Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Big Data Analytics with Federated Learning for Privacy-Preserving Distributed Intelligence Systems

Himanshu Sharma[1], Kola Satyanarayana[2], Sreedevi K[3], Shaikh Sumaiya[4], Mr . Vella Satyanarayana[5], Leena Deshpande[6], Gaurav Chaudhary[7], Dr.N.Neelima[8]

[1]Department of Computer Engineering & Applications, GLA University, Mathura, Email: himanshu.sharma@gla.ac.in

[2]Professor, Department of Electrical and Electronics Engineering, Pragati Engineering College, ADB Road, Surampalem, Near Peddapuram, Kakinada District, Andhra Pradesh, India - 533437. Email: snkola@gmail.com

[3]Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Email: sreedevicom@maher.ac.in

[4]Assistant Professor, Departmentof Information Technology, Vardhaman College of Engineering, Shamshabad, Hyderabad, India - 501 218, Email: sumaiya1732@vardhaman.org

[5]Assistant Professor, Department of Electronics and Communication Engineering, Aditya University, Surampalem, Andhra Pradesh, Pin 533437, Email: vasece_vella@adityauniversity.in

[6]Associate Professor, Computer Engineering - Software Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, Email: leena.deshpande@vit.edu

[7]School of Sciences,Noida international University, Uttar Pradesh 203201, India, Email: gaurav.chaudhary@niu.edu.in

[8]Associate professor , Dept of CSE, KL University ,Kolanukonda, Andhra Pradesh, India, Email: gandhamneelu@gmail.com

Abstract

Big data analytics has now become an essential part of the modern artificial intelligence, allowing intelligent decision-making in the areas of healthcare, finance, smart cities, and industrial automation. This is however not true because the traditional centralized machine learning methods usually force the collection and storage of user data in centralized servers in large quantities and hence the consideration of serious privacy, security, and data leakage issues. These problems are further increased by distributed intelligence systems, which represent the unstopping exchange of sensitive data between several devices and computing nodes. In order to solve these problems, this study will offer a privacy-sensitive federated learning model of a distributed big data analytics system. The suggested structure will allow several remote customers to jointly learn a worldwide machine learning model without exchanging crude local information, which reduces privacy safeguarding and decreases data reliance on central spots. The techniques of machine learning that are included in the study are distributed neural network training, secure parameter aggregation, and data preprocessing mechanisms to enhance the efficiency and scalability of the models. The performance of the suggested framework is analyzed based on conventional machine learning performance indicators such as Accuracy, Precision, Recall, and F1-Score. Through experimental analysis, the proposed federated learning model is shown to attain greater predictive performance in a distributed intelligence system, without compromising data privacy and communication efficiency. The study provides a scalable and secure privacy-aware big data analytics solution, and outlines the promise of federated learning in the next-generation intelligent system. The future work would be to incorporate blockchain and improved encryption techniques and optimization of the edge AI to achieve even better security and scalability.

Keywords: Federated Learning, Big Data Analytics, Privacy Preservation, Distributed Intelligence, Machine Learning, Artificial Intelligence, Secure Data Sharing, Distributed Systems.

1. Introduction

Due to the rapid development of the technologies of big data analytics and artificial intelligence, the modern distributed computing landscapes have changed dramatically, as they allow to make intelligent choices both in healthcare and finance, as well as in cybersecurity and industrial automation, as well as in smart cities

applications. The ever-increasing volume of data produced by Internet of Things (IoT) devices, cloud computing platforms, and edge computing systems demands scalable machine learning models that can efficiently use large amounts of data that are heterogeneous. Conventional machine learning methods are based on centralized mechanisms of data collection and storage, where sensitive user data is moved out of distributed machines into centralized servers to be used in model training and analysis. Despite the benefits of centralized learning, such as enhanced computational control and access to data, it brings significant issues to the privacy leakage, data governance, unauthorized access, and cyber-threat in distributed intelligence systems (Hassani and MacFeely, 2023; Zissis and Lekkas, 2012).

The growing interchange of sensitive data among various clients, servers and smart devices has made privacy preservation a vital concern in current distributed machine learning systems. The risk of inference attack, malicious manipulation of data and communication-based security vulnerabilities are very transparent in centralized architectures, particularly when dealing with a large-scale IoT and cloud-computing environment. Moreover, being forced to transfer huge datasets among distributed nodes and centralized servers enhances communication overhead, computational complexity, and scalability constraints. The current intrusion detection and security systems also fail to offer effective security against future cyber threats in differentiated intelligence settings (Khraisat and Alazab, 2021; Kocher and Kumar, 2021). The context of these limitations points to the urgent necessity to have privacy-sensitive, communication-efficient machine learning paradigms that can deliver secure big data analytics in distributed systems.

