Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 2
Open Access

A Federated Learning Approach for Non-Co-Located Datasets: Enhancing Data Governance and Privacy

Jyoti L. Bangare[1], Nilesh P. Sable[2], Parikshit N. Mahalle[3], Gitanjali R. Shinde[4]

[1]Ph.D. Scholar, Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India.

Email: jyoti.bangare@cumminscollege.in

[2]Associate Professor, Department of Computer Science & Engineering (Artificial Intelligence), BRACT's, Vishwakarma Institute of Information Technology, Pune, India.

Email: drsablenilesh@gmail.com

[3]Dean R&D, Professor and Head, Department of Artificial Intelligence and Data Science, Vishwakarma Institute of Information Technology, Pune, India.

Email: aalborg.pnm@gmail.com

[4]Associate Professor, BRACT's Vishwakarma Institute of Information Technology, Pune, India. Email: gr83gita@gmail.com

Article Info

Article type:

Research

Article History:

Received: 2024-03-18

Revised: 2024-05-10

Accepted: 2024-06-21

Keywords:

Federated Learning, Data Governance, Privacy Preservation, Transfer Learning, Non-Co-Located Datasets, Aggregation Methods

A B S T R A C T

In the contemporary landscape of data-driven decision-making, safeguarding data privacy and adhering to stringent data governance regulations are paramount. This research explores a federated learning approach for non-co-located datasets, aiming to enhance data governance and privacy. The study investigates the performance of federated models employing transfer learning techniques under different data split scenarios and aggregation methods. Initially, datasets are partitioned using IID (Independent and Identically Distributed) and non-IID (non-Independent and Identically Distributed) splits to simulate varied data distribution environments. Subsequently, two prominent federated aggregation methods, FedAvg and FedProx, are applied to aggregate the local models. Transfer learning models, specifically Federated MobileNet and Federated Inception, are utilized to leverage pre-trained networks for improved learning efficiency and accuracy. The experimental results demonstrate that, under IID dataset splits, the Federated MobileNet model achieves accuracies of 78.33% and 81.33% with FedAvg and FedProx respectively, while the Federated Inception model records accuracies of 71% and 74% with the same methods. Conversely, under non-IID dataset splits, Federated MobileNet attains an accuracy of 85.86% with FedAvg and 79% with FedProx, whereas Federated Inception consistently yields an accuracy of 66.67% for both aggregation methods. These findings underscore the efficacy of federated learning in managing non-co-located datasets, with a notable impact on model performance depending on the data split and aggregation method. The research highlights the potential of federated learning combined with transfer learning to enhance data privacy and governance, providing a robust framework for future applications in distributed environments.

1. INTRODUCTION

In the contemporary era of big data and advanced machine learning, the protection of data privacy and adherence to stringent data governance regulations have become paramount concerns[1], [2]. The exponential growth of data generation across various domains necessitates innovative solutions that can effectively manage and utilize this data while safeguarding individual privacy and ensuring compliance with regulatory standards. Federated learning emerges as a promising approach, offering a framework that allows for collaborative model training without the need

to centralize data. This approach addresses critical privacy concerns and regulatory requirements, enabling the utilization of non-co-located datasets[3], [4].

The primary objective of this research is to enhance data governance and privacy through a federated learning approach. This study investigates the efficacy of different data split strategies—specifically, "Independent and Identically Distributed" (IID) and "non-Independent and Identically Distributed" (non-IID) splits. It also examines the impact of two prominent aggregation methods, FedAvg and FedProx, on model performance. To further optimize the learning process, transfer learning models, namely Federated MobileNet and Federated Inception, are integrated within the federated learning framework[5], [6].

We utilize three datasets for this approach:

- Tuberculosis (TB) Chest X-Ray Dataset: This dataset, sourced from Kaggle, provides a comprehensive collection of chest X-ray images for TB diagnosis.
- Chest X-Ray Pneumonia Dataset: Sourced from IEEE Dataport, this dataset includes chest X-ray images used to detect pneumonia.
- Chest X-Ray COVID-19 Pneumonia Dataset: Available on Kaggle, this dataset comprises chest X-ray images to diagnose COVID-19 pneumonia.

These datasets are loaded and organized into training, testing, and validation directories to facilitate effective model training and evaluation. The data is then partitioned using IID and non-IID splits to simulate varied data distribution environments. This process allows us to assess the robustness of our federated learning approach under different conditions[7].

Federated learning enables multiple entities to collaboratively train a model while keeping their data localized. This study focuses on two aggregation methods within federated learning: FedAvg and FedProx. FedAvg, or Federated Averaging, aggregates local model updates by averaging them:

$$w_t = \frac{1}{N}\sum_{i=1}^{N} w_t^i$$

where w_t represents the global model weights at iteration t and w_t^i denotes the local model weights from the i^{th} client.

FedProx extends FedAvg by incorporating a proximal term to address heterogeneity among clients:

$$w_{t+1} = w_t - \eta(\nabla F_i(w_t) + \mu(w_t - w_t^i))$$

Here, η is the learning rate,$(\nabla F_i(w_t)$ is the gradient of the local objective function, and $\setminus\ \mu$ is the proximal term coefficient.

Transfer learning models, such as Federated MobileNet and Federated Inception, are employed to leverage pre-trained networks for improved learning efficiency and accuracy. The use of these models aims to enhance the federated learning process by transferring knowledge from previously trained models to new tasks.

The experimental results of our study highlight the performance of federated models under different scenarios. For IID dataset splits, the Federated MobileNet model achieves accuracies of 78.33% with FedAvg and 81.33% with FedProx, while the Federated Inception model records accuracies of 71% and 74% respectively. In contrast, for non-IID dataset splits, the Federated MobileNet attains an accuracy of 85.86% with FedAvg and 79% with FedProx, whereas the Federated Inception model consistently yields an accuracy of 66.67% for both aggregation methods.

Our contributions in this research are multifaceted:

- We demonstrate the viability of federated learning for managing non-co-located datasets while ensuring robust data governance and privacy.
- We provide a comprehensive comparison of the impact of IID and non-IID data splits on model performance.
- We evaluate the effectiveness of FedAvg and FedProx aggregation methods in a federated learning context.

This research highlights the potential of federated learning combined with transfer learning to address the challenges of data privacy and governance in distributed environments. The insights gained from this study pave the way for future applications and improvements in federated learning methodologies.

