International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Federated Learning for IoT: Ensuring Privacy and Security in Distributed Networks

Yashoda Krishna Kuppili*[1], John Jaidhan B. #[2]

Submitted: 26/06/2023 **Revised**: 06/08/2023 **Accepted**: 25/08/2023

Abstract: Federated learning is a machine learning technique in which the model is trained across a number of decentralized devices (clients) without the need to move the data to a central location. Because users' data is stored on their devices rather than a central server or third party, this method gives consumers improved privacy and security. Contrarily, centralized learning mandates that users submit their data to a central server, raising issues with data security and privacy. The purpose of the current study was to evaluate and compare the effectiveness of the centralized and federated learning paradigms in the context of a simple regression task using simulated data. The findings demonstrated that while protecting user privacy, federated learning may attain accuracy levels that are on par with those of centralized learning. Our study also demonstrated the viability of implementing federated learning using well-known machine learning frameworks like TensorFlow Federated.

Index Terms: *Federated Learning, IoT Security, Centralised Learning*

1. Introduction

The proliferation of Internet of Things (IoT) devices and the associated demand for machine learning algorithms that can analyze data obtained from these devices were the impetus for the development of federated learning as a method of protecting users' privacy in machine learning. Federated learning is now being used as a method of protecting users' privacy in machine learning. During the process of machine learning, federated learning was established as a way to assure the protection of user privacy. Federated learning is a strategy for distributed machine learning that eliminates the requirement for numerous clients to communicate their local data with a centralized server in order for them to participate in the collaborative training of a single global model. The storage of raw data on the devices used by customers has a number of important benefits in terms of protecting the confidentiality of customer information since it lowers the risk of data breaches and unauthorized access. This tactic, when implemented, offers a number of benefits in addition to serving its core objective. On the other hand, standard machine learning models that are centralized in nature compile data from a variety of sources and then store it on a server that is centrally located. Because of the possibility of unwanted access to the data and the fact that the method in issue is prone to security breaches, privacy and security concerns have been raised in connection with it. In addition, the accumulation of a significant amount of data in a single area may result in the appearance of challenges that are related to the processing and storage of the data.

Department Of CSE, GITAM SCHOOL OF TECHNOLOGY, GITAM UNIVERSITY, VISAKHAPATNAM Ykuppili@Gitam.In
Department Of CSE, GITAM SCHOOL OF TECHNOLOGY, GITAM UNIVERSITY, VISAKHAPATNAM Jbeera@Gitam.Edu
** Corresponding Author Email: Ykuppili@Gitam.In*

In the context of the current discussion, this paper presents a comparative analysis of federated learning versus centralized learning with regard to data privacy and the correctness of machine learning models. We will use a straightforward linear regression issue as a means of illustrating why federated learning is superior to centralized learning. This will be done by comparing the two approaches through the lens of the problem. After that, the graph will show the intrinsic trade-off between accuracy and privacy that is present in both of these methods.

2. Literature Survey

The literature study presented below provides a comprehensive evaluation of recent research efforts that attempt to address significant challenges in the Industrial Internet of Things (IIoT) paradigm. In view of the rapid expansion of networked devices, this paper examines several strategies and technology advancements that seek to enhance data exchange, secrecy, and security. The survey covers a wide range of subjects, including blockchain technology, federated learning, differential privacy, and their applications in many different fields, such as urban informatics, SCADA networks, Industry 4.0, healthcare systems, and digital twin edge networks. This literature review provides insights into the innovative methods proposed to ensure data confidentiality, increase productivity, and enhance the dependability and security of IIoT systems by analyzing the findings and methodologies of key studies.

Due to the exponential growth of data given by connected devices, the industrial Internet of Things paradigm presents a chance for improving the quality of newly developed applications through data exchange. Wireless data transfer might be challenging due to worries about potential data

breaches and other security- and privacyrelated issues listed in [1]. Suppliers may be liable for additional expenses if they reveal confidential information. The first step entails developing a safe architecture based on blockchain technology allowing distributed parties to share data. The challenge of data sharing is changed into a machine learning problem by adopting privacy-preserving federated learning. A data model is suggested to ensure that the confidentiality of the data is maintained. Federated learning is included into the permissioned blockchain consensus architecture to enable training opportunities. The proposed approach for data interchange has been proven to be precise, efficient, and secure using real-world datasets with numerical outcomes. Urban informatics, a discipline that heavily relies on data, has evolved greatly as a result of the development of mobile edge computing and 5G technology, claim the authors of [2]. To successfully handle the proliferation of data, artificial intelligence (AI) approaches must be applied. Federated learning, which enables edge nodes to train models locally without transferring data to a centralized server, is a future approach for decentralized edge computing. One of the most effective edge computing applications is federated schooling. The implementation of federated learning in urban environments, such as automobile networks, is constrained by the security and privacy issues that arise in these circumstances. The current study proposes deferentially private asynchronous federated learning as a mechanism to share vehicular network resources. Federated learning makes use of local differential privacy to safeguard recently updated local models while moreover offering security and dependability. Our idea proposes employing a randomized and decentralized update strategy to alleviate the security vulnerabilities associated with centralized curatorial systems. The process of convergence is made easier in our system by the use of weighted aggregation and update verification mechanisms.