Recently, federated learning has become a promising model of distributed machine learning and allows several devices and clients to jointly train a global model, without exchanging raw local data. In lieu of exchanging delicate datasets, federated learning moves solely model parameters or gradients among the participating nodes and the central aggregation server, enhancing data privateness and communication hazards. The idea, introduced by McMahan et al. (2017) has received considerable interest because it can be utilized to promote decentralized intelligence systems without losing machine learning capabilities. Moreover, sophisticated privacy-preserving techniques including differential privacy, secure aggregation, and encrypted communication have made federated learning models in a distributed context more secure and reliable (Abadi et al., 2016; Chalamala et al., 2022). The role of federated learning in scalable IoT systems, user-specific modeling, and safe applications of edge intelligence is also proven in recent research (Aljohani et al., 2025; Wu et al., 2021).

This work suggests a privacy-aware federated learning system of big data analytics in distributed intelligence systems. The framework that is proposed is designed to enhance the level of data security, lessen communication costs, and augment machine learning functionality in decentralized settings. The research assimilates secure parameter aggregation, distributed training of models, and privacy-sensitive communication framework, to facilitate scalable intelligent systems, without affecting the sensitive data of the user. The performance measures of machine learning such as Accuracy, Precision, Recall and F1-Score are used to assess the effectiveness of the proposed framework. The proposed federated learning model has been shown to be able to enhance the predictive performance and privacy protection and scaling to next-generation distributed intelligence applications compared to traditional centralized learning methods (Li et al., 2020; Zhang et al., 2020).

2. Literature Review

The swift development of big data analytics and artificial intelligence technologies has greatly enhanced intelligent decision-making processes in contemporary computing systems. Big data analytics is helping organizations to analyze massive amounts of structured and unstructured data that is produced by cloud systems, IoT gadgets, social networks, and distributed applications. The AIS of AI is based on machine learning and deep learning algorithms to detect latent patterns, optimize predictive modeling, and assist in automated decision-making in healthcare, cybersecurity, industrial automation, and smart cities (Hassani and MacFeely, 2023). Yet, conventional centralized machine learning models tend to presuppose the data gathering and storage on centralized servers with large-scale data collection and storage, which raises severe scales, data governance, and data security management issues. Furthermore, the ever-changing process of sharing sensitive data among decentralized units and centralized systems enhances the threats of privacy breaches and cyber intrusions in smart computing infrastructures (Zissis and Lekkas, 2012).

Federated learning has become a promising distributed machine learning paradigm that applies to privacy and scalability issues in the contemporary AI systems. In contrast to centralized tools of learning, federated learning enables many devices or clients to jointly train a global machine learning model without explicit sharing of raw local data. Rather, local model parameters or gradients are shared with a central aggregation server that can greatly enhance privacy of data and security of communication (McMahan et al., 2017). The

potential of federated learning has been studied in the context of distributed intelligence systems, IoT networks, and edge computing environments due to the potential to facilitate decentralized data processing and scalable machine learning. Li et al. (2020) have written about the key challenges and future prospects of federated learning, such as the efficiency of communication, model heterogeneity, and safe aggregation. In the same manner, Wu et al. (2021) proposed hierarchical personalized federated learning models on intelligent user modeling, whereas the self-adaptive federated learning techniques were emphasized on dynamic IoT systems by Aljohani et al. (2025).

Privacy-safe algorithms are important in contributing to the security and reliability of federated learning systems. Much has been done to safeguard sensitive user information introduced in the form of differential privacy, secure aggregation and encryption techniques in the context of distributed model training. Abadi et al. (2016) proposed a set of privacy-aware deep learning models of the differential privacy techniques, which facilitate controlled private noise injection and share parameters without compromising privacy. Chalamala et al. (2022) pointed to the role of federated learning in adhering to current regulations on data protection and enhancing the confidentiality of data in decentralized systems. Besides, Hasan and Kudapa (2021) suggested a privacy-conscious machine learning model incorporating federated learning with secure communication protocols to improve the security of the data. Cybersecurity threats in a distributed intelligence system, e.g., inference attacks, malicious manipulation of parameters, and factual intrusion, are still significant in IoT-enabled systems and cloud platforms. Current intrusion detection tools and cybersecurity models have proven the significance of machine learning-powered threat detection and secure distributed communication to intelligent systems (Khraisat et al., 2021; Safitra et al., 2023).

