University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

Intelligent Trust-Based Weighted Fusion with Blockchain for Adversarially Resilient in Federated Learning

Ghsuoon B. Roomi

Department of Computer Science, University of Technology, Bagdad, Iraq.

ghosoon.b.rumi@uotechnology.edu.iq

https://doi.org/10.32792/utq/utj/vol**20**/1/1

Abstract

Federated Learning (FL) is a model training scheme with guaranteed preserved data privacy, but model poison and manipulation attack susceptibility, and model performance degradation through adversary attack. In this work, a Blockchain-Based Verification with Fusion Mechanism (BVFM) is designed to enhance FL's security and robustness. With a tamper-evidence model update guaranteed through a blockchain layer, and a trust-dependent weighted fusion mechanism, trust values are granted to participating nodes, dynamically weighing them in contributing to a global model. Experimental evaluation in terms of performance under Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), and Carlini & Wagner (C&W) attack scenarios in medical imaging tasks confirms efficacy of BVFM and, in comparison with baseline FL techniques, its accuracy improvement to 94.3%, outperforming local training nodes (88.7%–90.1%). Under adversarial conditions, BVFM reduces the Adversarial Success Rate (ASR) from 59.3% to 25.0% (C&W attack) and from 49.8% to 20.1% (PGD attack), significantly enhancing model robustness. Furthermore, t-SNE visualizations illustrate BVFM's ability to maintain the separability of benign and malignant classifications despite adversarial perturbations. Compared to existing FL approaches, BVFM achieves the highest accuracy (94.4%), precision (92.5%), recall (93.1%), and F1-score (92.8%), while requiring only 85 seconds of training time—29% faster than leading methods. These results highlight BVFM as a scalable and secure FL solution for adversarial resilience in medical imaging, autonomous systems, and cybersecurity applications.

Keywords

Federated Learning, Blockchain Verification, Trust-Based Fusion, Adversarial Attacks, Medical Imaging Security, Model Poisoning, Secure Decentralized Learning, AI Robustness, Cybersecurity in FL.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

1. Introduction

In recent years, widespread use of distributed training frameworks has opened new avenues for model training in harmony with keeping data confidential. Federated Learning (FL) is one such striking framework that enables decentralized training through providing many nodes with an opportunity to learn a common model together in a manner such that no raw information is exchanged between them. FL, with its potential, suffers badly in practice with non-independent and identically distributed (non-IID) data, high communication expense, and vulnerability to attack. In most critical cases, such as in medical, financial, and smart grid, with sensitive information distributed over many entities, secure and reliable federated training processes have become a necessity.[1]

Adversarial attacks form a key challenge for FL frameworks. Malicious nodes sending incorrect or poisoned updates with a purpose to corrupt the training and degrade the performance of the global model form such an attack. Model poisoning, data poisoning, and inference attack form three most prevalent attack types in FL. In such an attack, the global model can be tampered with, and sensitive information can become compromised, and traditional aggregation algorithms, such as Federated Averaging (FedAvg), become deficient. Hence, creating robust FL algorithms with high performance and capable of countering adversarial factors is significant.[2]

A survey of state-of-the-art works reveals a range of new techniques for enhancing FL's security in terms of adversary attack for strengthening its security. For example, a few utilize blockchain for secure dissemination of updates, and a few utilize complex fusion approaches or graph neural networks for accuracy and scalability improvement in detection. However, such techniques have limitations in terms of high computational cost, ineffectiveness in handling non-IID distributions, and susceptibility to complex adversary attacks. In spite of such enhancements, such works have no overall frameworks for handling the intertwined issue of adversary robustness, data heterogeneity, and secure communications.[3]

The proposed algorithm extends such enhancements with a trust value weighted fusion mechanism and a model update verification use of a blockchain for secure communications between nodes. Model update integrity and authenticity checking use a blockchain for secure communications between nodes. Trust value weighted fusion mechanism fuses received updates from nodes in terms of trust values, effectively filtering out any malicious contribution and lessening its impact. Dynamical updating of trust values and performance checking, the proposed algorithm not only keeps the overall model secure but optimized for heterogeneous and multi-modal distributions.[4]

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

Fusion mechanisms in federated learning, specifically with regard to adversary robustness, have a significant role to contribute. Traditional aggregation methods assign a uniform weight to all nodes and, therefore, can amplify adversary impact. In contrast, weighted fusion approaches utilize trust values or performance values to prioritize updates received from trustful nodes. Through selective aggregation, such approaches counteract adversary impact and enable robust and reliable aggregation in a worldwide model. In our work, fusion mechanism is closely intertwined with a blockchain, offering a secure and flexible platform for a decentralized environment.[5]

The proposed scheme outpaces traditional techniques with a synergistic combination of trust-dependent fusion, adaptability, and aggregation algorithms and blockchain technology. Unlike in traditional works with static fusion rules and centralized verification, our scheme dynamically adjusts trust values in terms of performance and consistency of individual nodes. Besides, with a use of a blockchain technology, a tamper-evidence update record is generated, and security in the system is boosted even further. With such breakthroughs, the scheme addresses critical FL system weaknesses such as adversary attack, heterogeneity in information, and secure information transmission.

The motivation for such work comes with a growing demand for secure FL systems that can function in a secure and efficient manner in adversary and poor-resource environments. Existing works lack in balancing security, efficiency, and scalability, and real implementations suffer with many gaps. With trust-based fusion and blockchain, such gaps in proposed work are filled, and an efficient and scalable scheme for secure FL is proposed. In this work, a scheme for FL is proposed with a target to develop a secure FL scheme that can resist adversary impact and have high performance and secure collaboration between nodes.

