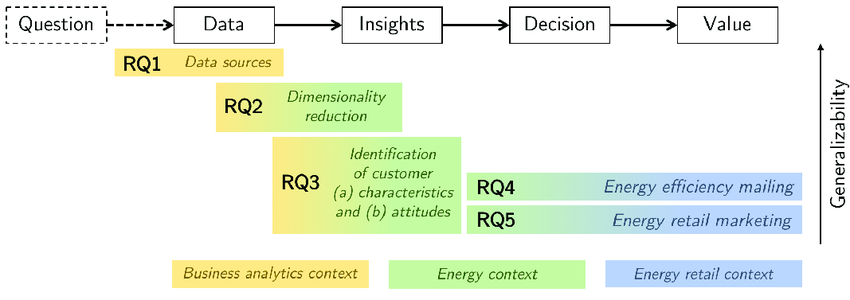
**Abstract:**

**Background:** Federated learning (FL) is a distributed machine learning framework that allows clients to train a shared model without having to share their data with a central server. This is achieved by having the clients collaboratively train the model on their local data, and then periodically sending updates to the central server. The central server then aggregates these updates and sends back a new model to the clients. One of the challenges of FL is that it can be inefficient in terms of communication and privacy. This is because the clients need to communicate with the central server frequently, and the central server needs to store the updates from all of the clients.

**Methods:** In this paper, we study several approaches for optimizing communication and privacy in FL. We first propose a method for clustering the clients based on their data distributions. This allows us to reduce the amount of communication between the clients and the central server, as well as the amount of data that needs to be stored by the central server. Analyzing some databases in the field of machine learning, we systematically examined some of them. We found 40 publications that used qualitative methods.

**Results:** Of the 40 publications, 20 studies address on several datasets, including the MNIST dataset, the CIFAR-10 dataset, and the FEMNIST dataset. Our finding methods can significantly improve the efficiency of FL in terms of communication and privacy.

**Conclusions:** We have studied several approaches for optimizing communication and privacy in federated learning (FL). Our findings suggest that clustered FL is a promising approach for improving the efficiency and privacy of FL. We believe that this work will be of interest to researchers and practitioners who are working on FL.

****

**Research Questions:**

* How can clustering be used to optimize communication in federated learning?
* How can clustering be used to improve privacy in federated learning?
* How can clustering be used to reduce the burden on the central server in federated learning?
* What are the trade-offs between communication, privacy, and central server burden in clustered federated learning?

**Keywords:** Communication efficiency, Privacy preservation, Differential privacy, Data distribution, Model training

**Inclusion/Exclusion criteria:**

The following rationale for searching databases was agreed upon during a research consultation. We used one primary search terms and three cross-references:

Search terms:

* Clustered federated learning

Cross-references for clustered federated learning:

* Communication efficiency: This refers to the amount of data that needs to be exchanged between the clients and the central server during the training process.
* Privacy preservation: This refers to the protection of the privacy of the individual clients' data during the training process.
* Central server burden: This refers to the amount of computational and storage resources that are required by the central server during the training process.

1. Current structure 1:Many relationship

Prosed Model:

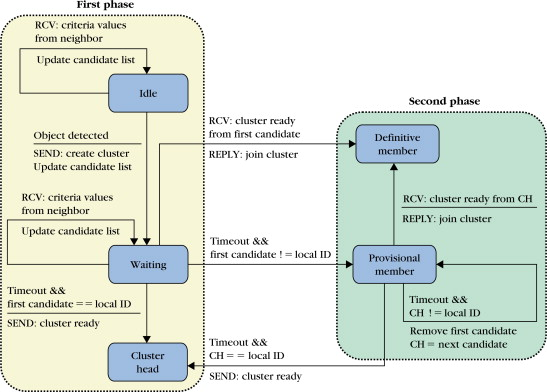
1. Tree Like structure Model -> Leaves -> branch -> Root

a. Reduce communication overload

b. logN times communication load will be reduced

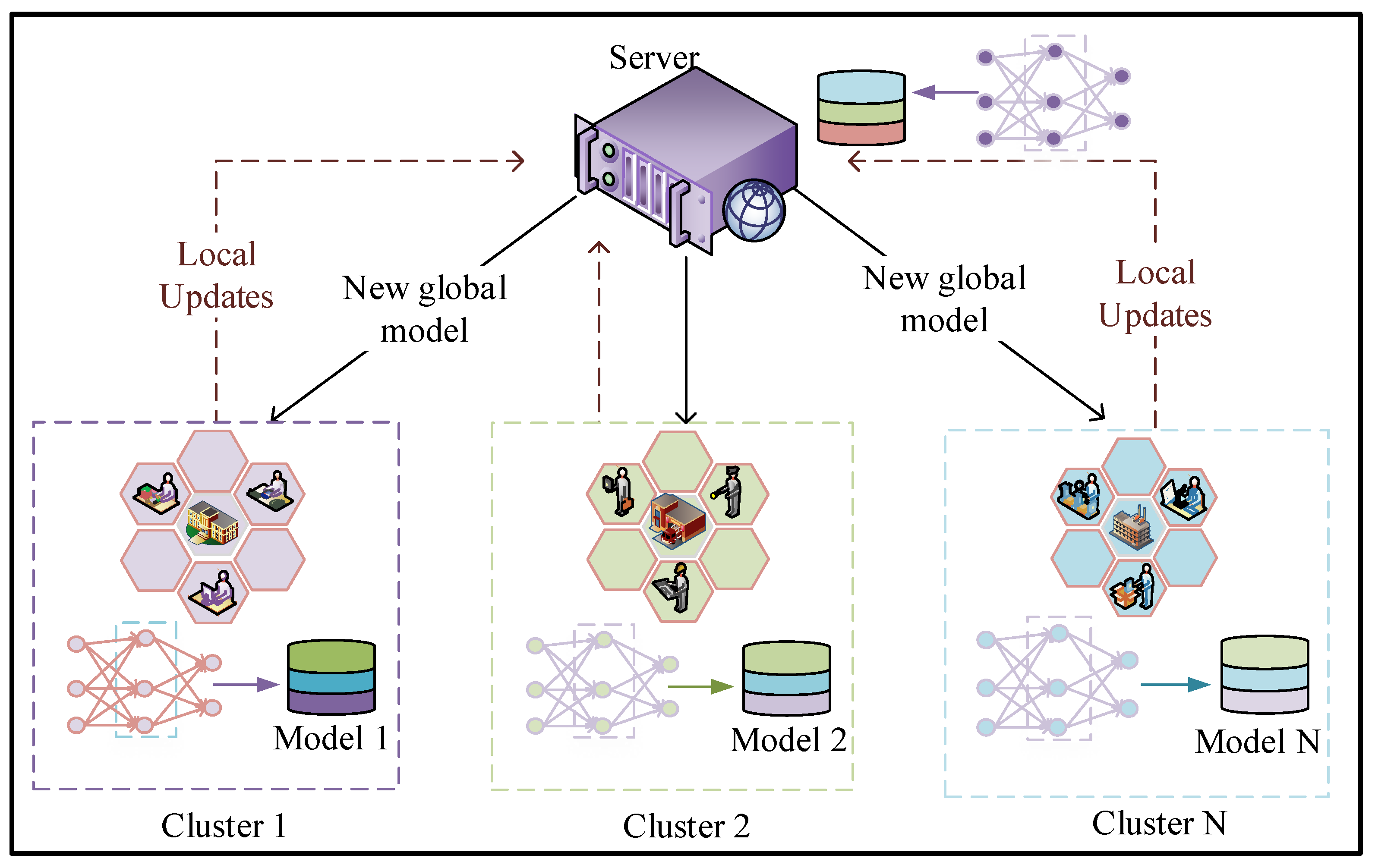
2. Incentive model needed otherwise why someone wants to be a cluster head

