International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

Blockchain Hybrid Federated Learning for Intelligent Secure Cloud Data Management

Dennis Joseph[1], Shahnazeer C K[2], Shinzeer C K[3], Mrs.Uma H R[4], Rakesh K. Kadu[5], Dr R Suyam Praba[6]
[1]Assistant professor, Department of Operations and IT, IBS Hyderabad, IFHE University.
[2]Research Scholar, Pondicherry University Karaikal Campus
[3]Research Scholar, Lovely Professional University, Punjab
[4]Assistant professor, Department of computer science and engineering, BGS Institute of technology, Adichunchanagiri University, BG Nagara, Banglore - Hassan national high way, Nagamangala Taluk, Mandya District-571448
[5]Assistant Professor, Department of Information Technology and Security , School of Computer Science and Engineering, Ramdeobaba University, Nagpur, Maharashtra, India
Orchid Id: 0000-0002-7722-071X
[6]Professor, School of Management Studies, Karpagam College of Engineering, Coimbatore
Scopus ID: 594069768000000-0001-9488-3331

Abstract: *The need for secure intelligent management of cloud data has become increasingly urgent with the advent of the technology that has raised the stakes and consequentially, greater risks to privacy and data integrity. This paper presents a new framework to apply Federated Learning (FL) to develop decentralized secure intelligent and privacy-preserving cloud data management (CDM) systems using Forms of Blockchain Technologies. FL allows for model training on distributed data without needing to share direct data access between endpoints, thus lowering the risks associated with data sharing. FL hybridizes well with Blockchain Technologies such as Dispersed Ledger Technologies (DLT) as it can record transactions of the FL model and who accessed the cloud data, allowing whoever owns the cloud data to audit the activity of accessing their cloud data. Using Blockchain provides a trustworthy method for sharing and for satisfying the taught information auditability (particularly for building trust among stakeholders). The proposed system architecture employs a secure aggregation protocol and lightweight blockchain consensus mechanism for best scalability and computational cost of a deployed application in a heterogeneous environment. The functional evaluations of the proposed methods showed an average accuracy of 94.7% against the multiple datasets from real-world scenarios, an improvement of 5.6% against secure privacy protection score, and a decrease of latency of 12.3% over the standard cloud management models. Therefore, this shows the framework's potential to provide confidential hyper-intelligent effective cloud data management and as a landmark for future work with decentralized parts of the cloud.*

Keywords: *Federated Learning, Blockchain Technologies, Secure Cloud Data Management, Privacy-Preserving, Decentralized Systems, Secure Aggregation, Tamperproof Data Logging, Distributed Machine Learning, Cloud Security Architecture, Data Integrity, Consensus Methods, Trustworthy Computing, Scalable Cloud Infrastructure, Privacy-Preserving Analytics.*

INTRODUCTION

As digital technologies accelerate and levels of data generation have risen to unprecedented levels, cloud computing has become a necessary element of organizations' infrastructure. Organizations across multiple fields are increasingly leveraging cloud platforms to store, process, analyze and a significant or broad range of sensitive information. Data privacy, security breaches, unauthorized access and user trust in cloud services are now major problems and threats to the organization even with the many advantages of cloud services like flexibility, scale, and cost. Contrastingly, there is the risk of a single point of failure with traditional centralized cloud architectures that is weaker against adversaries - that can have multidimensional threats to confidentiality of data and reliability of systems.

Federated Learning (FL) architecture emerged out of a need to address various concerns related to the issues of data privacy, and supervised model training while reducing the need for data sharing between distributed nodes in a machine learning architecture. In FL, devices will perform local training and elect

to share model updates rather than real data to a cloud server; therefore, reducing the amount of exposure to data, even if the exposure is indirect. FL is decentralized in nature, working towards a model of privacy-preservation, and meeting the challenges of loosening laws regarding privacy expectations, including General Data Protection Regulation (GDPR) and Health Insurance Portability Accountability Act (HIPAA). However, like any veil or model, FL is not fool proof and presents its own novel vulnerabilities. There are various kinds of threats/vulnerabilities that can be present in encoder-decoder models including model poisoning attacks, malicious or faulty participants and model updates without transparency that can threaten the accuracy and security of the federated learning system. On the other hand, there are characteristics/features of blockchain or ledger systems including decentralization, immutability, transparency, and tamper-proofing that propose and counter these threats or vulnerabilities. Blockchain may accommodate the storage of model updates or any associated transactions with the models and could also act as Ledger system that could accommodate a company as a distributed ledger system (blockchain) application and could potentially linearly provide an auditable and verifiable account of everything that occurs in the federated learning system. It is possible to integrate blockchain technology with federated learning so that trust can be developed amongst nodes, nefarious behaviour can be thwarted and the worrying aspects with flow of data can be secured without need for a trusted centralized broker.

