Device Selection Algorithm for Federated Learning

A B. Tech Project Report Submitted
in Partial Fulfillment of the Requirements
for the Degree of

Bachelor of Technology

by

Kartik Kailas Mouli
(2001CS35)

under the guidance of

Dr. Satendra Kumar

to the

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY PATNA
PATNA - 800013, BIHAR

CERTIFICATE

*This is to certify that the work contained in this thesis entitled "**Device Selection Algorithm for Federated Learning**" is a bonafide work of **Kartik Kailas Mouli** (**Roll No. 2001CS35**), carried out in the Department of Computer Science and Engineering, Indian Institute of Technology Patna under my supervision and that it has not been submitted elsewhere for a degree.*

Supervisor: **Dr. Satendra Kumar**

Assistant/Associate Professor,

Department of Computer Science & Engineering,

Indian Institute of Technology Patna, Bihar.

May, 2024

Patna.

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Satendra Kumar Sir, for his guidance and support throughout this project. His expertise and insights were invaluable to me, and I am grateful for his patience and encouragement.

I am also grateful to my parents for their unconditional love and support in whatever I do.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

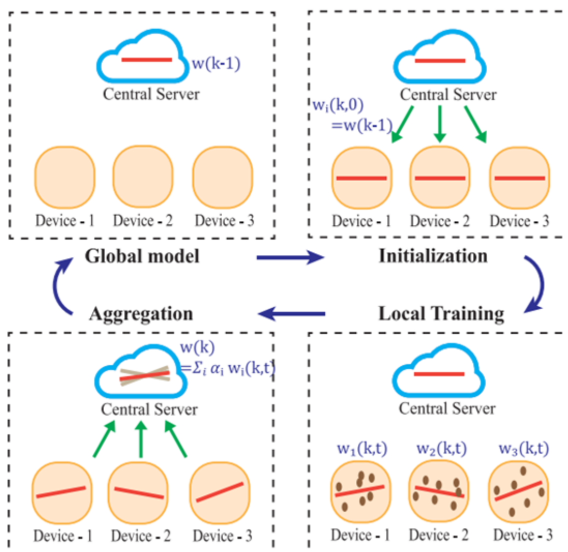

Fig. 1.1 Architecture of conventional federated learning

Over the past decade, cloud computing has emerged as a dominant force, offering efficient processing and storage capabilities. This rise has coincided with a surge in demand for AI algorithms capable of achieving near-human accuracy. This demand has been fuelled by deploying high-resolution sensors in various devices, including IoT devices, smartphones, and autonomous vehicles, resulting in vast data. However, transmitting this data to the cloud significantly strains network bandwidth, raising concerns about data privacy. Consequently, there has been a shift towards edge computing, which leverages the computational power of local devices to perform data processing tasks.[WZKA19]

Within edge computing, machine learning (ML), particularly federated learning, has gained considerable traction as a privacy-preserving and decentralized collaborative methodology. In federated learning, edge devices autonomously update models using their local datasets and periodically share these updates with a central coordinating node for aggregation into a global model. This approach offers several advantages, including reduced communication overhead and scalability across edge devices. As a result, federated learning has found applications in various domains, such as industrial operations, vehicular systems, military operations, financial services, and healthcare IoT applications.[ea]

Despite its promise, federated learning faces challenges involving numerous edge devices, primarily stemming from differences in hardware capabilities and network connectivity. The intelligent selection methods aim to enhance performance while minimizing resource demands.

Our methodology focuses on optimizing the selection of edge devices by considering factors such as computational power, memory, energy efficiency, and network access time. We aim to improve the overall performance of federated learning models by maximizing efficiency and accuracy metrics. However, it is essential to recognize that increasing the fraction of participating devices can improve performance but may also lead to increased resource consumption, including computational time and processing costs. Consequently, a nuanced approach to device selection is necessary to achieve optimal model performance in federated learning scenarios.

