Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 17 No. 17 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijim.v17i17.42835

PAPER

P2FLF: Privacy-Preserving Federated Learning Framework Based on Mobile Fog Computing

B. Ankayarkanni[1](✉), Niroj Kumar Pani[2], M. Anand[3], V. Malathy[4], Bhupati[5]

[1]Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

[2]Department of Computer Science Engineering and Applications, Indira Gandhi Institute of Technology, Sarang, Odisha, India

[3]Department of ECE, Dr M.G.R Educational and Research Institute, Chennai. Tamil Nadu, India

[4]Department of ECE, SR University, Warangal, Telangana, India

[5]Department of IoT, KL Deemed to be University, Vaddeswaram, Guntur, Andhra Pradesh, India

ankayarkanni.cse@sathyabama.ac.in

ABSTRACT

Mobile IoT devices provide a lot of data every day, which provides a strong base for machine learning to succeed. However, the stringent privacy demands associated with mobile IoT data pose significant challenges for its implementation in machine learning tasks. In order to address this challenge, we propose privacy-preserving federated learning framework (P2FLF) in a mobile fog computing environment. By employing federated learning, it is possible to bring together numerous dispersed user sets and collectively train models without the need to upload datasets. Federated learning, an approach to distributed machine learning, has garnered significant attention for its ability to enable collaborative model training without the need to share sensitive data. By utilizing fog nodes deployed at the edge of the network, P2FLF ensures that sensitive mobile IoT data remains local and is not transmitted to the central server. The framework integrates privacy-preserving methods, such as differential privacy and encryption, to safeguard the data throughout the learning process. We evaluate the performance and efficacy of P2FLF through experimental simulations and compare it with existing approaches. The results demonstrate that P2FLF strikes a balance between model accuracy and privacy protection while enabling efficient federated learning in mobile IoT environments.

KEYWORDS

privacy-preserving, Federated Learning (FL), fog computing, security, mobile IoT

1 INTRODUCTION

The mobile Internet of Things (IoT) has revolutionized the way we interact with the digital world by linking various devices and sensors to the Internet [1]. These IoT devices create a massive amount of data every day [2], offering a rich source of information that can be leveraged for various applications, including machine learning. Machine learning algorithms have shown great potential in analyzing and extracting insights from mobile IoT data, enabling intelligent decision-making and predictive analytics [3]. However, despite the immense benefits of

Ankayarkanni, B., Pani, N.K., Anand, M., Malathy, V., Bhupati. (2023). P2FLF: Privacy-Preserving Federated Learning Framework based on Mobile Fog Computing. *International Journal of Interactive Mobile Technologies (iJIM)*, 17(17), pp. 72–81. https://doi.org/10.3991/ijim.v17i17.42835

Article submitted 2023-05-10. Revision uploaded 2023-06-29. Final acceptance 2023-07-05.

utilizing mobile IoT data for machine learning, privacy concerns pose significant challenges. Mobile IoT devices collect sensitive information about individuals, such as personal preferences, location data, and behavior patterns. The transmission and storage of this sensitive data can lead to privacy breaches, data misuse, and security vulnerabilities [4]. Consequently, strict regulations and privacy demands have been imposed to protect users' sensitive information.

To overcome these challenges, privacy-preserving machine learning techniques have emerged as a promising solution. These techniques aim to enable effective machine learning while safeguarding the privacy of individuals' data. Federated learning is one strategy that facilitates collaborative training of machine learning models without revealing the underlying raw data of the participating parties [5]. Federated learning operates in a decentralized manner [6], where each participant trains a local model on their data and only shares model updates with a central server or among the participating parties [7]. This decentralized approach ensures that the sensitive data remains on the local devices and is not transmitted to a central server, thereby addressing privacy concerns associated with data transmission.

In the context of IoT environments, fog computing has gained attention as an efficient paradigm for processing and analyzing IoT data [8]. Fog computing involves deploying computing resources, such as fog nodes, at the network edge, nearer to the IoT devices. This enables data processing and analysis to occur near the data source, reducing latency and bandwidth requirements. Motivated by the potential of federated learning and the benefits of fog computing in IoT environments, we suggest a "privacy-preserving federated learning framework (P2FLF) in a fog computing setting". P2FLF aims to enable effective and privacy-preserving machine learning in IoT environments by leveraging federated learning and fog computing techniques. The P2FLF framework utilizes fog nodes deployed at the edge of the network to establish local federated learning environments. Each fog node acts as a coordinator for a group of IoT devices within its range. The IoT devices within each group participate in local model training using their data, while the fog node coordinates the federated learning process [9].

