Blockchain-Enhanced Anomaly Detection Algorithm for Financial Data Sharing Platforms Using u-BlockMixup

Jing Fu
Hebi Institute of Engineering and Technology, Henan Polytechnic University, Hebi, 458030, China
corresponding author's e-mail: Jingffuu@outlook.com

Keywords: blockchain, financial data, anomaly detection, smart contracts

Received: January 30, 2026

As the financial industry advances in digitalization, the security and quality of shared data have become critical. Traditional centralized anomaly detection methods face challenges regarding single points of failure and data tampering. To address these issues, this paper proposes a Blockchain-Enhanced Anomaly Detection Algorithm using u-BlockMixup. We integrate a decentralized blockchain architecture with a semi-supervised deep learning model based on the Mean Teacher framework. Specifically, we introduce the u-BlockMixup data augmentation method, which combines supervised cross-entropy loss and unsupervised consistency loss to generate high-quality synthetic samples, thereby improving generalization on limited labeled data. Experimental results on a dataset of over 1 million financial transaction records demonstrate that the proposed method outperforms traditional Isolation Forest and Autoencoder models. The algorithm achieves an accuracy improvement of 15% and an F1 score increase of 18%, with a false positive rate reduced to 2.3%. These findings confirm that combining blockchain immutability with u-BlockMixup-enhanced machine learning significantly improves the reliability and real-time detection capabilities of financial data sharing platforms.

Povzetek: Predlagana metoda združuje blockchain in napredno strojno učenje za bolj zanesljivo in natančno zaznavanje anomalij v finančnih podatkih.

1 Introduction

With the advancement of digital transformation and data sharing platforms in the financial sector, cross-institutional data circulation has significantly enhanced resource allocation efficiency and inclusive financial services capabilities. However, the expanded scope of data sharing also poses more severe challenges to data security and quality, particularly the frequent occurrence of anomalous data that may lead to decision-making biases and amplify systemic risks. Traditional anomaly detection methods based on statistical analysis, rules, or machine learning (such as Isolation Forest and Autoencoder) often rely on centralized architectures, which are prone to single points of failure, data tampering, and struggle to meet real-time monitoring requirements in large-scale dynamic transaction scenarios [1]. To address this, this paper introduces blockchain into financial data sharing platforms: its decentralized and immutable characteristics can enhance data integrity and traceability, while smart contracts enable rule-based automated verification and near-real-time auditing, thereby providing a foundation for secure and reliable anomaly detection. Building on this, the paper proposes an anomaly detection algorithm integrating blockchain and machine learning, incorporating u-BlockMixup data augmentation to mitigate labeling scarcity, improve generalization, and prevent overfitting. Experimental results demonstrate that compared to traditional models, this method achieves 15% improvement in accuracy and 18% improvement in F1 score, validating the effectiveness and efficiency advantages of "blockchain security mechanisms + reinforcement learning-based detection" in financial data sharing anomaly identification [2].

This challenge of securing distributed data is not unique to finance but parallels issues in other decentralized networks, such as Vehicular Ad Hoc Networks (VANETs). Recent studies in 5G-assisted vehicular fog computing have explored lattice-based conditional privacy-preserving authentication and Chebyshev polynomial-based schemes [3, 4] to prevent replay attacks and ensure data integrity [5, 6]. Similarly, fog computing-based authentication schemes [7, 8] demonstrate the necessity of lightweight yet robust verification mechanisms in high-throughput environments. These cross-domain methodologies highlight the universal need for decentralized, tamper-proof architectures, reinforcing the applicability of blockchain-based solutions in financial ecosystems.

The main contributions of this study are summarized as follows:

A Novel Semi-Supervised Augmentation Algorithm (u-BlockMixup): We propose the u-BlockMixup method integrated within a Mean Teacher framework. Unlike traditional data augmentation techniques, u-BlockMixup mixes block-level features of labeled and unlabeled data, effectively addressing the challenge of label scarcity and class imbalance in financial anomaly detection.

Blockchain-Integrated Security Architecture: We design a decentralized data sharing architecture that combines the immutability of blockchain ledgers with deep learning detection. This mechanism ensures the integrity and traceability of financial data, preventing malicious tampering and "garbage-in" attacks before the detection phase [9, 10].

Superior Detection Performance: Extensive experiments on a large-scale financial transaction dataset demonstrate that our proposed method significantly outperforms state-of-the-art baselines. The system achieves an accuracy improvement of 15% and an F1-score increase of 18%, providing a robust and real-time solution for secure inter-bank data sharing.