1.1 Motivation

Randomly limiting the participation of edge devices can negatively impact the performance of the global model in federated learning. Fig[1.2] depicts the potential heterogeneity condi-

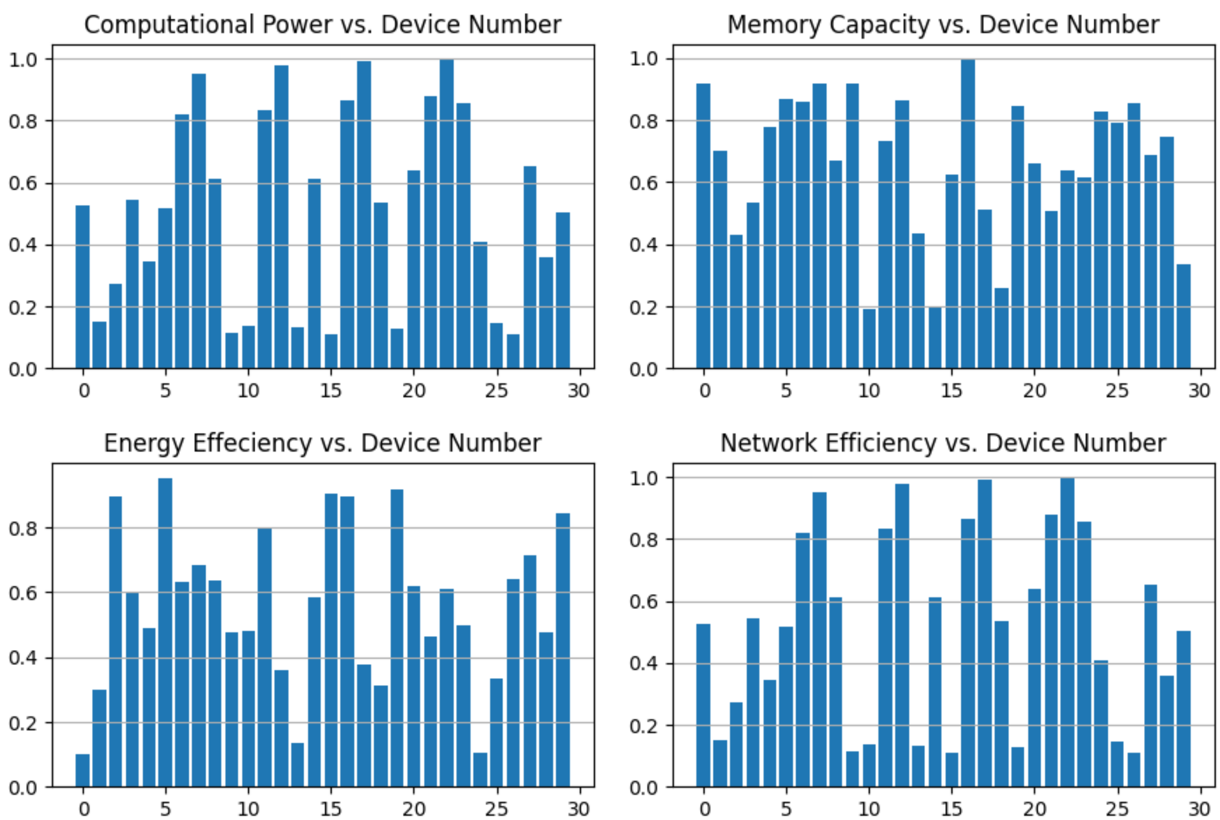

Fig. 1.2 Illustration of device heterogeneity

tions that may arise in an experimental setting based on the IRIS dataset distributed across edge devices. These figures highlight variations in device properties such as computational power, energy efficiency, network access efficiency, and memory capacity.