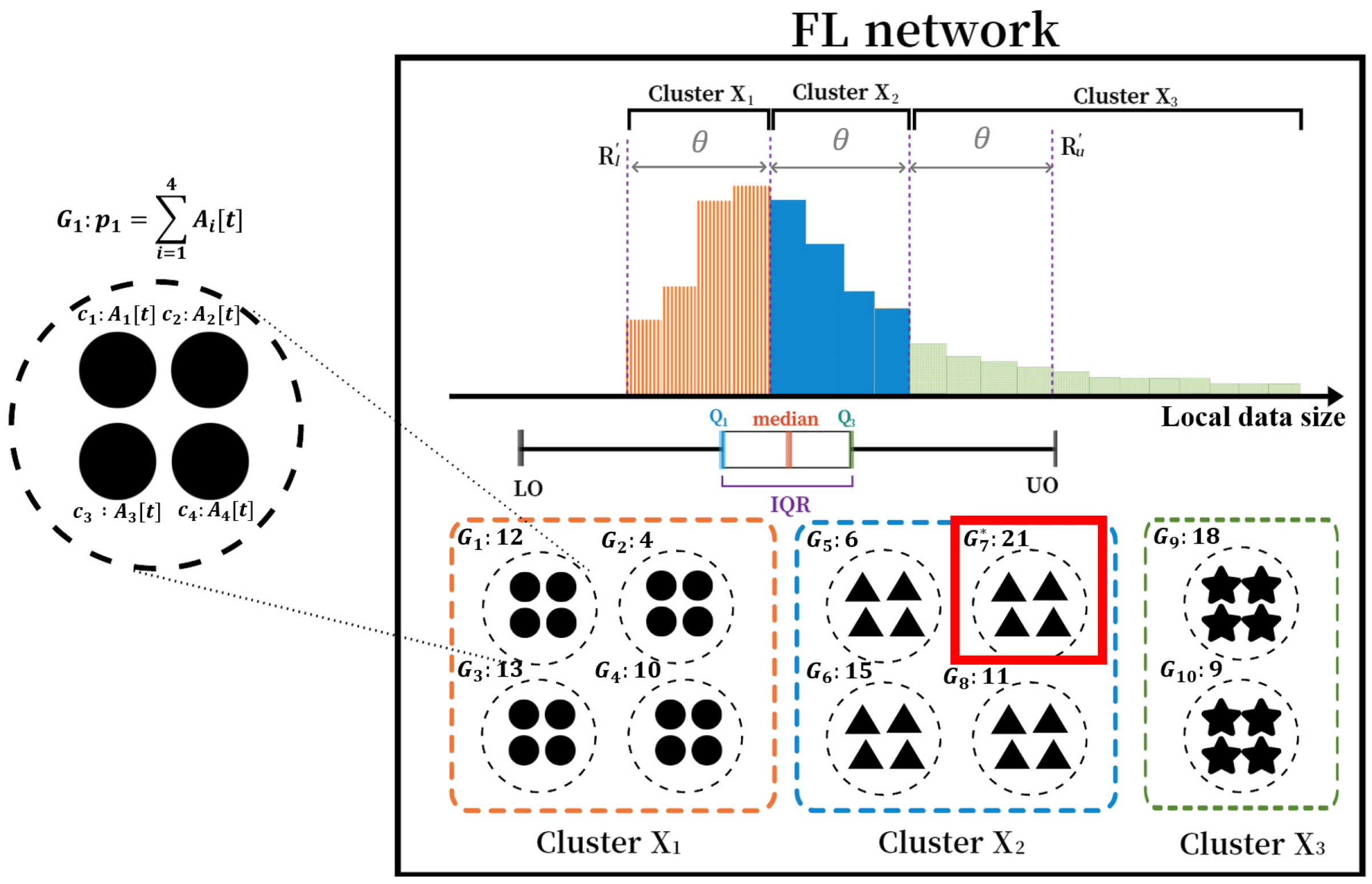
3. How to choose a cluster head -> Ring and Bully Algorithm

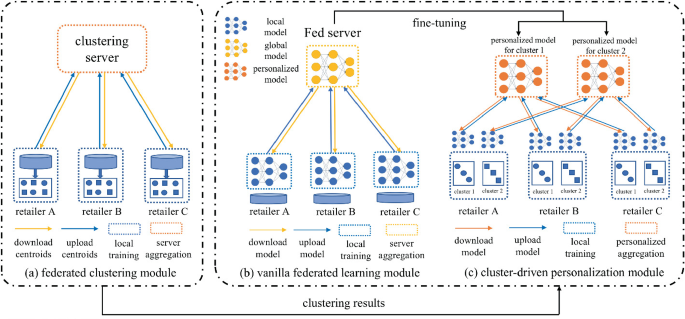


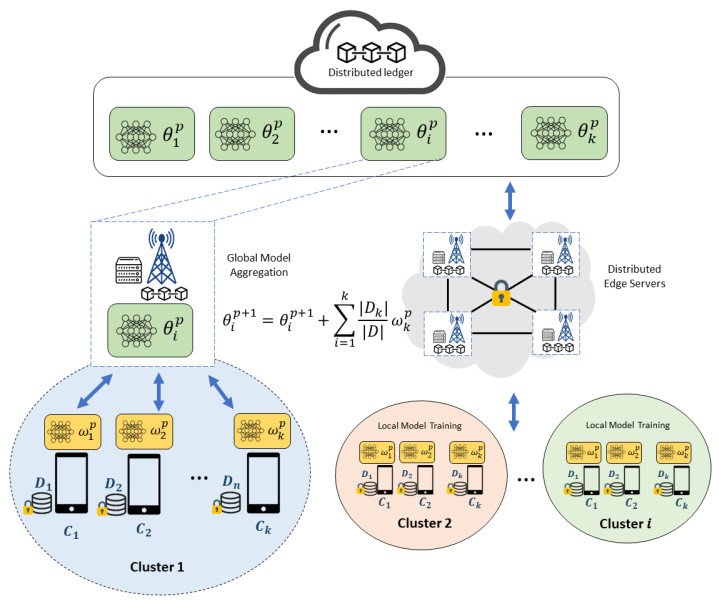
## [Cluster-Based Object Tracking by Wireless Camera Networks](https://www.sciencedirect.com/science/article/pii/B9780123746337000252)