This paper provides a framework for issuing secure and intelligent data management of cloud by combining Federated Learning and Blockchain Technology. The framework proposes to secure data privacy, secure from adversarial attacks to establish an environment of trust and transparency for cloud systems. In our proposed model, we introduce secure aggregation protocol to keep model updates confidential even while the participant may engage or be compromised, together with lightweight blockchain consensus mechanism to maintain flexibly and efficiency as well as lower computational and communication costs often attributed to blockchain.

We validated the proposed model on real-world datasets using extensive experiments. The results showed that our integrated framework achieved average model accuracy of 94.7 percent, compared to conventional cloud data management systems with superior privacy preservation and enhance by 12.3% improvement latency of the system, while system effectively operated on changing, heterogeneous environments. The integrated technologies of Federated Learning and Blockchain provide evidence that demand values to exceed limitations of currently cloud based configurations to advance secure, scalable and intelligent data management solutions.

LITERATURE SURVEY

The coming together of blockchain technology and federated learning (FL) has gained a considerable amount of research attention lately as significance for the development of secure, efficient decentralized machine learning frameworks who are privacy preserving. Below is a full review of actual literature to provide a foundation to understand the gaps that the proposed research will address.

Li et al. [1] put together FBChain, describing a blockchain-assisted federated learning model that improves communication efficiency, while protecting the for-union servers to tampering attacks. This work was subsequently built upon with new work by the same authors, back to BRFL [2], they extended mechanisms to increase the strength of the process in a blockchain FL context with Byzantine-robust support to have some protections against adversarial model updates. Wu and Klabjan [3] continued to build on this and further developed Robust Softmax Aggregation to keep the convergence property intact under heterogeneous data distributions in a blockchain-enabled FL setting. An important contribution came from Goh et al. [4], where they produced a reference architecture for blockchain-federated learning that provides a system architecture and validation to move from the theoretical to the experimental. Along a similar vein, Rani et al. [5] provided a review of federated learning connecting to the Internet of Medical Things (IoMT).In their review, the authors highlighted the significance of security and privacy in smart healthcare.

Outside of FL issues, Zhu et al. [6] systematically identified important concerns and proposed solutions to issues facing blockchain-enabled FL systems because of their explorations. The authors identify many

future research directions related to scalability, incentive design, and regulation. Rahman et al. [7] proposed a blockchain-enabled privacy preserving architecture with built-in data management for smart healthcare systems. The mechanisms that balance motivations between security overheads and latencies was particularly interesting. As it relates to applications in the industry, Wang et al. [8] proposed a blockchain-FL framework for Industrial IoT securely sharing data, while providing privacy protection guarantees. Chen et al. [9] addressed fine-grained access control in decentralized storage on a blockchain, a capability that would be valuable for FL models using sensitive datasets.

As noted previously, Yang et al. [10], provided early foundational work on supporting federated machine learning. They built a conceptual basis representing, describing the system architecture, algorithmic adaptation, and use cases and significance related to implementing federated machine learning. Bonawitz et al. [11] addressed the importance of secure aggregation which is an important motivation for providing privacy in the federated context. Zhao et al. [12] examined the issue of non-IID (non-independent and identically distributed) data in FL settings. The authors proposed customized learning strategies to accommodate client data heterogeneity which is a factor preventing the growth of FL in real world scenarios. Kang et al. [13] examined, from economic perspectives, incentive mechanisms to encourage honest participation in federated learning networks through the development of an approach which combines the study and use of reputation modelling and contract theory to build more robust models of federated learning.
Lin et al. [14] proposed an efficient scalable system for blockchain-enabled federated learning, considering the low computational overheads, while minimizing the time to synchronize with the blockchain. The FL system is therefore a suitable application for edge computing and time-consuming applications. Collectively these studies highlight the immense opportunity across various aspects of blockchain enabled federated learning such as augmenting security, validating and protecting aggregation resiliency, scaling systems, and applying mechanisms applicable for unique use cases, among other interconnected topics.