To ensure privacy preservation, P2FLF integrates various privacy-preserving techniques, including differential privacy and encryption. Differential privacy adds noise to the model updates to prevent the leakage of individual information, while encryption techniques are used to secure communication between fog nodes and central servers. In this paper, we evaluate the performance and effectiveness of the P2FLF framework through extensive experimental simulations. We compare the results of P2FLF with existing approaches in terms of model accuracy, privacy preservation, and communication overhead. The evaluation aims to demonstrate the capability of P2FLF in striking a balance between model accuracy and privacy protection while enabling efficient federated learning in IoT environments.

The subsequent sections of this paper are structured as follows: Section 2 provides an overview of the existing literature in the realm of privacy-preserving machine learning and federated learning, providing valuable insights into the research conducted in this area. Section 3 presents the methodology. Section 4 presents the experimental setup and evaluation results. Finally, Section 5 concludes the paper and outlines future research directions. Overall, this research aims to contribute to the development of privacy-preserving machine learning frameworks for IoT environments. The proposed P2FLF framework offers a novel approach to address the privacy challenges associated with IoT data and provides a foundation for secure and collaborative machine learning.

2 LITERATURE REVIEW

[10] proposed a privacy-preserving federated learning approach specifically designed for fog computing. In this scheme, each fog node assumes the role of a participant, enabling it to gather data from Internet of Things (IoT) devices and perform the learning task within the framework. Their design incorporates a hybrid approach, combining both centralized and distributed training modes, to enhance practicality. This approach effectively addresses the issue of ineffective training caused by the substantial variations in data volume and computing power among IoT devices.

[11] introduced PPFchain, a new federated learning-capable blockchain-based structure, to protect the "security and privacy of sensor-IoT-based infrastructure" using sampled ECS data. To ensure privacy for users and their data residing in off-chain fog nodes, we employed the federated model in the architecture while taking performance into account. Furthermore, the findings demonstrate that the PPFchain delivers precision, effectiveness, and heightened protection.

Although federated learning (FL) has been utilized extensively in the Internet of Vehicles (IoV), there are issues in its efficiency and privacy protection when employed with fog computing. To address these issues, [12] proposed Galaxy, a practical framework for privacy-preserving FL in non-cloud-assisted fog computing. Galaxy utilizes secure multi-party computation (MPC) to enable collaboration between N fog nodes and multiple users while resisting collusion and dropout. Additionally, it manages poor-quality data while safeguarding user-related data. Our system exhibits great scalability, processing effectiveness, and low resource overhead, making it appropriate for IoV based on fog computing. Numerous tests support its performance.

[13] developed a novel privacy protection-based federated deep learning (PP-FDL) framework that achieves good classification rates from non-i.i.d data while protecting data against privacy-related GAN assaults. This fog-based model enables collaborative training of the FDL model, ensuring data privacy among contributors and protecting class probabilities with unique private identities. Empirical evidence demonstrates PF-DFL's superior performance compared to other state-of-the-art models, validating its data protection capabilities.

[14] suggested a secure aggregation technique for fog computing (FC) to guarantee FL efficiency while safeguarding the privacy of input data from devices. The foundation of this protocol is effective additive secret sharing. They first utilize a fog node (FN) as an intermediary processing unit to deliver local services that can assist the cloud server aggregate the sum throughout the training process. Second, we create a simple Request-then-Broadcast mechanism to make sure our protocol is resilient to clients who drop out.

[15] provided a secure and verifiable federated learning framework that enables federated deep learning and produces verifiable learning outcomes. To ensure the anonymity of users' local gradients during federated learning, they initially suggest a double-masking technique. The cloud server is therefore needed to give each user proof that its aggregated results are accurate.