Henry Medeiros, Johnny Park, in





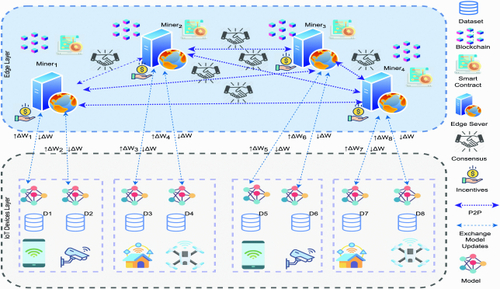




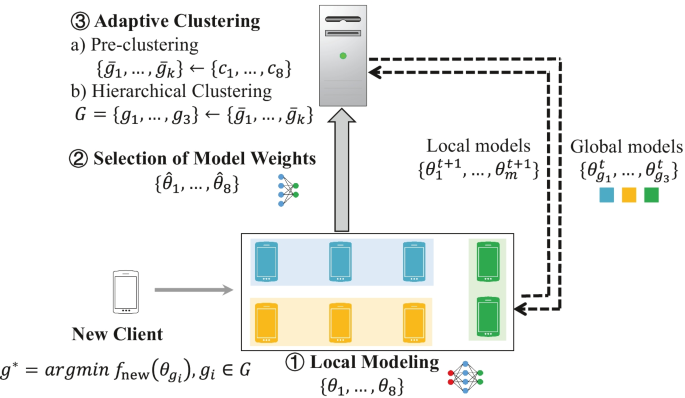
[**https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.aimspress.com%2Farticle%2Fdoi%2F10.3934%2Fmbe.2023475%3FviewType%3DHTML&psig=AOvVaw2Qfyeq9dgMYyfSh2A6\_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABBI**](https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.aimspress.com%2Farticle%2Fdoi%2F10.3934%2Fmbe.2023475%3FviewType%3DHTML&psig=AOvVaw2Qfyeq9dgMYyfSh2A6_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABBI)

[**https://www.google.com/url?sa=i&url=https%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs11036-022-01978-8&psig=AOvVaw2Qfyeq9dgMYyfSh2A6\_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABBA**](https://www.google.com/url?sa=i&url=https%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs11036-022-01978-8&psig=AOvVaw2Qfyeq9dgMYyfSh2A6_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABBA)

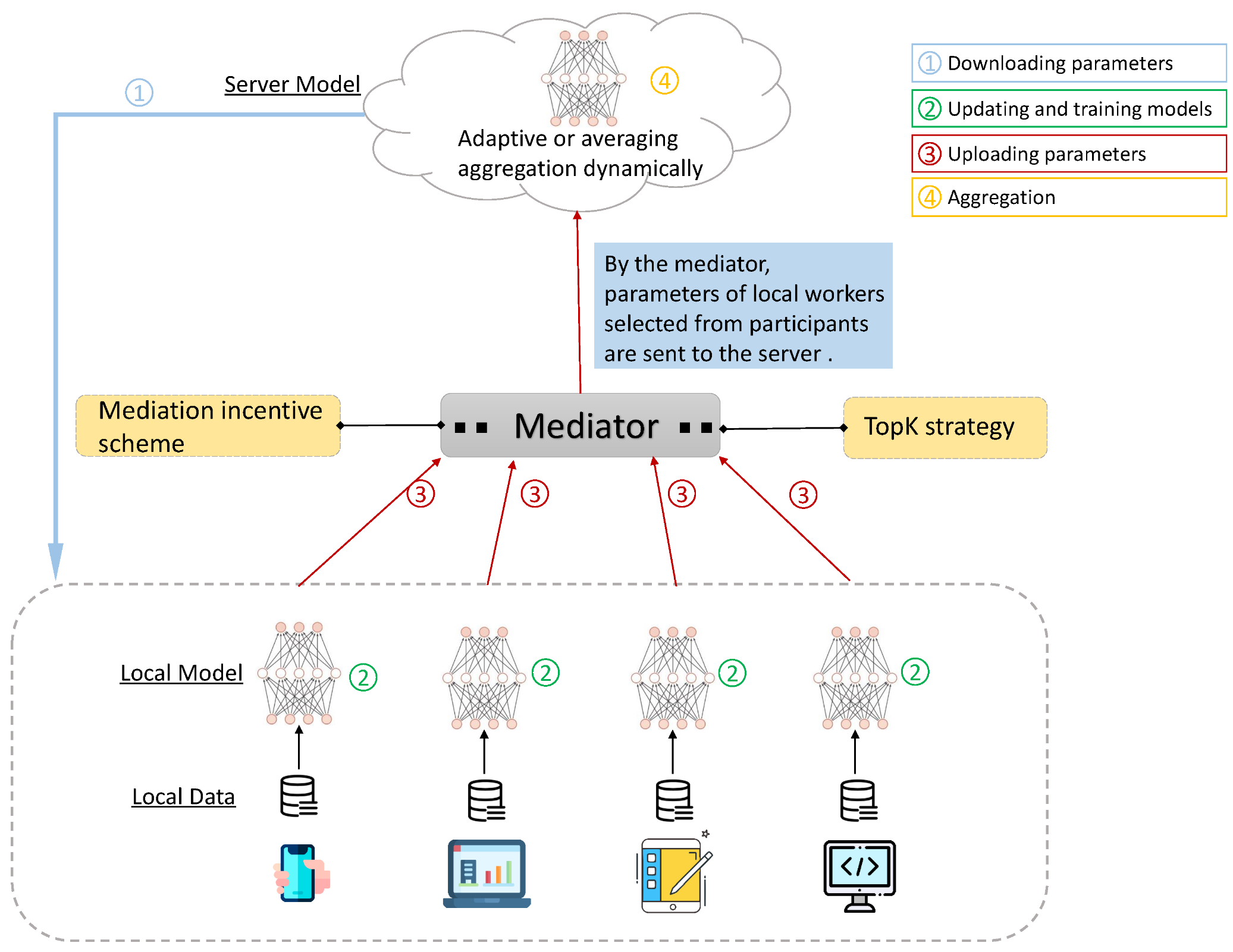
[**https://www.google.com/url?sa=i&url=https%3A%2F%2Fdl.acm.org%2Fdoi%2F10.1145%2F3560816&psig=AOvVaw2Qfyeq9dgMYyfSh2A6\_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABBg**](https://www.google.com/url?sa=i&url=https%3A%2F%2Fdl.acm.org%2Fdoi%2F10.1145%2F3560816&psig=AOvVaw2Qfyeq9dgMYyfSh2A6_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABBg)



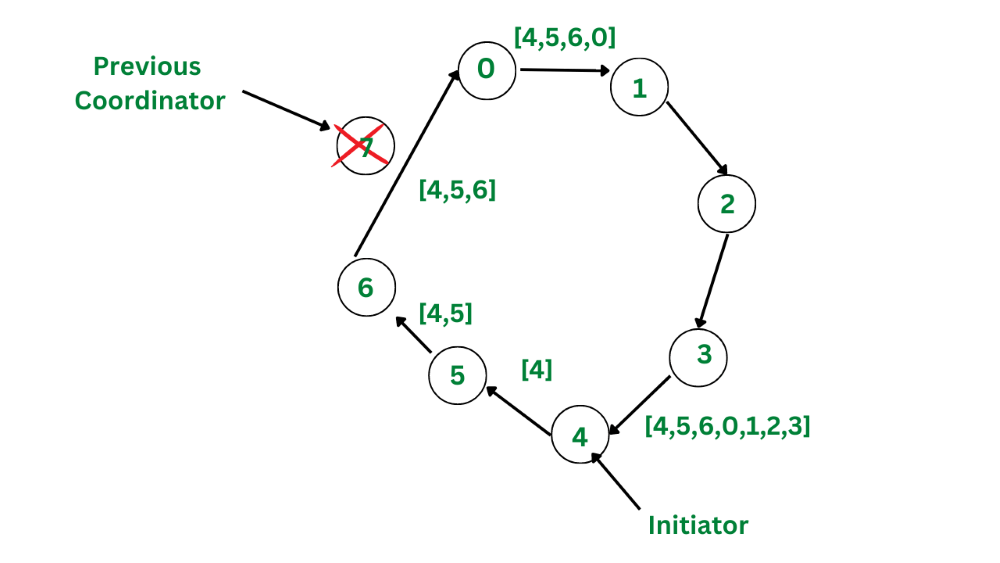
[**https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FThe-architecture-of-federated-learning-based-data-sharing-for-IoV-with-blockchain-52\_fig2\_350836342&psig=AOvVaw2Qfyeq9dgMYyfSh2A6\_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABB4**](https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FThe-architecture-of-federated-learning-based-data-sharing-for-IoV-with-blockchain-52_fig2_350836342&psig=AOvVaw2Qfyeq9dgMYyfSh2A6_C44&ust=1701493641381000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCNDNwPK77YIDFQAAAAAdAAAAABB4)