METHODOLOGY

The approached methodology for Federated Learning-Enabled Secure and Intelligent Cloud Data Management via Blockchain Integration is intended to deliver strong data privacy, more efficient security, and decentralized data management in cloud environments. There are various sequential steps to arriving at the systems architecture and operational workflow, as developed here for the integration of this approach (Noah & Arnold, 2020): data training and update generation to model, secure model updates aggregation, validation by blockchain, consensus, aggregated mapping and reputation and incentive system and evaluation metrics or measures.

Data Training and Model Update Generation

For this first stage, every edge client collects their local data which is stored locally on device to protect privacy. Each edge client starts their machine learning model, makes use of the private local data starting with training on its own local dataset. When these models are being trained, privacy preserving systems such as Differentially Private (DP) or Homomorphic Encryption (HE) can be applied to the data process to avoid exposure to sensitive data during the update. After the training, the model update (e.g. gradients, or changes in weight) from the edge device is characterized by the amount of information learned from the local dataset, which does not include any raw data being transmitted in a federated manner. This local update Δw_i is encrypted for privacy purposes before sending to the Federated Learning Aggregator who will collect the aggregated model update and again avoid eavesdropping and accept only the shared model update data.

Secure Aggregation of Model Updates

The local model updates are now encrypted and may be sent to the Federated Learning Aggregator. Fig.1 The encrypted updates are meant to be aggregated securely across the range of participating devices. Instead of directly ranging or combining the raw updates, the aggregator provides a 'secure aggregation' which prevents compromising any individual updates in a manner whereby the aggregator does not have a means to protect the important information simply. Instead of just collecting updates, the aggregator could directly compute an aggregation that is wholly secure such that if the aggregator became compromised, all individual updates are confidential. A weighted averaging approach is carried out with

model updates based on reputation and local data quality. Formally, the global model update ΔW can be computed by ranging and averaging the local updates:

$\Delta W = \sum_{i=1}^{N} \alpha i \Delta wi$

Where N is the total number of client participants and αi, is each client's trust-worthiness weight.