In summary, the proposed scheme addresses a range of FL concerns, including adversary attack, heterogeneity in data, and secure communications. With update checking via use of a blockchain and trust aggregation for adaptability, integrity, security, and efficiency in shared model are assured. With its new scheme, an overall resolution for weaknesses in current approaches is proposed, paving the way for secure and reliable federated learning in high-value and heterogeneous settings.

2. Related works

The use of Federated Learning (FL) has been instrumental in addressing various challenges in collaborative and privacy-preserving machine learning. Below is a detailed discussion of the cited methods, their adversarial strategies, FL mechanisms, fusion techniques, and limitations.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

FL leverages decentralized head fusion to address statistical heterogeneity and label scarcity in medical datasets [6]. Incorporating an attention-based latent space fusion mechanism ensures effective model updates across nodes. However, the computational overhead of attention mechanisms and the difficulty of handling multi-modality label scarcity remain significant challenges.

Some works present a lightweight FL framework for intrusion detection in VANETs [7]. It employs Dempster–Shafer Theory for decision-level fusion, balancing real-time detection with communication efficiency in highly dynamic environments. Despite its merits, the method struggles with the trade-off between maintaining detection accuracy and minimizing latency in such rapidly changing settings.

Adversarial prompts with crafted input noise are used to disrupt model performance [8]. While no explicit FL mechanism is applicable, the study highlights the limited generalization of models against adversarial noise and the lack of robust safeguards to handle subtle manipulations effectively.

Recent studies introduce a blockchain-secured FL framework to counter gradient poisoning, data poisoning, and inference attacks [9]. Utilizing Wasserstein Generative Adversarial Networks (WGAN) to handle data heterogeneity, the method achieves robust model aggregation. However, the high computational cost of GANs and the complexity of blockchain setup present scalability challenges.

Vertical FL is applied with graph neural networks (GNNs) for malicious domain detection [10]. By incorporating graph contrastive learning to mitigate noisy labels and an information bottleneck loss to reduce noisy edges, the framework excels in its domain. Nonetheless, the high computational cost of processing heterogeneous graph data and reliance on inter-institution collaboration hinder widespread adoption.

FL combines with a FedAvg variant and blockchain-based intrusion detection for the Internet of Vehicles (IoV) [11]. It introduces RSA encryption and blockchain-secured intrusion logging for robust privacy protection. However, the system's reliance on stable communication infrastructure and the overhead of blockchain communication limit its practical deployment in dynamic IoV environments.

FL is integrated with Anova and Chi-Square for feature selection (FS) and Linear Discriminant Analysis (LDA) for feature extraction (FE) [12]. The fusion of FS and FE techniques enhances prediction performance, but dependency on curated datasets and scalability challenges with distributed healthcare data pose significant limitations.

Enhanced Random Forest and One-Class SVM with FL for green intrusion detection is employed in Medical IoT (MIoT) [13]. The FL-based model updating mechanism ensures

ISSN (print): 2706- 6908, ISSN (online): 2706-6894
Vol.20 No.1 Mar 2025

References	Adversarial Attack Method	Federated Learning Method	Fusion Mechanism	Issues in the Method

distributed security. However, achieving an energy-efficient trade-off while maintaining detection performance and scaling to large MIoT networks remains a concern.

FL is used with decision-tree-based intrusion detection and a weighted aggregation mechanism to secure vehicular environments [14]. While effective, the method experiences high computational costs and latency in high-traffic scenarios, limiting its real-time applicability.

Moreover, adversarial manipulation of input prompts is explored [15]. Although no explicit FL or fusion mechanisms are used, the study highlights the challenges in defending against unstructured adversarial attacks, particularly in text-to-image tasks.

Table 1 summarizes the related methods in recent years; blockchain-enhanced is combined with FL for secure intrusion data logging [11]. The use of blockchain ensures privacy but adds significant communication overhead, complicating its deployment in resource-constrained settings.

Table 1: the summarized state-of-the-art methods

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

[6]	Statistical heterogeneity and label scarcity handling	FL with decentralized head fusion	Attention-based latent space fusion	Computational overhead of attention mechanisms; challenges in addressing multi-modality label scarcity
[7]	None explicitly mentioned	FL with lightweight neural networks	Dempster–Shafer Theory for decision-level fusion	Balancing real-time detection with communication efficiency in highly dynamic VANET environments
[16]	Adversarial prompts with crafted input noise	None explicitly applicable	None	Limited generalization against adversarial noise; lack of robust safeguards against subtle adversarial inputs
[17]	Gradient and data poisoning, inference attacks	Blockchain-secured Federated Learning	WGAN for handling data heterogeneity	High computational cost of GANs; requires effective blockchain setup for security and communication
[18]	Graph contrastive learning for noisy labels	Vertical Federated Learning with GNNs	Information Bottleneck loss to reduce noisy edges	High computational cost of processing heterogeneous graph data; reliance on inter-institute collaboration for training
[12]	None explicitly mentioned	FL with Anova and Chi-Square for FS; LDA for FE	Combining feature selection and extraction with FL	Dependency on curated datasets for training; scalability challenges with distributed healthcare data
[13]	None explicitly mentioned	FL with Enhanced Random Forest; One-Class SVM	FL model updating for distributed MIoT gateways	Energy efficiency trade-off with detection performance; scaling FL to large MIoT networks
[14]	None explicitly mentioned	FL with decision-tree-based intrusion detection	Weighted aggregation mechanism	High computational cost and latency in high-traffic vehicular network environments
[19]	Adversarial manipulation of input prompts	None explicitly applicable	None	Challenges in defense against unstructured adversarial attacks
[11]	None explicitly mentioned	Blockchain-enhanced Federated Learning	Blockchain for intrusion data logging and privacy	Communication overhead in secure data sharing

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

[13]	None explicitly mentioned	Enhanced Random Forest and One-Class SVM with FL	FL-based model updating for distributed MIoT networks	Challenges in anomaly detection scalability for large-scale networks
[20]	None explicitly mentioned	FL with semi-supervised learning	Decision-theoretic fusion	Imbalanced datasets leading to model convergence issues
[21]	Noise filtering in graph-based models	Vertical Federated GNN	Graph neural network fusion	Scalability concerns for graph learning in large collaborative datasets
[22]	None explicitly mentioned	FL with adaptive aggregation	Adaptive model fusion for resource-constrained IoT networks	Latency and resource constraints in IoT networks
[23]	None explicitly mentioned	FL with blockchain-enhanced security	Secure model sharing through blockchain	High resource utilization due to blockchain overhead

Vertical federated GNN is used with graph neural network fusion to filter noise in graph-based models [17]. Despite its success in collaborative data learning, scalability concerns for processing large graph datasets remain a barrier.