While previous studies have primarily focused on reducing resource utilization or improving model training performance, our work aims to address the limited resources of devices while simultaneously enhancing the performance of the global model. We propose a resource-efficient framework for optimal device selection that intelligently considers device resources while safeguarding user privacy. Our methodology optimizes parameters such as computational power, energy efficiency, network access efficiency, and memory capacity when selecting edge devices for processing. In doing so, we must balance the tradeoff between resource utilization and accuracy metrics.

Based on the setup depicted in Fig[1.1] and Fig[1.2], increasing the fraction of participating devices can improve performance and lead to increased resource consumption. It is crucial to recognize that selecting devices based solely on one parameter may result in the suboptimal performance of the global model.

Our work contributes in the following ways:

1. To navigate the accuracy versus resource tradeoff, we formulate device selection in federated learning as an optimization problem.

2. We develop a resource-efficient algorithm to address this optimization.

3. Evaluation results obtained with our methodology demonstrate the optimal balance between resource consumption and global model performance.

1.2 Organization of The Report

Chapter 1: This chapter introduces the reader to the problem. It describes the motivation behind it and briefly describes our approach to it.

Chapter 2: This chapter discusses some works related to this thesis.

Chapter 3: This chapter describes the mathematical model for the federated learning framework.

Chapter 4: This Chapter describes the proposed federated learning framework for optimal device selection. A detailed discussion of the proposed optimal device selection algorithm and model aggregation is subsequently presented.

Chapter 5: This chapter presents a comprehensive analysis of the performance of the proposed optimal device selection methodology within a federated learning framework.

Chapter 6: This chapter concludes the thesis.

Chapter 2

Review of Prior Works

Various approaches have been proposed to manage limited resources on edge devices, including partial device selection [CWCS], energy-efficient ML models, model compression, partial model training, and hierarchical aggregation techniques. However, this work concentrates on partial device selection to tackle communication and computation resource constraints. Traditionally, device sampling for federated learning has relied on either random device selection schemes [SJL20] or selection schemes based on data quantity[SPKS20].

In [CWCS], the focus is on sampling to address resource constraints. Here, devices are chosen and scheduled for sequential model uploading to the server, concurrently optimizing resource costs and training time. Additionally, federated client selection, described in [NY19], involves requesting random devices to share resource information. Subsequently, devices are selected based on these resources to maximize the number of selected devices.

2.1 Conclusion

These methods often entail partial or complete local training, leading to inefficient resource utilization. Therefore, there is a need to devise an effective edge device selection scheme that optimizes network resource usage without compromising model training performance or data privacy.

Chapter 3

System Model

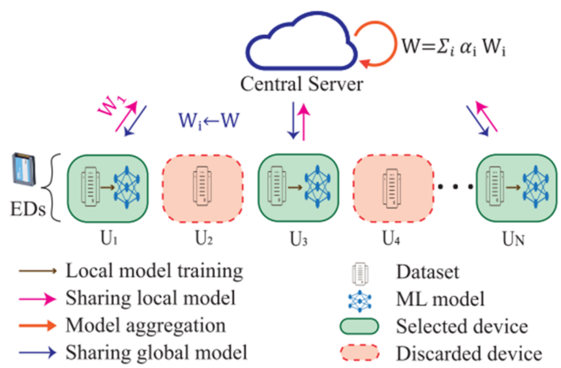

Fig. 3.1 High-level illustration of device selection in federated learning.

3.1 Problem Definition

Consider a network comprising N edge devices $N = \{1, 2, \ldots, N\}$ and a server, as illustrated in Fig. [3.1]. In this network, the edge devices collaboratively train a machine learning (ML) model characterized by a parameter matrix, $\mathbf{W}$. each device n possesses a local dataset $|D_{\mathrm{n}}|$.

. Each data sample in D_{n} takes the form (x, y), where x represents an $M \times 1$ vector, and y denotes a scalar label, particularly in the case of supervised learning.