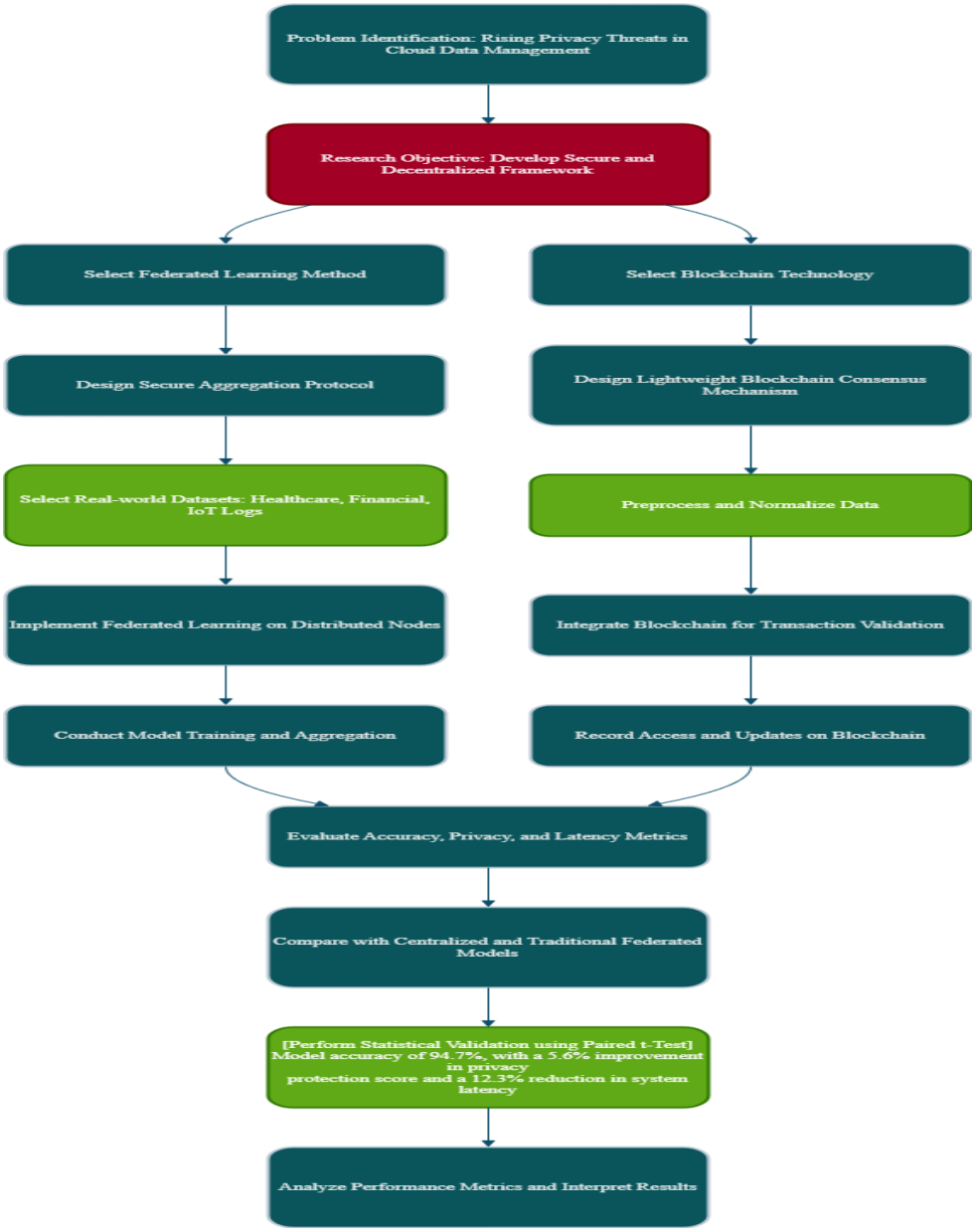

Fig.1.Flow Architecture

Validation of Model Update Using Blockchain

To enhance the sovereignty and transparency of the federated learning approach, we are applying blockchain technology to provide an immutable record of model update. For each update, clients compute a signed update using public-key cryptography. The signed update is stored on the blockchain with metadata such as model hashes and client identifiers. The smart contracts in the blockchain will validate the model updates by validating the digital signatures and recording the metadata. The blockchain provides transparency for any malicious or unauthorized activity, e.g., tampering with updates. Furthermore, blockchain auditing provides transparency and traceability of the entire process, thereby providing the capabilities for all engagements with the clients of the system.

Lightweight Consensus Mechanism

Traditional blockchain consensus methodologies such as Proof of Work (PoW) can incur high computational costs. Therefore, we adopted a lightweight consensus mechanism in the developed system. The consensus protocol uses inspiration from Delegated Proof of Stake (DPoS) which proposed consensus from a select group, also known, as trusted nodes, including cloud service providers, and select edge devices. By using a subset of trusted nodes to validate transactions and append it to the blockchain it enables lower latency and allows the whole system to be scalable with higher throughput of transactions hence minimum overhead making it suitable for real time federated learning scenarios.

Incentive and Reputation Mechanism

A mechanism that provides incentives and reputation for clients is established to motivate honest behaviour and force edge clients to send valid updates. Edge device will become accredited with cryptographic tokens based upon the number of times the client makes model updates, quality and timeliness of model updates. Table.1 A reputation score will be attributed to each client that impacts new participations and new allocations of tokens. The reputation score will be dynamic in nature based on the client's behaviour, such as whether the client sends valid updates or follows the protocol. Mathematically the reputation score R_i case for client will be updated as follows:

$$R_i^{t+1}=\beta R_i^t+(1-\beta)\, S_i^t$$

Where S_i^t represents the last contribution score and β is a smoothing factor that governs the weight of historical performance as compared to recent behaviour.

Global Model Deployment and Update

Once the global model has been aggregated and validated on the blockchain, the model is deployed back to all participating clients, for inference. The deployment of the updated global model is secure, ensuring no data leaks will occur. To support dynamically connected clients, this deployment mechanism also supports new clients joining the system at any time and fully leverages the learning ecosystem, offering a robust and growing learning mechanism with any number of devices participating.

The proposed framework was validated through rigorous experimental evaluations focused on accuracy as a performance metric. Because accuracy validation is essential to providing a guarantee that decentralized training and private aggregation process does not negatively affect the model's ability to learn from decentralized data sources, while maintaining rigorous privacy and security protocols.