**Mediator**



[**https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.mdpi.com%2F1424-8220%2F20%2F14%2F4048&psig=AOvVaw0fzs5SwgkINvC1AGqetDwn&ust=1701845645441000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCMCKsZ3b94IDFQAAAAAdAAAAABAE**](https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.mdpi.com%2F1424-8220%2F20%2F14%2F4048&psig=AOvVaw0fzs5SwgkINvC1AGqetDwn&ust=1701845645441000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCMCKsZ3b94IDFQAAAAAdAAAAABAE)



[**https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.geeksforgeeks.org%2Fdifference-between-ring-and-bully-algorithm%2F&psig=AOvVaw02rK9EAmisgRUL8sNDTHZ8&ust=1701848887214000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCMDSjKfn94IDFQAAAAAdAAAAABAE**](https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.geeksforgeeks.org%2Fdifference-between-ring-and-bully-algorithm%2F&psig=AOvVaw02rK9EAmisgRUL8sNDTHZ8&ust=1701848887214000&source=images&cd=vfe&opi=89978449&ved=0CBQQjhxqFwoTCMDSjKfn94IDFQAAAAAdAAAAABAE)

**Apply ring and bully algorithm for Distributed database concept we are going to apply here to choose intermediate clustering parameters aggregator. Hence, number of aggregators increases it will have more opportunity to reduce communication overhead. The aggregator node can be given an incentives like points which later can be converted to monetary values.**

**Analysis:**

Our search identified 50 relevant publications after eliminating duplicates and applying inclusion/exclusion criteria. Using word, the results were summarized. We summarized and tabulated each selected publication. Each publication was evaluated and discussed by the research team on a regular basis in order to assess potential biases.

According to the subgroups of participants, the studies were divided into two categories ( First one is studies with different dataset, accuracy and how accuracy measured and second one is Studies with different clustering algorithm and what problem solved by using these algorithm ).

Figure 1: The review process

**Discussion:**

Clustering can reduce the communication overhead and computation required on the central server by allowing groups of similar devices to collaborate on training a model locally, instead of transmitting their updates to the central server. Clustering can also help to improve the scalability and performance of the federated learning system by reducing the number of devices that need to communicate with the central server. According to particular characteristics, such as their physical location, network connectivity, or hardware capabilities, clustering entails putting similar devices in one group. The devices can then cooperate to train the model locally, sending only updates to the central server. In federated learning environments where devices may have limited network access or processing capabilities, this might lessen the amount of communication that is necessary between devices and the central server.

Table-1: Studies with different dataset, accuracy and how accuracy measured

|  |  |  |  |
| --- | --- | --- | --- |
| **Ref.** | **Dataset Used** | **Accuracy** | **How accuracy Measured** |
| (Ghosh et al., 2020) [1] | Rotated MNIST  Rotated CIFAR | Test accuracies (%) ± std on Rotated MNIST (k = 4): IFCA 95.25±.40 and Rotated CIFAR (k = 2): 81.51 ± 1.37 | * Cluster models * Calculating the difference between the predicted labels and the ground truth labels |
| (Mishra et al., 2023) [2] | MNIST  HAR,  CIFAR-10  SHL | Accuracy at (k =5)   * MNIST dataset is **97.73**% * HAR (93.54%) * CIFAR-10 (91.01%) * SHL (90.27%) | * Comparing the predicted output of a model with the actual ground truth values |
| (Zeng et al., 2023) [3] | MNIST | 95% on MNIST and  91% on FEMNIST | * Predicted labels of the global model with the true labels of the data samples. |
| (Ruan & Joe-Wong, 2022) [4] | EMNIST  CIFAR-10 | Test accuracy 71.9 % of centers for the mixture of 4 distributions with original and 90◦ -rotated letter images. | * Trained models on their respective data distributions * Comparing predictions with actual labels |
| (Briggs et al., 2020) [6] | MNIST  FEMNIST | Using FedAvg, they need 50 rounds to converge, Using FL+HC trains 80% clients and achieved 99% test set accuracy | * Applying FL+HC & mini batch SGD, * Using agglomerative hierarchical clustering algorithm, * Return a number of cluster similar to one another. |
| (Shlezinger et al., 2020) [7] | MNIST  CIFAR 100  **Fashion-MNIST** | * MNIST- 99.1% * CIFAR 100 – 73.5% * Fashion-MNIST- 89.5% * Reduce communication cost up to 50% | * The CA-FL improved non-IID settings by clustering the users based on the similarity of their training sets * CA-FL used a communication-efficient aggregation algorithm to reduce the communication overhead |
| (Long et al., 2023) [8] | FEMNIST  MNIST | The clustering algorithm will converge very fast and didn’t take more than 10 iterations to converge | * Used hierarchical clustering to cluster the clients based on the similarity of their local model updates * Parallel training on specialized models in independent cluster. |
| (Duan et al., 2021) [9] | MINIST  FEMNIST | * +10 .6% on FEMNIST compared to FedAvg , * +3 .5% on FashionMNIST compared to FedProx * +8 .4% on MNIST compared to FeSEM. | * The FCFL improved n non-IID settings by clustering the clients based on the similar local data distributions * achieved more accuracy than tradition FL |
| (Kandati & Gadekallu, 2022) [10] | Covid19 dataset | * **F1**- 0.9365, * **recall**- 0.9366, * **loss**-0.2758, * **precision**- 0.9362, * **accuracy**-0.9187. | * Genetic CFL involves grouping edge devices based on hypertuned parameters and modifying the parameters cluster-wise genetically |
| (Kim et al., 2020) [11] | Fashion-MNIST  CIFAR-10. | Fashion-MNIST (84.3%) and CIFAR-10 (50.4%) | * Cluster models * Used ARI |
| (Luo, Y., Liu..,2021) [12] | Handover data set | algorithm improves prediction accuracy in wireless network handovers by 43% | Not Mention |