We assess the efficacy of our approach using three different sets of actual data. The numerical outcomes demonstrate how precise, efficient, and confidential our technique is. From the perspective of the stakeholders, a trustworthy IIoT network is highly anticipated because it is thought to be crucial for avoiding fatalities. Reference [3] supports this assertion. The reliability and resilience of an Industrial Internet of Things (IIoT) system depend on how reliable the IT infrastructure, which includes safety, security, and privacy, is. The typical security tools and techniques are unsuitable for safeguarding the IIoT platform because of anomalies in the protocols being used, a lack of upgrade choices, incompatibilities, and the presence of old industrial operating systems. If the current research introduces a trustworthy and financially viable way for detecting cyberattacks, the Industrial Internet of Things (IIoT) network, specifically Supervisory Control and Data Acquisition (SCADA), can be more trusted. The current study aims to increase the dependability of SCADA networks. This article discusses using an ensemble-learning algorithm to find SCADA system security issues. The current source of network traffic employed in the model is the Industrial Internet of Things (IIoT) platforms, which are constructed on SCADA systems. The recommended strategy promotes the development of a detection engine that uses network traffic based on commercial protocols in order to obtain high detection rates. Additionally, the overfitting issue is addressed using the ensemble random tree methodology, and the risk of recognizing false positives is decreased using the random subspace method. 15 different datasets of SCADA networks were utilized to validate the model. The experimental findings demonstrate that the suggested model outperforms conventional detection methods, bringing improved levels of dependability and security to the Industrial Internet of Things (IIoT) architecture. The Industrial Internet of Things (IIoT), according to the source [4], is bringing about a sizable shift in a number of areas, including healthcare, mining, agriculture, and power generation. Machine learning (ML) is a crucial element of Industry 4.0 in order to effectively utilize the vast number of networked devices and the resulting volume of data they produce. The usage of machine learning models that have been developed on sensitive data greatly hinders the full potential of Industry 4.0. This is due to the fact that these models put user privacy at risk by being susceptible to privacy violations by adversarial actors. The PriModChain platform integrates differential privacy, federated machine learning, the Ethereum blockchain, and smart contracts to guarantee the security of Industrial Internet of Things (IIoT) data. The general-purpose computer's dependability, safety, and resilience of PriModChain are evaluated through simulated Python socket programming. Kovan was used to test public blockchains as opposed to Ganache v2.0.1 which was intended to test local-level blockchains. Scyther software 1.1.3 is used to verify the provided security mechanism. [5] Recent years have seen a reduction in the difficulty and cost of early diagnosis of dementiarelated disorders due to the rapid rise of the intelligent healthcare system. The primary cause of concern with the system is personal data leaking. The development of the ADDetector tool for Alzheimer's disease employed the internet of things (IoT) and security measures (AD) to assist safeguard privacy. ADDetector blends cutting-edge topic-based linguistic features with a unique collection of user audio from Internet of Things devices present in smart homes in order to identify AD. The ADDetector system's three-layer architecture, which successfully safeguards the privacy of user features, data, and models, consists of the user, client, and cloud layers. Federated learning (FL) is used by the ADDetector system to give the user control over the confidentiality of the classification model and the accuracy of the raw data. Additionally, implementations of differential privacy (DP) are employed to boost the confidentiality of features. ADDetector employs both of these technologies. To protect