2. LITERATURE REVIEW

The growing importance of data privacy and governance in the field of machine learning has led to extensive research on privacy-preserving techniques and frameworks. Federated learning, in particular, has gained significant attention due to its ability to train models across distributed datasets without the need for data centralization. This review examines various studies that contribute to the understanding and development of federated learning and related privacy-preserving methodologies, highlighting existing gaps that the proposed approach aims to address.

Guan et al.[8] proposed a data-driven simulation approach for two-dimensional cross-correlated random fields using joint sparse representation, showcasing the potential of leveraging limited measurements for reliable data modeling. Chen et al.[9] explored the association between primary and behavioral health integrated care and pediatric mental disorder treatment, emphasizing the need for effective data integration and privacy considerations in healthcare.

Li et al.[10] discussed the challenges and solutions for preserving data privacy via federated learning, highlighting key obstacles such as data heterogeneity and communication overheads. Truong et al.[11] provided an insightful survey on privacy preservation in federated learning from the GDPR perspective, underscoring the regulatory challenges and the need for compliant data processing methods. Chen[12] focused on model compression techniques for efficient and privacy-preserving federated learning, proposing methods to reduce computational complexity while maintaining privacy.

Salomons et al.[13] investigated the impact of an in-home co-located robotic coach on exercise behavior, illustrating the practical applications of privacy-preserving technologies in personal health management . Zhao et al.[14] addressed energy-efficient and fair IoT data distribution in decentralized federated learning, presenting strategies to balance computational load and energy consumption. Lu et al.[4] integrated blockchain with federated learning for privacy-preserved data sharing in industrial IoT, demonstrating a novel approach to secure data transactions.

Wang et al.[15] proposed federated transfer learning for cross-domain prediction in smart manufacturing, indicating the effectiveness of combining federated and transfer learning to enhance predictive capabilities across different domains. Lee et al.[16] explored privacy-preserving data mining for open government data from heterogeneous sources, emphasizing the importance of secure data processing in public sector applications. Cuzzocrea et al.[17] discussed supporting privacy-preserving big data analytics on temporal open big data, highlighting the challenges of maintaining privacy in large-scale data environments .

Zapechnikov[18] examined privacy-preserving machine learning as a tool for secure personalized information services, showcasing the role of secure machine learning models in providing tailored services without compromising user privacy. Javid et al.[19] introduced a hybrid-security model for privacy-enhanced distributed data mining, proposing a multi-faceted approach to enhance data security and privacy in distributed systems .

Research Gap and Proposed Approach

Despite the advancements in federated learning and privacy-preserving techniques, several gaps remain. Current research often overlooks the practical challenges of implementing federated learning in environments with non-co-located datasets, particularly in terms of data governance and privacy compliance. Additionally, the impact of different data split strategies (IID vs. non-IID) and the effectiveness of various aggregation methods (FedAvg vs. FedProx) have not been extensively explored. The integration of transfer learning models within the federated learning framework also requires further investigation to understand its potential benefits fully.

In light of these gaps, this research proposes a federated learning approach for non-co-located datasets, focusing on enhancing data governance and privacy. By investigating different data split strategies and aggregation methods, and incorporating transfer learning models such as Federated MobileNet and Federated Inception, this study aims to provide a comprehensive solution to the challenges of distributed data management and privacy preservation. The findings are expected to contribute significantly to the development of robust, privacy-preserving federated learning frameworks applicable across various domains.

3. METHODOLOGY

3.1. Dataset used - For this research, we utilized three datasets: the Tuberculosis (TB) Chest X-Ray Dataset from Kaggle, the Chest X-Ray Pneumonia Dataset from IEEE Dataport, and the Chest X-Ray COVID-19 Pneumonia Dataset from Kaggle. These datasets were loaded and organized into training, testing, and validation directories

to facilitate effective model training and evaluation. The structured approach to data preparation ensures that the datasets are ready for subsequent processing and analysis within the federated learning framework.

3.2. Dataset preprocessing

To prepare the datasets for training, several preprocessing steps were applied. Firstly, all images were resized to a standard dimension of 224x224 pixels to ensure uniform input size for the models:

$$Resized\ Image=Resize(I,(224,224)$$

where I is the original image.

Next, the images were converted from RGB to grayscale to simplify the data and reduce computational complexity:

$$Gray\ Image=0.2989 R+0.5870 G+0.1140 B$$

where R, G, and B represent the" red, green, and blue" color channels respectively.

Finally, the datasets were shuffled to ensure a random distribution of data points, which helps in preventing any bias during the training phase:

$$Shuffled\ Dataset=Shuffle(D)$$

where D is the dataset. This preprocessing pipeline ensures that the data is standardized, simplified, and randomized, facilitating more effective and unbiased training of the models.

3.3. Perform Data Split

3.3.1. IID Split

In an "*Independent and Identically Distributed*" (IID) split, the dataset is divided such that each subset (training, validation, and test) maintains the same distribution of data. This ensures that each subset is a representative sample of the entire dataset. Let D be the entire dataset and let D_{train}, D_{val} and D_{test} be the training, validation and test sets resp., the IID split will represent as:

$$D_{train} \cup D_{val} \cup D_{test} = D$$
$$P(D_{train}) \approx P(D_{val}) \approx P(D_{test})$$

where P represents the probability distribution of the datasets. Each subset is drawn randomly and independently from D.

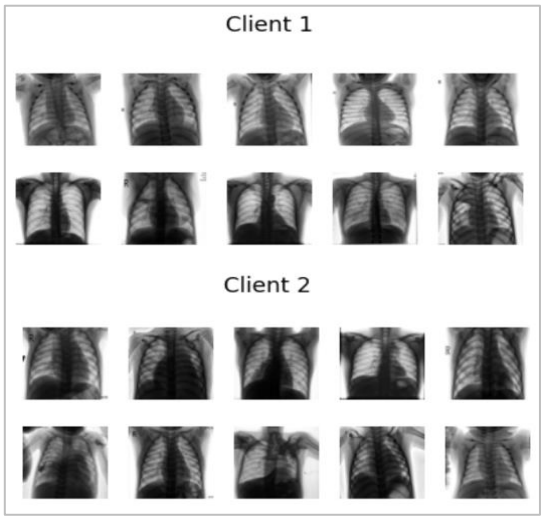


Figure. Data Split

3.3.2. Non-IID Split

In a "*Non-Independent and Identically Distributed*" (non-IID) split, the data distribution varies across different subsets. This means that the subsets may not have the same statistical properties, simulating real-world scenarios where data distribution is often uneven. Let D be the entire dataset and let D_{train}, D_{val} and D_{test}be the training, validation and test sets resp., the IID split will represent as:

$$D_{train} \cup D_{val} \cup D_{test} = D$$
$$P(D_{train}) \neq P(D_{val}) \neq P(D_{test})$$

where *P* represents the probability distribution of the datasets. In this case, the data points in each subset are selected in a way that they follow different distributions, reflecting the non-homogeneous nature of data in practical applications.