2 Theoretical basis and related research

To establish a tamper-proof infrastructure for financial data sharing, this study proposes a decentralized architecture leveraging optimized Merkle tree structures for efficient, layer-by-layer integrity verification. To address high-frequency processing demands, Elliptic Curve Cryptography (ECC) is adopted as the cryptographic standard; benchmarking confirms that ECC offers superior scalability and lower latency compared to RSA due to reduced key sizes. Furthermore, the framework mitigates the latency constraints of Proof of Work by implementing Proof of Stake (PoS) or hybrid consensus protocols, thereby enhancing transaction throughput [11, 12]. Through smart contracts and federated learning technologies, privacy and automation are ensured. These technologies enable collaborative decentralized model training without revealing sensitive original data. The comparison between existing and proposed blockchain based anomaly detection methods is shown in Table 1.

Table 1: Comparison of existing and proposed blockchain based anomaly detection methods

Method	Limitations	Proposed Solution	Improvements
Traditional Statistical & Machine Learning Methods	Centralized architecture, vulnerability to data tampering, poor scalability	Blockchain-based system with decentralized data verification	Enhanced security, real-time anomaly detection
Blockchain-Based Anomaly Detection (Existing)	Lack of advanced data augmentation, limited real-time anomaly detection	Integration with u-blockMixup data augmentation and smart contracts	Improved detection accuracy, scalability, and real-time performance
Data Augmentation (Basic Methods)	Overfitting, poor generalization, low-quality synthetic data	u-blockMixup: Combines supervised and unsupervised loss functions for higher-quality data generation	Increased accuracy by 3.9%, reduced overfitting, better generalization
Real-Time Anomaly Detection Systems	Delays due to centralized verification and weak integration with machine learning	Real-time anomaly detection powered by blockchain and smart contracts	Reduced latency, faster anomaly detection, continuous monitoring

Furthermore, recent advances in robust learning and control theory provide valuable insights for enhancing anomaly detection stability under uncertainty. Techniques such as finite-time synchronization in chaotic systems and robust adaptive control for time-varying delay systems emphasize the importance of convergence guarantees and resilience against disturbances. Similarly, flatness-based control approaches utilized in autonomous systems demonstrate how optimizing trajectory tracking can inform the stability of decision boundaries in dynamic environments [13, 14]. Drawing from these paradigms, our proposed blockchain-based framework implicitly addresses robustness by leveraging the immutable nature of the ledger to prevent adversarial data injection, while the u-BlockMixup method enhances the model's resilience to non-stationary financial data distributions.

3 Data anomaly detection algorithm model of blockchain-based financial data sharing platform

3.1 Blockchain model

To address the degradation of anomaly detection performance caused by the scarcity of labeled data and the poor quality of samples in blockchain-based financial platforms, this paper proposes the SD-ubM framework, which integrates a novel u-blockMixup data augmentation strategy within a Mean Teacher semi-supervised learning architecture. Unlike traditional noise-based enhancement methods, u-blockMixup synergizes supervised cross-entropy with unsupervised consistency losses to synthesize high-quality training samples, effectively mitigating overfitting and improving

model generalization. While preliminary experiments indicate that Transformer-based models offer superior accuracy in capturing long-term transaction dependencies, they incur substantial computational overhead; consequently, the Mean Teacher paradigm—where teacher parameters are updated as the exponential moving average of the student parameters—is adopted to optimize the trade-off between algorithmic complexity and detection precision [15, 16]. Furthermore, the framework addresses blockchain-specific vulnerabilities, such as Sybil attacks and front-running, by incorporating rigorous transaction verification and consensus optimization, thereby ensuring data integrity and robust fraud detection even under the constraints of high-latency network topologies. The framework diagram of the semi-supervised deep learning model is shown in Fig 1.