In the conventional federated learning framework, a single global model is derived by training local models across all participating devices. However, this work explores a framework where only a subset of participating devices engages in local model training and subsequent aggregation. Illustrated in Fig. [3.1], this framework depicts the schematic of federated learning utilizing selected devices. Specifically, a subset of edge devices, denoted as $S \subseteq N$ partakes in both the training and aggregation processes.

Overall, federated learning aims to minimize the local model loss, denoted as $L(W_{\mathrm{n}}, D_{\mathrm{n}})$, where W_{n} represents the model parameter matrix associated with device n's local dataset D_{n}.

$$\min_{W} \sum_{n \in S} L(W(n), D(n)) \tag{3.1}$$

The model loss for device n is defined as follows:

$$L(W_{\mathrm{n}}, D_{\mathrm{n}}) = \frac{1}{D_{\mathrm{n}}} \sum_{(x_{\mathrm{d}}, y_{\mathrm{d}}) \in D_{\mathrm{n}}} l_{\mathrm{n}}(\mathbf{W}_{\mathrm{n}}, x_{\mathrm{d}}, y_{\mathrm{d}}) \tag{3.2}$$

Where l_n is the loss function (e.g., categorical crossentropy loss in the case of multiclass classification) for data $(x_{\mathrm{d}}, y_{\mathrm{d}}) \in D_{\mathrm{n}}$. The formulation of federated learning in (Eqn 3.1) aims to minimize the local loss to improve the accuracy of the global model, irrespective of the resource requirement. To address this, we aim to maintain accuracy above the threshold in this work by minimizing the model loss. To do so, we formulated a function that maximizes efficiency by comparing the heterogeneity of the device's properties using a heuristic-based approach. Formally, our optimization problem is as follows:

$$\max F = f(\phi) + H(S) \tag{3.3}$$

where $f(\phi)$ is the threshold accuracy of a global model. $H(S)$ is a heuristic function that considers device properties.

3.2 Heuristic Function

We propose a function to optimize resource requirements to enhance performance, denoted as $H(S)$, where S represents the subset of participating devices. This function can be expressed as follows:

$$H(S) = \alpha.f_1(S) + \beta.f_2(S) + \gamma.f_3(S) + \delta.f_4(S) \tag{3.4}$$

Here, f_1,f_2,f_3, and f_4 represent coefficients that account for various aspects of device heterogeneity, including computational power, energy efficiency, memory capacity, and network efficiency, respectively, within the subset of devices S.

Device heterogeneity can significantly impact performance, as the optimization function aims to maximize efficiency across these diverse parameters.

Formally, each subfunction is an N dimension vector $\mathbf{f}_{\mathrm{i}} = [f_{\mathrm{i},1}, \ldots, f_{\mathrm{i,N}}]^T$

for $i = 1, 2, \ldots, 4$ where $f_{\mathrm{i,n}}$ is the value of i^{th} subfunction for device n. For device selection status, we define an N dimension binary vector $\mathbf{s} = [s_1, \ldots, f_{\mathrm{N}}]^T$. $s_{\mathrm{n}} = 1$, if, $n \in S$, 0 otherwise.

The sub-function is defined as follows:

1) $f_1(S)$ Computational power: It quantifies the processing capability of each device within the subset S. Devices with higher computational power contribute more significantly to the optimization process, allowing for faster model training and aggregation.

$$f_1(S) = \frac{1}{|S|} s^T \mathbf{f}_1(S) \tag{3.5}$$

2) $f_2(S)$ Energy efficiency: This factor accounts for the energy consumption profile of each device in S. Devices with higher energy efficiency contribute positively to the optimization function, as they consume less energy during model training and aggregation, thereby reducing overall resource consumption.

$$f_2(S) = \frac{1}{|S|} s^T \mathbf{f}_2(S) \tag{3.6}$$

3) $f_3(S)$ Memory capacity: This factor reflects the storage capacity available on each device within S. Devices with larger memory capacities can accommodate larger datasets and store intermediate model parameters more efficiently, thus enhancing the overall optimization.