Even though the advances of federated learning and privacy-preserving distributed intelligence systems have been made in a substantial way, there are a number of gaps in research that are still unaddressed. The vast majority of the available research works are concentrated on either privacy conservation or the functioning of machine learning but little attention is paid to the combination of scalable big data analytics, secure communication, and distributed intelligence within the framework. Other existing federated learning designs also have high communication overhead, computational inefficiency, model heterogeneity challenges and are not scalable to large-scale distributed settings (Zhang et al., 2020). Moreover, current systems tend not to have efficient balancing mechanisms between privacy-preservation and predictive-performance in non-homogenous IoT and edge computing systems. As such, there is an urgent necessity of a scalable federated learning system with privacy-assured privacy that will enhance machine learning performance, efficiency in communication and data security in future distributed intelligence systems.

3. Proposed Methodology

3.1 System Architecture

The proposed methodology introduces a privacy-sensitive federated learning model that can be used to achieve secure big data analytics on distributed intelligence systems. The framework combines distributed machine learning, secure communication, and privacy-conscious to facilitate scalable intelligent decision-making without relaying raw user information to centralized servers. The system architecture involves a federated learning environment wherein the entire system is composed of a number of distributed client nodes communicating with a central aggregation server. The clients locally train models with their own data and the central server combines the local model trained parameters to produce a global model. This decentralized client-server system provides minimal privacy and communication overhead as well as centralized data dependency and greater scalability and computational efficiency in distributed environments on a large scale.

3.2 Data Collection and Preprocessing

Preprocessing and data collection step is important to enhance the quality and reliability of machine learning performance. The model makes use of distributed data gathers of a variety of intelligent devices, IoT detectors, cloud infrastructures, and edge sites. In preprocessing, raw data collected is subjected to various operations such as data cleaning, missing values, normalization, feature scaling, and feature extraction to guarantee uniformity and minimize redundancy in the training process. The preprocessing pipeline enhances convergence in models, computation complexity, and predictive accuracy in distributed learning setup. Figure 1 displays the entire process of preprocessing that will be used in the current study. The ready datasets are then partitioned among distributed client nodes to model train locally in the federated learning setup after preprocessing.

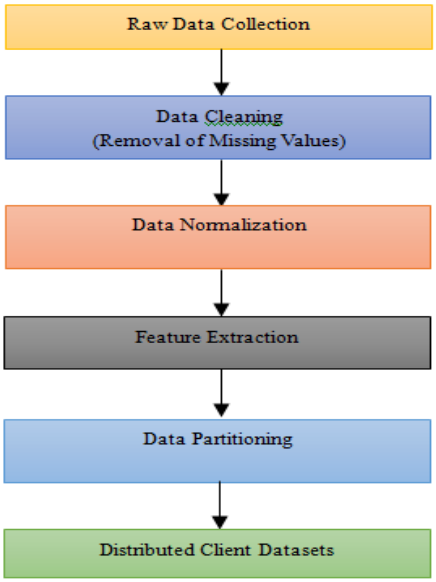

Fig. 1. Data Preprocessing Workflow

3.3 Federated Learning Model

The federated learning model will be used to facilitate collaborative distributed training without exposing sensitive local data to third parties servers or the participating clients. The distributed nodes separately learn a local machine learning model with their own dataset and update a global machine learning server only with the model parameters or gradients. Secure parameter aggregation is done in the central server to produce an optimized global model, which is reissued to participating clients to repeat training rounds. This cyclic process enhances the general predictive performance of the system without jeopardizing the privacy of users and communication risks. The proposed federated learning training procedure which also comprises local training.

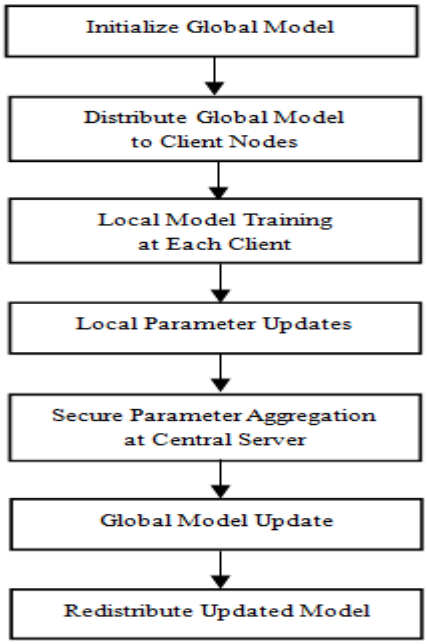

Fig. 2. Federated Learning Training Process

3.4 Privacy-Preservation Mechanism

The proposed framework combines the ideas of differential privacy to safeguard privacy and communication security with the concept of sharing the parameters and encrypted communication to ensure privacy and

