Journal of Electronic Research and Application, 2025, Volume 9, Issue 6
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

# A Hybrid Encryption Medical Information Security Collaboration Scheme Based on Blockchain and Federated Learning

**Jia Sun**[1, 2], **Peng Zhao**[1, 2,*]

[1] School of Computer Science and Technology, Taiyuan Normal University, Jinzhong, Shanxi 030619, China
[2] Shanxi Key Laboratory of Intelligent Optimization Computation and Blockchain Technology, Taiyuan Normal University, Jinzhong, Shanxi 030619, China

***Author to whom correspondence should be addressed.**

**Abstract:** In current medical data sharing practices, the tension between data privacy protection and cross-institutional collaboration efficiency has become increasingly prominent. To address existing security challenges in healthcare data sharing, we propose a collaborative data cooperation model based on blockchain and federated learning, Through federated learning technology, data is made "usable but not visible" by enabling medical institutions to share only encrypted model parameters, thereby preventing the leakage of raw data. Meanwhile, blockchain technology is introduced to establish a decentralized trust mechanism, utilizing smart contracts to automate data access management and track training processes. In addition, the dual security protection strategy is designed, where differential privacy and Paillier homomorphic encryption technology are adopted to resist member reasoning attacks and ensure secure storage and sharing of information. Through security analysis and experimental validation, the scheme has been proven to have good security and usability.

**Keywords:** Data sharing; Differential privacy; Paillier; Consortium blockchain; Federal learning

**Online publication:** December 16, 2025

## 1. Introduction

In the digital transformation of healthcare, data sharing serves as a pivotal driver for enhancing diagnostic accuracy and accelerating drug development. While this data-sharing mechanism is crucial for elevating clinical standards and advancing medical research, it faces multiple practical challenges. Medical data contains highly sensitive patient information, making traditional centralized sharing models prone to data breaches and failing to meet data protection regulations like GDPR and HIPAA. The "data silos" phenomenon between healthcare institutions, stemming from inadequate trust mechanisms and collaborative frameworks, results in inefficient cross-institutional data exchange. Furthermore, reliance on third-party intermediaries in current sharing solutions not only increases administrative costs but may also lead to disputes over data ownership and benefit distribution. These factors severely hinder the full utilization of medical data's potential and collaborative applications.

In medical data sharing scenarios, blockchain technology addresses data credibility and privacy traceability through distributed ledger technology and smart contracts [1]. However, the inherent conflict persists between the requirement for original data to remain within its domain and the need for cross-institutional collaborative modeling. Federated learning, a distributed machine learning framework, enables collaborative model training among participants (such as hospitals) using local data, with only gradients or weight parameters being uploaded [2]. This approach effectively prevents privacy breaches caused by direct data transmission while fulfilling the dual requirements of data domain retention and cross-institutional collaboration. Kim *et al.* proposed integrating device-side federated learning with blockchain by storing locally computed gradients in blocks after node verification, achieving traceability throughout the training process [3]. However, this method fails to resolve privacy leakage issues during parameter aggregation in federated learning.

Majeed *et al.* developed a secure federated learning platform on Ethereum, though their single-key homomorphic encryption technique shows limitations in protecting data model transfers [4]. Zhang *et al.* created a blockchain-based federated learning framework for healthcare, incorporating differential privacy noise to safeguard patient data [5]. However, excessive noise addition reduces global model accuracy and leaves the system vulnerable to backdoor attacks. Qu *et al.* designed a hybrid identity mechanism combining digital signatures and encryption protocols to prevent attackers from stealing blockchain-stored data [6]. However, when the number of iterations of federated learning is large, this mechanism will bring huge computational overhead.

Therefore, a single privacy-enhancing technique struggles to achieve optimal balance among "data privacy", "model utility", and "computational security". To address this, this paper proposes a medical information security sharing scheme that integrates differential privacy with Paillier homomorphic encryption, built on a federated learning and blockchain-based infrastructure. The scheme aims to establish an end-to-end, multi-layered security framework, where Paillier homomorphic encryption first encrypts gradients during data transmission and aggregation phases to ensure confidentiality during computations. Then, after decrypting aggregated results, noise compliant with differential privacy requirements is applied to defend against inference attacks on the final model. Finally, the blockchain network records the hash evidence of the encryption gradient and aggregation operation to ensure the traceability and tamper resistance of the process.