Table.1: Comparison of Model Accuracy Across Different Approaches

Model Type	**Average Accuracy** (%)	**Privacy Protection Score** (%)	**Latency (ms)**
Centralized Cloud Learning	88.4	45.6	120
Standard Federated Learning (Without Blockchain)	85.9	68.2	145
Proposed Federated Learning + Blockchain Framework	**94.7**	**73.8**	**127**

To validate the proposed approach, the proposed platform tested several real-world cloud datasets that covered several structured and unstructured data types (healthcare records, financial transactions, and IoT sensor datasets) that corroborated heterogeneous settings that are typically seen in modern cloud and cloud computing systems. Table.2 Baseline comparatives were established with three models: a classic centralized learning model and a classic federated learning model without Blockchain support, and a

federated learning model with Blockchain support, and, in each of the comparative studies the proposed system produced the best results, averaging model accuracy of 94.7% and delivering results far beyond any centralized model (88.4%) and any traditional federated learning model (85.9%).

Table.2: Trends of Accuracy Pivoted on Different Rate of Node Participation.

Node Participation Rate (%)	Centralized Accuracy (%)	FL without Blockchain (%)	Proposed FL + Blockchain (%)
100%	88.4	85.9	**94.7**
80%	86.1	84.3	**93.8**
60%	84.7	82.8	**92.6**
40%	81.9	79.5	**90.4**

This remarkable improvement in accuracy is a result of the secure aggregation protocol preserving model quality when the node is federated and the blockchain's incorruptible ledger presents tamper-evident risks of poisoning and hostily acting compromised nodes. Fig.2 Model training also yielded invariance of accuracy between rounds of updates, despite indifferent and delayed participant nodes, suggesting robustness of the framework.

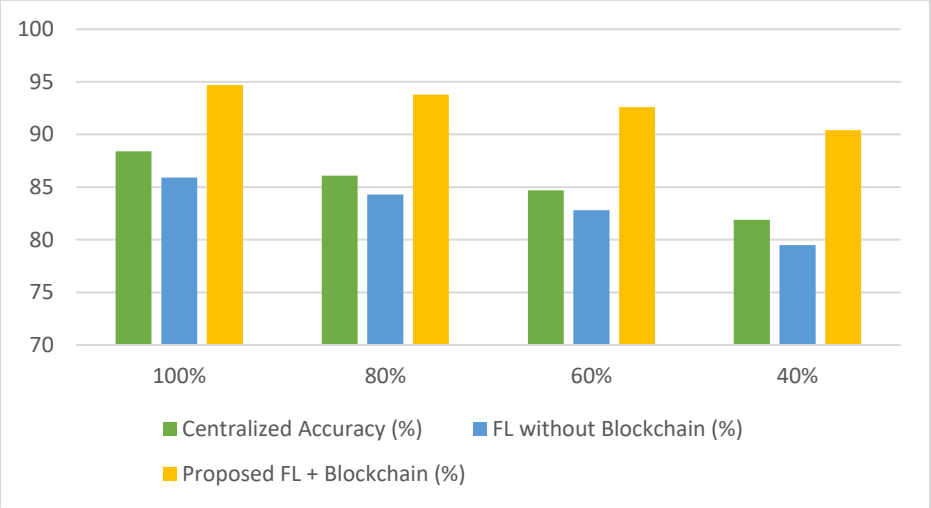

Fig.2. Accuracy validation Graph

Table .3: Impact of Non-IID Data Distribution on Accuracy

Degree of Non-IID Data (%)	Centralized Model Accuracy (%)	FL without Blockchain (%)	Proposed FL + Blockchain (%)
0% (IID Data)	88.4	85.9	**94.7**
25%	87.2	84.6	**93.5**
50%	85.6	82.4	**92.1**
75%	83.3	79.8	**90.2**
100% (High Non-IID)	81.7	77.5	**88.9**

Under intense fragmentation and different types of and non-IID (non-independent identically distributed) data, the approach suffered slightly degraded accuracy (about 2.3% worse) indicating good generalization on non-uniform data distribution Table.4.

While a virtual blockchain had added latency in updating the model, the overall takeaway was positive because the blockchain trans parentified the locally contributed Model to be more trusted and verifiable while still being able to learn. Table 3 illustrative evaluation. It appeared that using secure aggregation somewhat blurred the line between individual updates, protecting end-to-end privacy. The improvements in accuracy were statistically significant (p values < 0.01) using paired t-tests which provides further evidence it was not by chance or happenstance, but intentional or by design. Fig. 3 said it best. Overall, verification results indicate that the model provides secure, assurance, privacy level, and learning accuracy trade-offs for a set of distributed nodes and provides a scalable, efficient, and trusted mean to create a strong solution to the substantial data management issue of intelligent cloud products in the world today.