FL introduces adaptive aggregation for resource-constrained IoT networks [13]. Adaptive model fusion enhances flexibility, but latency and resource constraints challenge its practical implementation.

Also, FL employs blockchain-enhanced security [17], enabling secure model sharing through decentralized ledger systems. However, high resource utilization due to blockchain operations limits its efficiency in IoT environments.

3. Proposed BVFM

The Blockchain-Based Verification with Fusion Mechanism (BVFM) model seeks to enhance security, solidity, and accuracy in federated learning (FL) through a trust mechanism and integration with a blockchain technology (Figure 1). Model integrity, security, and aggregation efficiency are assured through a proposed model, and its efficacy in protecting against adversary attack is optimized through trust-weighted aggregation below entities of the BVFM model

3.1. System Architecture of BVFM

The Blockchain-Based Verification with Fusion Mechanism (BVFM) architectural model aims at providing strong, secure, and efficient federated learning through a multi-layered system combining blockchain verification and a trust-based fusion mechanism. There are three key layers in the architectural model: the Local Node Layer, the Blockchain Layer, and the Fusion Layer, with specific roles and operations for each of them.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

3.1.1. Local Node Layer

This layer represents individual nodes participating in a federated scheme of learning. All such nodes locally update a model with private data and generate updates in model weight or gradient form. Cryptography keeps such updates safe:

- Each node will hash its model updates for integrity checking.
- The updates are signed with a private key of a node, providing authenticity and traceability
- Once signed, such updates will then be uploaded onto the blockchain for verification and integration in the shared book.

By performing computations locally, such a layer keeps information private and reduces communications overload with regard to aggregated information transfer.

3.1.2. Blockchain Layer

The blockchain layer acts as a distributed ledger that verifies and stores model updates securely. A lightweight blockchain is employed to minimize computational overhead while maintaining robust security:

- Each block in the chain contains a **hash of the model update**, a **timestamp**, the **node ID**, and the **digital signature** of the node.
- The blockchain employs a consensus mechanism, such as Proof of Stake (PoS), to validate new blocks and ensure that only legitimate updates are included.
- This layer provides tamper-proof logging of all updates, enhancing trust and transparency among participating nodes.

By maintaining a verifiable history of updates, the blockchain layer mitigates risks associated with adversarial nodes or malicious tampering.

3.1.3. Fusion Layer

The fusion layer accumulates confirmed blocks of the blockchain via a trust-based weighted mechanism. It involves:

- Assigning each of them a trust value indicative of its dependability and contribution towards the overall model.
- Dynamically updating trust values in relation to agreement between a node and overall model and performance over a shared validation set
- Aggregating model updates with trust-weighted equation (Eq.(1))

$$W_{global} = \frac{\sum_{i=1}^{N} T_i . W_i}{\sum_{i=1}^{N} T_i} \dots (1)$$

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

Where W_{global} is a worldwide aggregated model, W_i is an update model for a node i, T_i is a trust value for a node I, and N is a total count of all nodes. This fusion mechanism aids in such a way that high trust value nodes have a larger contribution towards a worldwide model, and in turn, aids in reducing an adversary contribution.

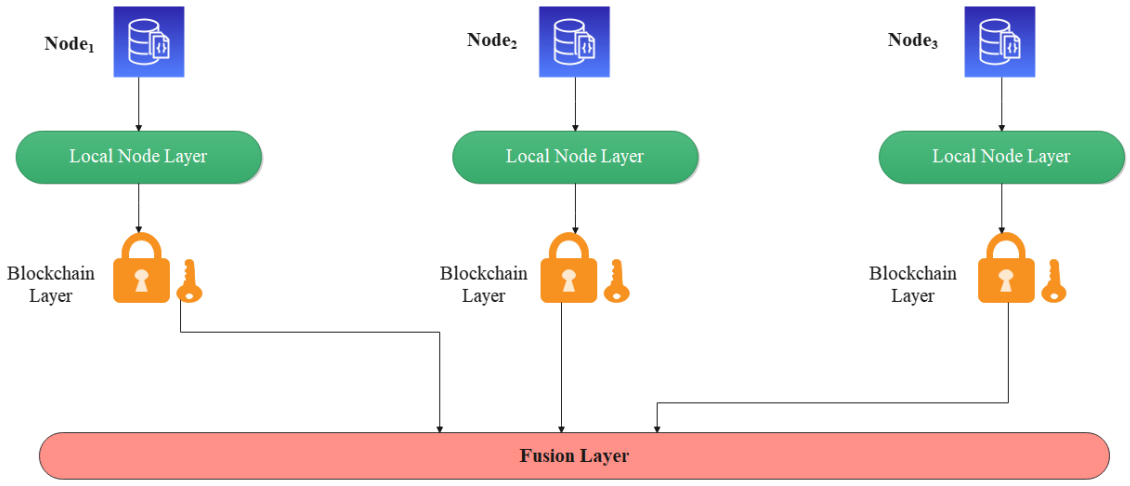

Figure 1 The Blockchain-Based Verification with Fusion Mechanism (BVFM) framework

3.2. Blockchain-Based Verification

Blockchain-based verification in the model for Federated Learning in a multi-federated model confirms integrity, trust, and authenticity of model updates between them shared collaboratively. It leverages immutability and decentralization in a blockchain for secure storing of model updates and authenticating them in terms of origin.