Table-1: Studies with different dataset, accuracy and how accuracy measured (continued)

|  |  |  |  |
| --- | --- | --- | --- |
| **Ref.** | **Dataset Used** | **Accuracy** | **How accuracy Measured** |
| (Wang et al., 2022) [13] | MNIST  CIFAR-10. | Global model accuracy higher under strong social effects in FL datasets. | The global model is measured in the experiments conducted on the MNIST and CIFAR-10 datasets. |
| (Lu et al., 2021)[14] | MNIST  CIFAR-10, | Client selection scheme using gradient clustering for federated learning accuracy improvement. | Based on convergence rate and performance on benchmark datasets like MNIST and CIFAR-10. |
| Zhang et al.,(2021) [15] | MNIST  FEMNIST | MNIST dataset, the accuracy of the FedLab Cluster algorithm was improved by 6.5% and 6.9%.  FEMNIST dataset, accuracy was improved by 15% and 28.6% | Comparing performance with a baseline algorithm |
| ([Joel Wolfrath](https://ieeexplore.ieee.org/author/37088760931), [Nikhil Sreekumar](https://ieeexplore.ieee.org/author/37088764328), 2022)[16] | MNIST  CIFAR 10  FEMNIST | MNIST – 90.83  CIFAR – 85.90  FEMNIST – 80.50 | FEMNIST 2400 iterations**.**  MNIST, CIFAR for 280 and 2400 iterations |
| (Harshvardhan et al., 2022) [17] | MNIST  CIFAR 10  FEMNIST  Shakespeare | MNIST – 92.83  CIFAR – 88.90  FEMNIST – 84.50  Shakespeare- 47.50 | FEMNIST and Shakespeare for 1000 and 2400 iterations**.**  Rotated MNIST, Inverted MNIST and Rotated CIFAR10 for 250, 280 and 2400 iterations |
| ([Pu Tian](https://ieeexplore.ieee.org/author/37088908022) et al.,2022) [18] | MNIST  CIFAR  FEMNIST  Shakespeare | MNIST – 92.83  CIFAR – 88.90  FEMNIST – 84.50  Shakespeare- 47.50 | FEMNIST and Shakespeare for 1000 and 2400 iterations**.**  Rotated MNIST, Inverted MNIST and Rotated CIFAR10 for 250, 280 and 2400 iterations |
| (Cho et al., 2021) [19] | MNIST  CIFAR | MNIST – 92.83  CIFAR – 88.90 | MNIST, CIFAR10 for 250,280 iterations |
| (Albaseer et al., 2021) [22] | FEMNIST | When c=15, test accuracy is 81.8% | Evaluating the test accuracies achieved by specialized machine learning models for different client groups |
| (Taik et al., 2022) [23] | MNIST  Fashion MNIST | The data accuracy 95.5% | MNIST is 80% ± 10%for vanilla FL, it reaches and average of 82% ± 9% |

Table-2: Studies with different clustering algorithm and what problem solved by using these algorithm

|  |  |  |
| --- | --- | --- |
| **Ref.** | **What Problem Solved** | **Future Work** |
| (Ghosh et al., 2020) [1] | * Cluster identity estimation * Gradient descent optimization | * Exploring different loss functions * Examining convergence rates in various scenarios * Studying the impact of initialization |
| (Mishra et al., 2023) [2] | * Improving the performance of lightweight models * Minimize the error caused by participants' heterogeneity. | * Communication delays * Limited computational resources |
| (Zeng et al., 2023) [3] | * Non-IID issues in federated learning | * Incorporating different architectures * Knowledge transfer techniques * Implementing a dynamic join-leave mechanism |
| (Ruan & Joe-Wong, 2022) [4] | * Slow model training * Improving the accuracy of cluster models * Producing accurate local models | * Analyzing convergence properties * Comparing with other algorithms * Improving accuracy of local models |
| (Tao et al., 2023) [5] | * A distributed system * Robust estimation in a clustering setting * Estimating parameter values in a clustering model * Improving error rate | * Proving the accuracy of the j-th cluster estimate * Analyzing the impact of Byzantine machines in distributed learning * Improving the error rate in a theorem |
| (Briggs et al., 2020) [6] | * The distance metric and linkage hyperparameters are tuned * improve the performance of the clustering algorithm | * Identify malicious clients * The effect of the noisy client updates on the ability of FL+HC to find good clustering of clients. * The effect of compression methods |
| (Shlezinger et al., 2020) [7] | * CACFL is effective massively distributed data * Handle heterogeneity and communication overhead | * Privacy preservation will be observed further * The clustering step can be computationally expensive |
| (Long et al., 2023) [8] | * Global models trained from data as the cluster centers * Optimal matching between users and centers * produce accurate specialized models by cluster | * **Studying the performance of multi-Center FL in different settings** * **investigating the use of other clustering algorithms** |
| (Duan et al., 2021) [9] | * Handle client-level data distribution shift. * Improve the accuracy of FL in non-IID settings. * scalable and used with large datasets | * **Studying the security and privacy implications of FlexCFL** * **Investigating the use of federated learning for other machine learning tasks** |
| (Kandati & Gadekallu, 2022) [10] | * A novel hybrid algorithm named genetic clustered FL * Groups edge devices based on the hypertuned parameters * modifies the parameters cluster wise genetically | * Use genetic CFL model on scalable and real-time datasets. * time-sensitive techniques will also be used |
| (Kim et al., 2020) [11] | * Local data across different devices are not independent * Identically distributed (non-IID). | * Investigating the performance of CAFL on Real-world scenarios. * Impact of different hyperparameters on the performance of CAFL. |

Table-2: Studies with different clustering algorithm and what problem solved by using these algorithm (Continued)

|  |  |  |
| --- | --- | --- |
| **Ref.** | **What Problem Solved** | **Future Work** |
| (Luo, Y., Liu..,2021) [12] | * Handover prediction accuracy is degraded by non-independent data.   Distributed data in the resource management of wireless networks | * Proposed algorithm's performance on other tasks. * Datasets to evaluate its generalizability.   Algorithm in other domains beyond wireless networks. |
| (Wang et al., 2022) [13] | Privacy leakage in federated learning. | Not Given |
| (Lu et al., 2021)[14] | * Data heterogeneity and energy consumption imbalance in mobile edge computing systems. | * **Clustering algorithm performance evaluation** * Scalability of proposed method for large datasets and complex models |
| Zhang et al.,(2021) [15] | Impact of Non-Independent and Identically Distributed (Non-IID) data on the prediction accuracy  Performance of federated learning | Optimization and refinement of the algorithm.  FedLab Cluster algorithm on other datasets and compare its performance with others FLA |
| ([Joel Wolfrath](https://ieeexplore.ieee.org/author/37088760931), [Nikhil Sreekumar](https://ieeexplore.ieee.org/author/37088764328),2022)[16] | HACCS addresses efficient client selection for faster and more effective federated learning, considering device heterogeneity through clustering. | Optimizing cluster formation, adaptive heterogeneity management, and scalability |
| (Harshvardhan et al., 2022) [17] | 1. Reducing the amount of data transmitted  2. Federated Learning can suffer from slower convergence. | 1. To enhance communication efficiency, convergence speed, or scalability.  2. Investigating ways to strengthen privacy and security |
| ([Pu Tian](https://ieeexplore.ieee.org/author/37088908022) et al.,2022) [18] | The "WSCC" paper addresses non-IID data challenges in federated learning by clustering clients based on weight similarity for improved convergence. | Future work may focus on enhancing cluster initialization, addressing cluster imbalance, and extending the approach to different data distributions. |
| (Cho et al., 2021) [19] | Personalized federated learning framework called PerFed-CKT that enables clients to use heterogeneous. | Exploring different clustering techniques for determining co-distillation weights, conducting experiments |
| (Nelus et al., 2021) [20] | Source-dominated microphone clusters in acoustic sensor networks | Enhancing additional ASN-deployable tasks (wake word detection or event classification) |
| (Sattler et al., 2019) [21] | Analyzing the behavior of a federated learning system and bounding the similarity between updates from different client | Investigate the utility of weight-updates for distribution similarity estimation |
| (Albaseer et al., 2021) [22] | Minimizing the training time and achieving satisfying performance for each client with an optimal fit model for its local data distribution | Finding the optimal threshold values for splitting the clusters |
| (Taik et al., 2022) [23] | Optimizing federated learning in vehicular networks using clustered approaches for improved efficiency. | Explore adaptive clustering and dynamic participation for improved efficiency and scalability. |