3 METHODOLOGY

The proposed framework is separated into three layers "IoT Devices, Fog Computing, and Cloud Computing". In protection of privacy, the fog nodes play a

main role. Fog computing-based privacy-preserving federated learning frameworks seek to enable collaborative machine learning while protecting the confidentiality of sensitive data. These frameworks take advantage of the fog computing principle, which entails deploying computing resources closer to edge devices or fog nodes, hence minimizing the need for data transfer to a centralized cloud server. Fog nodes are purposefully placed closer to the data-generating devices at the network edge. To enable local data processing and analysis, these fog nodes serve as a bridge between the edge devices and the centralized cloud server. To ensure that no single fog node has full access to all the data, the sensitive data from edge devices are divided into subsets.

This division may be determined by a number of factors, including location, data type, and user identity. Utilizing its specific data subset, each fog node separately does local model training. Depending on the requirements of the particular application, different machine learning algorithms and approaches may be used throughout this training phase. Without sending the raw data to a centralized server, the local model is trained to utilize the data present on the fog node. The fog nodes communicate with one another or a central location, such as the cloud server, to aggregate their local models after local training. Depending on the privacy-preserving techniques used, the fog nodes may share model updates, model parameters, or encrypted model representations. The performance of the entire model is improved by model aggregation, which makes sure that the collective knowledge from many fog nodes is merged. Figure 1 is the proposed framework of privacy-preserving federated learning.

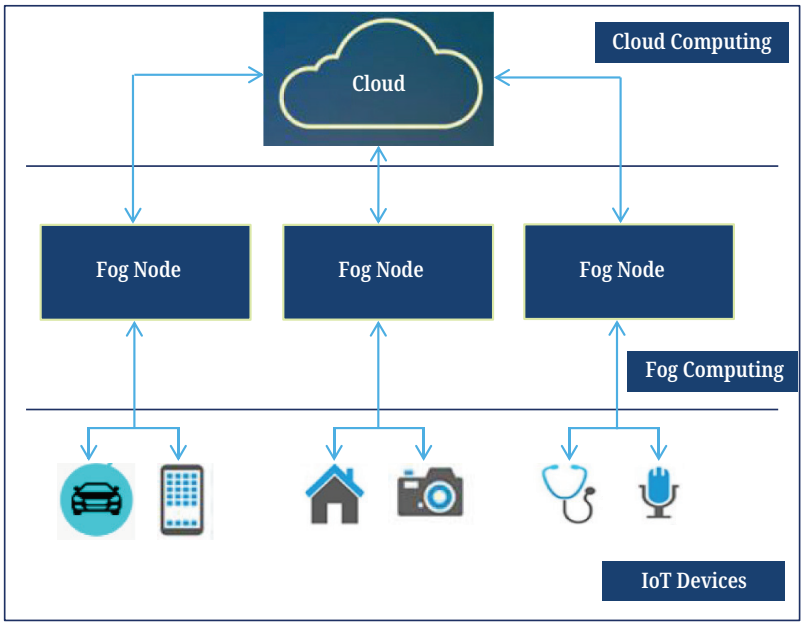

Fig. 1. Proposed framework of P2FLF based on fog computing

Multiple iterations of local training and model aggregation are commonly used in the federated learning architecture. The fog nodes work together to increase the overall model accuracy and performance by iteratively improving their local models on the basis of fresh data. Depending on the exact deployment option, the trained model can either be installed on the fog nodes themselves or on the edge devices

after the federated learning process is over. Then, the inference is carried out locally to reduce the need for data transmission and maintain privacy.

Edge devices, such as smartphones, IoT devices, or sensors, are located at the network edge and produce the data needed to train machine learning models. The task of gathering and preparing data before transmitting it to the fog nodes is performed by edge devices, which often have low computational resources. With this method, machine learning applications can process data more effectively at the network edge, with less connection overhead, and with improved privacy. Before the training process, the initialization should be done in the cloud server. The parameter and blinding server act as a pillar of the initialization process. Prior to training the initial model, the parameter server will gather some data sets from reliable users. The key pairs of the blinding factors and Paillier homomorphic encryption are generated by the blinding server. Then, each fog node receives them from the blinding server.

3.1 Initialization

In the initialization process, the blinding server first chooses "blinding factors (β_n)" and is expressed as,

$$\beta_1.\beta_2.,\ldots\ldots\beta_n = \frac{1}{v^i} \tag{1}$$

Where,

n – number of fog nodes, "*i*" stands for the current iteration's round number and $v^i \in V$. Then v^i will be directed to the parameter server.