3.4. Create Clients

To simulate a federated learning environment, we need to create multiple clients each with its own subset of the data.

- **Provide name to clients**: Assign unique identifiers to each client for easy reference like "Client_1", "Client_2".
- **Divide input data into number of clients**:
  - **Determine the no**. of data points each client will receive.

$$Size = \frac{len(data)}{num\_clients}$$

- **Distribute data to each client :**
  - Assign the calculated no. of data points to each client. This ensures that each client has an approximately equal portion of the dataset.

$$client_i\ data = data\ [i\ \times size : (i+1) \times size$$

- **Separate shared data into data and labels lists**: For each client separate the data points and their corresponding labels into 2 distinct list.
  - This help in organizing the data such that each client has access to it subset of the input data and the corresponding labels for training and evaluation.

3.5. Apply Aggregating Method

3.5.1. FedAvg

Federated Averaging (FedAvg) is a method for aggregating model updates from multiple clients. The core idea is to average the updates (e.g., model weights) from each client to form a global model. Let N be the no. of clients and w_t^i be the model weights of the ith client at iteration t. The global weight w_{t+1} are updated by averaging the weights from all clients:

$$w_{t+1} = \frac{1}{N}\sum_{i=1}^{N} w_t^i$$

where, w_{t+1} is "new global model weights after aggregation"; this method ensures that each client's contribution is equally weighted.

3.5.2. FedProx

Federated Proximal (FedProx) extends FedAvg by adding a proximal term to address heterogeneity among clients. This term helps to keep the local updates close to the global model, which can be particularly useful when the clients' data distributions are non-IID. The update rule for FedProx is as follows:

Let η be the learning rate, $\nabla F_i(w_t)$ be the gradient of the local objective function for the ith client, μ be the proximal term coefficient, w_t be the global model weights at iteration t. The updated rule for the local model weights w_t^i is

$$w_{t+1}^i = w_t^i - \eta\left(\nabla F_i\left(w_t^i\right)\right) + \mu(w_t^i - w_t))$$

The global model weights w_{t+1} are then updated as in FedAvg:

$$w_{t+1} = \frac{1}{N}\sum_{i=1}^{N} w_{t+1}^i$$

3.6. Transfer Learning Model

3.6.1. Federated MobileNet

MobileNet is a lightweight deep learning model designed for mobile and embedded vision applications. In the context of federated learning, Federated MobileNet involves using a pre-trained MobileNet model and fine-tuning it on the federated dataset distributed across multiple clients.

- **Pre-trained model**: Let θ_{pre} be the parameters of the pre-trained MobileNet model.
- **Local Fine-Tuning**: Each client I fine-tunes the model parameters θ_i based on its local data. The update rule for client *i* is:

$$\theta_i^{t+1} = \theta_i^t - \eta\nabla F_i(\theta_i^t)$$

where, η is the learning rate, $\nabla F_i(\theta_i^t)$ is the gradient of the local loss funtion for the client i.

- **Federated Averaging:** The global model parameters θ are then updated by averaging the parameters from all clients:

$$\theta^{t+1} = \frac{1}{N}\sum_{i=1}^{N} \theta_i^{t+1}$$

3.6.2. Federated Inception

Inception is a deep learning model known for its inception modules, which allow for more efficient computation. Federated Inception involves using a pre-trained Inception model and adapting it to the federated learning framework as shown in figure-1.

- **Pre-trained model**: Let $\emptyset_{pre}$ be the parameters of the pre-trained inception model.
- **Local Fine-Tuning**: Each client I fine-tunes the model parameters $\emptyset_i$ based on its local data. The update rule for client *i* is:

$$\emptyset_i^{t+1} = \emptyset_i^t - \eta\nabla F_i(\emptyset_i^t)$$

where, η is the learning rate, $\nabla F_i(\emptyset_i^t)$ is the gradient of the local loss funtion for the client i.

- **Federated Averaging:** The global model parameters $\emptyset$ are then updated by averaging the parameters from all clients:

$$\emptyset^{t+1} = \frac{1}{N}\sum_{i=1}^{N} \emptyset_i^{t+1}$$

```
InceptionModel(
  (stem): Sequential(
    (0): Conv2d(3, 32, kernel_size=(3, 3), stride=(2, 2), padding=(1, 1))
    (1): BatchNorm2d(32, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): ReLU(inplace=True)
    (3): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (4): BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (5): ReLU(inplace=True)
    (6): MaxPool2d(kernel_size=3, stride=2, padding=1, dilation=1, ceil_mode=False)
  )
  (inception_block): Sequential(
    (0): Conv2d(64, 128, kernel_size=(1, 1), stride=(1, 1))
    (1): BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (2): ReLU(inplace=True)
    (3): Conv2d(128, 192, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
    (4): BatchNorm2d(192, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (5): ReLU(inplace=True)
    (6): Conv2d(192, 256, kernel_size=(1, 1), stride=(1, 1))
    (7): BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
    (8): ReLU(inplace=True)
  )
  (fc_layers): Sequential(
    (0): AdaptiveAvgPool2d(output_size=1)
    (1): Flatten(start_dim=1, end_dim=-1)
    (2): Dropout(p=0.5, inplace=False)
    (3): Linear(in_features=256, out_features=4, bias=True)
  )
```

Figure 1 Inception Model Architecture

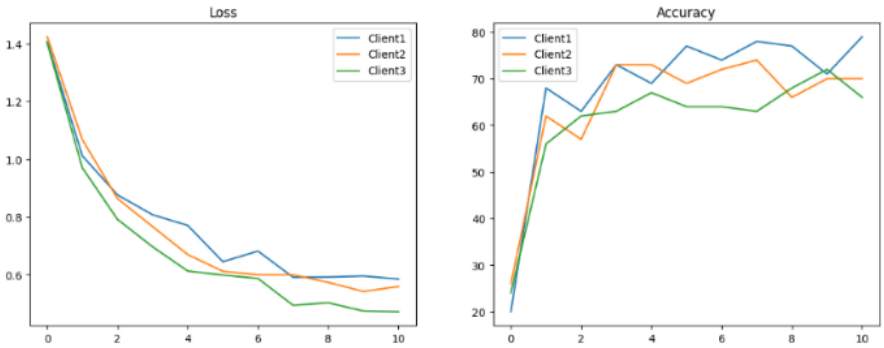
In both Federated MobileNet and Federated Inception, the transfer learning approach leverages pre-trained models, which are fine-tuned in a federated manner to adapt to the specific data available on each client, enhancing learning efficiency and model performance.