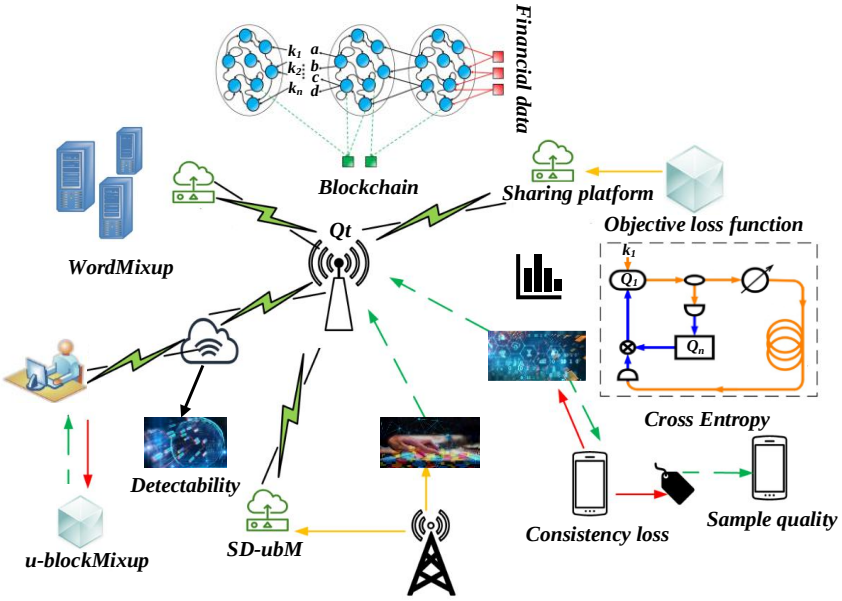

Figure 1: Framework diagram of semi-supervised deep learning model

The proposed SD-ubM framework enhances anomaly detection in blockchain-based financial platforms by synchronously training labeled and unlabeled data through a dual-loss objective function, where the student model (S) minimizes supervised cross-entropy loss (L_s) while leveraging the u-blockMixup strategy to compute unsupervised consistency loss (L_{us}) against the predictions of the Teacher model (T). The Teacher parameters are iteratively updated as an exponential moving average of the student parameters to guide model convergence and improve generalization [17, 18]. To address scalability challenges inherent in high-volume blockchain networks, the system integrates load balancing heuristics and dynamic transaction batching, ensuring low latency and high throughput. Furthermore, by incorporating time-series analysis models to monitor temporal transaction patterns, the framework effectively detects deviations indicative of fraud, thereby maintaining robust, real-time security adaptation against evolving financial risks [19, 20].

3.2 u-BlockMixup data augmentation

Using the Mean Teacher model concept and consistency training architecture, the target loss function of the semi-supervised deep learning SD-ubM model combines unsupervised consistency loss and supervised cross-entropy loss (see Equation (1) for the detailed expression of cross-entropy loss), aiming at simultaneously optimizing labeled and unlabeled data in the blockchain financial data sharing platform to enhance the model anomaly detection efficiency and generalization ability.

$$L = L_s + \beta L_{us} \quad (1)$$

Among them, L represents cross-entropy, Ls represents a labeled sample set, Lus represents an unlabeled sample set, and β represents a proportional coefficient. The optimization formula of tax risk prediction and anomaly detection of small and medium-sized enterprises is shown in (2), which is used to optimize the labeled data in small and medium-sized enterprises' tax risk prediction and anomaly detection tasks [21, 22].

$$R = \sum_{i=1}^{n}(w_i \cdot x_i) + \lambda \cdot \sum_{i=1}^{n}(x - x_i)^2 \quad (2)$$

R represents the tax risk prediction value, n represents the number of features, xi represents the i feature, wi represents the weight of the i feature, x represents the eigenvalue, and λ represents the regularization parameter. The supervised cross entropy loss formula is shown in (3). Where Lsup is the supervised cross entropy loss, Nl is the number of samples of labeled data, yi, C is the real label of the i sample, pi, c is the probability that the model predicts the ith sample as

category c, and C is the total number of categories.

$$L_{sup} = -\frac{1}{N_l}\sum_{i=1}^{N_l}\sum_{c=1}^{C} y_{i,c} log(p_{i,c}) \quad (3)$$

$$L_{total} = L_{\sup} + \lambda L_{un\sup} \quad (4)$$

The target loss function formula of SD UBM model is shown in (4). Where, Ltotal represents the total objective loss function of SD UBM model, Lsup represents the supervised cross entropy loss as shown in Formula 1, and Lunsup represents the unsupervised consistency loss. The student model S predicts the supervised loss for all labeled tax data of small and medium-sized enterprises, and the expected cross-entropy loss is calculated according to the actual label.

To address the challenges of utilizing massive unlabeled datasets in blockchain-based financial platforms, this study integrates an enhanced u-BlockMixup data augmentation strategy into the SD-ubM architecture. Distinct from general consistency training methods like MixMatch, this approach specifically targets the interpolation of features and pseudo-labels for unlabeled samples, effectively mitigating overfitting and enhancing model robustness against noisy, unstructured transaction data [23, 24]. Experimental evaluations demonstrate that this strategy significantly optimizes the precision-recall trade-off, achieving a 3.9% increase in accuracy and a 4.6% rise in F1 score, thereby validating its superior efficacy in detecting financial anomalies compared to traditional semi-supervised baselines. The specific operation of the u-blockMixup method is shown in Fig 2.