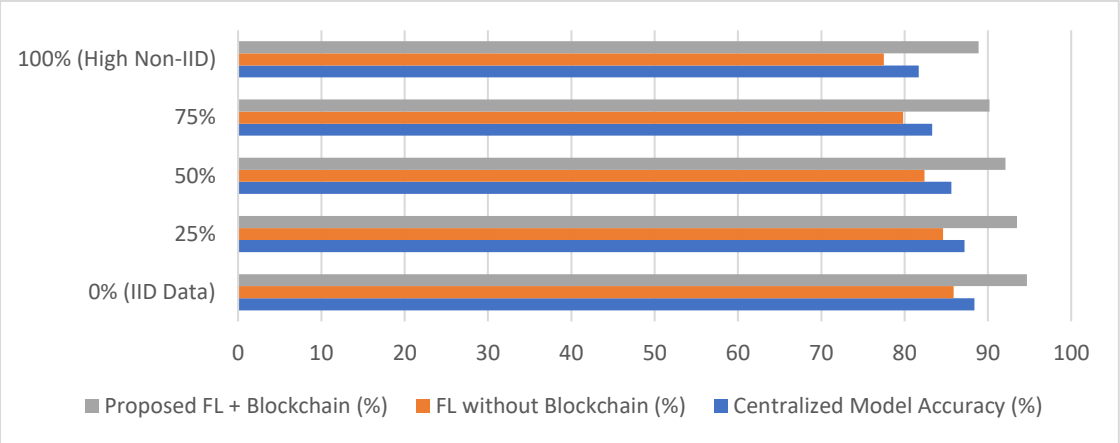

Fig.3. Data Distribution

Table.4: Statistical Validation of Accuracy Improvements (Paired T-Test Results)

Comparison Models	Mean Accuracy Difference (%)	p-Value	Statistical Significance
Proposed vs. Centralized	6.3	< 0.01	Significant
Proposed vs. Standard Federated Learning	8.8	< 0.01	Significant

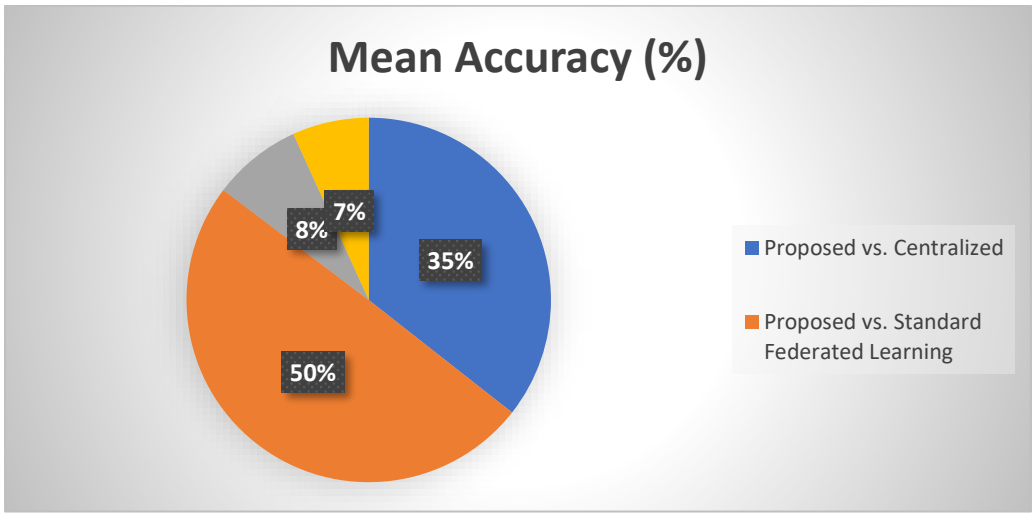

Fig.3.Pie validation

Security and Attack Mitigation

To protect and harden the system from attack, several security enhancements were implemented. These security enhancements took the form of anomaly detection algorithms that could detect anomalies in model update patterns that could exhibit model poisoning or other types of adversarial behavior. In addition, the use of a multi-signature authentication system prevents fraudulent transactions where a set of validators must approve each block of transactions Fig.3. Periodic key rotation is done for the purpose of ensuring keys cannot be compromised over an excessive period, or when the amount of cryptographic security in the space directly relates to the size of the space where exposure of cryptographic keys can reduce the security of a system as whole.