$$f_3(S) = \frac{1}{|S|} s^T \mathbf{f}_3(S) \tag{3.7}$$

4) $f_4(S)$ Network efficiency: It evaluates the communication efficiency of each device in S, considering factors such as network bandwidth and latency. Devices with higher network efficiency contribute more effectively to the optimization function, facilitating faster and more reliable data transmission during the model training and aggregation phases.

$$f_4(S) = \min_n [s.\mathbf{f}_4(S)]_\mathrm{n} \tag{3.8}$$

Chapter 4

Device Selection Algorithm in Federated Learning

This section outlines the proposed federated learning framework for optimal device selection. Initially, we present a comprehensive framework description and a detailed discussion of the optimal device selection algorithm and model aggregation process.

The protocol for optimal device selection within a federated learning framework is a step-by-step process reiterated at each learning round. These steps are enumerated as follows:

1. Device Information: At the outset, available edge network devices exchange information about their properties. This information is diverse information about properties and aids in subsequent device selection.

2. Device Selection: Leveraging the information obtained from the previous step, devices are intelligently selected based on various factors such as computational power, energy efficiency, memory capacity, and network efficiency.

3. Model Distribution: Following device selection, the server disseminates the global model parameters to the devices chosen in the previous step.

4. Local Training: The selected devices proceed to perform local updates on the global

parameters received from the server. Subsequently, they return the updated parameters and model weights to the server.

5. Model Aggregation: Finally, the server aggregates the local parameters received from the devices in the previous step and updates the global model accordingly. This aggregation process ensures that the global model reflects the collective knowledge of participating devices.

By following these systematic steps, the proposed federated learning framework facilitates the optimal selection of devices and the aggregation of model parameters, thereby enhancing the efficiency and performance of the federated learning process.

4.1 Device Selection Using Heuristic Function

In this section, we address the optimal device selection problem utilizing the heuristic function and then detail the algorithm to solve this optimization challenge. The problem formulation aims to maintain the accuracy above the threshold simultaneously, and the heuristic function to identify the optimal devices and optimize model parameters. Within environments, accuracy calculation entails updating global models across all devices in each round, leading to increased resource consumption. Instead of exact accuracy values, training quality can be anticipated by analyzing local devices to mitigate excessive computation. Therefore, optimal device selection revolves around maximizing the heuristic function, defined as a subproblem of (Eqn. 3.3).

4.2 Device Selection Algorithm

A common approach to selecting optimal devices is set-based device selection, where the set with the maximum $F(S)$ among all possible sets of size $|S|$ is identified as optimal. However, this method's search space size, initially increasing exponentially with N, becomes symmetric and decreases after reaching a maximum at $|S| = N/2$. This exponential growth in search space size with N and $|S|$ leads to significant time complexity and resource consumption.

Algorithm 1 Heuristic-Based Device Selection

Input: Network model N; fraction of devices selected ϵ; diversity information f_i^n $\forall i, n$. parameters $\alpha, \beta, \delta, \gamma$

Initialization: $|S| = [\epsilon N]$

  for $j = 1$ to N **do**

      $S_{\text{j}} = j$

    for $k = 1$ to $|S| - 1$ **do**

      $S^{\sim} = N/S_{\text{j}}$

      $n_{\max} = \text{arg max } H(S^{\sim})$

      $S_j \cup n_{\max}$

    end for

  end for

      return $S^* = \text{arg } max_j F(S_j)$

To address this, we propose an iterative improvement method in which devices that increase the heuristic are successfully added to the preexisting set until $|S|$ devices are selected. This iterative approach offers quicker candidate set selection, simplifying the evaluation of possible solutions for each device as the initial device. For a specific round, k, let, S^* be the optimal set of $|S|$ devices. At the j^{th} iteration, S_{j} represents the optimal set when device j is taken as the initial device. Subsequently, the remaining $|S|$ -1 devices are selected one after another, with the following device chosen from the remaining devices (N/S_{j}) to maximize the objective function F when added to the preexisting set S_{j}. The overall search space size of the iterative improvement method is then computed as $N \sum_{k=1}^{|S|-1}$, which, on average, is much smaller than that of the set-based selection method.