Table-3: Studies with different Federated Clustering Algorithm with their strength and weakness

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref** | **Algorithm** | **Description** | **Strength** | **Weakness** |
| (Ghosh et al., 2020) [1] | Iterative Federated Clustering Algorithm (IFCA) | It is an iterative clustering algorithm that estimates cluster identities and improves the clustering of worker machines. It uses two variations: gradient averaging and model averaging. | * Near-optimal statistical error rate * High success probability * Efficient implementation without expensive computations * Faster convergence | * SGD is a relatively weak assumption * Parameters need to be satisfied for the algorithm |
| (Mishra et al., 2023) [2] | Fed-RAC (Federated learning with Resource Aware Clustering) | It includes steps such as gathering information about participants' devices and networking resources, performing resource aware clustering using Dunn Indices. | * Efficient training and communication * Participants assignment optimization * Deriving expressions for error and intra-heterogeneity | * Lack of efficient clustering techniques * Discarding stragglers reduces generalization * Slower aggregation process |
| (Zeng et al., 2023) [3] | Stochastic Clustered Federated Learning algorithm( StoCFL) | The StoCFL model could train better-generalized models by clustering federated clients with different unknown distributions. It can handle a variety of FL systems, such as those that vary in size and complexity. | * Effectively addresses Non-IID issues * Handle an arbitrary proportion of client participation * Effective use of cosine similarity | * Not work well with non-IID data distributions * Does not address the issue of data heterogeneity among clients |
| (Ruan & Joe-Wong, 2022) [4] | FedSoft | Algorithm outperform existing FL implementations in both global cluster models for future users and personalized local models for participating clients. | * Estimates importance weights for each client * Aggregation weights based on importance weights | * Computational complexity * Limited consideration for the impact of different distributions |
| (Tao et al., 2023) [5] | Byzantine-Robust Iterative Federated Clustering Algorithm | Uses coordinate-wise trimmed mean and coordinate-wise median aggregation methods to make the framework robust against adversarial attacks. | Optimize the models learned by each cluster in the presence of Byzantine machines.  Strongly convex loss functions. | Scalability and computational efficiency. |
| (Briggs et al., 2020) [6] | Agglomerative hierarchical clustering algorithm | separate clusters of clients based on the similarity of their local updates to the global joint model, once the clusters are formed, specialized models are trained independently and in parallel on these clusters | * Address the challenge posed by non iid data distribution and competing objectives of clients   Suggest good hyperparameters to promote superior performing specialized models | * High communication overhead * Communication costs   when there is a concept shift in the data distribution across clients, this assumption may be not hold true |
| (Shlezinger et al., 2020) [7] | Communication -aware CFL algorithm | CACFL takes into account the statistical heterogeneity of the data and the communication constraints to balance the ability of the clients to learn a proper model and the accuracy in aggregating these models into a global inference rule | * Based on multi-source adaptation methods which allows for balancing performance measures of individual cluster models * Accuracy of aggregating models into a global inference | * Marginal distribution of each cluster may not be always available * Additional complexity and computational overhead * More evaluation is needed for large scale and real time dataset |
| (Long et al., 2023) [8] | Multi-center aggregation algorithm | The algorithm worked by Clustering clients, Optimal matching, Optimization problem, iterative procedure. | * addresses the statistical heterogeneity and cluster clients based on their models' parameters * earns multiple global models from data, allowing for better personalization in decision-making | * If the number of clusters is too much small or too much large |

Table-3: Studies with different Federated Clustering Algorithm with their strength and weakness (Continued)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref** | **Algorithm** | **Description** | **Strength** | **Weakness** |
| (Duan et al., 2021) [9] | A flexible clustered federated learning (CFL) framework named FlexCFL | Achieves challenges by grouping clients based on similarities in optimization directions, implementing an efficient newcomer device could start mechanism, and flexibly migrating clients to handle client-level data distribution shift | * Encounter challenges of non iid, imbalance and distributed shifted training data * Less communication rounds | * Not scalable * Computational complexity * Need more computational resource |
| (Kandati & Gadekallu, 2022) [10] | Generic CFL | The algorithm groups edge devices based on hypertuned parameters and modifies the parameters cluster-wise genetically | * Optimize hyperparameters * Address privacy concern | * Not applicable for scalability * Real time application scenario |
| (Kim et al., 2020) [11] | Causal Adjustment for Feedback Loops (CAFL) | CAFL algorithm breaks feedback loops in recommender systems by estimating causal effects and adjusting user behaviour data to remove system influence. | * Any recommendation algorithm * Easy to implement large datasets * Provably effective in breaking feedback loops | * User behavior can be computationally expensive. * Causal effect estimation imperfectly can cause bias in adjusted data. |
| (Luo, Y., Liu..,2021) [12] | LSTM | LSTM is a powerful algorithm for machine translation, speech recognition, and time series forecasting. | * Long-term dependencies than traditional RNNs * **Less susceptible to the vanishing gradient problem** | * **Computationally expensive** * **Requires more memory** * **Prone to overfitting** |
| (Wang et al., 2022) [13] | Stochastic coded federated learning (SCFL) | It is mitigates the straggler effect by using coded computing techniques. It effectively reduces straggler effect and improves convergence speed in federated learning. | * Achieve faster convergence * **Enhanced privacy protection** * **Reduced communication overhead** | * **More complex implementation** * **Requires more computation** * **Not suitable for all datasets** |
| (Lu et al., 2021)[14] | Cluster-based clients selection  Auction-based clients selection | Cluster-based clients selection algorithm selects subset of clients for federated learning task based on similarity.  Auction-based clients selection algorithm selects through high bids. | * Balances the energy consumption of clients * Mitigates the impact of data heterogeneity on model convergence * Selects the most efficient clients for FL * Adaptive Resource Allocation | * Communication Imbalance * May not fully utilize the resources * More complex to implement * Clients' strategic bidding may manipulate selection chances. |
| Zhang et al.,(2021) [15] | FedLabCluster | FedLab Cluster is a federated learning algorithm that clusters clients based on their data sample labels to improve accuracy and robustness in non-IID data. | * A centralized server that manages the overall experiment * Cloud-based platform simplifies large-scale federated learning experiments deployment. | * Communication overhead * It does not provide any built-in security features * It is not as scalable as some other federated learning frameworks |
| ([Joel et](https://ieeexplore.ieee.org/author/37088760931) al.,2022)[16] | HACCS: Heterogeneity-Aware Clustered Client Selection for Accelerated Federated Learning | HACCS is a method or technique designed to enhance federated learning by taking into consideration the heterogeneity of clients (devices or servers) that participate in the training process. enhancement for broader application in federated learning. | * Innovative Approach * Performance Improvement * Heterogeneity Consideration | * Evaluation Rigor * Limited Generalizability * Scalability * Overhead * Lack of Comparison |