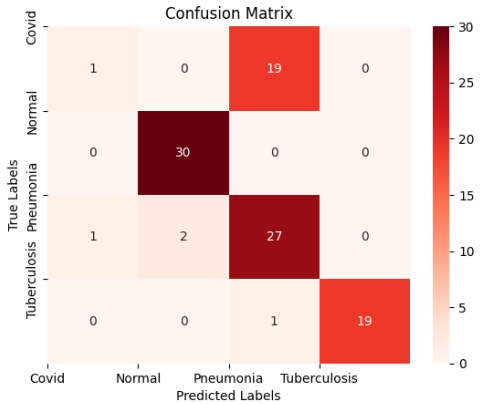
4. RESULTS, OUTPUT AND DISCUSSION

4.1. Inception Model

4.1.1. IID Split

- ***FedAvg (Local Accuracy and Local Loss Comparison)***

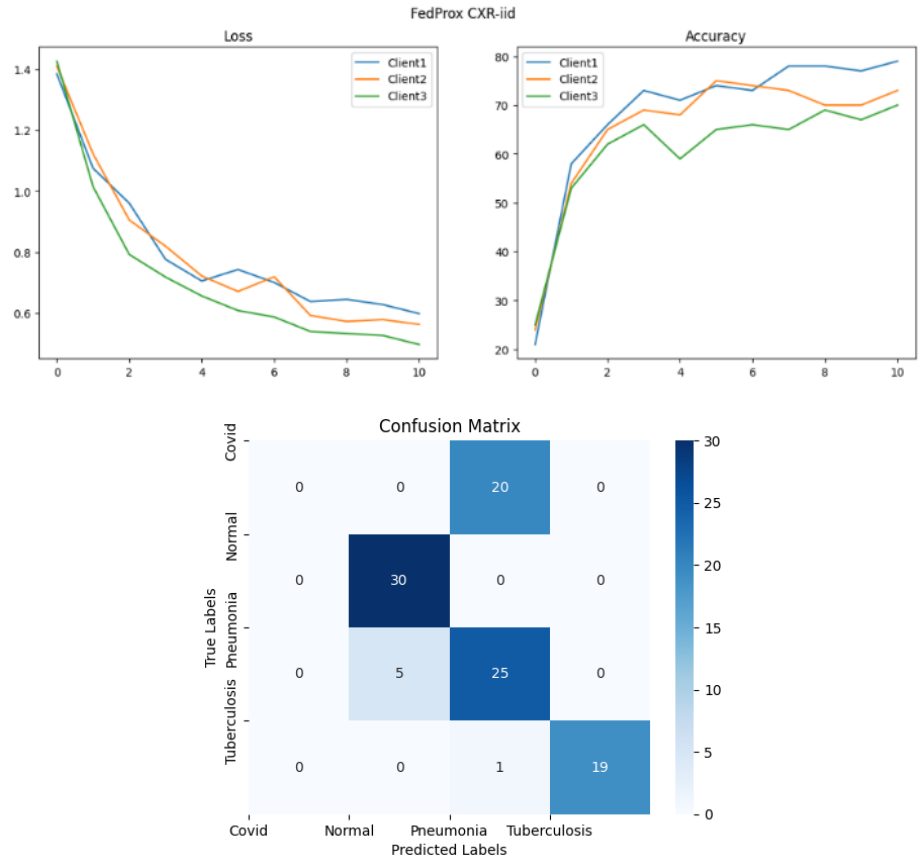




The figure- demonstrates the performance of the FedAvg aggregation method under an IID data split. The loss graphs show a consistent decrease in loss for all three clients, indicating that the model is converging well across all clients. The accuracy graphs reveal that each client achieves a high level of accuracy, with Client2 slightly

outperforming the others. This consistent performance across clients reflects the effectiveness of FedAvg in an IID environment.

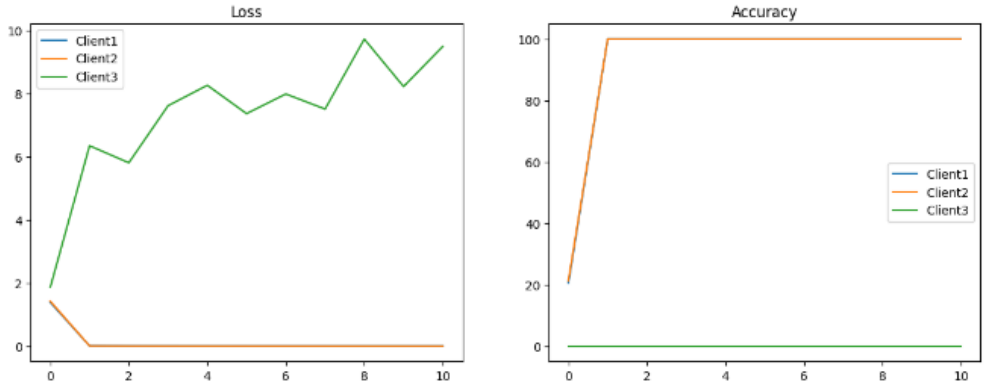
- ***FedProx (Local Accuracy and Local Loss Comparison)***