3.2 IoT device data collection

After the process of initialization, the IoT device data will be collected. Depending on the number of IoT devices, every fog node allocates a privacy budget (α) to each one. The IoT device creates Laplacian noise after that and includes it to its data set based on α and sensitivity *Sf*. The probability density function of the Laplacian distribution with scale *c* represents the distribution as:

$$Lap\left(\frac{x}{c}\right) = \frac{1}{2c}\exp\left(-\frac{|x|}{c}\right) \tag{2}$$

IoT devices can go offline after uploading without using up their limited computational resources on training tasks. The IoT gadget keeps its local sensitivity a secret from other devices and the fog node.

The weight calculation

$$W_g^l = \sum_{cl=1}^{N_{cl}} W_m^l \cdot \frac{1}{N_{cl}} \tag{3}$$

Where,

W_g^l *and* W_m^l – weight mediums of the l-th layer

N_{cl} – *number of clients participating in the training*

3.3 P2FLF algorithm

Initialization

i. The FL framework's settings and the dimensionality of privileged class IDs are agreed upon by fog nodes.
ii. Each fog node p generates a random class identifier S_p^c for each class c in the applicable classes.
iii. During adversarial training, each attacker fog node p creates one or more additional fake classes C_p^{fake} and the corresponding class identifier S_p^{attack} attack to throw an adversarial attack for any given class identification.

Training phase

Construct a global model and set the weights Wg to any values.
Set the training epoch count E
Decide how many clients N_{cl} will attend the training.
Set the client side's desired number of training rounds Tr
Regarding Epoch = 1 to E;
Parallel executing
Provide the global model to every client taking the training
Regarding round = 1 to Tr:
Local model weights updated Wm
End for
End
Clients receive models
Add the weights together using Equation (3)
W_g global model weights update
End for

The algorithm summarizes the primary steps of the proposed framework. Commencing the training process earlier leads to an increased number of epochs being trained, resulting in a higher level of precision in the training.

4 RESULTS AND DISCUSSION

This section presents the experimental findings obtained from evaluating the proposed P2FL framework utilizing fog computing. The primary objective is to establish the feasibility of our system and demonstrate the effectiveness of the P2FL approach in comparison to a centralized algorithm based on Gradient Boosting Decision Trees (GBDT). The time analysis aimed to assess the efficiency of privacy-preserving federated learning with fog computing and traditional gradient boosting decision trees.

The experiments involved a distributed dataset across multiple fog nodes, and the training process consisted of iterative rounds of model updates and aggregations. The results revealed that the proposed approach significantly reduced the overall training time compared to conventional federated learning methods. The integration of fog computing played a crucial role in achieving this improvement. By allowing local model updates and reducing the need for frequent communication with a central server, fog computing facilitated parallelization and accelerated the training process.

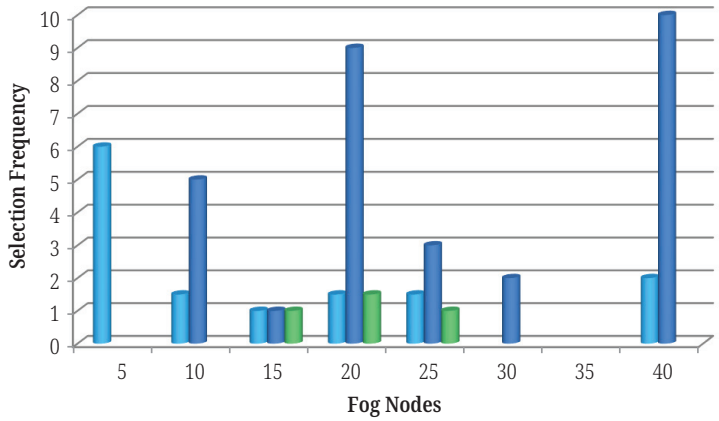

Fig. 2. Fog node selection graph

The outcome in Figure 2 illustrates how frequently various fog nodes are chosen as a global aggregator node at each round. By choosing various fog nodes with the least amount of workload and latency at each round, it makes sure that the suggested P2FLF approach brings robust to the selection process. In the worst case scenario, the cloud server will be in charge of performing the global aggregation if none of the fog nodes are able to match the workload and latency requirements.