Table-3: Studies with different Federated Clustering Algorithm with their strength and weakness (Continued)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref** | **Algorithm** | **Description** | **Strength** | **Weakness** |
| (Harshvardhan et al., 2022) [17] | An Improved Algorithm for Clustered Federated Learning | Reducing the amount of data transmitted. Federated Learning can suffer from slower convergence. | * comprehensive experimentation * clear problem definition * Relevance to real-world applications. | * To enhance communication efficiency, convergence speed, or scalability. * Investigating ways to strengthen privacy and security |
| ([Pu Tian](https://ieeexplore.ieee.org/author/37088908022) et al.,2022)[18] | WSCC: A Weight-Similarity-Based Client Clustering Approach for Non-IID Federated Learning | The "WSCC" paper proposes a client clustering technique for non-IID federated learning, grouping similar clients based on weight similarity to enhance convergence and accommodate varying data distributions. | * Improved Convergence * Heterogeneity Handling * Reduced Communication * Privacy Benefits | * Cluster Formation Sensitivity * Centralized Cluster Selection * Cluster Imbalance * Complexity and Scalability |
| (Cho et al., 2021) [19] | Personalized Federated Learning for Heterogeneous Clients | Personalized federated learning framework called PerFed-CKT that enables clients to use heterogeneous. | * Personalization * Clustered Knowledge Transfer * Heterogeneity Handling * Adaptation to Diversity | * Scalability * Complexity * Cluster Quality * Resource Allocation |
| (Taik et al., 2022) [23] | Optimizing federated learning in vehicular networks | Optimizing federated learning in vehicular networks using clustered approaches for improved efficiency. | * innovative methods * comprehensive experimentation * clear problem definition * Relevance to real-world applications. | Assumptions that limit applicability, lack of comparison to existing methods, or insufficient validation |

Table-4: Studies with different clustering algorithm

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ref.** | **IFCA** | **Fed-RAC** | **StoCFL** | **FedSoft** | **BRI-CFL** | **FL+HC** | **CA-CFL** | **FeSEM** | **FlexCFL** | **Generic CFL** | **LSTM** | **SCFL** | **SR-FCA** | **CFL** | **HACCS** | **FedLab** | **WSCC** |
| (Ghosh et al., 2020) [1] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Mishra et al., 2023) [2] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Zeng et al., 2023) [3] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Ruan & Joe-Wong, 2022) [4] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Tao et al., 2023) [5] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Briggs et al., 2020) [6] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Shlezinger et al., 2020) [7]  (Kim et al., 2020) [11] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Long et al., 2023) [8] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Duan et al., 2021) [9] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Kandati & Gadekallu, 2022) [10] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Luo, Y., Liu..,2021) [12] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Wang et al., 2022) [13] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Harshvardhan et al., 2022) [17]  (Cho et al., 2021) [19] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (Sattler et al., 2019) [21]  (Albaseer et al., 2021) [22] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ([Joel et](https://ieeexplore.ieee.org/author/37088760931) al.,2022)[16] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Zhang et al.,(2021) [15] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ([Pu Tian](https://ieeexplore.ieee.org/author/37088908022) et al.,2022)[18] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Results and Discussion:**

To optimize communication in federated learning the following are some of the previous algorithms that have used clustering:

* Iterative Federated Clustering Algorithm (IFCA): This algorithm iteratively estimates the cluster identities of the users and optimizes model parameters for the user clusters via gradient descent.
* Fed-RAC (Federated learning with Resource Aware Clustering): This algorithm takes into account the resources available to each user when clustering the users.
* StoCFL (Stochastic Clustered Federated Learning algorithm): This algorithm is a stochastic version of IFCA. It can be more efficient than IFCA, but it may not converge as quickly.

The following are some of the previous algorithms that have been used to improve privacy in federated learning using clustering:

* FedSoft: This algorithm is a federated learning algorithm that uses a soft clustering approach. This can be more robust to noise and outliers than hard clustering approaches.
* Byzantine-Robust Iterative Federated Clustering Algorithm: This algorithm is robust to Byzantine failures, which are failures that occur when a malicious user intentionally sends incorrect data to the server.

Here is a more detailed discussion of the trade-offs for each of the algorithms:

* **IFCA:** IFCA is a centralized algorithm, so the central server bears the full burden of clustering and communication. However, IFCA can be more efficient than decentralized algorithms, especially when the number of users is large.
* **Fed-RAC:** Fed-RAC is a decentralized algorithm, so the burden of clustering and communication is shared among the users. This can reduce the burden on the central server, but it can also make the algorithm less efficient.
* **StoCFL:** StoCFL is a stochastic algorithm, so it can be less accurate than deterministic algorithms. However, StoCFL can be more efficient, especially when the number of users is large.
* **FedSoft:** FedSoft is a soft clustering algorithm, so it can be more robust to noise and outliers than hard clustering algorithms. However, FedSoft can be less efficient than hard clustering algorithms.
* **Byzantine-Robust IFCA:** This algorithm is robust to Byzantine failures, which are failures that occur when a malicious user intentionally sends incorrect data to the server. However, this algorithm can be less efficient than non-Byzantine-robust algorithms.
* **Causal Adjustment for Feedback Loops (CAFL):** This algorithm is designed to handle feedback loops in federated learning. Feedback loops can occur when the model parameters are used to update the data, which can then be used to update the model parameters again. However, this algorithm can be less efficient than algorithms that do not handle feedback loops.
* **LSTM:** LSTM is a neural network architecture that can be used to handle temporal data. This can be useful in federated learning settings where the data is collected over time. However, LSTM can be more complex and computationally expensive than other neural network architectures.
* **Federated Learning with Hierarchical Clustering (FL+HC):** This algorithm uses hierarchical clustering to cluster the users. This can be more efficient than flat clustering, especially when there are a large number of users. However, FL+HC can be less accurate than flat clustering.
* **Stochastic coded federated learning (SCFL):** This algorithm uses coded communication to improve the communication efficiency of federated learning. This can reduce the communication costs, but it can also make the algorithm less secure.

**Conclusion:**

Clustered Federated Learning is a promising approach that aims to optimize the communication efficiency and privacy while reducing the burden on the central server in the context of distributed machine learning. With the success of federated learning, which allows training machine learning models on distributed data while preserving data privacy, communication efficiency has emerged as a central theme in this field. Communication efficiency is crucial in distributed optimization, and it has gained increasing attention due to the success of federated learning. Federated learning enables multiple users' edge devices to train a shared model without the need for their raw data to leave their devices, thereby preserving privacy and addressing concerns regarding data security. However, in traditional federated learning approaches with a central server orchestrating the training process, there can be scalability challenges and a high burden on the central server. To overcome these challenges, clustered federated learning introduces a new approach where clients are grouped into clusters, with each cluster having a cluster head responsible for coordinating the training process within the cluster. This approach reduces the communication overhead between the clients and the central server, as the cluster heads facilitate local model updates within their respective cluster.