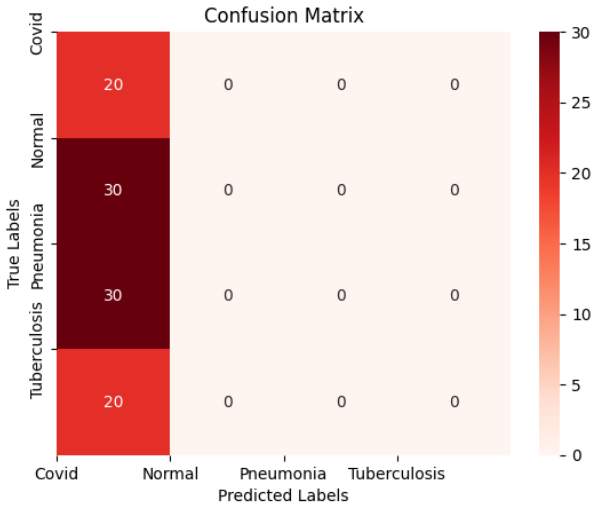


The second set of graphs displays the results for the FedProx aggregation method with IID data. The loss curves again show a smooth decline, similar to FedAvg, indicating effective model convergence. The accuracy graphs illustrate that FedProx also achieves high accuracy levels for all clients, with minor variations among them. FedProx's proximal term helps maintain stability and performance, showing comparable results to FedAvg in an IID scenario.

4.1.2. Non-IID Split

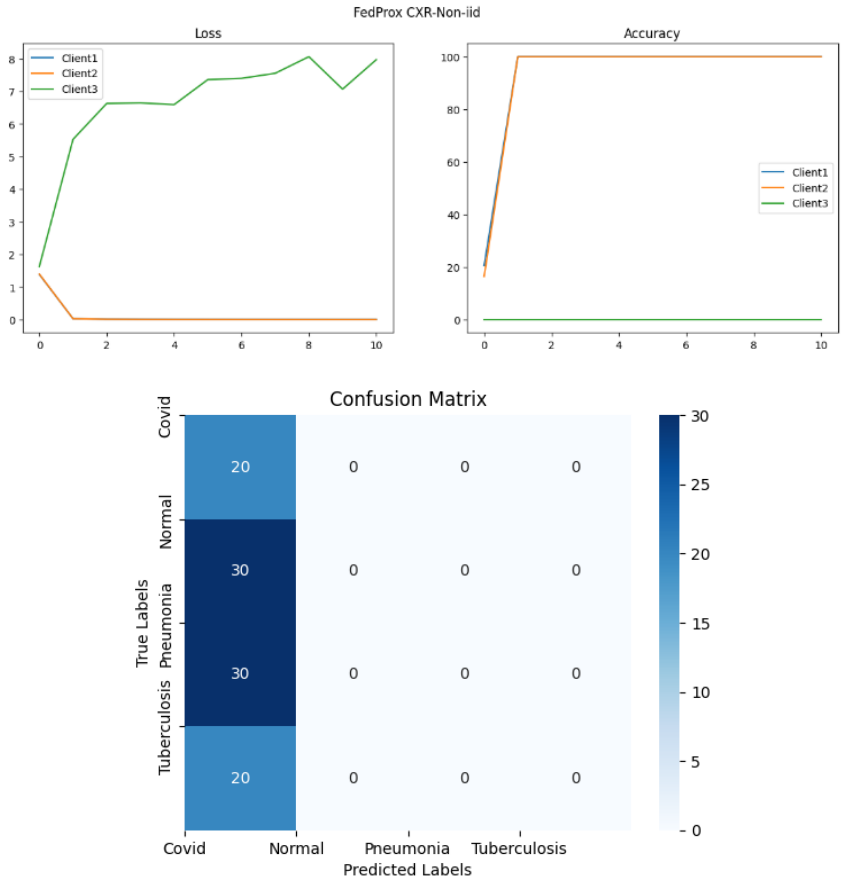
- ***FedAvg (Local Accuracy and Local Loss Comparison)***





The third set of graphs presents the FedAvg method's performance with non-IID data. The loss curves indicate a more varied convergence compared to the IID split, with some fluctuations reflecting the heterogeneous nature of the data. The accuracy graphs show that while the clients reach relatively high accuracy, there is more variability among them. This suggests that FedAvg can handle non-IID data but with some challenges in achieving uniform performance across all clients.

- ***FedProx (Local Accuracy and Local Loss Comparison)***



The fourth set of graphs highlights the FedProx method's performance under a non-IID split. The loss curves indicate a smoother convergence compared to FedAvg, likely due to the proximal term's regularizing effect. The accuracy graphs show that FedProx achieves high accuracy for all clients, with less variability compared to FedAvg. This

demonstrates FedProx's robustness in handling non-IID data, ensuring more consistent performance across different clients.