4.3 Model Aggregation

Upon acquiring the set of optimal devices for the k^{th} round of federated learning, denoted as $S^{(k)} = S^*(T_{k-1}+1)$, where T_{k1} represents the time of the $(k-1)^{th}$ aggregation, the global model parameters obtained from the previous aggregation, $\mathbf{W}^{k-1}$, are disseminated to the selected devices. Specifically, for all devices n within $S^{(k)}$, the parameters are updated as

$\mathbf{W}_n(T_{k-1}+1) \leftarrow \mathbf{W}^{k-1}$. Subsequently, these devices update the parameters using their local dataset and transmit the updated parameters to the server for aggregation. The updated parameters at each selected device for the k^{th} round are assigned model weights denoted as d_k^n. The k^{th} aggregation is executed as described in equation (4.1), resulting in the global model $\mathbf{W}^{(k)}$.

$$\mathbf{W}^{(k)} = \frac{1}{d^{(k)}} \sum_{n \in S^{(k)}} d_n^{(k)} \mathbf{W}_n(T_k) \tag{4.1}$$

where $d^{(k)} = \sum_{n \in S^{(k)}} d_n^{(k)}$ and $d_n^{(k)}$ is number of training data samples used by device n during k^{th} federated learning round.

Chapter 5

Performance Evaluation

This section presents a comprehensive analysis of the performance of the proposed optimal device selection methodology within a federated learning framework. It begins with a detailed description of the experimental setup, elucidating the parameters and conditions governing the experiments. Subsequently, experiments are conducted to evaluate and compare the efficacy of the proposed optimal device selection approach against the set based selection method. Accuracy assessments and resource utilization analyses are meticulously carried out through these experiments. This examination aims to provide valuable insights into the effectiveness and efficiency of the proposed methodology in selecting optimal devices for federated learning tasks.

5.1 Experimental Setup

1) Data Set and Machine Learning Model

We utilize the IRIS dataset, which consists of 150 samples of iris flowers, each containing four features: sepal length, sepal width, petal length, and petal width. This dataset is renowned in machine learning and statistics for classification tasks and is a benchmark dataset in the field. We train a single multilayer perceptron (MLP) with 32 neurons in the hidden layer for image classification using federated learning within the Tensor-

Flow Federated framework. Each user trains for five epochs, employing sparse categorical cross-entropy as the loss function. While alternative machine learning models can enhance accuracy across all scenarios, their relative performance remains consistent.

2) Evaluation Parameters

For federated learning, we engaged 30 edge devices alongside one server to strike the optimal balance between properties using a parameter value (of $\alpha, \beta, \gamma, \delta$) 0.25 in Eqn (3.4). Our proposed approach sought to identify optimal devices based on their properties, with performance comparisons across various parameter values. Additionally, we contrasted these results with scenarios where all devices partake in training and aggregation, as depicted in Fig. [5.1]