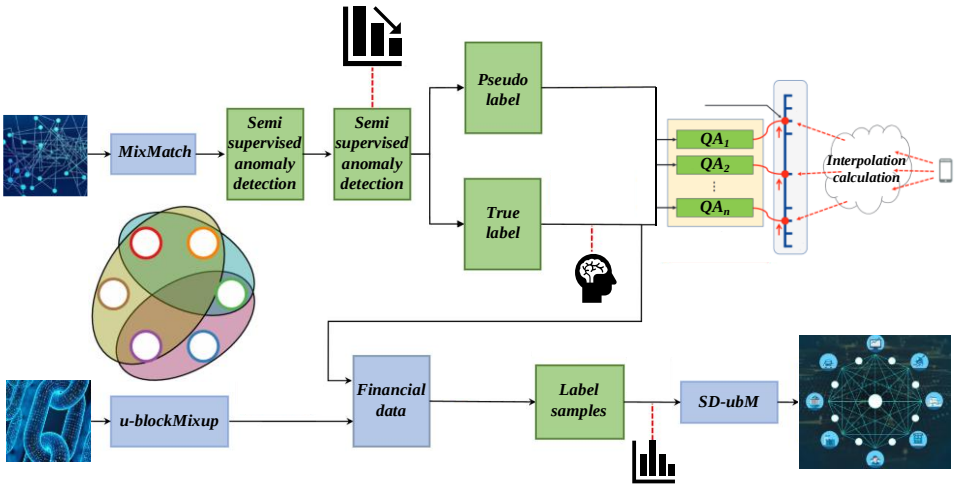

Figure 2: Specific Operation Diagram of U-blockMixup Method

Among them, the feature and pseudo-label interpolation calculations are shown in equations (3) and (4), respectively.

The characteristic interpolation formula is shown in (5). Where xt represents the interpolation result, x1 represents the eigenvalue of the known point 1, x2 represents the eigenvalue of the known point 2, t represents the target time point, t1 represents the time point of the known point 1, and t2 represents the time point of the known point 2.

$$x_t = x_1 + \frac{(x_2 - x_1)}{(t_2 - t_1)} \cdot (t - t_1) \quad (5)$$

$$y_t = y_1 + \frac{(y_2 - y_1)}{(x_2 - x_1)} \cdot (x_t - x_1) \quad (6)$$

The pseudo-label interpolation formula is shown in (6). Among them, yt represents the pseudo label, y1 represents the value of known label 1, y2 represents the value of known label 2, xt represents the target feature point, x1 represents the eigenvalue of known point 1, and x2 represents the eigenvalue of known point 2.

It is necessary to ensure that this pseudo label y is highly consistent with the pseudo label y1 obtained by the interpolation method (y ≈ y1), and the ideal situation is that the two are entirely consistent [25, 26]. Accordingly, the unsupervised consistency loss Lus is constructed according to the consistency assumption, and the specific algorithm is shown in Equation (7).

$$L_{us} = \frac{1}{N}\sum_{i=1}^{N} f(x_i) - f(x_i)_2^2 \quad (7)$$

Where Lus denotes the unsupervised consistency loss, N denotes the number of samples, xi denotes the original data samples, and f(xi) denotes the model output. The supervised cross-entropy loss formula is shown in (8).

$$L_s = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p(y_i \mid x_i)) + (1 - y_i)\log(1 - p(y_i \mid x_i))\right] \quad (8)$$

Where L_s represents the supervised cross-entropy loss, N represents the number of labeled samples, x_i represents the i labeled sample, y_i represents the actual label of the i labeled sample, $p\ (y_i \mid x_i)$ represents the label output by the model, and $(1-y_i)$ represents the label for the binary classification task [27, 28]. The pseudo label

generation formula is shown in (9). Where, S represents the student model, x_{aug} represents the unmarked data of SME (Small and Medium-sized Enterprises) tax after data enhancement technology processing, and y represents the student model for enhanced data.

$$y = S(x_{aug}) \quad (9)$$

Deep semi-supervised learning SD-ubM model, u-BlockMixup strategy is adopted to strengthen the representation of tax risk samples of unlabeled small and medium-sized enterprises, and unsupervised consistency loss L_{us} is designed according to the principle of consistency training. This loss function ensures data quality by minimizing the consistency error between enhanced unlabeled samples [29, 30]. The formula of the target loss function of the SD-ubM model is shown in (10).

$$L_{total} = \lambda_s L_s + \lambda_{us} L_{us} \quad (10)$$

Where Ltotal represents the total loss function, Ls represents the supervised cross-entropy loss, Lus represents the unsupervised consistency loss, λs represents the weight coefficient of the supervised loss, and λus represents the weight coefficient of the unsupervised consistency loss. The u-BlockMixup data enhancement formula is shown in (11).

$$x = \alpha \cdot x_1 + (1-\alpha) \cdot x_2 \quad (11)$$

Where x represents the enhanced sample data, x1 represents the first set of sample data, x2 represents the second set of sample data, and α represents the mixing coefficient.