4.2. Mobilenet Model

4.2.1. IID Split

- ***FedAvg (Local Accuracy and Local Loss Comparison)***

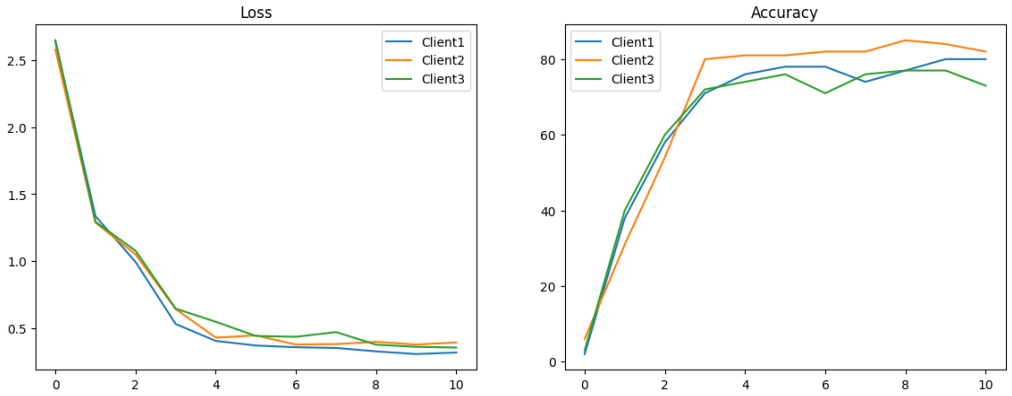


Figure 2 Local Accuracy and Local Loss Comparison

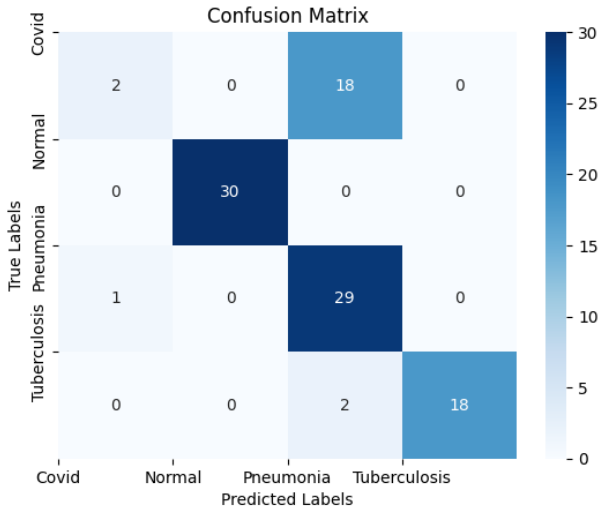


Figure 3 Confusion matrix

The figure-2 and 3 for the IID split with the MobileNet model using the FedAvg aggregation method shows a consistent decline in loss for all three clients, indicating effective convergence. The accuracy graphs reveal that all clients achieve high accuracy, with minor variations among them. This consistency demonstrates that FedAvg performs well under an IID split, ensuring each client benefits equally from the global model updates.

- ***FedProx (Local Accuracy and Local Loss Comparison)***

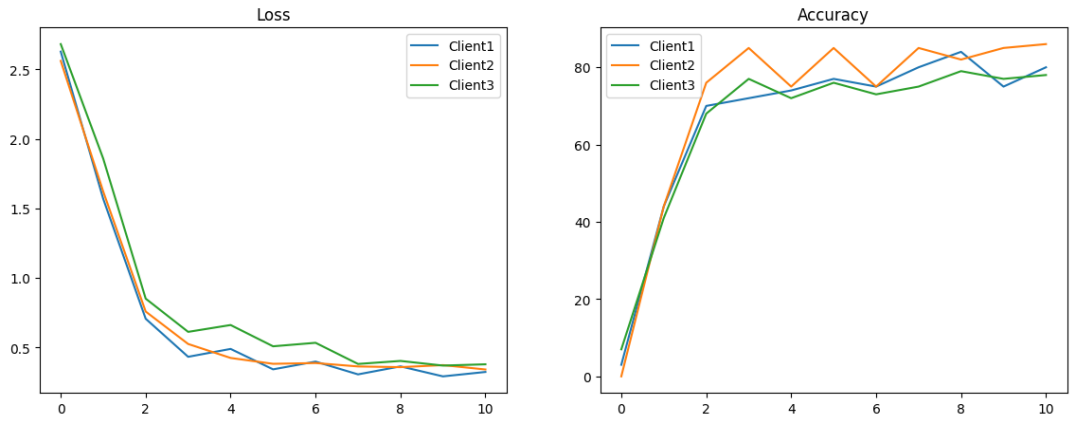


Figure 4 Local Accuracy and Local Loss Comparison

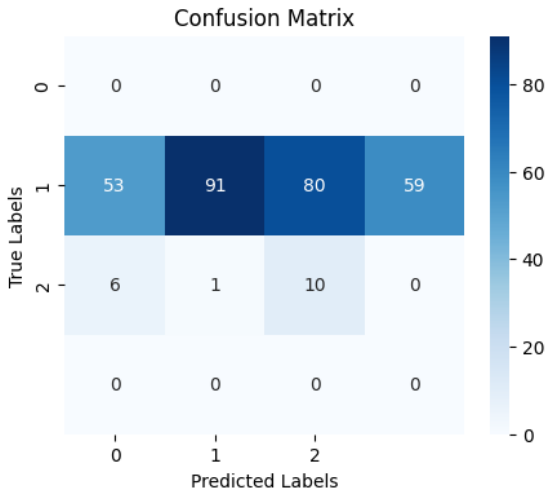


Figure 5 Confusion matrix

In the figure-4,5 for the IID split using the FedProx method, we observe a similar trend of decreasing loss across clients, though with a slightly smoother curve compared to FedAvg. The accuracy graphs indicate that clients achieve high and consistent accuracy levels, reflecting FedProx's ability to maintain stability and performance in an IID environment, likely due to its proximal term that helps regularize local updates.

4.3. NON IID Split

- ***FedAvg (Local Accuracy and Local Loss Comparison)***

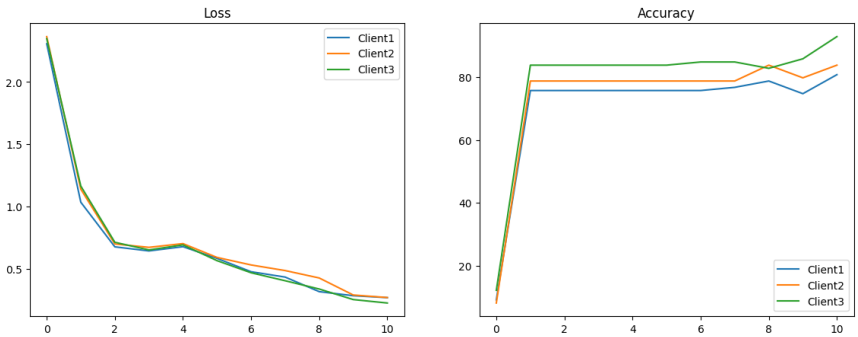


Figure 6 Local Accuracy and Local Loss Comparison

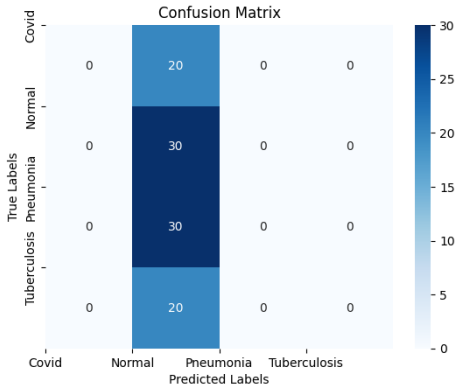


Figure 7 Confusion matrix

For the non-IID split using FedAvg, the figure-6,7 shows a more varied loss reduction, with some fluctuations reflecting the data heterogeneity. The accuracy graphs demonstrate that while clients reach high accuracy, there is noticeable variability among them. This indicates that FedAvg can handle non-IID data but with challenges in achieving uniform performance across clients.