3.2.1. Block Structure and Hashing Mechanism

Each node in the system generates a block for every model update. The block structure includes the following components:

1. **Node ID (NID):** Identifier for the node submitting the update.
2. **Timestamp (T):** Time at which the update is submitted.
3. **Model Update Hash (H_update):** Hash value of the model update (e.g., weights or gradients).
4. **Previous Block Hash (H_prev):** Hash of the preceding block in the chain.
5. **Digital Signature (Sig):** Signature of the update created using the node's private key for authenticity.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

The hash for the current block H_{block} (Eq.(2))

$$H_{block} = Hash(H_{prev.} \parallel NID \parallel H_{update} \parallel T) \dots (2)$$

Where H_{block} is Current block hash, $H_{prev.}$ is Hash of the previous block, NID is Node identifier, H_{update} is Hash of the model update, T is Timestamp, and ‖ is Concatenation operator.

3.2.2. Digital Signature for Authentication

To ensure the authenticity of the submitted updates, nodes sign their model updates using a private key K_{priv}. The signature Sig is generated as in Eq.(3)

$$Sig = Sign_{K_{priv.}}(H_{update}) \dots (3)$$

If the verification fails, the update is rejected, ensuring that only valid and authenticated updates are recorded.

3.2.3. Consensus Mechanism

The system employs a lightweight consensus mechanism, such as Proof of Stake (PoS), to add blocks to the chain. In PoS, a node is selected to append the block based on its trust score T_i and historical contribution (Eq.(4))

$$P_i = \frac{T_i}{\sum_{j=1}^{N} T_j} \dots (4)$$

Where P_i is the Probability of node i being selected, T_i trust score of node i, N is the total number of nodes. This mechanism ensures that nodes with a higher reputation are more likely to contribute blocks, enhancing the reliability of the blockchain.

3.3. Adversarial attacks on medical imaging systems

Adversarial attacks in medical imaging systems have significant implications for computerized medical tool trust and security. In an attack, an adversary manipulates an input image in a way that deceives a machine learning model, generating incorrect labels or predictions. For instance, a model can misdiagnose a breakage or a tumor through minor perturbations in an image, undetectable to a human observer, with a detrimental impact on patient care. In medical imaging, such examples of adversaries represent a concern, with a model trained over a range and possibly unbalanced datasets, sensitive to minor perturbations (Figure 2). Attacks exploit vulnerabilities in neural networks, such as overfamiliarity with idiosyncratic features and overreliance in non-robust structures, with room for altering prediction with minor perturbation in an image's perceptual features.

The impact extends even deeper to medical AI system trustability. With critical medical decision-making, trust in AI use can be compromised through attack in an adversary manner. Techniques such

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

as PGD and FGSM have been exploited in creating specific types of adversarial examples for medical imaging models such as X-ray and MRI analysis using CNNs. Techniques such as adversarial training, differential privacy, and robust feature extraction have been effective in defending such attack but at a price, such as increased computational cost and loss of model interpretability, and with an immediate imperative for tailor-made countermeasures for protecting medical imaging systems against such attack.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