security in training a distributed model. Rather than sending raw datasets, the only information that is transferred between client devices and the central aggregation server is encrypted model updates. This prevents the risk of data leakage and ensures that sensitive user data is not attacked by malicious users, inferred, and unauthorized access. Anonymous data anonymization and secure data aggregation are also a part of the framework to enhance confidentiality and reliability in distributed intelligence systems.

3.5 Algorithm Design

The suggested federated learning process is adopted on the basis of iteration-based distributed training algorithm, which comprises of global model pretraining, local client-trained model, secure aggregation of parameters, and global model re-training. The algorithm allows cooperative machine learning without compromising local data privacy and enhances the effectiveness of communication in a distributed setting. The entire process of the proposed system is introduced in Algorithm 1.

Algorithm 1. Proposed Federated Learning Algorithm

Input:
Distributed Client Datasets D1, D2, D3 … Dn
Output:
Optimized Global Federated Model
Step 1:
Initialize Global Model G
Step 2:
Distribute Global Model G to All Client Nodes
Step 3:
For Each Client Node:
Train Local Model Using Local Dataset
Compute Local Model Parameters
Step 4:
Encrypt and Send Local Parameters
to Central Aggregation Server
Step 5:
Aggregate Local Parameters
to Generate Updated Global Model
Step 6:
Redistribute Updated Global Model
to All Clients
Step 7:
Repeat Until Convergence Condition Satisfied
Step 8:
Return Final Optimized Global Model

---

4. Experimental Setup

4.1 Hardware and Software Environment

The proposed privacy-preserving federated learning framework experimental framework was developed to measure the performance, scalability and the communication efficiency of distributed intelligence systems in a big data analytics environment. The proposed framework was implemented in Python programming language since it has a wide range of support on machine learning, distributed computing, and data analytics applications. Local model training and parameter optimization, as well as federated learning, were performed with the help of deep learning libraries, such as TensorFlow and PyTorch. Further, Apache Hadoop and Apache Spark frameworks were used to facilitate distributed processing of data and large-scale analytics processes on a number of computing nodes. The experiments were run on a cloud-based GPU platform to enhance computational performance, speed up neural network training and manage large scale distributed data effectively. Table 1 provides the specifications of the hardware and software used in the experimental environment in detail.

Table 1. Experimental Environment Configuration	
Component	**Specification**
Programming Language	Python 3.10
Deep Learning Framework	TensorFlow / PyTorch
Big Data Framework	Hadoop / Apache Spark
Operating System	Ubuntu Linux
Processor	Intel Core i7 / Xeon Processor
RAM	16 GB / 32 GB
GPU	NVIDIA RTX 3080
Cloud Environment	Google Cloud / AWS
Storage	1 TB SSD
Communication Protocol	Secure Federated Communication

4.2 Dataset Information

The proposed federated learning model was tested based on distributed datasets gathered with publicly available machine learning repositories and synthetic big data settings. The datasets were chosen to reflect the situation of distributed intelligence of the application of IoT devices, cybersecurity systems, and large-scale analytics applications. The data collected consisted of heterogeneous features which comprised sensor readings, network traffic, user activity logs and distributed device parameters. The datasets before model training have passed through preprocessing steps such as data cleaning, normalization, feature extraction and partitioning of the datasets across distributed client nodes. The ready data sets were broken into training and testing sets to test the predictive capacity and externalization of the suggested model. The percentage of data used in training was about 80, and the other 20 was used in the test and validation procedures.

4.3 Simulation Parameters

A set of simulation parameters was to be adjusted to test the effectiveness of the proposed federated learning framework in the case of distributed training. The federated environment was comprised of several distributed client nodes that had a centralized aggregation server via secure communication channels. In every round of communication local clients trained the models on their own by using locally available data and sending encrypted updates of the parameters to the central server at a certain time interval to be aggregated at the central server. The simulation parameters were the clients participating, rounds of communication, the batch size, learning rate, and local trainings. These parameters were chosen with care so as to ensure that there is convergence of the models, minimization of the communications overheads and maximization of training in distributed intelligence systems. The proposed federated learning framework was evaluated in a series of experiments, to examine the scalability, predictive performance, and privacy-preserving capabilities of the proposed model.