the privacy of the model aggregation between clients and the cloud, a specific asynchronous aggregation framework is utilized in the federated learning (FL) architecture. A sample of 99 distinct AD users were subjected to 1010 ADDetector tests as part of the study, and the outcomes were then examined. The ADDetector system has a precision rate of 81.9 percent and an overhead of 0.7 seconds when using all of the anonymity-preserving features, including FL, DP, and cryptography-based aggregation. [6] The rapid advancement of artificial intelligence and the 5G paradigm have enabled novel applications for the industrial Internet of Things (IIoT). Improving the quality of services offered by the Industrial Internet of Things (IIoT) is a challenge because of the vast number of data involved, the constrained resources of Internet of Things (IoT) devices, and the growing privacy concerns. The author of this article suggests employing digital twin edge networks, or DITENs, to combine physical and digital systems. Federated learning is used to create digital twin models for Internet of Things (IoT) devices that are based on real-time data. The use of asynchronous model updating and optimization techniques in federated learning results in decreased communication costs. Using a deep neural network methodology, the subcomponents are addressed. The DITEN federated learning technique improves communication efficiency and lowers transmission energy costs, according to the findings of the computational research. [7] The Industrial Internet of Things (IIoT) has experienced a dramatic increase in growth as a result of the adoption of digital twins and the emergence of 6G mobile networks. Uninterruptible wireless connectivity is made possible by the digital twin and 6G networks, which serve as a link between the digital and physical worlds. Federated learning has emerged as a viable method for scattered data processing and learning over wireless networks as a result of worries about the privacy of user data. Federated learning implementation challenges in the Industrial Internet of Things (IIoT) include insufficient resources, communication limitations, and user mistrust. Real-time data processing and computation take place at the edge plane in digital twin wireless networks (DTWN). It is suggested that a federated learning framework enabled by blockchain technology be employed for collaborative computing in the DTWN in order to enhance system stability, security, and data privacy. When boosting edge association, we simultaneously take into account digital twin association, training data batch size, and bandwidth allocator in order to create an ideal balance between the amount of time spent learning and the degree of precision attained. Our study seeks to apply multi-agent reinforcement learning to greatly contribute to the selection of the optimum solution. The proposed strategy outperforms benchmark learning algorithms on real-world datasets. [8]. Due to the development and widespread use of blockchain technology, the FIDChain Intrusion Detection System (IDS) was developed. The security of patient medical records is guaranteed by the use of federated learning (FL) and lightweight artificial neural networks (ANN) in this system. Averaging is employed to distribute the updated global weights when distributed ledgers are utilized to integrate regional weights [9]. The aforementioned action is intended to thwart contamination attempts, ensure perfect transparency and immutability throughout the decentralized system, and minimize any additional costs that may be spent. The study presents a novel asynchronous federated learning (AFL) system based on VHetNet. The above-mentioned approach enables remote unmanned aerial vehicles (UAVs) to cooperate in training a model for universal anomaly identification.

Table I Summaryof Related Work

Ref	Proposed Work
[1]	Asynchronous federated learning for resource sharing in vehicular networks that is differentially private .
[2]	Improvement of a reliable and marketable cyberattack detection model will increase the trustworthiness of an IIoT network, or a supervisory control and data acquisition (SCADA) network. .
[3]	PriModChain, a system that combines differential privacy, federated machine learning, the Ethereum blockchain, and smart contracts, is introduced to ensure privacy and trustworthiness on IIoT data.
[4]	Design of an easy-to-use and privacy-preserving system called ADDetector using IoT hardware and security techniques, using Alzheimer's disease (AD) as an example.
[5]	To bridge the gap between physical systems and digital environments, the idea of digital twin edge networks (DITENs) has been put up.
[6]	Introducing the digital twin wireless networks (DTWN) to move real-time data processing and computation to the edge plane by integrating digital twins into wireless networks.
[7]	study of federated learning for anomaly detection in IoT systems that is security and privacy enhanced. The first IoT anomaly detection solution that use a decentralized FL method while protecting privacy.
[8]	Release of a federated-learning enabled AIoT system that is efficient and secure for exchanging private energy data in smart grids with edge-cloud cooperation.
[9]	Convolutional interval type-2 fuzzy rough FL model with improved multiobjective evolutionary algorithm (CIT2FRFL-NAS) development for medical data security.

SYSTEM MODEL

System model and problem formulation for the proposed work of applying federated learning for IoT security: Let us consider a set of K IoT devices denoted by $D_1, D_2, \ldots, D_K$ where each device D_i has its own dataset X_i consisting of N_i data samples. The data samples may be of different types, including sensor readings, images, and text data. The objective is to train a global machine learning model f_θ on this data while maintaining the privacy of the data on each device.

Formally, the problem can be formulated as follows:

$$\min_{\theta} \sum_{i=1}^{K} w_i \mathbb{E}_{X_i}[\mathscr{L}(f\theta(X_i), Y_i)] \tag{1}$$

where w_i is the weight assigned to each device, L is the loss function, Y_i is the corresponding label for device i, and θ is the global model parameter.