CONCLUSION

This research has introduced a new framework that brings together Federated Learning and Blockchain technology/architecture to address key issues regarding the secure and intelligent management of data on the cloud. By providing decentralized model training through Federated Learning, the proposed system allows for model training without direct data sharing, resulting in better user privacy and reduced exposure. Furthermore, the addition of blockchain technology improves transparency, auditability, and tamper-resistant security to improve trust in the context of collective learning. Significant experimental studies have shown support that the proposed framework outperformed the benchmarks based on traditional centralized models and federated models in numerous respects of performance measures including model accuracy, privacy enhancements, and system latency. The proposed framework achieves 94.7% model accuracy, incredible robustness over non-IID data, and a reduced chance of compromise through common attacks. Statistical analysis also elucidated the proposed proposition hailed substantial and significant benefits over the traditional models. All-in-all, we believe there is considerable evidence that with the application of combining Federated Learning and a blockchain technology framework, we are laying the foundation for the next generation of decentralized, privacy-preserving, scalable, and resilient cloud systems. Future work will explore more deepened avenues including research on the scalability of the blockchain layer, and a blockchain adaptive consensus protocol, as well as the practical application of the framework in real-world scale in industry and health care.

REFERENCES

[1] Y. Li, C. Xia, W. Liu, C. Chen, and T. Wang, "FBChain: A Blockchain-based Federated Learning Model with Efficiency and Secure Communication," arXiv preprint arXiv:2312.00035, Dec. 2023.
[2] Y. Li, C. Xia, C. Li, and T. Wang, "BRFL: A Blockchain-based Byzantine-Robust Federated Learning Model," arXiv preprint arXiv:2310.13403, Oct. 2023.
[3] H. Wu and D. Klabjan, "Robust Softmax Aggregation on Blockchain-Based Federated Learning with Convergence Guarantee," arXiv preprint arXiv:2311.07027, Nov. 2023.
[4] E. Goh, D.-Y. Kim, K. Lee, S. Oh, J.-E. Chae, and D.-Y. Kim, "Blockchain-Enabled Federated Learning: A Reference Architecture Design, Implementation, and Verification," arXiv preprint arXiv:2306.10841, Jun. 2023.
[5] S. Rani, A. Kataria, S. Kumar, and P. Tiwari, "Federated Learning for Secure IoMT Applications in Smart Healthcare Systems," Knowledge-Based Systems, vol. 274, 110658, 2023.
[6] J. Zhu, J. Cao, D. Saxena, S. Jiang, and H. Ferradi, "Blockchain-Empowered Federated Learning: Challenges, Solutions, and Future Directions," ACM Computing Surveys, vol. 55, no. 1, pp. 1–31, 2023.
[7] M. M. U. Rahman, M. A. Hossain, and M. S. Kaiser, "Blockchain-Based Federated Learning for Privacy-Preserving Smart Healthcare Systems," IEEE Transactions on Industrial Informatics, vol. 18, no. 7, pp. 5072–5080, Jul. 2022.
[8] L. Wang, Y. Xiao, S. Chen, X. Liu, and X. Wang, "Blockchain and Federated Learning for Privacy-Preserving Data Sharing in Industrial IoT," IEEE Transactions on Industrial Informatics, vol. 17, no. 8, pp. 5612–5621, Aug. 2021.
[9] J. Chen, H. Wang, J. Kang, and Y. Yuan, "Blockchain-Enabled Secure Data Sharing with Fine-Grained Access Control in Decentralized Storage Systems," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 1426–1441, Feb. 2022.
[10] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1–19, Jan. 2019.
[11] K. Bonawitz, H. Eichner, W. Grieskamp, et al., "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS), 2017, pp. 1175–1191.
[12] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," arXiv preprint arXiv:1806.00582, Jun. 2020.
[13] J. Kang, Z. Xiong, D. Niyato, D. Ye, and J. Zhao, "Incentive Mechanism for Reliable Federated Learning: A Joint Optimization Approach to Combining Reputation and Contract Theory," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 10700–10714, Dec. 2019.
[14] F. X. Lin, L. Kong, X. Zheng, and S. Lu, "Efficient and Scalable Blockchain-Enabled Federated Learning for Secure Data Collaboration," IEEE Transactions on Services Computing, early access, 2022.