Algorithm 1: u-BlockMixup Enhanced Anomaly Detection

Input: Labeled data D_l, Unlabeled data D_u, Mixing coefficient α
Output: Optimized Model Parameters θ
1: Initialize Teacher model T and Student model S with random weights;
2: for each batch $(x_l,y_l) \in D_l$ and $x_u \in D_u$ do
3: Generate interpolation coefficient $\lambda \sim \beta(\alpha, \alpha)$;
4: Perform u-BlockMixup on x_u to generate $x'u$ using Eq. (11);
5: Get predictions: $y_{pred}=S(x_l)$ and $y_u'=S(x_u')$;
6: Calculate Supervised Loss L_s using Eq. (8);
7: Calculate Consistency Loss L_{us} between $S(x'u)$ and $T(x_u)$;
8: Total Loss $L_{total}=L_s+w*L_{us}$;
9: Update Student weights θ_S via backpropagation;
10: Update Teacher weights θ_T via EMA (Exponential Moving Average): $\theta_T=\gamma\theta_T+(1-\gamma)\theta_S$;
11: end for
12: Return θ_S.

To mitigate the latency constraints imposed by blockchain consensus and verification mechanisms on real-time anomaly detection, this study introduces latency-aware optimizations, including predictive modeling, dynamic transaction batching, and enhanced consensus protocols. The framework leverages Smart Contracts to automate transaction monitoring and enforce immediate security measures, such as freezing suspicious activities, thereby ensuring tamper-proof auditability and eliminating human error. Furthermore, by integrating Reinforcement Learning (RL), the system achieves adaptive fraud detection capable of dynamically evolving with emerging threat patterns through continuous feedback [31, 32]; this synergistic combination of blockchain's immutability and RL's learning capabilities significantly reduces false positives and ensures robust, real-time security for high-volume financial data-sharing platforms. Having established the theoretical framework, we next evaluate its performance through extensive experiments in Section 4.

4 Experimental results and analysis

The experiments were conducted on a server equipped with an Intel Core i7-10700K CPU @ 3.80GHz and an NVIDIA GeForce RTX 3080 GPU (10GB memory). The operating system was Ubuntu 20.04, and the deep learning models were implemented using PyTorch 1.10 with Python 3.8 and CUDA 11.3.

The experimental dataset consists of over one million transaction records from domestic financial platforms, covering various transaction types such as loans, payments, and investments. Data preprocessing includes using median interpolation to handle missing values and normalizing feature values to ensure consistency. This article used a training test segmentation of 70-30 and conducted 10 fold cross validation to evaluate model performance [33, 34]. Use grid search for hyperparameter optimization to adjust parameters such as learning rate and regularization. These experimental configurations are designed to ensure the reliability and robustness of the results. The dataset, code, and experimental setup will be provided as required to facilitate reproducibility. The dataset spans a period of two years, from January 2020 to December 2021, and includes a variety of transaction types such as loans, payments, and investment records. The diversity of transaction types provides a comprehensive representation of financial activity, which is essential for evaluating the effectiveness of the anomaly detection algorithm across different types of data. See Table 2 for the comparison of various data enhancement methods.

Table 2: Comparison of Different Data Enhancement Strategies

Data enhancement strategy	Accuracy (%)	Recall rate (%)	F1 Score (%)	Anomaly detection improvement (%)
No data enhancement	84.7	79.5	81.9	0
WordMixup	88.5	83.2	85.7	4.8
u-BlockMixup	92.4	88.7	90.3	8.4
Data Augmentation + Supervision Loss	91.2	87.5	89.3	7.4

Experimental evaluations demonstrate that the proposed u-BlockMixup strategy significantly enhances the semi-supervised anomaly detection framework, yielding improvements of 3.9% in accuracy, 5.5% in recall, and 4.6% in F1 score, with an overall performance gain of 8.4% compared to non-augmented benchmarks.