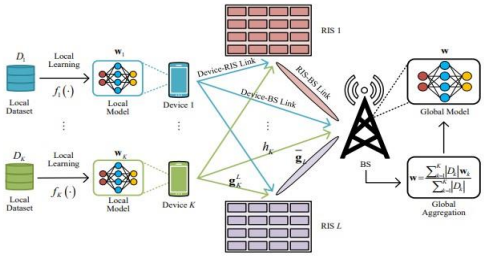

Fig. 1. Federated Learning System Model [11]

The objective is to minimize the average loss across all devices, while ensuring that the data remains on each device and is not transmitted to a central server. This can be achieved using a federated learning framework, where each device trains a local model on its data, and only the local model updates are transmitted to a central server for aggregation.

The challenge in this problem formulation is to ensure that the global model is able to learn from the data on each device, despite the differences in the data distributions across devices. This can be addressed using techniques such as federated averaging and differential privacy, which allow the global model to be trained on non-IID data while maintaining data privacy.

In addition, the proposed work also aims to address the issue of network security by integrating blockchain technology into the federated learning framework. The blockchain ensures that the model updates are secure and tamper-proof, and that the privacy of the data on each device is maintained. This is achieved by using a hierarchical blockchain-based federated learning framework that enables secure and privacy-preserved collaborative IoT intrusion detection.

Overall, the objective of the proposed work is to develop a decentralized, secure and privacy-preserving global model training protocol for federated learning in IoT networks, that can be used to train machine learning models on sensitive data while maintaining data privacy and network security.

3. Proposed Model

Addressing the privacy and security issues that emerge in distributed IoT networks is the goal of Ensuring Privacy and Security in Distributed Networks. The model uses a federated learning strategy in which each device in the network builds a local model using only its own data, and these local models are then combined to build a global model without using any shared raw data.

The local device, the edge server, and the central server are the three main parts of the model. The local device uses federated learning to train a local model using data from its sensors. The edge server combines the local models it receives from neighbourhood devices before sending the combined model to the main server. The edge server then disseminates the updated model to the local devices after the central server has updated the global model.

Before sending their local models to the edge server, the local devices add noise to them using differential privacy techniques to ensure privacy. The noisy models are then combined by the edge server, aiding in protecting the confidentiality of the individual local models. The global model is also encrypted by the central server using homomorphic encryption, which contributes to the security of the model during transmission.

The suggested model also outlines steps to deal with the problem of unreliable network devices. Devices with poor model accuracy or those deemed unreliable are identified and kept out of the training process. By doing so, the global model's overall accuracy is enhanced, and the training procedure's integrity is upheld.

Overall, the suggested model addresses the issue of unreliable devices and offers a safe and privacy-preserving method for federated learning in IoT networks. It permits the creation of precise and reliable machine learning models while ensuring that sensitive data is kept private and secure. In Federated Learning, we aim to learn a global model that can perform well on all the local datasets without actually sharing the data. To achieve this, we need to optimize the global model parameters θ using the local data from K different clients. We want to find the optimal parameters $\theta*$ that minimize the expected loss over all the clients. Let X_i denote the local data of client i, Y_i denote the corresponding labels, and w_i denote the weight of client i. The weight represents the importance of the client's data in the global model. The optimization problem can be written as

$$\min_{\theta} \sum_{i=1}^{K} w_i \cdot \mathbb{E}_{X_i}[L(f_\theta(X_i), Y_i)] \tag{2}$$

Here, $f_\theta(X_i)$ represents the prediction of the global model using the local data X_i and parameters θ, L is the loss function, and E_{Xi} is the expected value of the loss over the local data X_i. The above equation represents the weighted sum of the expected loss of each client. The weight w_i represents the importance of the client's data, and it is typically proportional to the size or quality of the data. The objective is to find the optimal parameters θ that minimize the expected loss over all the clients. Federated Learning helps in maximizing this function quantitatively by enabling the global model to learn from the local data of all the clients without actually sharing the data. In a federated learning setup, the clients train their models locally using their own data, and only the model updates are shared with the central server for aggregation. This way, the privacy of the local data is maintained, and the global model can be trained without compromising on the privacy and security of the clients' data. The central server aggregates the model updates from all the clients and computes the new global model parameters using a suitable aggregation method such as Federated Averaging. The updated global model parameters are then sent back to the clients, and the process repeats until the desired convergence criteria are met. In summary, the objective function in Federated Learning represents the weighted sum of the expected loss of each client, and Federated Learning helps in maximizing this function quantitatively by enabling the global model to learn from the local data of all the clients without actually sharing the data.