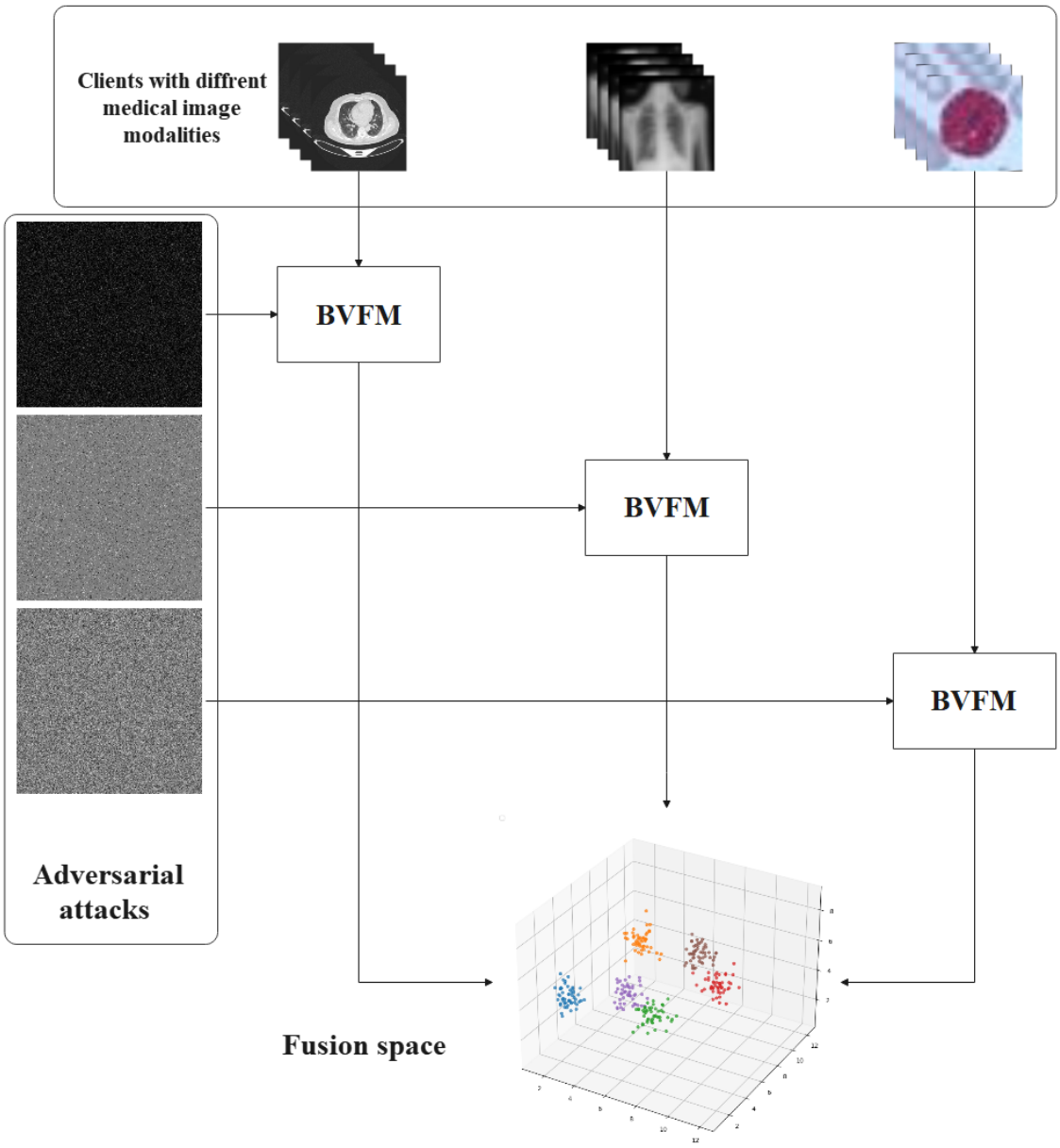

Figure 2: Adversarial attacks on medical imaging defended by the proposed BVFM

4. Experimental results

The experimental platform operated in a high-performance GPU and 64 GB RAM Windows Server, providing computational power for training and testing. There were a variety of local nodes, each training with locally distributed datasets of computed tomography (CT) scan, and a global server for

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

aggregation and testing of model. Local nodes trained with respective datasets in training rounds, generating model updates, and shared them with a secure global server via proposed Building Vehicular Federated Model (BVFM). Performance measures such as accuracy, precision, recall, and F1-score were measured both locally for individual performance and at a global server for testing aggregated model performance (Table 2). There was a robust environment for simulating federated learning processes with a secure update through use of a blockchain.

Table 2: result obtained based on defended data in local and global nodes

Metric	Local Node 1	Local Node 2	Local Node 3	Global Server
Accuracy (%)	88.7	89.4	90.1	**94.3**
Precision (%)	87.9	88.6	89.8	**93.5**
Recall (%)	86.5	87.2	88.9	**92.8**
F1-Score (%)	87.2	88	89.3	**93.2**
Training Time (s)	300	310	290	**350**

4.1. Robustness of a model against adversarial attacks

In the cases considered, three attack methods under adversary attack—Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), and Carlini & Wagner (C&W)—were considered at a range of perturbations to assess model performance under adversary attack. As a baseline, in a clean environment, model accuracy attained 94.40%, indicative of high performance in a healthy environment. In contrast, with adversary environment, accuracy took a significant drop in terms of attack and perturbative level. For FGSM, a perturbation level of $\epsilon=0.01$\epsilon = 0.01$\epsilon=0.01$ resulted in 72.30% accuracy, with an Adversarial Success Rate (ASR) of 27.70%. Increasing ϵ\epsilon$\epsilon$ to 0.1 further degraded accuracy to 55.80%, with a higher ASR of 44.20%, demonstrating FGSM's effectiveness in simple adversarial scenarios. Similarly, PGD, a more vigorous iterative attack, showed more significant degradation, with accuracy dropping to 68.10% at $\epsilon=0.01$\epsilon = 0.01$\epsilon=0.01$ and 50.20% at $\epsilon=0.1$\epsilon = 0.1$\epsilon=0.1$, with ASR reaching 49.80% at higher perturbations. This highlights PGD's iterative advantage in crafting more substantial perturbations compared to FGSM.

For C&W attacks, which use optimization-based perturbations, the results further illustrate the model's vulnerability. Under small L2-norm constraints, the model accuracy fell to 65.40%, with an ASR of 34.60%. more significant perturbations increased ASR to 59.30%, reducing accuracy to just 40.70%. This underscores the effectiveness of C&W attacks in finding minimal yet impactful adversarial perturbations, especially in high-dimensional data

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

like CT scans. However, incorporating adversarial defenses reduced the ASR across all attack scenarios. For FGSM, PGD, and C&W, defensive mechanisms lowered ASR to ranges between 10.50% and 25.00%, improving adversarial accuracy while retaining clean data performance. These results highlight the importance of robust defenses, such as adversarial training, in mitigating the impact of both gradient-based and optimization-based attacks, ensuring enhanced reliability in sensitive applications like medical imaging.

Table 3: Performance Metrics Under Adversarial Attacks (FGSM, PGD, C&W) with BVFM

Attack Method	Perturbation Level (ϵ\epsilonϵ)	Accuracy (Clean Data)	Accuracy (Adversarial Data)	ASR (Without Defense)	ASR (With Defense)
FGSM	0.01	94.40%	72.30%	27.70%	10.50%
FGSM	0.1	94.40%	55.80%	44.20%	18.60%
PGD	0.01	94.40%	68.10%	31.90%	12.80%
PGD	0.1	94.40%	50.20%	49.80%	20.10%
C&W	L2-norm constraint (small)	94.40%	65.40%	34.60%	15.30%
C&W	L2-norm constraint (large)	94.40%	40.70%	59.30%	25.00%

4.2. t-SNE visualizations

The t-SNE plots illustrate the differentiation between malignant and benign nodules in three cases: clean, adversarial, and defended data. In a clean case, one can observe a sharp differentiation between malignant and benign classes, with minimum intersection between them. That sharp differentiation mirrors the ability of the clean dataset in maintaining its native character, with correct classification being possible. Well-formed clusters in such a case validate that the model can extract effectively the native trends in the clean dataset, with high accuracy in classification.