## 2. Background knowledge

### 2.1. Blockchain

Blockchain is a decentralized distributed ledger technology that ensures data immutability, transparency, and traceability through cryptographic encryption and consensus mechanisms [7]. It organizes data into blocks, each containing transaction details or other information, which are linked via cryptographic algorithms to form an ever-expanding chain. The chain structure and node verification mechanism enable secure transactions and information sharing without requiring trust in intermediaries.

### 2.2. Federal learning

Federated learning is a distributed machine learning framework [8]. The basic workflow involves: participating clients train models on local datasets and upload updated parameters to a central server; the server aggregates updates from multiple clients to optimize the global model; the updated model parameters are then distributed back to the clients, forming an iterative cycle. In a federated learning system with K terminals, each terminal holds a local dataset $DB_i(1 \leq i \leq K)$. The central server's loss function is defined as the weighted average of each terminal's local loss functions, specifically:

$$f(w) = \sum_{i=1}^{K} \frac{n_i}{n} F_i(w) \tag{1}$$

Here, $n = \sum_{i=1}^{K} n_i$ represents the total number of data samples from all terminals. $F_i(w) = \frac{1}{n_i} \sum_{j \in DB_i} f_j(w)$ denotes the local loss function of the i-th terminal, $f_j(w) = l(x_j,y_j,w)$ represents the loss incurred by the model $w$ on the sample $(x_j,y_j)$. The system is usually optimized by random gradient descent, and the training process is terminated after reaching the preset number of iterations or model accuracy.

## 2.3. Paillier homomorphic encryption

The Paillier encryption algorithm is a form of homomorphic encryption. Its homomorphic property allows direct addition operations on two encrypted ciphertexts without decryption [9]. The algorithm's workflow is as follows:

(1) KeyGen Phase

(i) The key generation center randomly selects two large prime numbers p and q, each with k bits., ensuring $gcd(pq, (p\text{-}1)(q\text{-}1))$. Calculate simultaneously n=pxq, send $\lambda = Icm(p\text{-}1,q\text{-}1)$;

(ii) Send ,defined function $L(x)=(x\text{-}1)/n$ ;

(iii) Calculate private key $u=(L(g^{\lambda} modn^2))^{-1}\ mod\ n$;

(iv) Public key is $(n,g)$, Private key is $(\lambda,u)$;

(2) Encry Phase

(i) The encryptor randomly selects a random number $r$, Satisfy $0 < r < n$ and $r\epsilon Z_n^2$, $r$ and $n$ are coprime;

(ii) The encryptor encrypts the message m to be transmitted using the public key and outputs the ciphertext. $C$:$c=(g^m r^n) mod\ n^2$.

(3) Decrypt Phase: Enter the ciphertext C ,Use private key $(\lambda,u)$, Decrypt the ciphertext and output the message $m$:$m=L(C^{\lambda} mod\ n^2)\cdot\mu$.

The Paillier encryption algorithm satisfies additive homomorphism, that is: $DEC(C_1 \cdot C_2)= m^1+m^2$.

## 2.4. Differential privacy

Differential privacy is a cryptographic technique designed to ensure both the accuracy of statistical database queries and the maximum reduction of the probability of inferring specific records [10]. Its core principle stems from a rigorous mathematical definition: For two adjacent datasets D and D' differing by at most one record, if a privacy-preserving algorithm A satisfies the following inequality across all possible output ranges S:

$$P[A(D) \in S] \leq e^{\epsilon} P[A(D^{'}) \in S] + \delta \tag{2}$$

Algorithm A is said to satisfy $(\varepsilon,\delta)$-differential privacy. The probability in this inequality stems from the inherent randomness of the algorithm, directly reflecting the risk of privacy leakage. Here, $\varepsilon$ represents the privacy budget, the smaller its value, the stronger the privacy protection, but at the cost of reduced data utility. The parameter $\delta$ indicates the probability constraint that the algorithm violates strict $\varepsilon$-differential privacy, which is typically set to a minimal value.

For a query function $f: D \rightarrow R^d$, the maximum $L_p$ norm variation across all adjacent datasets $D$ and $D'$:

$$\Delta f = \max_{D,D'} \left\| f(D) - f(D') \right\|_p \qquad (3)$$

The maximum sensitivity defines the maximum impact a single record can have on the output, and thus directly determines the lower limit of the noise required to implement differential privacy. The lower the sensitivity, the less noise is required.