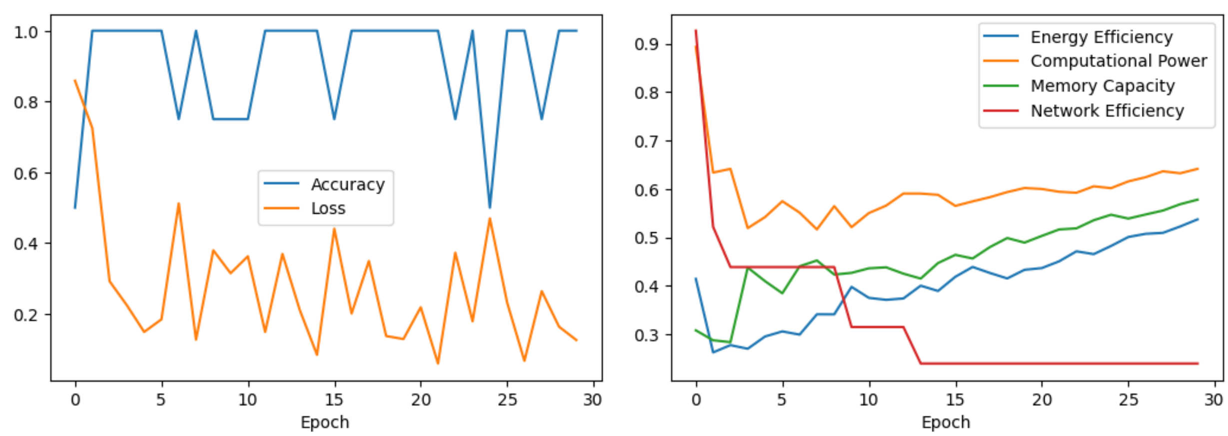

Fig. 5.1 Result for $\epsilon = 1$

5.2 Experiments on Optimal Device Selection

Experiments are conducted to analyze two device selection method. Set-based selection and Iterative improvement, in terms of computational time and optimality of the selected device set.A rigorous optimal device selection for $\epsilon = 0.5, 0.75$ using both selection methods is shown in Fig. [5.2].

A comparison of device selection methods based on the optimality of selected devices is conducted in terms of accuracy in finding the optimal set.