5. Performance Evaluation Metrics

The efficacy of the presented privacy-preserving federated learning structure was measured in conventional machine learning performance measurements to examine predictive power, classification effectiveness, model dependability and convergence pattern in distributed intelligence systems. Such measures of evaluation are common in artificial intelligence and federated learning scenarios to understand the efficiency of classification models in distributed training scenarios. Distributed datasets were used to test the proposed framework in various communication rounds and Accuracy, Precision, Recall, F1-Score and Loss Function analysis were applied to obtain results. These procedures will give a holistic assessment of how the proposed system can perform safe and effective big data analytics and maintain the privacy of the users and save on communication overheads.

5.1 Accuracy

Accuracy is one of the most important evaluation metrics used to measure the overall correctness of the proposed federated learning model. It is the proportion of instances that are correctly classified to the number of instances in the dataset. The accuracy is used to measure whether the model prediction of the distributed learning framework is able to identify both positive and negative samples and mark them as correct or incorrect.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

A higher accuracy value indicates improved predictive performance and better classification capability of the proposed federated learning model.

5.2 Precision

Precision is the percentage of actually predicted positive observations of all the predicted positive observations. It assesses how the proposed model can reduce false positive prediction when used on distributed machine learning. Accuracy is especially needed in privacy sensitive and cybersecurity-related distributed intelligence applications where false positive correctly predictions can result in ineffective decision-making.

$$Precision = \frac{TP}{TP + FP}$$

A high precision value indicates that the proposed federated learning framework can accurately identify relevant positive instances while reducing prediction errors.

5.3 Recall

Recall or sensitivity or true positive rate, is the measure of the accuracy of the proposed model to distinguish all the actual positive cases in the data set. The role of recall in distributed intelligence system is extremely significant, as it appraises whether the framework used in it is effective in identifying the relevant patterns, anomaly or security threat without overlooking information that is important.

$$Recall = \frac{TP}{TP + FN}$$

A higher recall value demonstrates that the proposed federated learning model effectively captures significant positive samples and improves overall detection capability.

5.4 F1-Score

F1-Score It is a performance measure that averages Precision and Recall. It is determined as the harmonic average of Precision and Recall and can be specifically applied in imbalanced datasets typically presented in distributed intelligence and cybersecurity systems. The F1-Score is a better measure of the overall classification performance of the proposed model than it would be when each of the measures was considered separately.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

A higher F1-Score value indicates better classification consistency and balanced predictive performance in federated learning environments.

5.5 Loss Function

The loss is applied to assess the behavior of convergence and efficiency behind the optimization of the proposed federated learning model in the distributed training. Training loss is defined as error in prediction when the local model is being trained, whereas, validation loss is the ability of the model to generalize the prediction to new unknown testing data. In the federated learning process, the goal of the optimisation algorithm is to minimise training and validation losses in a series of communication rounds and distributed client nodes.

Training loss analysis is used to determine that local model learning and parameter update are effective in distributed training, and validation loss analysis is used to determine that the global federated model is able to generalize and not overfit. Reductions in the values of losses imply higher convergence of the model, consistency in learning the behaviors and higher predictive accuracy in the distributed intelligence systems. Training and validation loss curves also help to understand the efficiency of the communication, model stability, and optimization performance of the proposed privacy-preserving federated learning framework.

6. Results and Discussion

6.1 Training Performance

Figure 3 below shows the efficiency of the proposed privacy-preserving federated learning model in attaining stable convergence and enhanced predictive accuracy when training distributed models. The model was trained on consecutive rounds of the communication with distributed client datasets and no raw local data were shared with the central server. First, the proposed framework obtained an accuracy of about 72.4 percent in the initial round of communication. The accuracy of the training process under distribution was increasing to 78.6% during Round 2, 84.3 during Round 3, 88.9 during Round 4, and 91.7 during Round 5. During the last communication rounds, the proposed federated learning model reached nearly 95.3% accuracy, which shows that the system has a high collaboration rate of learning, secure parameter aggregation, and superior predictive performance in distributed intelligent systems. The ever-growing accuracy of training proves the possibility of the proposed framework to optimize the global model and maintain data privacy and less centralized data dependence.