The Gaussian mechanism achieves *(ε, δ)*-differential privacy by adding Gaussian noise to the output of the function *f*, where the noise scale is determined by the sensitivity of the function *f*, as shown in Equation (4).

$$A(D) = f(D) + N\left(0, (\Delta f \sigma)^2 I\right) \qquad (4)$$

## 3. Related technical architecture

### 3.1. Overall architecture design of the solution

The proposed decentralized secure medical collaboration system model consists of four core components: task publishers, verification committees, medical institution participants, and a secure key management center. As shown in **Figure 1**, the framework outlines the responsibilities of each component as follows:

(1) Task publisher: Responsible for initializing federated learning tasks, including model architecture definition, training hyperparameter settings, and task scheduling. It stores the intermediate state of the global model and coordinates training rounds across medical institutions. The Task publisher does not directly participate in data aggregation, only providing task publishing and progress management functions;

(2) Verification committee: Composed of multiple decentralized consensus nodes, this committee is responsible for validating the model gradients or parameters uploaded by medical institutions. It achieves trusted aggregation of model updates through smart contracts and stores the final aggregated results on the blockchain for certification. This mechanism ensures traceability, transparency, and immutability throughout the training process;

(3) Healthcare institution participant: As the data holder, the healthcare institution is responsible for training sub-models using local data. In each training round, it submits encrypted model updates to the blockchain validation network, retrieves the latest aggregated model to initiate the next local training round, and this process repeats until the model converges;

(4) Security key management center: As a trusted third party, it is responsible for generating, distributing, and managing the system's encryption keys. It generates unique public-private key pairs for each medical institution and distributes the private keys through secure channels. The public keys and system security parameters are recorded on the blockchain via smart contracts for global verification.

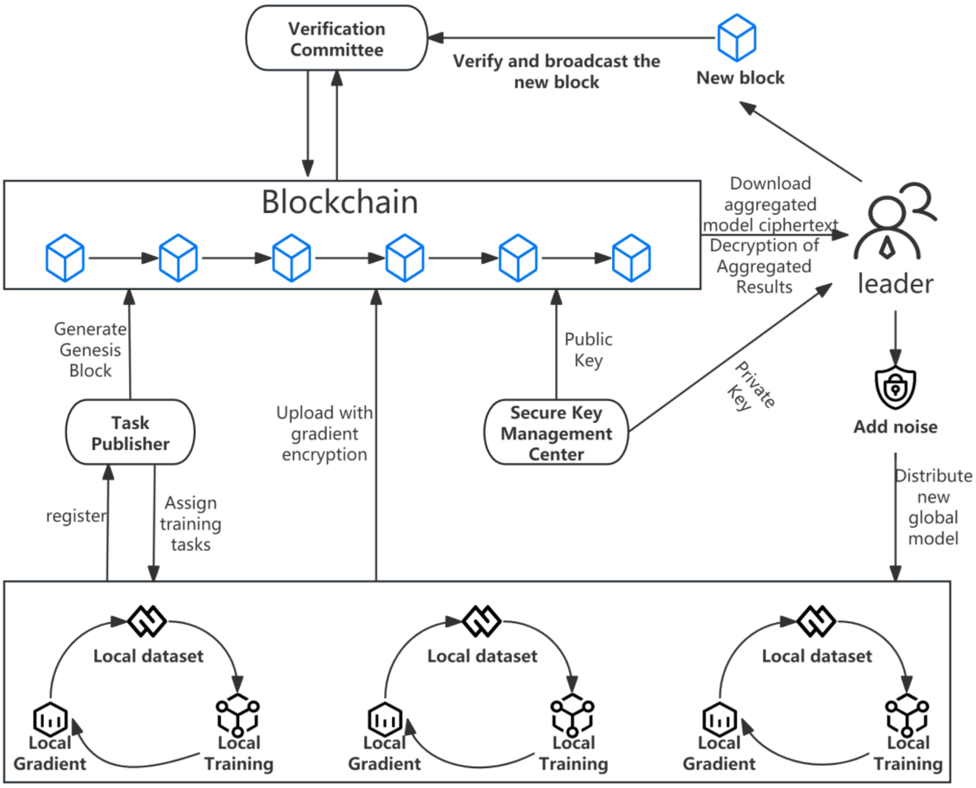

**Figure 1.** Solution architecture.

## 3.2. Implementation process

The system consists of *K* medical institutions (local devices) and multiple edge computing nodes (miners). These institutions (e.g., hospitals and clinics) serve as data holders, possessing local medical datasets $D_i$ (with sample sizes $n_i$). The consortium chain nodes are responsible for gradient verification and blockchain consensus. The process is as follows:

(1) Step 1: System initialization and registration

Medical institutions and miners apply for registration with the task publisher, submitting data set size $n_i$ and proof of computational capability. The Key management center assigns each entity as outlined:

(i) Signature key pair, $(PK_i^{sig},\ SK_i^{sig})$ (For transaction signing);

(ii) Homomorphic encryption key pair, $(PK_i^{enc},\ SK_i^{enc})$ (Paillier encryption);

The task publisher creates a genesis block based on the training task and distributes it to all nodes via a secure link (TLS+SGX) to initialize the model. The genesis block contains the following information:

(i) Initial model parameters $W_0$ and total training rounds T;

(ii) Public keys for all entities $\{PK_i^{sig},PK_j^{enc}\}$;

(iii) The dataset size $n_i$;

(iv) Differential privacy parameters (ϵ, δ) and gradient clipping threshold S.

(v) Random seed $seed_0$ (used to generate seedt for leader election in subsequent training rounds based on $seed_{t-1}$);

(vi) Initial token allocation (for incentives);

(2) Step 2: Local model training and noise addition

Medical institutions train machine learning models on their local medical datasets. Based on the sample size $n_i$ of the local dataset, $n_i$ iterations of training are performed. The gradient is clipped to a threshold $S$, and the clipped gradient is encrypted using the system public key $PK$ through the Paillier algorithm, enabling computation of encrypted data without decryption.

(3) Step 3: Data upload and homomorphic encryption

Medical institutions package the ciphertext gradient, local computation time (ensured by Intel SGX trusted hardware's time-of-use authentication mechanism), and digital signature into a single transaction. They then upload the encrypted data, computation time, and digital signature to the associated miner for verification and packaging.

(4) Step 4: Miner verification and malicious update detection

To prevent data tampering, miners first use the public key of the medical institution to verify the validity of the digital signature. After confirming the authenticity of the identity, the miners will verify the reliability of the gradient, and finally screen for malicious updates through the formed verification committee.

(5) Step 5: Leader election and global gradient calculation

The system utilizes a verifiable random function to elect a leader based on the pre-set random seed from the Genesis block and each miner's token quantity. This leader collects homomorphic gradient data uploaded by medical institutions and performs aggregation computations directly on encrypted data using homomorphic properties, generating a global gradient. The entire computational process remains encrypted throughout, effectively safeguarding data privacy.

(6) Step 6: Generate and verify new blocks

The leader packages the computed global gradient (in encrypted form), relevant verification information, and their digital signature to generate a new block. The validation committee verifies the legitimacy of the new block, including the leader's signature and the accuracy of the global gradient calculation. The validated block is broadcast to the entire network, where all nodes synchronize and update their ledgers to address potential threats in the network, ensuring data consistency and security.

(7) Step 7: Secure decryption and differential privacy noise

The leader obtains a private key from the secure key management center, authenticates it, and retrieves the key via a secure channel. It then decrypts the aggregated gradient using this private key, generates Gaussian noise based on predefined differential privacy parameters, and finally adds the noise to the decrypted gradient.

(8) Step 8: Model update and next round of training

The global model is updated using an aggregated gradient with injected noise and distributed to all participating parties. When each medical institution synchronizes and updates its local model, a new iteration cycle begins (starting from step 2). This process repeats across the federated system until model convergence is achieved or the training rounds exceed the threshold $T$.

## 4. Plan analysis

### 4.1. Safety analysis Paillier

Homomorphic encryption ensures the confidentiality of gradients during transmission and aggregation, allowing only intermediaries to process ciphertexts. Differential privacy noise is then added to the aggregated results, providing quantifiable mathematical guarantees against inference attacks from model outputs. Meanwhile, blockchain provides tamper-proof audit trails, while Intel SGX verifies the authenticity of local computations. This architecture effectively resists various threats including gradient inversion, member inference, model poisoning, and collusion attacks, achieving secure and reliable multi-party collaborative modeling while safeguarding patient privacy.

The safety comparison between this scheme and other literature is shown in **Table 1** [11–14].