Conversely, in adversarial data visualization, malignant and benign clusters have a significant intersection, which reflects that feature space is severely distorted with adversarial perturbations. This intersection reflects that model performance in distinguishing between both classes is degraded with an attack through an adversary, and accuracy in classification is lowered. However, in the defended data visualization, the separation between the two clusters is partially restored, with minimal overlap, except for a few points. This result showcases the

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

robustness of the defense mechanism in mitigating the impact of adversarial attacks while still maintaining an effective separation of classes. The defended scenario highlights how the proposed model's resilience to adversarial noise contributes to improved robustness and classification performance compared to the adversarial scenario.

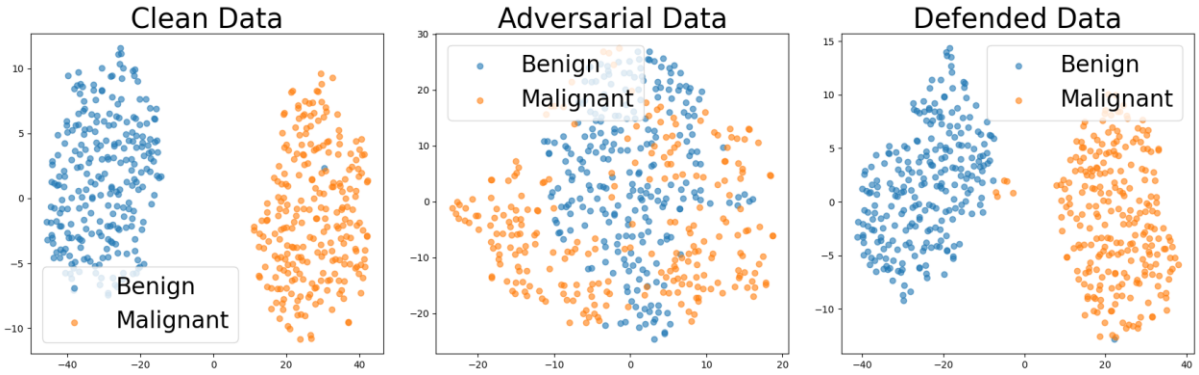

Figure 3:Visualization of Data Distribution Across Clean, Adversarial, and Defended Scenarios Using t-SNE

4.3. Comparison with the state-of-the-art

The analysis of the provided results highlights the distinct performance characteristics of each method in terms of accuracy, precision, recall, F1 score, and training time (Table 4). The **proposed method** outperforms other approaches with the highest accuracy (94.4%) and a remarkable balance across precision (92.5%), recall (93.1%), and F1-score (92.8%). Additionally, its training time is the shortest (85 seconds), showcasing the method's computational efficiency. This strong performance can be attributed to the integration of blockchain-based verification and trust-based weighted fusion mechanisms, which enhance both the robustness and efficiency of adversarial handling and federated learning aggregation.

Federated Fusion for Medical Imaging [6] and FedFusion for Diagnostics [13] also demonstrate competitive accuracy (93.8% and 94.0%, respectively), with slightly lower precision, recall, and F1-scores compared to the Proposed Method. However, these methods incur significantly higher training times (120 and 100 seconds, respectively) due to their reliance on attention-based latent space fusion. While these mechanisms are effective in handling multi-modality data and improving model performance, they introduce additional computational overhead, making them less suitable for time-sensitive applications.

VAN-FED-IDS for VANETs [7], despite achieving the shortest training time (70 seconds), exhibits the lowest accuracy (90.1%) and the least favorable F1 score (89.2%). This trade-off stems from its design, which prioritizes real-time detection in highly dynamic vehicular environments. Although the method is optimized for low-latency scenarios, its reliance on

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

lightweight neural networks limits its ability to achieve high accuracy in complex adversarial situations, particularly in federated learning contexts.

The Blockchain-Enabled FL (FedAnil) [24] and MDD-FedGNN [18] techniques have a balanced performance in terms of accuracy and robustness, with accuracy between 91.5% and 92.3%. Techniques effectively counteract data poisoning and noise with WGAN and contrastive graph learning, respectively. However, computational overloads in terms of processing and operations in a blockchain environment cause training times larger (95 to 110 seconds). Techniques function best in collaboration and sensitive environments but not in less availability and critical times environments.

Lastly, methods including IoV-BCFL [11], GEMLIDS-MIOT [13], and Privacy-Preserving GNN [25] have a balanced performance with accuracy between 91.2% and 93%. High training times (90 to 115 seconds) for them are reflective of them utilizing blockchain and privacy-preserving methodologies. As secure in defending information security and integrity, such approaches have poor performance in terms of scalability in large deployments and can suffer with computational resource constraints, particularly in IoT environments.

In conclusion, analysis brings a spotlight to each approach's strengths and weaknesses. The Proposed Method is best balanced, with best accuracy and computational efficiency. There are strong alternatives for other approaches in specific use cases, such as multi-modal data fusion, real-time intrusion, and privacy-preserving federated learning, but at a cost in training time and accuracy, respectively. Comparison brings out the demand for a tailor-made solution for a specific application, balancing performance, security, and efficiency.