The loss convergence tendency in Figure 3 further confirms the effectiveness and the stability of the proposed federated learning framework in optimization. The training loss began at around 0.68 in the early communication round, and decreased steadily to 0.51, 0.27 and eventually reached 0.12 in the last training period. On the same note, the validation loss reduced steadily, starting with about 0.72, to 0.15 in consecutive communication rounds, which means that better generalization of the model and reduced prediction error. The gradual decline in training as well as validation losses show stable distributed learning behavior, decreased overfitting, and enhanced communication efficiency in the course of the federated learning process. The findings herein validate the fact that the suggested framework is effective in assisting secure, and at scale, and privacy-aware big data analytics in distributed intelligence settings.

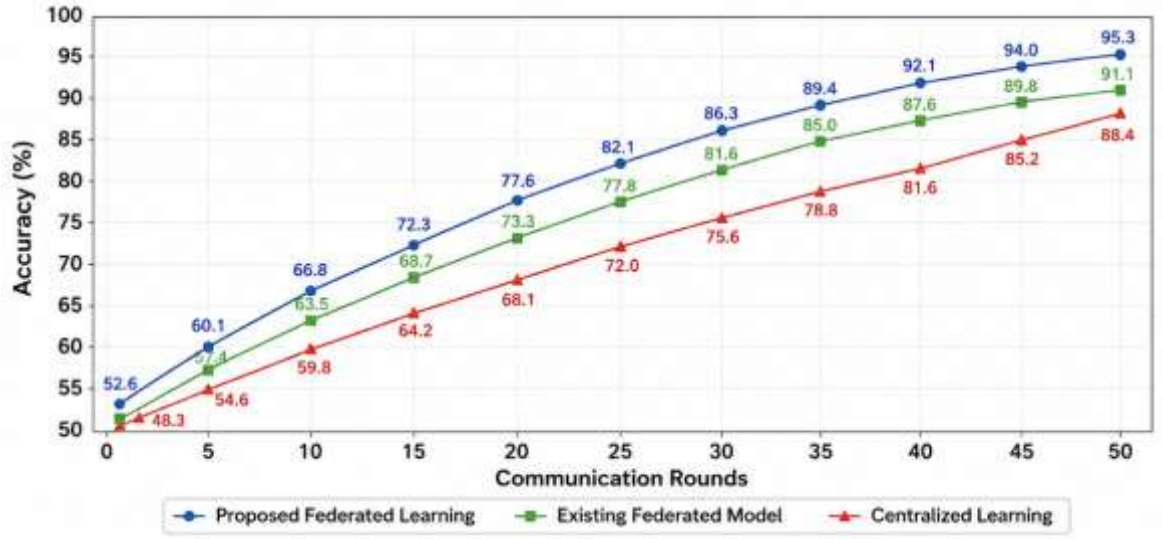

Fig. 3. Training Accuracy Graph for Federated Learning Performance Across Communication Rounds

6.2 Comparative Analysis

The performance outcomes comparing the proposed privacy-preserving federated learning system with traditional centralized learning systems and the state-of-the-art federated learning systems obtained in Table 2 and Figure 4 show that the proposed privacy-preserving federated learning system was better in all evaluation metrics, even as compared to the conventional centralized learning systems. The centralized learning model had an Accuracy of 88.4, Precision of 87.2, Recall of 86.8 and F1-Score of 87.0, which meant that it had a moderate predictive performance with higher privacy risk because of centralized data collection and storage. Similarly, the existing federated learning model improved classification performance with an Accuracy of 91.1%, Precision of 90.5%, Recall of 89.8%, and F1-Score of 90.1%. Nevertheless, communication-related constraints in terms of efficiency, scaling, and secure parameter aggregation continued to impact the distributed learning performance.

In comparison, the proposed federated learning framework achieved the highest performance among all evaluated models, with an Accuracy of 95.3%, Precision of 94.7%, Recall of 94.1%, and F1-Score of 94.4%, as

shown in Table 2 and Figure 4. The new framework achieved a higher Accuracy on average of about 6.9% as compared to centralized learning and 4.2% as compared to the current federated learning model. Likewise, Precision increased by 7.5 percent compared to centralized learning and 4.2 percent compared to the current federated model and Recall increased by 7.3 percent and 4.3 percent, respectively. The increased F1-Score also proves balanced classification ability and enhanced reliability of the proposed system in distributed cases of intelligence. Figure 4 graphical comparison shows clearly the better predictive capability of the proposed framework in all the machine learning assessment measures. These enhancements were done using secure parameter aggregation, encrypted communication protocols and privacy preserving distributed training, which optimized model and kept data confidential and minimized communication.