**Table 1.** Comparison of safety of different protocols

| Attack type | Literature [11] | Literature [12] | Literature [13] | Literature [14] | This article's solution |
|---|---|---|---|---|---|
| Gradient inversion attack | √ | √ | √ | × | √ |
| Member inference attack | × | √ | × | × | √ |
| Model inverse attack | × | × | × | × | √ |
| Conspiracy attack | × | √ | × | √ | √ |
| Model integrity protection | √ | √ | × | × | √ |

## 4.2. Performance testing

The experiment was conducted on a Windows 11 system with hardware specifications including an Intel i7-9700K CPU, GTX 1080T GPU, and 16 GB RAM. Blockchain-related functionalities were implemented in Go, while PyTorch 2.5.0 was employed for model training and noise injection. Key configurations included: MNIST and CIFAR10 datasets with initial clipping thresholds C set to 4 and 3 respectively, and privacy budgets $\varepsilon$ of 3 and 1. To evaluate the algorithm's impact on federated learning accuracy, it was compared with the original federated learning algorithm and the differential privacy-based federated learning algorithm from reference [14].

As shown in **Figure 2** and **Figure 3**, implementing differential privacy introduces data perturbations into the extracted dataset features, resulting in accuracy degradation. While the original federated learning algorithm achieves high accuracy, it offers inadequate privacy protection. In contrast, This approach only applies differential privacy to specific parts of the aggregated results, significantly reducing the impact of noise on accuracy while maintaining robust security.