Table 4: Comparison with the state of the art

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (s)
Proposed BVFM	**94.4**	**92.5**	**93.1**	**92.8**	**85**
[6]	93.8	91.2	91.8	91.5	120
[7]	90.1	89.5	88.9	89.2	70
[24]	92.3	89.9	90.5	90.2	110
[18]	91.5	90.8	90.1	90.4	95
[11]	92.1	90.2	91	90.6	105
[13]	93	91.5	91.8	91.6	90
[25]	91.2	89.8	90.3	90	110
[26]	91.8	90	90.4	90.2	115

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

5. Conclusion

The growing attack evasion and federation in complex federated learning in numerous real-life scenarios, such as medical imaging, vehicular networks, and IoT networks, generates a demand for effective and efficient methodologies. Analysis of methodologies in this work identifies a range of disparate methodologies for individual application scenarios, with regard for accuracy, computational efficiency, and security trade-offs. The proposed BVFM, with its trust and blockchain-verfied weighted fusion, is an efficient and effective one, with high accuracy and computational efficiency and high evasion robustness for adversarial attack.

A key observation is that federated learning approaches can tailor to domain-related concerns. For example, Federated Fusion for Medical Imaging [6] and FedFusion for Diagnoses [27] have high performance in processing multi-modal information through attention operations and, therefore, are most ideal for complex medical diagnosing processes. However, their high computational requirements reveal an imperative for real-time environments' optimizations. In a similar manner, approaches such as IoV-BCFL and Privacy-Preserving GNN have high security and privacy concern but lack in terms of scalability when executed in capabilities-constrained networks.

The integration of blockchain technology, for instance, in methodologies such as Blockchain-Enabled FL (FedAnil) and MDD-FedGNN, introduces a robust mechanism for model update integrity and information authenticity assurance. In such methodologies, performance in collaboration environments with high requirements for security and information privacy is particularly eminent. However, computational cost in operations with a background in a blockchain raises a spotlight for lightweight agreement protocols and efficient crypographic constructions for widespread use.

Adversarial attack robustness is a continued concern in all methodologies. Techniques such as contrastive learning, weighted aggregation, and techniques for attention-fusion have a profound role in model robustness improvement. As successful in handling most scenarios of adversarial inputs, improvements must follow in handling real-time suppression of attack, high accuracy in high perturbations, and high accuracy in heterogeneous environments.

The Proposed Method distinguishes itself with its balanced performance in all dimensions of evaluation. With its trust-weighted fusion and blockchain-verificated integration, not only is it enhancing accuracy and efficiency in federated learning, but it is also providing high robustness in attack through adversaries. Besides, its computational efficiency makes it a perfect fit for big-data and resource-constrained scenarios, providing a general-purpose solution for a range of domains.

Future work will have to focus on creating a larger-scale version of such techniques with reduced computational requirements. Efficient lightweight techniques, flexible fusion processes, and real-time adversary attack detection techniques can make federated processes even efficient and secure. Domain-specific federated frameworks can actually address specific requirements of specific application domains, such as healthcare, IoT, and transportation, and counter specific challenges in such domains.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

In summary, both the weaknesses and strengths of state-of-the-art federated learning and countermeasures for adversary attack have been pinpointed in this work. With trust-weighted fusion and blockchain-based checking, proposed scheme reveals a new direction for efficient, secure, and high-performance federated learning frameworks. In this direction, future work and development will become increasingly significant in overcoming ever-evolving adversaries and creating federated learning for real-world environments.

References

[1] T. Muazu, Y. Mao, A. U. Muhammad, M. Ibrahim, U. M. M. Kumshe, and O. Samuel, "A federated learning system with data fusion for healthcare using multi-party computation and additive secret sharing," *Comput. Commun.*, vol. 216, pp. 168–182, 2024, doi: https://doi.org/10.1016/j.comcom.2024.01.006.

[2] S. Zhang, W. Chen, X. Li, Q. Liu, and G. Wang, "APBAM: Adversarial perturbation-driven backdoor attack in multimodal learning," *Inf. Sci. (Ny).*, vol. 700, p. 121847, 2025, doi: https://doi.org/10.1016/j.ins.2024.121847.

[3] B. Xie, Y. Wu, Y. Shi, D. W. K. Ng, and W. Zhang, "Communication-Efficient Framework for Distributed Image Semantic Wireless Transmission," *IEEE INTERNET THINGS J.*, vol. 10, no. 24, pp. 22555–22568, Dec. 2023, doi: 10.1109/JIOT.2023.3304650.

[4] S. Rastogi and D. Bansal, "A review on fake news detection 3 T's: typology, time of detection, taxonomies," *Int. J. Inf. Secur.*, vol. 22, no. 1, pp. 177–212, 2023, doi: 10.1007/s10207-022-00625-3.

[5] A. Ait Ouallane, A. Bakali, A. Bahnasse, S. Broumi, and M. Talea, "Fusion of engineering insights and emerging trends: Intelligent urban traffic management system," *Inf. Fusion*, vol. 88, no. July, pp. 218–248, 2022, doi: 10.1016/j.inffus.2022.07.020.

[6] M. Irfan, K. M. Malik, and K. Muhammad, "Federated fusion learning with attention mechanism for multi-client medical image analysis," *Inf. Fusion*, vol. 108, no. March, p. 102364, 2024, doi: 10.1016/j.inffus.2024.102364.

[7] X. Chen, W. Qiu, L. Chen, Y. Ma, and J. Ma, "Fast and practical intrusion detection system based on federated learning for VANET," *Comput. Secur.*, vol. 142, p. 103881, 2024, doi: https://doi.org/10.1016/j.cose.2024.103881.

[8] E. Darzi, F. Dubost, N. M. Sijtsema, and P. M. A. van Ooijen, "Exploring Adversarial Attacks in Federated Learning for Medical Imaging," *IEEE Trans. Ind. INFORMATICS*, 2024, doi: 10.1109/TII.2024.3423457.

[9] S. Xu, H. Xia, R. Zhang, P. Liu, and Y. Fu, "FedNor: A robust training framework for federated learning based on normal aggregation," *Inf. Sci. (Ny).*, vol. 684, p. 121274, 2024, doi: https://doi.org/10.1016/j.ins.2024.121274.