- ***FedProx (Local Accuracy and Local Loss Comparison)***

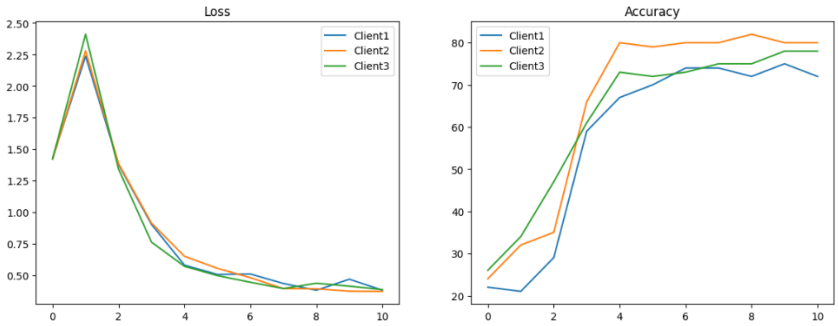


Figure 8 Local Accuracy and Local Loss Comparison

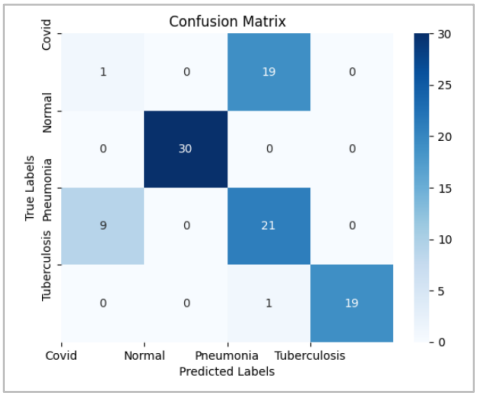


Figure 9 Confusion matrix

The figure-8,9 for the non-IID split using FedProx shows a smoother loss convergence compared to FedAvg, suggesting better handling of data heterogeneity. The accuracy graphs reveal high accuracy for all clients, with reduced variability compared to FedAvg. This underscores FedProx's robustness in non-IID settings, providing more consistent performance due to its regularizing effect.

4.4. Comparison of federated model using Transfer learning approach with IID and Non-IID dataset split

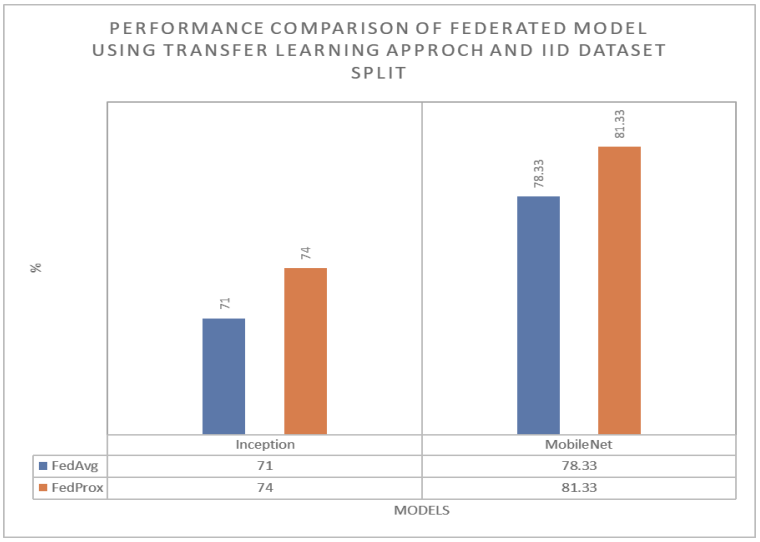


Figure 10 Performance Comparison of Federated Model Using Transfer Learning Approach and IID Dataset Split

The performance of federated models using transfer learning with an IID dataset split as shown in figure-10 shows distinct differences between the two aggregation methods, FedAvg and FedProx. For the Inception model, FedAvg achieves an accuracy of 71%, while FedProx improves this slightly to 74%. The MobileNet model performs better overall, with FedAvg reaching an accuracy of 78.33% and FedProx further enhancing it to 81.33%. These results indicate that FedProx consistently provides better performance than FedAvg when dealing with IID data, likely due to its ability to better handle variations and stabilize the training process through its proximal term.

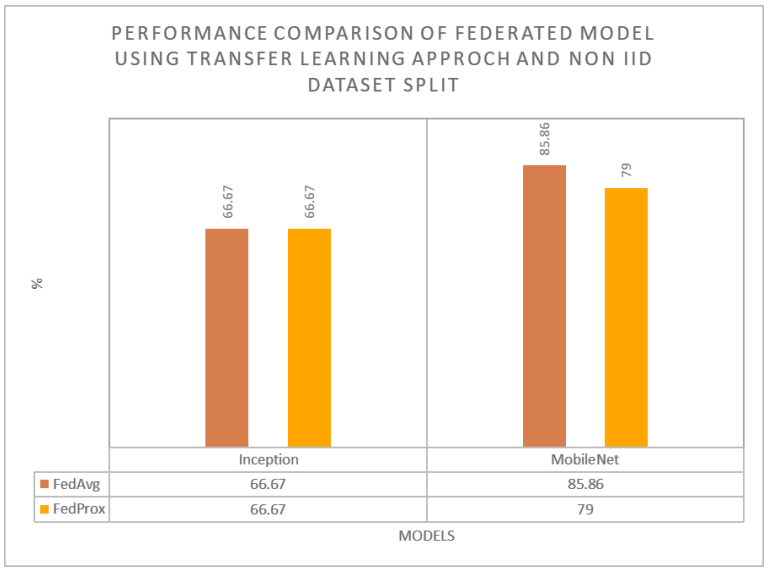


Figure 11 Performance Comparison of Federated Model Using Transfer Learning Approach and Non IID Dataset Split

When using a non-IID dataset split as shown in figure-11, the performance dynamics change noticeably. For the Inception model, both FedAvg and FedProx achieve the same accuracy of 66.67%, suggesting that neither method has a significant advantage in this non-IID context for this specific model. However, the MobileNet model shows a stark contrast: FedAvg achieves a higher accuracy of 85.86%, whereas FedProx results in a lower accuracy of 79%. This indicates that while FedProx generally helps in stabilizing training in IID

scenarios, it may not always offer superior performance in non-IID settings compared to FedAvg, particularly for certain models like MobileNet.