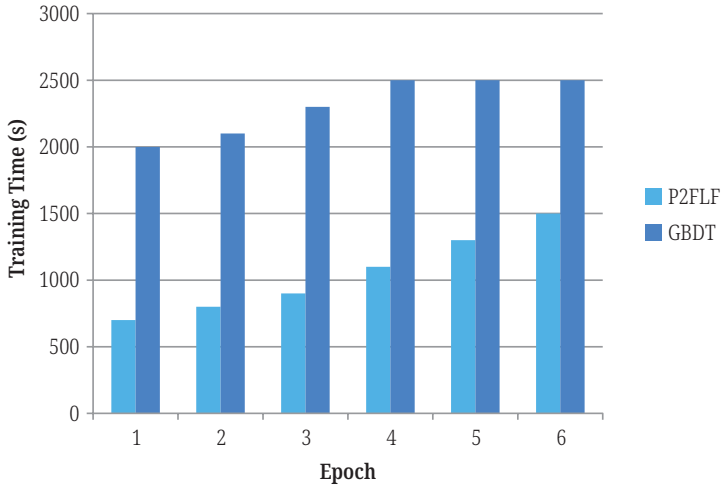

Fig. 3. Training time comparison between P2FLF and GBDT

The proximity of fog nodes to the data sources further contributed to reduced latency and improved response times during model updates. Leveraging the computational capabilities of fog nodes for local training and aggregation reduced the burden on the network and improved overall time efficiency. Figure 3 demonstrates the comparison of training time between the proposed privacy-preserving federated learning framework and GBDT. The results revealed that the proposed approach consumes less time for every epoch to process than the compared model.

The accuracy analysis aimed to evaluate the performance of the privacy-preserving federated learning system in terms of the resulting model's accuracy. The experiments assessed the model's performance on both the training data and a separate test dataset to gauge its generalization capabilities. The findings indicated that the proposed approach achieved competitive accuracy compared to conventional federated learning techniques. By utilizing traditional gradient boosting decision trees as the base

model, the federated learning system demonstrated robust and accurate results. The ensemble nature of gradient boosting helped mitigate the challenges posed by data heterogeneity across fog nodes and facilitated the creation of a reliable global model.

The privacy-preserving mechanisms employed in the federated learning process, such as secure aggregation and encryption techniques, did not significantly compromise the model's accuracy. The experimental results highlighted that it was feasible to incorporate effective privacy preservation measures without sacrificing the overall performance and accuracy of the federated learning system. The accuracy estimation formula is represented in given equation 4.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{4}$$

Figure 4 demonstrates the accuracy comparison of the proposed framework with the GBDT model and results revealed that the P2FLF accuracy percentage constantly increases for every epoch rather than the GBDT.

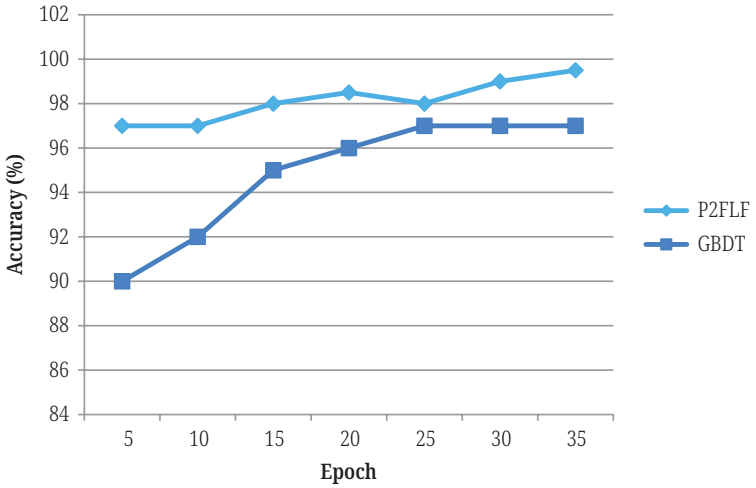

Fig. 4. Accuracy Comparison between P2FLF and GBDT

The utilization of fog computing in the proposed framework reduced training time by enabling local computations and minimizing communication overheads. Furthermore, the employment of traditional gradient boosting decision trees as the base model offered improved accuracy compared to other commonly used models in federated learning, such as logistic regression or neural networks. These findings provide promising avenues for enhancing both time efficiency and model accuracy in distributed learning scenarios.