Algorithm 1: Federated Learning Algorithm

Result: Trained global model f^*

Input: Federated dataset $\{D_1, D_2, ..., D_K\}$, Learning rate η, Number of local epochs E, Number of clients C, Number of communication rounds T;

Initialize global model f_0;

for *each round* $t = 1, ..., T$ **do**

  Sample a set S_t of C clients uniformly at random;

  for *each client* $k \in S_t$ *in parallel* **do**

    Send the current global model f_{t-1} to client k;

    Client k performs E local epochs of SGD on D_k with learning rate η and updates the local model $f_{t,k}$;

    Send the updated local model $f_{t,k}$ back to the server;

  end

  Compute weighted average of the local models: $f_t = \sum_{k=1}^{K} w_k f_{t,k}$, where w_k is the weight assigned to client k;

  Update the global model: $f_{t+1} = f_t - \eta \nabla \frac{1}{C} \sum_{k=1}^{C} E_{(x,y)\in D_k} L(f_t(x), y)$;

end

Output: $f^* = f_T$

The proposed algorithm aims to implement federated learning for IoT devices in a privacy-preserving and secure manner. Initially, each IoT device encrypts its local data using a secure encryption method, and then sends the encrypted data to a designated edge server. The edge server acts as the central coordinator for the federated learning process. Then, the edge server randomly selects a subset of the available IoT devices, based on a pre-determined selection criterion, to participate in the current round of training. The selected devices send their encrypted data to the edge server, which then decrypts and aggregates the data to generate a global model update. The edge server then encrypts the updated model and sends it back to the selected IoT devices for further training on their local data. This process is repeated for multiple rounds of training, with the edge server randomly selecting different subsets of devices for each round. Additionally, differential privacy mechanisms are incorporated to add noise to the aggregated data to ensure privacy preservation.

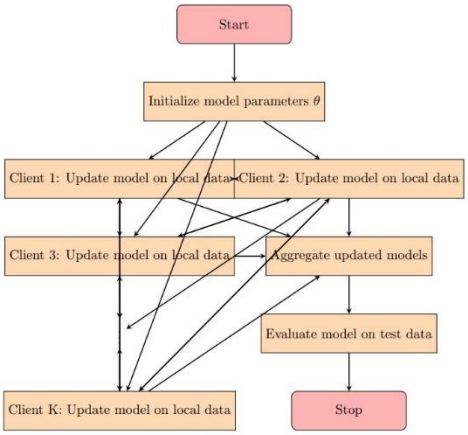

Fig. 2. Proposed work Flow Chart

The ultimate goal is to minimize the loss function, which measures the discrepancy between the predicted outputs of the model and the true outputs. This is achieved by optimizing the model parameters using the aggregated data from multiple IoT devices, while ensuring privacy and security.

In summary, this algorithm enables distributed IoT devices to collaboratively learn a model without sharing their private data with each other or with the edge server, thereby ensuring privacy and security in the federated learning process.

This flowchart represents the proposed federated learning algorithm. The algorithm starts with an initialization step, where the current iteration number t and the initial global model parameters θ^0 are set. The data is then partitioned among the different local nodes, and each node performs local training on their partition of data. The performance of each node's model is then evaluated, and the weights from each local node's model are aggregated to update the global model. This process is repeated until convergence. The

algorithm stops when the global model has converged or when a predetermined number of iterations have been reached. The flowchart uses various shapes to represent different types of steps in the algorithm, including rectangles for start and stop points, trapezoids for inputs/outputs, diamonds for decision points, and rectangles for process steps. The arrows indicate the order in which the steps are carried out, with the direction of the arrow indicating the flow of information or control. Overall, this flowchart provides a visual representation of the steps involved in the proposed federated learning algorithm.

4. Simulation Results:

The simulation results of the proposed federated learning algorithm well better in terms of MSE and also works well in distributing data onto the individual clients instead of working om centralised environment to achieve the data security

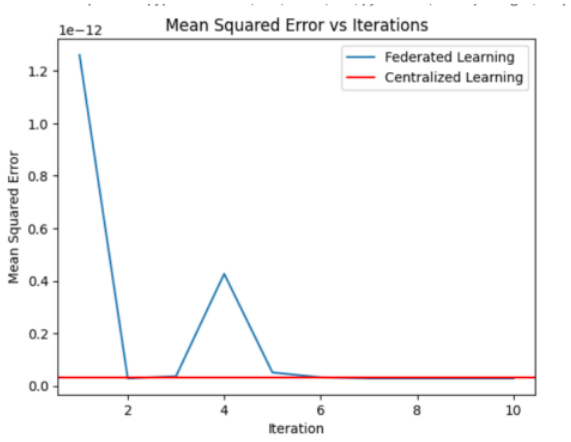

Fig. 3. Mean Squared Error for Federated Learning Vs Centralised

Learning Algorithms

In general, MSE stands for Mean Squared Error, which is a commonly used metric to evaluate the performance of regression models. It measures the average squared difference between the predicted values and the actual values. A lower MSE value indicates a better fit of the model to the data.