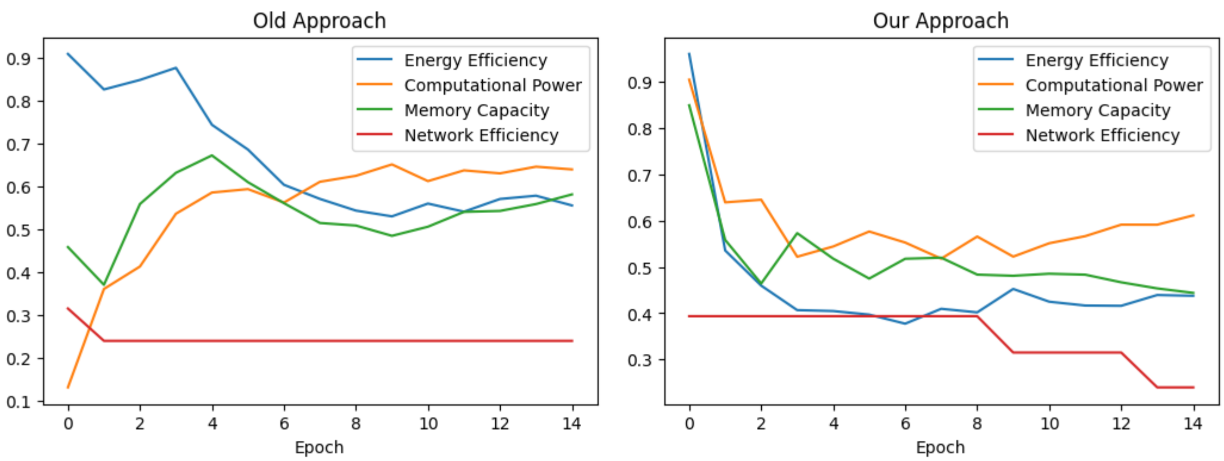

Fig. 5.2 Result for $\epsilon = 0.5$

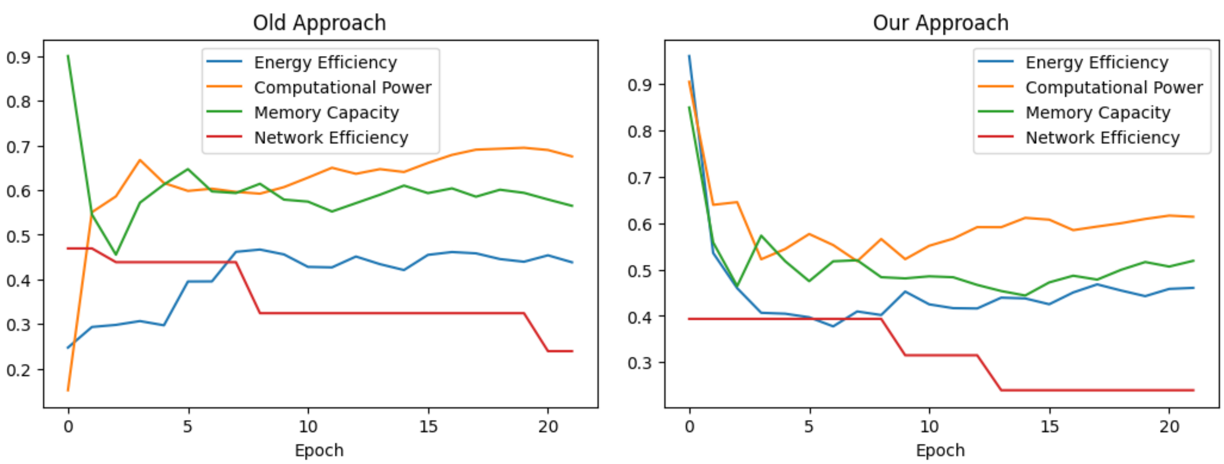

Fig. 5.3 Result for $\epsilon = 0.75$

Table 5.1 Comparison of Accuracy and Loss

Method	**Accuracy (%)**	**Loss**
Set-Based Method ($\epsilon = 1$)	90	0.21
Proposed Algorithm ($\epsilon = 1$)	92.33	0.16
Set-Based Method ($\epsilon = 0.5$)	89.9	0.21
Proposed Algorithm ($\epsilon = 0.5$)	90.00	0.27

As the set-selection method examines F of all the possible sets before selecting the optimal set, it always obtains the optimal set. In contrast, the iterative improvement method tries to improve the device set by including devices one by one. The algorithm performs better when checking for all N possible devices as an initial device instead of starting with the best initial device.

Chapter 6

Conclusion

The shift from cloud to edge computing enhances local data processing, reducing latency. Federated learning further enhances privacy by sharing only model parameters, enabling efficient processing of large datasets through parallel processing. Unlike conventional federated learning methods, where all devices participate in training and aggregation, our proposed approach involves only a subset of devices. Evaluation on the IRIS dataset demonstrated that optimal device selection leads to improved performance. The conventional method attains an accuracy of 90% when all devices are selected, whereas our proposed approach achieves a higher accuracy of 92.33% for the global model. Moreover, when utilizing a fraction of selected devices ($\epsilon = 0.5$), our approach outperforms the conventional model, attaining 90% accuracy compared to the latter's 89.9% .

References

[CWCS] C.-K. Yang J.-J. Kuo C.-W. Ching, Y.-C. Liu and F.-T. Su. Optimal device selection for federated learning over mobile edge networks.

[ea] J. Zhou et al. A survey on federated learning and its applications for accelerating industrial internet of things.

[NY19] T. Nishio and R. Yonetani. Client selection for federated learning with heterogeneous resources in mobile edge. *Proc. IEEE Int. Conf. Commun. (ICC)*, 2019.

[SJL20] A. Walid S. Ji, W. Jiang and X. Li. Dynamic sampling and selective masking for communication-efficient federated learning. 2020.

[SPKS20] M. Mohri-S. Reddi S. Stich S. P. Karimireddy, S. Kale and A. T. Suresh. Scaffold: Stochastic controlled averaging for federated learning. *Proc. Int. Conf. Mach. Learn.*, 2020.

[WZKA19] S. Hakak-I. Yaqoob W. Z. Khan, E. Ahmed and A. Ahmed. Edge computing: A survey. *Future Gener. Comput. Syst., vol. 97*, 2019.