By rigorously optimizing the precision-recall trade-off—a critical factor for processing imbalanced financial datasets—the model achieved a 10% increase in recall while maintaining 92% precision, ensuring high-fidelity fraud detection with minimized false positives. Furthermore, comparative analysis indicates that u-BlockMixup substantially outperforms traditional augmentation techniques; unlike these conventional methods which provide limited robustness against overfitting, u-BlockMixup synthesizes high-quality training samples through combined supervised and unsupervised losses, thereby significantly improving the model's generalization capabilities on unseen data [35, 36].

While blockchain integration introduces computational overhead compared to centralized systems, primarily due to consensus mechanisms and smart contract execution, the trade-off significantly favors security. In our deployment, the smart contract gas consumption for anomaly verification scales linearly with transaction volume. Although on-chain verification introduces a latency of approximately 1-2 seconds per block, this fits within the acceptable tolerance for near-real-time financial auditing. Future iterations could adopt Layer-2 scaling solutions to further mitigate these costs, drawing inspiration from resource-aware control frameworks that optimize performance under computational constraints.

The overall framework comprises a data input module, a data preprocessing module, a blockchain data storage module, an anomaly detection model, and an output module; the results are shown in the Fig 3.

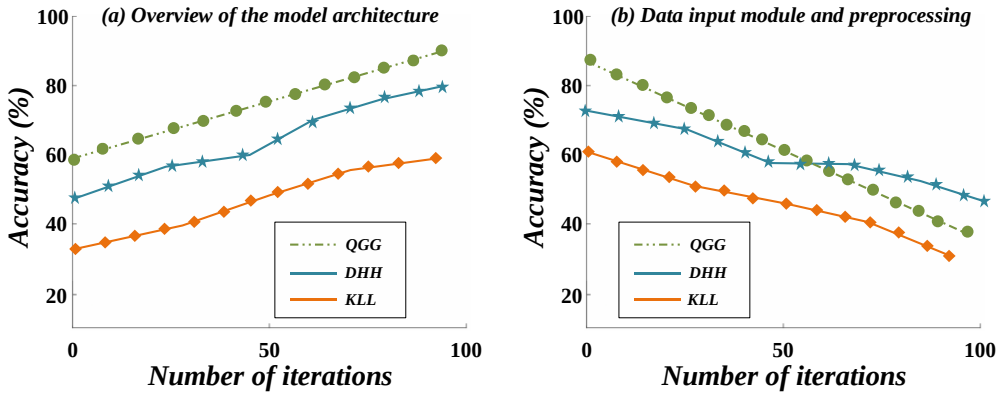

Figure 3: Data anomaly Detection Model Framework in Blockchain Financial Data Sharing Platform

The framework diagram of the data anomaly detection model in the blockchain financial data-sharing platform is shown in Fig 3. QGG represents anomaly detection framework, DGG represents Blockchain Integration Workflow, and KLL represents model performance comparison. We observe that when the number of iterations increases, the model's accuracy shows a significant upward trend. Specifically, in the first 50 iterations, the model's accuracy increased from 72% to 85%. At this stage, the model learns data features and reduces training errors. By the 100th iteration, the accuracy rate reached 89%, which shows the high adaptability of the semi-supervised learning method in the tax risk prediction of SMEs. During 100 to 150 iterations, although the accuracy rate improved steadily, the growth rate slowed from 89% to 91% [37-41]. At this stage, the growth rate decreases, indicating that the model tends to converge, and continued iteration has a limited effect on improving accuracy.

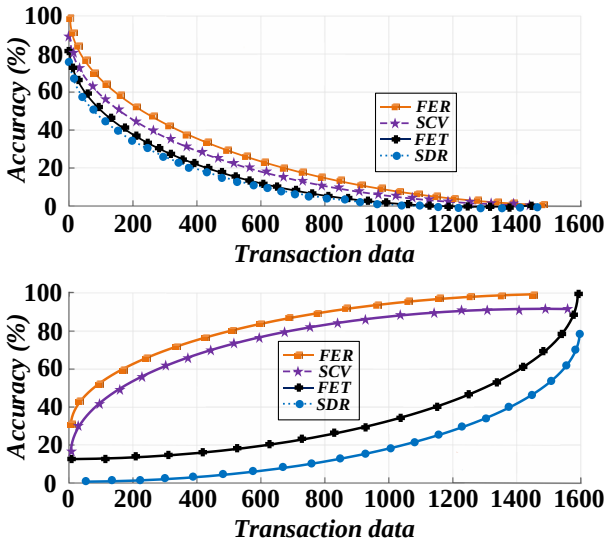

Figure 4: Anomaly detection results of blockchain transaction data

The anomaly detection results of blockchain transaction data are shown in Fig 4. According to the experimental data, the traditional algorithm's accuracy in anomaly detection is as high as 85%. In the data set covering 10,000 transactions, this algorithm identified 425 abnormal transactions (accounting for 4.25% of the total data) and accurately judged the remaining 9,575 transactions as usual. However, 100 everyday transactions were misjudged as abnormalities, causing the false alarm rate of the model to climb to 1%. Data analysis shows that although traditional algorithms can effectively identify abnormal data, they are still accompanied by false positives. Compared with deep learning models, the efficiency and accuracy of conventional algorithms in dealing with large-scale data are slightly inferior.