5 CONCLUSION

In this research, we proposed a privacy-preserving federated learning framework based on fog computing, known as P2FLF (Privacy-Preserving Federated Learning Framework). The framework aimed to address the privacy concerns associated with traditional federated learning while leveraging the benefits of fog computing. Through our research, we successfully demonstrated the feasibility and effectiveness of the P2FLF framework. By utilizing fog computing, we were able to distribute the computational load and training process across fog nodes, reducing the reliance on a central server and minimizing the privacy risks associated

with data transfer. This decentralized approach enhanced the privacy guarantees for participants in the federated learning process. The P2FLF framework integrated various privacy-preserving techniques such as secure aggregation and encryption methods. These mechanisms ensured that sensitive data and model parameters remained protected throughout the learning process, safeguarding individual privacy and maintaining data confidentiality. Furthermore, our experimental results indicated that the P2FLF framework achieved competitive model accuracy compared to traditional federated learning approaches. The fog computing infrastructure facilitated local model updates and aggregations, reducing communication overhead and latency, which ultimately led to improved model performance.

6 REFERENCES

[1] Chaouchi, H., & Bourgeau, T. (2018). Internet of Things: Building the New Digital Society. *IoT*, *1*(1), 1. https://doi.org/10.3390/iot1010001

[2] Samann, F. E. F., Abdulazeez, A. M., & Askar, S. (2021). Fog Computing Based on Machine Learning: A Review. *International Journal of Interactive Mobile Technologies (iJIM)*, *15*(12), pp. 21–46. https://doi.org/10.3991/ijim.v15i12.21313

[3] Mahdavinejad, M. S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2018). Machine Learning for Internet of Things Data Analysis: A survey. *Digital Communications and Networks*, *4*(3), 161–175. https://doi.org/10.1016/j.dcan.2017.10.002

[4] Subahi, A., & Theodorakopoulos, G. (2019). Detecting IoT User Behavior and Sensitive Information in Encrypted IoT-App Traffic. *Sensors*, *19*(21), 4777. https://doi.org/10.3390/s19214777

[5] Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., & Zhou, Y. (2019, November). A Hybrid Approach to Privacy-Preserving Federated Learning. In *Proceedings of the 12th ACM workshop on artificial intelligence and security* (pp. 1–11). https://doi.org/10.1145/3338501.3357370

[6] Tabassum, N., Ahmed, M., Shorna, N. J., Ur Rahman Sowad, MD M., & Zabir Haque, H. M. (2023). Depression Detection Through Smartphone Sensing: A Federated Learning Approach. *International Journal of Interactive Mobile Technologies (iJIM)*, *17*(01), pp. 40–56. https://doi.org/10.3991/ijim.v17i01.35131

[7] Lyu, L., Yu, J., Nandakumar, K., Li, Y., Ma, X., Jin, J., & Ng, K. S. (2020). Towards Fair and Privacy-Preserving Federated Deep Models. *IEEE Transactions on Parallel and Distributed Systems*, *31*(11), 2524–2541. https://doi.org/10.1109/TPDS.2020.2996273

[8] Tuli, S., Basumatary, N., Gill, S. S., Kahani, M., Arya, R. C., Wander, G. S., & Buyya, R. (2020). HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments. *Future Generation Computer Systems*, *104*, 187–200. https://doi.org/10.1016/j.future.2019.10.043

[9] Wang, X., Wang, Y., Javaheri, Z., Almutairi, L., Moghadamnejad, N., & Younes, O. S. (2023). Federated Deep Learning for Anomaly Detection in the Internet of Things. *Computers and Electrical Engineering*, *108*, 108651. https://doi.org/10.1016/j.compeleceng.2023.108651

[10] Zhou, C., Fu, A., Yu, S., Yang, W., Wang, H., & Zhang, Y. (2020). Privacy-Preserving Federated Learning in Fog Computing. *IEEE Internet of Things Journal, 7*(11), 10782–10793. https://doi.org/10.1109/JIOT.2020.2987958