These comparisons highlight the importance of selecting appropriate aggregation methods based on the data distribution and the specific model being used. FedProx generally enhances performance in IID splits but may not always outperform FedAvg in non-IID splits. MobileNet consistently outperforms Inception in both IID and non-IID scenarios, underscoring its robustness and effectiveness as a transfer learning model in federated learning contexts.

5. CONCLUSION AND FUTURE SCOPE

In conclusion, this research underscores the potential of federated learning to revolutionize data governance and privacy preservation for non-co-located datasets. By systematically examining the effects of IID and non-IID data splits and applying different aggregation methods such as FedAvg and FedProx, this study provides valuable insights into optimizing federated learning frameworks. The integration of transfer learning models, specifically Federated MobileNet and Federated Inception, demonstrates significant improvements in model performance, particularly in handling diverse and distributed data environments. These findings not only validate the efficacy of federated learning but also pave the way for its broader adoption in various sectors requiring stringent data privacy and governance measures.

Looking ahead, the future scope of this research is expansive. One promising direction is the exploration of advanced aggregation techniques that can further enhance the robustness and efficiency of federated learning models. Additionally, extending this approach to include more complex and larger-scale datasets will be crucial in assessing the scalability and generalizability of the proposed methods. The integration of blockchain technology with federated learning also holds significant promise for enhancing security and trustworthiness in data transactions. In the end, continued research and innovation in this field will be essential to fully harness the transformative potential of federated learning in promoting secure, efficient, and privacy-preserving data management practices.

REFERENCES

[1] T. Alam and R. Gupta, "Federated Learning and Its Role in the Privacy Preservation of IoT Devices," *Future Internet*, vol. 14, no. 9. 2022, doi: 10.3390/fi14090246.

[2] M. Ali, F. Naeem, M. Tariq, and G. Kaddoum, "Federated Learning for Privacy Preservation in Smart Healthcare Systems: A Comprehensive Survey," *IEEE J. Biomed. Heal. Informatics*, vol. 27, no. 2, pp. 778–789, 2023, doi: 10.1109/JBHI.2022.3181823.

[3] Q. Li *et al.*, "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3347–3366, 2023, doi: 10.1109/TKDE.2021.3124599.

[4] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT," *IEEE Trans. Ind. Informatics*, vol. 16, no. 6, pp. 4177–4186, 2020, doi: 10.1109/TII.2019.2942190.

[5] S. Singh, S. Rathore, O. Alfarraj, A. Tolba, and B. Yoon, "A framework for privacy-preservation of IoT healthcare data using Federated Learning and blockchain technology," *Futur. Gener. Comput. Syst.*, vol. 129, pp. 380–388, 2022, doi: https://doi.org/10.1016/j.future.2021.11.028.

[6] Q. Yang, "Toward Responsible AI: An Overview of Federated Learning for User-centered Privacy-preserving Computing," *ACM Trans. Interact. Intell. Syst.*, vol. 11, no. 3–4, 2021, doi: 10.1145/3485875.

[7] W. Zhang and X. Li, "Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions," *Struct. Heal. Monit.*, vol. 21, no. 4, pp. 1329–1344, Jul. 2021, doi: 10.1177/14759217211029201.

[8] Z. Guan and Y. Wang, "Data-driven simulation of two-dimensional cross-correlated random fields from limited measurements using joint sparse representation," *Reliab. Eng. Syst. Saf.*, vol. 238, p. 109408, 2023, doi:

https://doi.org/10.1016/j.ress.2023.109408.

[9] H. Chen, N. Upadhyay, N. Lyu, and P. J. Rowan, "Association of Primary and Behavioral Health Integrated Care Upon Pediatric Mental Disorder Treatment," *Acad. Pediatr.*, vol. 21, no. 7, pp. 1187–1194, 2021, doi: https://doi.org/10.1016/j.acap.2021.05.021.

[10] Z. Li, V. Sharma, and S. P. Mohanty, "Preserving Data Privacy via Federated Learning: Challenges and Solutions," *IEEE Consum. Electron. Mag.*, vol. 9, no. 3, pp. 8–16, 2020, doi: 10.1109/MCE.2019.2959108.

[11] N. Truong, K. Sun, S. Wang, F. Guitton, and Y. Guo, "Privacy preservation in federated learning: An insightful survey from the GDPR perspective," *Comput. Secur.*, vol. 110, p. 102402, 2021, doi: https://doi.org/10.1016/j.cose.2021.102402.

[12] Y. Chen, "Learned Model Compression for Efficient and Privacy-Preserving Federated Learning," pp. 1–12.

[13] N. Salomons, T. Wallenstein, D. Ghose, and B. Scassellati, "The Impact of an In-Home Co-Located Robotic Coach in Helping People Make Fewer Exercise Mistakes," in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2022, pp. 149–154, doi: 10.1109/RO-MAN53752.2022.9900722.

[14] J. Zhao *et al.*, "Energy-Efficient and Fair IoT Data Distribution in Decentralised Federated Learning," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 3, pp. 1352–1363, 2023, doi: 10.1109/TNSE.2022.3185672.

[15] K. I.-K. Wang, X. Zhou, W. Liang, Z. Yan, and J. She, "Federated Transfer Learning Based Cross-Domain Prediction for Smart Manufacturing," *IEEE Trans. Ind. Informatics*, vol. 18, no. 6, pp. 4088–4096, 2022, doi: 10.1109/TII.2021.3088057.

[16] J. S. Lee and S. P. Jun, "Privacy-preserving data mining for open government data from heterogeneous sources," *Gov. Inf. Q.*, vol. 38, no. 1, p. 101544, 2021, doi: 10.1016/j.giq.2020.101544.

[17] A. Cuzzocrea, C. K. Leung, A. M. Olawoyin, and E. Fadda, "Supporting Privacy-Preserving Big Data Analytics on Temporal Open Big Data," *Procedia Comput. Sci.*, vol. 198, no. 2021, pp. 112–121, 2021, doi: 10.1016/j.procs.2021.12.217.

[18] S. Zapechnikov, "Privacy-Preserving Machine Learning as a Tool for Secure Personalized Information Services," *Procedia Comput. Sci.*, vol. 169, no. 2019, pp. 393–399, 2020, doi: 10.1016/j.procs.2020.02.235.

[19] T. Javid, M. K. Gupta, and A. Gupta, "A hybrid-security model for privacy-enhanced distributed data mining," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3602–3614, 2022, doi: 10.1016/j.jksuci.2020.06.010.