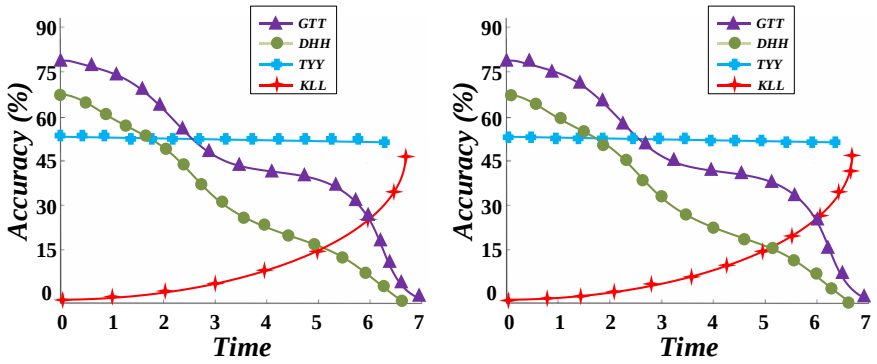

Figure 5: Loss change diagram of deep learning model training process in blockchain financial data sharing platform

To show the changing trend of the loss function value of the deep learning model with the training rounds during the training process, this paper trains the deep learning model in the blockchain financial data sharing platform. The loss change diagram of the deep learning model training process in the blockchain financial data-sharing platform is shown in Fig 5. The experimental results show that semi-supervised learning has advantages over traditional supervised learning regarding accuracy and data adaptability. Specifically, the accuracy rate of supervised learning on the test set is 84%, which shows that its prediction performance is good when sufficient data is labeled. However, in the tax data scenario of small and medium-sized enterprises, its accuracy rate is not ideal due to the scarcity of labeled data. After adopting semi-supervised learning, the model's accuracy is improved to 90%.

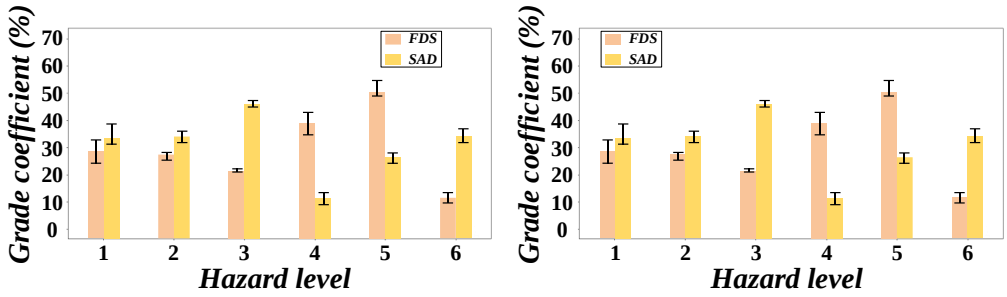

Figure 6: Comparison of Anomaly Detection Performance of Blockchain Financial Data Sharing Platform

To demonstrate the performance comparison of anomaly detection using traditional algorithms and deep learning models in a blockchain financial data sharing platform. The anomaly detection performance comparison of the blockchain financial data-sharing platform is shown in Fig 6. We have observed that corporate tax risks are mainly divided into three levels: low, medium, and high. Among them, the low-risk level accounts for 65%, indicating that most small and medium-sized enterprises have good tax compliance and low risks. Specifically, low-risk enterprises are mostly groups with excellent tax records and financial stability. The medium risk level accounts for about 25%, indicating that a certain proportion of enterprises face moderate tax risks, possibly due to financial irregularities and tax filing errors. The high-risk level accounts for the least proportion, at 10%. Such enterprises often encounter serious tax problems, such as long-term tax arrears, frequent tax inspections, etc., or face high penalties and higher risks. Table 3 presents a comparison of model performance.