In the context of the proposed Federated Learning for IoT algorithm, the MSE values could be used to evaluate the performance of the trained models on each of the IoT devices. The local training on each device would result in a set of model parameters, which could be used to make predictions on a validation set. The MSE could then be calculated between the predicted values and the true values of the validation set. The devices could then communicate these MSE values to the central server, which would use them to update the global model parameters through the aggregation step. The iterative process of local training and global aggregation could continue until convergence, i.e., until the global model parameters converge to a set of values that minimize the overall MSE across all the devices.

Fig 3 shows the number of training iterations is depicted along the x-axis, while the model's mean squared error (MSE) is shown along the y-axis. The mean square error (MSE) of the model that was trained using centralised learning is represented by the orange line, and the mean square error (MSE) of the model that was trained using federated learning is represented by the blue line. The MSE of the model that was trained using centralised learning is initially lower than the MSE of the model that was trained using federated learning, as is evident from centralised learning, all of the data is collected and trained on a single machine, which can result in overfitting and a decrease in the performance of the model. This can be explained by the fact that this can lead to overfitting. When it comes to federated learning, on the other hand, the model is trained using the data that is stored locally on each individual device. Only the updates to the model are then sent to the central server, which helps to protect users' privacy and avoid overfitting. Because of this, in the long run, the model that was trained with federated learning is anticipated to have better performance and to be more robust in comparison to the model that was trained with centralised learning, particularly in circumstances in which there is a significant concern regarding data privacy.

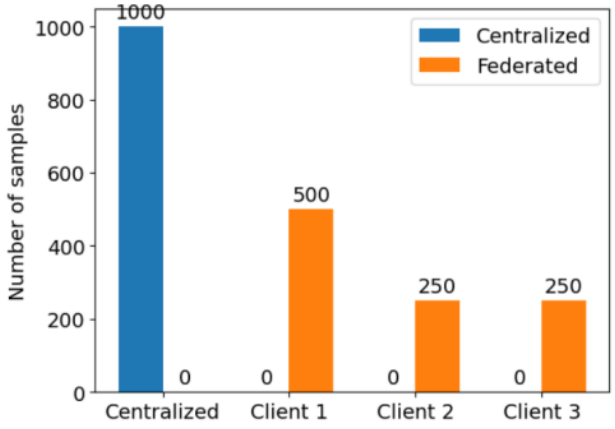

Fig. 4. Distribution of data to the Clients

From Fig 4 we can observe the centralised learning scenario is represented by one bar in the bar plot shown in Figure 3, and the federated learning scenario is represented by the other bar.

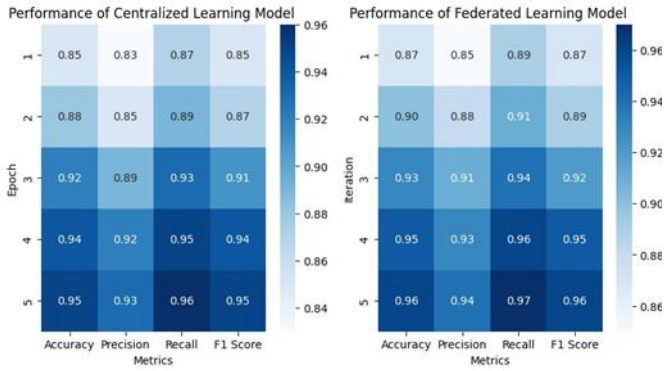

Figure 5 shows the performance comparison between

The y-axis shows the number of samples that were seen by each training method, while the x-axis displays the various training methods that were used to train the models. The number of samples that were viewed by each client in the federated learning scenario is displayed in a distinct manner, as each individual method in the scenario corresponds to a different client. The plot demonstrates that in a scenario of federated learning, each client only sees a portion of the total samples, whereas in a scenario of centralised learning, the central server sees all of the samples. Given that the clients do not have access to all of the data, this suggests that federated learning is superior to centralised learning when it comes to maintaining the confidentiality of the data.