[11] Sezer, B. B., Turkmen, H., & Nuriyev, U. (2023). PPFchain: A Novel Framework Privacy-Preserving Blockchain-Based Federated Learning Method for Sensor Networks. *Internet of Things*, *22*, 100781. https://doi.org/10.1016/j.iot.2023.100781

[12] Li, Y., Li, H., Xu, G., Xiang, T., & Lu, R. (2022). Practical Privacy-Preserving Federated Learning in Vehicular Fog Computing. *IEEE Transactions on Vehicular Technology*, *71*(5), 4692–4705. https://doi.org/10.1109/TVT.2022.3150806

[13] Abdel-Basset, M., Hawash, H., Moustafa, N., Razzak, I., & Abd Elfattah, M. (2022). Privacy-Preserved Learning from Non-iid Data in Fog-Assisted IoT: A Federated Learning Approach. *Digital Communications and Networks*. https://doi.org/10.1016/j.dcan.2022.12.013

[14] Liu, Y., Dong, Y., Wang, H., Jiang, H., & Xu, Q. (2022). Distributed Fog Computing and Federated-Learning-Enabled Secure Aggregation for IoT Devices. *IEEE Internet of Things Journal*, *9*(21), 21025–21037. https://doi.org/10.1109/JIOT.2022.3176305

[15] G. Xu, H. Li, S. Liu, K. Yang, and X. Lin, Verifynet: Secure and Verifiable Federated Learning, *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 911–926, 2020. https://doi.org/10.1109/TIFS.2019.2929409

7 AUTHORS

Dr. B. Ankayarkanni completed her B.E. (ECE) in 2001 from Manonmaniam Sundaranar University, M.E. (CSE) in 2003 from Manonmaniam Sundaranar University and Ph.D. in 2019 from Sathyabama Institute of Science and Technology. She has got 20 years of teaching experience and presently she is working at Sathyabama Institute of Science and Technology as Professor in the Department of Computer Science and Engineering. Her area of interest includes Machine learning, Image Processing and Artificial Intelligence and has published more than 30 publications in Scopus and Web of Science journals.

Niroj Kumar Pani received his M.Tech in Computer Science with specialization in Information Security from National Institute of Technology, Rourkela, India, in 2009 and Ph.D. in Computer Science from Utkal University, Bhubaneswar, India, in 2018. He had worked as an Assistant Professor in Indian Institute of Science and Information Technology and as a senior analyst in ProcessMAP Corporations, India. At present, he is working as Assistant Professor in the Department of Computer Science Engineering and Applications in Indira Gandhi Institute of Technology, Sarang, India. His research interests include network security, wireless ad hoc and sensor networks, cloud computing, IoT, and machine learning. He has authored in more than 10 international journals and books in his field of expertise (email: nirojpani@gmail.com).

Dr. M. Anand is working as a professor in the department of ECE, Dr M.G.R Educational and Research Institute. He has totally 18 years of experience both in teaching and research. He published more than 100 articles in various journals and holds four patents Thirteen students completed their Ph.D. under his guidance and currently he is guiding 7 students. His area of interest is Embedded Systems, Artificial Intelligence, Image Processing, Internet of Things, Quantum Computing, Networking (email: anand.ece@drmgrdu.ac.in).

Dr. V. Malathy is working in SR University, Warangal, Telangana State. She has completed 15 years of service in teaching field and has industrial experience also. She graduated from Thiagarajar College of Engineering, Madurai. She has completed M.E. and Ph.D. from Anna University, Chennai. Her area of interest is Image Processing and has published more than twenty journals. She has published 8 books in "The Charulatha Publications" (email: malathy.v@sru.edu.in).

Bhupati is assistant professor in Department Internet of Things (IoT) at K L Deemed to be University and has a research focus on Internet of Things. He brings expertise in areas such as Embedded Systems, IoT, and Cloud Computing. With a passion for both research and teaching, Bhupati pioneered many aspects that enhance the fault tolerance of many mission-critical IoT-based applications. He has published 2 SCI papers, 4 SCOPUS journals and 5 patents. Bhupati has conducted many workshops in different areas like Embedded Systems and IoT (email: bhupati@kluniversity.in).