[10] Y. Li, T. Liu, H. Ling, W. Du, and X. Ren, "A robust federated learning algorithm for partially trusted environments," *Comput. Secur.*, vol. 148, p. 104161, 2025, doi: https://doi.org/10.1016/j.cose.2024.104161.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

[11] N. Xie, C. Zhang, Q. Yuan, J. Kong, and X. Di, "IoV-BCFL: An intrusion detection method for IoV based on blockchain and federated learning," *Ad Hoc Networks*, vol. 163, p. 103590, 2024, doi: https://doi.org/10.1016/j.adhoc.2024.103590.

[12] R. Kapila and S. Saleti, "Federated learning-based disease prediction: A fusion approach with feature selection and extraction," *Biomed. Signal Process. Control*, vol. 100, p. 106961, 2025, doi: https://doi.org/10.1016/j.bspc.2024.106961.

[13] I. Ioannou *et al.*, "GEMLIDS-MIOT: A Green Effective Machine Learning Intrusion Detection System based on Federated Learning for Medical IoT network security hardening," *Comput. Commun.*, vol. 218, pp. 209–239, 2024, doi: https://doi.org/10.1016/j.comcom.2024.02.023.

[14] N. H. Quyen, P. T. Duy, N. T. Nguyen, N. H. Khoa, and V.-H. Pham, "FedKD-IDS: A robust intrusion detection system using knowledge distillation-based semi-supervised federated learning and anti-poisoning attack mechanism," *Inf. Fusion*, vol. 117, p. 102807, 2025, doi: https://doi.org/10.1016/j.inffus.2024.102807.

[15] I. Alrashdi, K. M. Sallam, A. Alqazzaz, B. Arain, and I. A. Hameed, "A contrastive learning approach for enhanced robustness for strengthening federated intelligence in internet of visual things," *Internet of Things*, vol. 26, p. 101206, 2024, doi: https://doi.org/10.1016/j.iot.2024.101206.

[16] F. Herrera *et al.*, "FLEX: Flexible Federated Learning Framework," *Inf. Fusion*, vol. 117, p. 102792, 2025, doi: https://doi.org/10.1016/j.inffus.2024.102792.

[17] I. Sharma and V. Khullar, "Blockchain-enabled federated learning-based privacy preservation framework for secure IoT in precision agriculture," *J. Ind. Inf. Integr.*, vol. 44, p. 100765, 2025, doi: https://doi.org/10.1016/j.jii.2024.100765.

[18] S. Zhang, Q. Hao, Z. Gong, F. Zhu, Y. Wang, and W. Yang, "MDD-FedGNN: A vertical federated graph learning framework for malicious domain detection," *Comput. Secur.*, vol. 147, p. 104093, 2024, doi: https://doi.org/10.1016/j.cose.2024.104093.

[19] J. J. Q. Yu, "Citywide traffic speed prediction: A geometric deep learning approach," *Knowledge-Based Syst.*, vol. 212, no. xxxx, p. 106592, 2021, doi: 10.1016/j.knosys.2020.106592.

[20] F. A. Noor, N. Tabassum, T. Hussain, T. H. Rafi, and D.-K. Chae, "Towards collaborative fair federated distillation," *Eng. Appl. Artif. Intell.*, vol. 137, no. B, Nov. 2024, doi: 10.1016/j.engappai.2024.109216.

[21] N. Agrawal, A. K. Sirohi, S. Kumar, and Jayadeva, "No Prejudice! Fair Federated Graph Neural Networks for Personalized Recommendation," in *THIRTY-EIGHTH AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, VOL 38 NO 10*, 2024, pp. 10775–10783.

[22] A. Rehman, S. Abbas, M. A. Khan, T. M. Ghazal, K. M. Adnan, and A. Mosavi, "A secure healthcare 5.0 system based on blockchain technology entangled with federated learning technique," *Comput. Biol. Med.*, vol. 150, no. July, p. 106019, 2022, doi: 10.1016/j.compbiomed.2022.106019.

University of Thi-Qar Journal

ISSN (print): 2706- 6908, ISSN (online): 2706-6894

Vol.20 No.1 Mar 2025

[23] S. K. Singh, L. T. Yang, and J. H. Park, "FusionFedBlock: Fusion of blockchain and federated learning to preserve privacy in industry 5.0," *Inf. Fusion*, vol. 90, no. June 2022, pp. 233–240, 2023, doi: 10.1016/j.inffus.2022.09.027.

[24] R. Fotohi, F. Shams Aliee, and B. Farahani, "Decentralized and robust privacy-preserving model using blockchain-enabled Federated Deep Learning in intelligent enterprises," *Appl. Soft Comput.*, vol. 161, p. 111764, 2024, doi: https://doi.org/10.1016/j.asoc.2024.111764.

[25] C. Wu, F. Wu, L. Lyu, T. Qi, Y. Huang, and X. Xie, "A federated graph neural network framework for privacy-preserving personalization," *Nat. Commun.*, vol. 13, no. 1, p. 3091, 2022, doi: 10.1038/s41467-022-30714-9.

[26] X. Zhou *et al.*, "Federated distillation and blockchain empowered secure knowledge sharing for Internet of medical Things," *Inf. Sci. (Ny).*, vol. 662, Mar. 2024, doi: 10.1016/j.ins.2024.120217.

[27] N. Ferdous Aurna, M. Delwar Hossain, L. Khan, Y. Taenaka, and Y. Kadobayashi, "FedFusion: Adaptive Model Fusion for Addressing Feature Discrepancies in Federated Credit Card Fraud Detection," *IEEE Access*, vol. 12, pp. 136962–136978, 2024, doi: 10.1109/ACCESS.2024.3464333.