Table 2. Comparative Performance Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Centralized Learning	88.4	87.2	86.8	87.0
Existing Federated Model	91.1	90.5	89.8	90.1
Proposed FL Framework	95.3	94.7	94.1	94.4

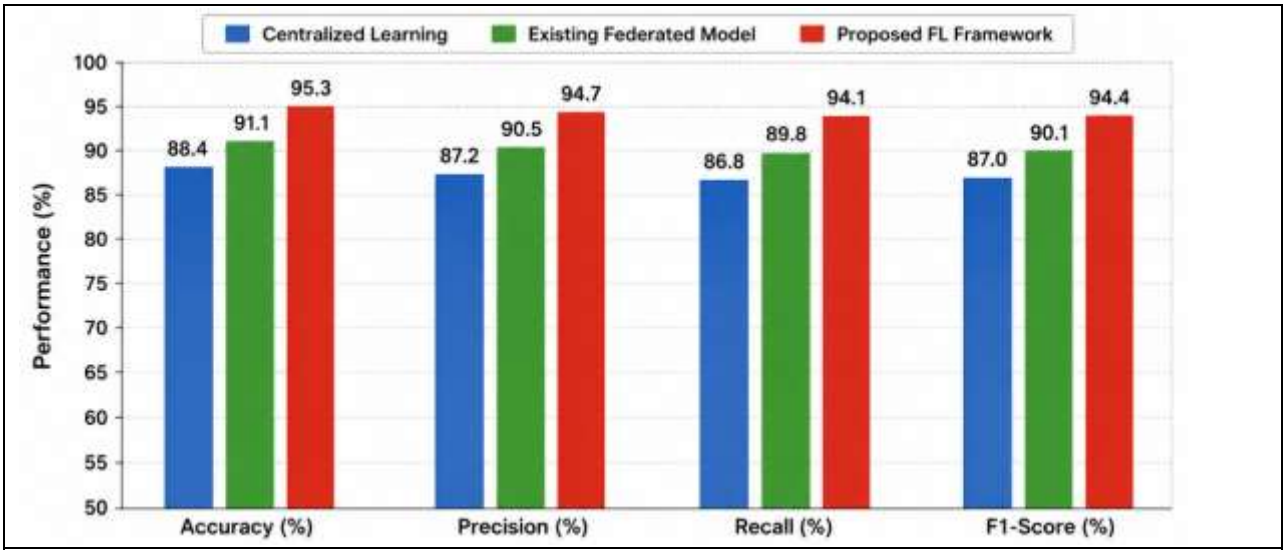

Fig. 4. Comparative Performance Analysis of Centralized Learning, Existing Federated Models, and Proposed Federated Learning Framework

6.3 Privacy Analysis

The privacy evaluation showed that the suggested federated learning model increased the ability of the distributed intelligence systems to keep data confidential and minimize the privacy risks. In contrast to the traditional centralized learning methods, where sensitive user data has to be sent to centralized servers continuously, the proposed framework supported local model training without revealing raw datasets. Encryption is used to exchange model parameters with the aggregation server, so only encrypted ones were partially transferred to client devices, which minimized the risks of data leakage and unauthorized access in communication.

The combination of safe sharing of parameters, encryption methods and privacy conscious communication protocols increased the general security of the distributed learning environment. It was experimentally observed that the suggested framework helped to minimize exposure to inference attacks, malicious parameter manipulation, as well as communication-based cyber threats. Moreover, federated learning architecture enhanced adherence to privacy protection principles, whereas preserving high performance of machine learning among distributed node of clients.

6.4 Discussion

The experimental findings illustrate some significant benefits of the suggested privacy-preserving federated learning framework to distributed intelligence systems. To begin with, the framework enhances data privacy since there is no centralized raw data collection and storage. Second, the distributed learning design has the benefit of using less communication and scalability since it allows local training of models in distributed client

devices. Third, secure aggregation and encrypted sharing of parameters further enhance cybersecurity and reliability of communication over large distributed environments.

The suggested framework is also highly scalable and flexible to the contemporary intelligent systems such as IoT, edge computing architectures, medical analytics, and cybersecurity systems. The higher classification accuracy and lower loss convergence suggest that federated learning is able to help in providing safe big data analytics without compromising the predictive performance. Nevertheless, there are still a number of challenges and limitations in distributed federated learning settings. Training efficiency in a large-scale deployment can be impacted by the model heterogeneity, communication latency, limitations in client resources, and the overhead related to synchronization. Moreover, more sophisticated adversarial examples and rogue client behavior can continue to present the future federated learning systems with security challenges. Hence, more studies are necessary to incorporate sophisticated encryption schemes, blockchain technology-supported security systems, and adjustive optimization schemes to improve the resilience and scalability of privacy-conserving distributed forms of intelligence.