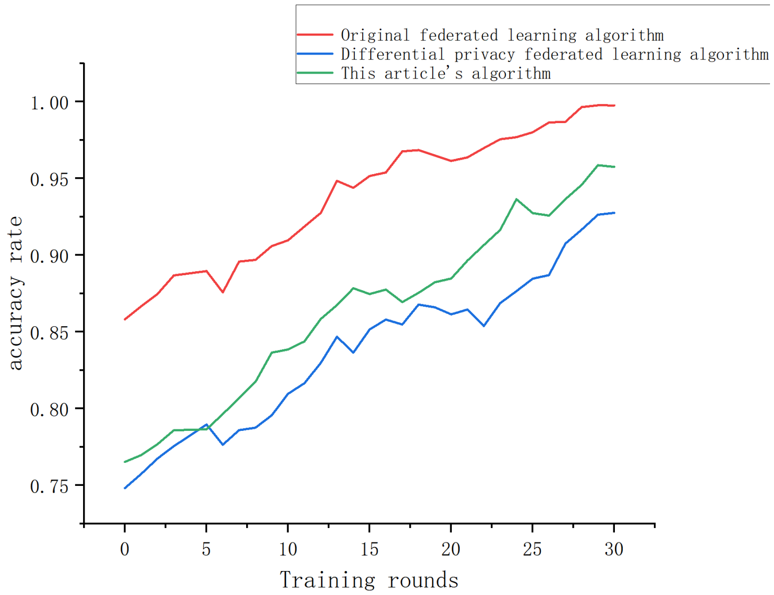

**Figure 2.** Accuracy of MNIST dataset.

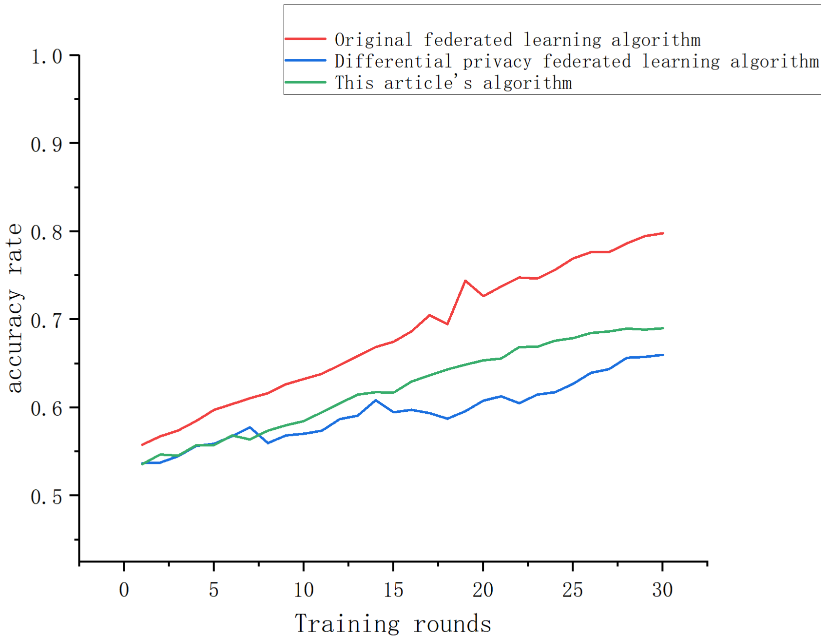

**Figure 3.** Accuracy of CIFAR-10 dataset.

## 5. Conclusion

This paper proposes a hybrid encrypted medical information security collaboration framework based on blockchain federated learning. The innovative solution integrates differential privacy with Paillier's homomorphic encryption, establishing a comprehensive defense system that spans the entire data lifecycle through a collaborative mechanism of "edge-side encryption, ciphertext aggregation, and global noise addition". Supported by blockchain's trusted traceability and SGX's integrity verification, this framework effectively addresses privacy leakage and trust challenges in cross-institutional collaboration. Future work will focus on optimizing Paillier's computational efficiency to enhance system practicality, exploring adaptive differential privacy budget allocation strategies for finer-grained privacy-utility balancing, and introducing zero-knowledge proofs to further strengthen computational verifiability.

## Disclosure statement

The authors declare no conflict of interest.

## References

[1] Li Y, Li Z, Li Q, et al., 2025, A Searchable Cryptographic Electronic Medical Record Sharing Scheme in Blockchain. Computer Engineering and Science, 47(8): 1399–1407.

[2] Zhang M, Gu C, Zhang Y, et al., 2025, Privacy Protection Scheme for Blockchain Federated Learning Based on Dynamic Group Signatures. Computer Application Research, 42(11): 1–8.

[3] Kim H, Park J, Bennis M, et al., 2020, Blockchained On-Device Federated Learning. IEEE Communications Letters,

24(6): 1279–1283.

[4] Majeed U, Khan L, Yousafzai A, et al., 2021, St-bfl: A Structured Transparency Empowered Cross-Silo Federated Learning on the Blockchain Framework. IEEE Access, 2021(9): 155634–155650.

[5] Zhang H, Li G, Zhang Y, et al., 2021, Blockchain-Based Privacy-Preserving Medical Data Sharing Scheme using Federated Learning. International Conference on Knowledge Science, Engineering and Management, 634–646.

[6] Qu Y, Gao L, Luan T, et al., 2020, Decentralized Privacy using Blockchain-Enabled Federated Learning in Fog Computing. IEEE Internet of Things Journal, 7(6): 5171–5183.

[7] Zeng S, Huo R, Huang T, et al., 2020, Survey of Blockchain: Principle, Progress and Application. Journal on Communications, 41(1): 134–151.

[8] Zhang Q, Ding Q, Zhu J, et al., 2021, Blockchain Empowered Reliable Federated Learning by Worker Selection: A Trustworthy Reputation Evaluation Method. IEEE Wireless Communications and Networking Conference Workshops (WCNCW), 1–6.

[9] Wang Q, Guo Y, Wang X, et al., 2020, AI at the Edge: Blockchain-Empowered Secure Multiparty Learning with Heterogenous Models. IEEE Internet of Things Journal, 7(10): 9600–9610.

[10] Geyer R, Klein T, Nabi M, et al., 2017, Differentially Private Federated Learning: A Client Level Perspective, arXiv, https://doi.org/10.48550/arXiv.1712.07557

[11] Guan G, Tao Z, Zhi T, et al., 2024, Privacy Protection Methods for Decentralized Federated Learning in Healthcare Settings. Computer Applications, 44(S2): 112–117.

[12] Niu S, Wang N, Zhou X, et al., 2024, A Secure Federated Learning Scheme for Smart Healthcare Based on Secret Sharing and Homomorphic Encryption. Computer Engineering, 12(8): 1–13.

[13] Wang B, Li H, Wang J, et al., 2023, Privacy-Preserving Federated Learning Architecture for Medical Data. Journal of Xi'an University of Electronic Science and Technology, 50(5): 166–177.

[14] Wen Y, Chen M, 2022, A Medical Data Sharing Solution that Integrates Federated Learning with Blockchain. Computer Engineering, 48(5): 145–153.

**Publisher's note**

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.