**References:**

[1] A. Ghosh, J. Chung, D. Yin, and K. Ramchandran, “An Efficient Framework for Clustered Federated Learning,” in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020, pp. 19586–19597. Accessed: Jul. 10, 2023. [Online]. Available: <https://proceedings.neurips.cc/paper_files/paper/2020/hash/e32cc80bf07915058ce90722ee17bb71-Abstract.html>

[2] R. Mishra, H. P. Gupta, and G. Banga, “Resource Aware Clustering for Tackling the Heterogeneity of Participants in Federated Learning.” arXiv, Jun. 07, 2023. Accessed: Jul. 10, 2023. [Online]. Available: <http://arxiv.org/abs/2306.04207>

[3] D. Zeng, X. Hu, S. Liu, Y. Yu, Q. Wang, and Z. Xu, “Stochastic Clustered Federated Learning.” arXiv, Mar. 01, 2023. Accessed: Jul. 10, 2023. [Online]. Available:<http://arxiv.org/abs/2303.00897>

[4] Y. Ruan and C. Joe-Wong, “FedSoft: Soft Clustered Federated Learning with Proximal Local Updating.” arXiv, Mar. 22, 2022. Accessed: Jul. 10, 2023. [Online]. Available:<http://arxiv.org/abs/2112.06053>

[5] Z. Tao, K. Yang, and S. R. Kulkarni, “Byzantine-Robust Clustered Federated Learning.” arXiv, Jun. 01, 2023. doi: [10.48550/arXiv.2306.00638](https://doi.org/10.48550/arXiv.2306.00638).

[6] C. Briggs, Z. Fan and P. Andras, “Federated learning with hierarchical clustering of local updates to improve training on non-IID data” arXiv:2004.11791v2, 6 May 2020

[7] N. Shlezinger, S. Rini, and Y. C. Eldar, “the Communication-Aware Clustered Federated Learning Problem” IEEE Xplore, September 06,2020

[8] G. Long1, M. Xie1, T. Shen1, T. Zhou, X. Wang1, J. Jiang, “Multi‑center federated learning: clients clustering for better personalization” Network Data Scenarios and Applications, 9 June 2022

[9] M. Duan, D. Liu, X. Ji, Y. Wu, L. Liang, X. Chen, Y. Tan, “Flexible Clustered Federated Learning for Client-Level Data Distribution Shift” the IEEE International Symposium on Parallel and Distributed Processing with Applications, 22Aug 2021

[10] D. R. Kandati and T. R. Gadekallu, “Genetic Clustered Federated Learning for COVID-19 Detection” Licensee MDPI, Basel, Switzerland, 29 August 2022

[11] Kim, Y., Hakim, E. A., Haraldson, J., Eriksson, H., Da Silva, J. A., & Fischione, C. (2021). *Dynamic Clustering in Federated Learning*.<https://doi.org/10.1109/icc42927.2021.9500877>​

[12] Luo, Y., Liu, X., & Xiu, J. (2021). *Energy-efficient Clustering to Address Data Heterogeneity in Federated Learning*.<https://doi.org/10.1109/icc42927.2021.9500901>​

[13] Wang, Y., Su, Z., Pan, Y., Luan, T. H., Li, R., & Yu, S. (2022). Social-Aware Clustered Federated Learning with Customized Privacy   Preservation. arXiv (Cornell University). <https://doi.org/10.48550/arxiv.2212.13992>​

[14] Auction Based Clustered Federated Learning in Mobile Edge Computing System       [**https://doi.org/10.48550/arXiv.2103.07150**](https://doi.org/10.48550/arXiv.2103.07150)

[15] Zhang, J., & Lv, S. (2021). FedLabCluster: a clustered federated learning algorithm based on data sample label. 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS). <https://doi.org/10.1109/eiecs53707.2021.9588126>

[16] [Joel Wolfrath](https://ieeexplore.ieee.org/author/37088760931), [Nikhil Sreekumar](https://ieeexplore.ieee.org/author/37088764328), [Dhruv Kumar](https://ieeexplore.ieee.org/author/37088761318), [Yuanli Wang](https://ieeexplore.ieee.org/author/37088761165), [Abhishek Chandra](https://ieeexplore.ieee.org/author/37332527200),” HACCS: Heterogeneity-Aware Clustered Client Selection for Accelerated Federated Learning

”,IEEE,2022,doi:  [**https://doi.org/**10.1109/IPDPS53621.2022.00100](https://doi.org/10.3390/electronics11172714)

[17] [Harshvardhan](https://arxiv.org/search/stat?searchtype=author&query=Harshvardhan), [Avishek Ghosh](https://arxiv.org/search/stat?searchtype=author&query=Ghosh%2C+A), [Arya Mazumdar](https://arxiv.org/search/stat?searchtype=author&query=Mazumdar%2C+A),”An Improved Algorithm for Clustered Federated Learning”,arXiv,2022,doi: <https://doi.org/10.48550/arXiv.2210.11538>

[18] [Pu Tian](https://ieeexplore.ieee.org/author/37088908022),[Weixian Liao](https://ieeexplore.ieee.org/author/37085438161), [Wei Yu](https://ieeexplore.ieee.org/author/37279841200), [Erik Blasch](https://ieeexplore.ieee.org/author/37295205900),” WSCC: A Weight-Similarity-Based Client Clustering Approach for Non-IID Federated Learning

”,IEEE,2022,doi: [https://doi.org/10.1109/JIOT.2022.3175149](https://doi.org/10.48550/arXiv.2006.04088)

[19] [Yae Jee Cho](https://arxiv.org/search/cs?searchtype=author&query=Cho%2C+Y+J), [Jianyu Wang](https://arxiv.org/search/cs?searchtype=author&query=Wang%2C+J), [Tarun Chiruvolu](https://arxiv.org/search/cs?searchtype=author&query=Chiruvolu%2C+T), [Gauri Joshi](https://arxiv.org/search/cs?searchtype=author&query=Joshi%2C+G),” Personalized Federated Learning for Heterogeneous Clients with Clustered Knowledge Transfer”, arXiv,2021,doi: <https://doi.org/10.48550/arXiv.2109.08119>

[20]A. Nelus, R. Glitza, and R. Martin, “Unsupervised Clustered Federated Learning in Complex Multi-source Acoustic Environments.” arXiv, Jun. 07, 2021. Accessed: Aug. 01, 2023. [Online]. Available: <http://arxiv.org/abs/2106.03671>

[21] F. Sattler, K.-R. Müller, and W. Samek, “Clustered Federated Learning: Model-Agnostic Distributed Multi-Task Optimization under Privacy Constraints.” arXiv, Oct. 04, 2019. Accessed: Aug. 01, 2023. [Online]. Available: <http://arxiv.org/abs/1910.01991>

[22] A. Albaseer, M. Abdallah, A. Al-Fuqaha, and A. Erbad, “Client Selection Approach in Support of Clustered Federated Learning over Wireless Edge Networks.” arXiv, Aug. 16, 2021. Accessed: Aug. 01, 2023. [Online]. Available: <http://arxiv.org/abs/2108.08768>

[23] [Afaf Taik](https://arxiv.org/search/cs?searchtype=author&query=Taik%2C+A), [Zoubeir Mlika](https://arxiv.org/search/cs?searchtype=author&query=Mlika%2C+Z), [Soumaya Cherkaoui](https://arxiv.org/search/cs?searchtype=author&query=Cherkaoui%2C+S),” Clustered Vehicular Federated Learning: Process and Optimization”,Thu,27 jan2022,dio: <https://doi.org/10.48550/arXiv.2201.11271>