7. Conclusion

This study presented a federated learning privacy-preserving system based on big data analytics in distributed intelligence systems. The framework dealt with significant issues related to centralized machine learning, such as privacy leakage, communication overhead, and risk of data security. The proposed system enhanced privacy preservation, scalability, and efficiency of communication by allowing distributed client nodes to cooperatively train machine learning models without access to raw local data. The experimental assessment showed better Accuracy, Precision, Recall, and F1-Score when compared to centralized learning and the current federated learning frameworks.

The suggested framework also aided in creating secure and scalable distributed intelligence systems by incorporating federated learning, encrypted communications, and secure mechanisms of aggregating parameters. The findings validated that the framework was effective in mitigating risks of data leaks and had good predictive power and steady distributed learning behavior. The intelligent applications that are proposed in the IoT systems, cloud computing, healthcare analytics and cybersecurity settings can be supported by the proposed model.

Further work efforts will be required to incorporate blockchain technology, optimization of edge AI and development of more efficient privacy protection techniques, scalability, and distributed learning. Other studies can also be considered to study adaptive federated optimization and adversarial defenses to very secure next generation intelligent systems.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308-318).
2. Aljohani, A., Rana, O., & Perera, C. (2025). Self-adaptive federated learning in internet of things systems: A review. ACM Computing Surveys, 57(10), 1-36.
3. Chalamala, S. R., Kummari, N. K., Singh, A. K., Saibewar, A., &Chalavadi, K. M. (2022). Federated learning to comply with data protection regulations. CSI Transactions on ICT, 10(1), 47-60.
4. Hasan, M. T., &Kudapa, S. P. (2021). Data privacy-aware machine learning and federated learning: A framework for data security. American Journal of Interdisciplinary Studies, 2(03), 01-34.
5. Hassani, H., &MacFeely, S. (2023). Driving excellence in official statistics: unleashing the potential of comprehensive digital data governance. Big Data and Cognitive Computing, 7(3), 134.
6. Khraisat, A., &Alazab, A. (2021). A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. Cybersecurity, 4(1), 18.
7. Kocher, G., & Kumar, G. (2021). Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges. Soft Computing, 25(15), 9731-9763.
8. Le Jeune, L., Goedeme, T., &Mentens, N. (2021). Machine learning for misuse-based network intrusion detection: overview, unified evaluation and feature choice comparison framework. Ieee Access, 9, 63995-64015.
9. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE signal processing magazine, 37(3), 50-60.

10. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). Pmlr.
11. Safitra, M. F., Lubis, M., &Fakhrurroja, H. (2023). Counterattacking cyber threats: A framework for the future of cybersecurity. Sustainability, 15(18), 13369.
12. Sun, Yunchuan, et al. "Data security and privacy in cloud computing." International Journal of Distributed Sensor Networks 10.7 (2014): 190903.
13. Wu, J., Liu, Q., Huang, Z., Ning, Y., Wang, H., Chen, E., ... & Zhou, B. (2021, April). Hierarchical personalized federated learning for user modeling. In Proceedings of the Web Conference 2021 (pp. 957-968).
14. Zaman, U., Imran, Mehmood, F., Iqbal, N., Kim, J., & Ibrahim, M. (2022). Towards secure and intelligent internet of health things: A survey of enabling technologies and applications. Electronics, 11(12), 1893.
15. Zhang, X., Yin, W., Hong, M., & Chen, T. (2020). Hybrid federated learning: Algorithms and implementation. arXiv preprint arXiv:2012.12420.
16. Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. Future Generation computer systems, 28(3), 583-592.
17. Metahun Lemeon, JinfeRegash. (2026). Thermal-Electromagnetic Co-Simulation Framework for Performance Optimization of High-Power Electric Machine Drives. *National Journal of Electrical Machines & Power Conversion*, 52-63.
18. Fahad Al-Jame. (2025). Environment-Adaptive Multi-Agent Learning for Sustainable Edge-Oriented Services. *Transactions on Internet Security, Cloud Services, and Distributed Applications*, 47–55.
19. Deepika J. (2025). Machine Learning-Based Optimization of High-Frequency Injection Amplitude in Sensorless PMSM Drives. *Journal of Wireless Sensor Networks and IoT*, *2*(2), 85-91.