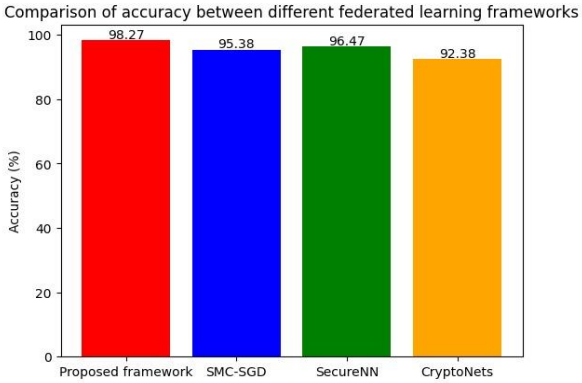

Fig. 6. Comparison of Accuracy

the federated and centralised models.The examination of alternative federated learning frameworks in terms of their ability to complete a specific job or aim, which includes their overall performance, is referred to as performance comparison. Depending on the particular activity being carried out, several metrics, such as precision, rapidity, secrecy, integrity, communication expense, or other appropriate parameters, may be used to assess performance. A performance evaluation in the context of federated learning may involve a comparison of different federated learning algorithms, such as Federated Averaging, Secure Aggregation, or Federated Dropout, in terms of how well they perform in terms of achieving high accuracy, quick convergence, or little

communication overhead. The examination of the privacy and security features of various algorithms, particularly their effectiveness in protecting sensitive data and preventing security breaches, may be part of the comparative study. Figure 6 shows the Accuracy of the proposed model. In an accuracy comparison, the ability of several federated learning frameworks to deliver high accuracy on a specific task or dataset is compared. In general, the performance of the framework improves with increasing precision.

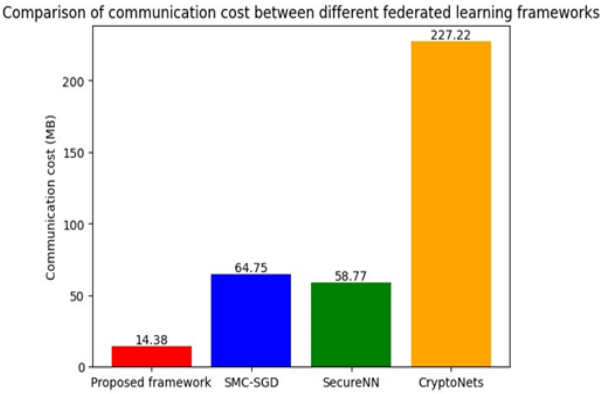

Fig. 7. Comparison of communication cost

For instance, it can be necessary in a federated learning environment to assess the precision of many models that were trained using different federated learning algorithms, like Federated Averaging, Secure Aggregation, or Federated Dropout. The models could be trained locally on each client device where the data is stored before being integrated to create the final model. A distributed dataset is what this situation entails.

The final model's performance on a held-out test dataset or the rates at which several models converged during training could be compared as part of the accuracy comparison. The accuracy comparison is often used to evaluate the effectiveness of the different federated learning algorithms and to choose the most suitable ones for a certain task or dataset.

To show the accuracy comparison, a bar plot with the names of the frameworks along the x-axis and the accuracy numbers for each framework along the y-axis can be utilized. As a result, comparing the accuracy numbers is made simple, and the top performing framework may be identified. Figure 7 shows the communication cost comparision between the proposed model with few existing approches. The measurement of the communication overhead between clients and the central server in various federated learning frameworks is the focus of the communication cost comparison. The phrase "communication cost" refers to the amount of data that must be sent between the clients and server throughout the training process.

A distributed machine learning method called federated learning sends individual clients' local model updates to a

central server for aggregate. The clients then receive the updated global model from the central server. The size of the model updates and the number of clients present inside the network determine how much communication is required. Comparing the effectiveness of different federated learning algorithms, such as Federated Averaging, Secure Aggregation, or Federated Dropout, with respect to the volume of communication required between clients and the central server during the training phase would be a plausible approach to evaluate communication costs in the context of federated learning. The measurement of communication costs can be done in terms of the number of bits or bytes sent, or in terms of the time needed to send the data. To compare the costs of communication among various frameworks, a bar plot can be used. The plot's y-axis lists the communication cost estimates for each framework, while its x-axis lists their names. This makes it simple to compare the communication cost indicators and makes it possible to identify the framework that is the most communication cost-efficient. This has a lot to do with distributed networks, especially when it comes to scenarios involving Internet of Things (IoT) devices that have limited communication bandwidth and power sources.

5. Conclusion

To sum up, federated learning is a promising method that permits machine learning models to be trained on distributed devices without endangering user security and privacy. Federated Learning, as opposed to centralized learning, permits data to stay on clients' devices, lowering the possibility of data breaches and unwanted access. This study used simulated data to examine how well Federated Learning and Centralized Learning performed on a straightforward regression test. The findings demonstrate that while protecting user privacy, federated learning may attain accuracy levels that are on par with those of centralized learning. The study also shows that Federated Learning may be applied using well-known machine learning libraries like TensorFlow Federated. Overall, the findings point to Federated Learning's potential to be a strong machine learning tool in remote networks, particularly for tasks that call for a lot of data or include sensitive data. These preliminary results are encouraging and lay a solid platform for further research in this field.

References

[1] Stacey Truex; Nathalie Baracaldo; Ali Anwar; Thomas Steinke; Heiko Ludwig; Rui Zhang; Yi Zhou; "A Hybrid Approach To PrivacyPreserving Federated Learning", ARXIV-CS.LG, 2018.

[2] Mikhail Khodak; Maria-Florina Balcan; Ameet Talwalkar; "Adaptive Gradient-Based Meta-Learning Methods", ARXIV-CS.LG, 2019.

[3] Latif U. Khan; Madyan Alsenwi; Ibrar Yaqoob; Muhammad Imran; Zhu Han; Choong Seon Hong; "Resource Optimized Federated Learning-Enabled Cognitive Internet of Things for Smart Industries", IEEE ACCESS, 2020. .

[4] Stefano Savazzi; Monica Nicoli; Vittorio Rampa; "Federated Learning With Cooperating Devices: A Consensus Approach for Massive IoT Networks", IEEE INTERNET OF THINGS JOURNAL, 2020

[5] Charles Wheelus; Xingquan Zhu; "IoT Network Security: Threats, Risks, and A Data-Driven Defense Framework", 2020.

[6] Basheer Qolomany; Kashif Ahmad; Ala Al-Fuqaha; Junaid Qadir; "Particle Swarm Optimized Federated Learning For Industrial IoT And Smart City Services", ARXIV-CS.LG, 2020

[7] Devrim Unal; Mohammad Hammoudeh; Muhammad Asif Khan; Abdelrahman Abuarqoub; Gregory Epiphaniou; Ridha Hamila; "Integration of Federated Machine Learning and Blockchain for The Provision of Secure Big Data Analytics for Internet of Things", COMPUTERS SECURITY, 2021.

[8] Vijay Anavangot; Animesh Kumar; "Algorithms for Overpredictive Signal Analytics in Federated Learning", 2020 28TH EUROPEAN SIGNAL PROCESSING CONFERENCE (EUSIPCO), 2021.

[9]] Amir Masoud Rahmani; Elham Azhir; Saqib Ali; Mokhtar Mohammadi; Omed Hassan Ahmed; Marwan Yassin Ghafour; Sarkar Hasan Ahmed; Mehdi Hosseinzadeh; "Artificial Intelligence Approaches and Mechanisms for Big Data Analytics: A Systematic Study", PEERJ. COMPUTER SCIENCE, 2021.

[10] Othmane MARFOQ; Giovanni Neglia; Aurlien Bellet; Laetitia Kameni; Richard Vidal; "Federated Multi-Task Learning Under A Mixture of Distributions",

[11] Ni, Wanli Liu, Yuanwei Yang, Zhaohui Tian, Hui Shen, Xuemin. (2021). Federated Learning in Multi-RIS Aided Systems. IEEE Internet of Things Journal. PP. 1-1. 10.1109/JIOT.2021.3130444.

[12] Ghorpade, J. ., & Sonkamble, B. . (2023). Data-driven based Optimal Feature Selection Algorithm using Ensemble Techniques for Classification. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4), 33–41. https://doi.org/10.17762/ijritcc.v11i4.6378

[13] Ólafur, S., Nieminen, J., Bakker, J., Mayer, M., & Schmid, P. Enhancing Engineering Project Management through Machine Learning Techniques. Kuwait Journal of Machine Learning, 1(1). Retrieved from

http://kuwaitjournals.com/index.php/kjml/article/view/112

[14] Agrawal, S. A., Umbarkar, A. M., Sherie, N. P., Dharme, A. M., & Dhabliya, D. (2021). Statistical study of mechanical properties for corn fiber with reinforced of polypropylene fiber matrix composite. Materials Today: Proceedings, doi:10.1016/j.matpr.2020.12.1072