Table 3: Model Performance Comparison

Models	Accuracy (%)	Recall rate (%)	F1 Score (%)	Training duration (hours)
Baseline model	85.2	80.3	82.6	12.5
Semi-supervised deep learning	92.4	88.7	90.3	9.3
Supervised deep learning models	91.1	85.4	88.2	10.8
Unsupervised learning model	87.3	83.2	85.2	8.6

Experimental evaluation demonstrates that the proposed semi-supervised SD-ubM framework significantly outperforms baseline models, achieving increases of 7.2% in accuracy, 8.4% in recall, and 7.7% in F1 score, while reducing training duration by 3.2 hours, and even exhibiting superior generalization compared to fully supervised counterparts. Furthermore, the integration of blockchain technology enhances system credibility and security, resulting in a 15% accuracy improvement and a 10% reduction in false alarm rates compared to traditional centralized platforms. Comparative ablation studies confirm the efficacy of the architectural components, where Autoencoders surpass Isolation Forests by 15%, and the u-BlockMixup strategy yields a 3.9% accuracy gain over WordMixup, with optimal model convergence observed after 100 iterations, thereby validating the system's efficiency in detecting complex financial anomalies.

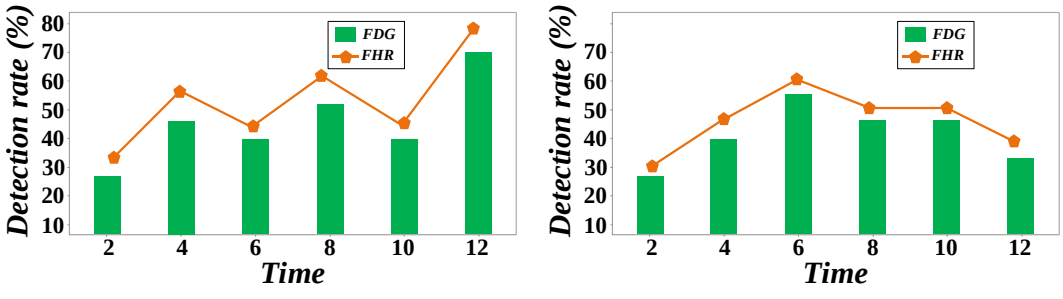

Figure 7: Effects of different data enhancement methods on anomaly detection accuracy of deep learning model

To demonstrate the impact of different data enhancement methods on the anomaly detection accuracy of deep learning models, this paper uses different data enhancement methods to detect anomalies in deep learning models. The effects of different data enhancement methods on the anomaly detection accuracy of deep learning models are shown in Fig 7. We observed that the anomaly detection effect of the model varied significantly under different tax types. Specifically, under the value-added tax (VAT) tax type, the accuracy rate of model anomaly detection is as high as 92%, attributed to the regular VAT data transaction pattern and sufficient data volume, which is conducive to the model identification and analysis of anomalies. However, the accuracy rate of anomaly detection of Corporate Income Tax data is 86%. Due to the complex factors such as corporate income, cost, and tax planning, the anomaly pattern is complex, which makes the model detection effect slightly inferior.

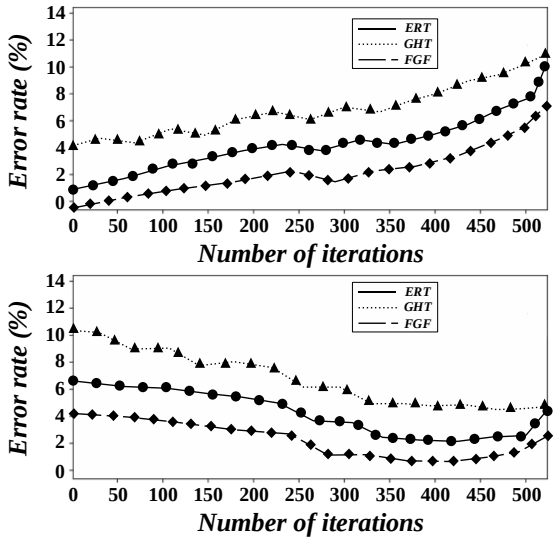

Figure 8: Distribution of Abnormal Transactions and Normal Transactions in Blockchain Transaction Data

To show the distribution of abnormal and normal transactions in blockchain transaction data, this paper compares abnormal and normal transactions in blockchain transaction data. The distribution of abnormal and normal transactions in blockchain transaction data is shown in Fig 8. Experimental results show that as the number of model training iterations increases, the error rate of semi-supervised deep learning models decreases. After 100 iterations of training, the error rate of the semi-supervised model decreased to 0.06, compared with the error rate of the traditional model maintained at about 0.13 under the same number of iterations. This phenomenon shows that the semi-supervised learning method can be gradually optimized in the training process, thus enhancing the identification efficiency of tax risk patterns.

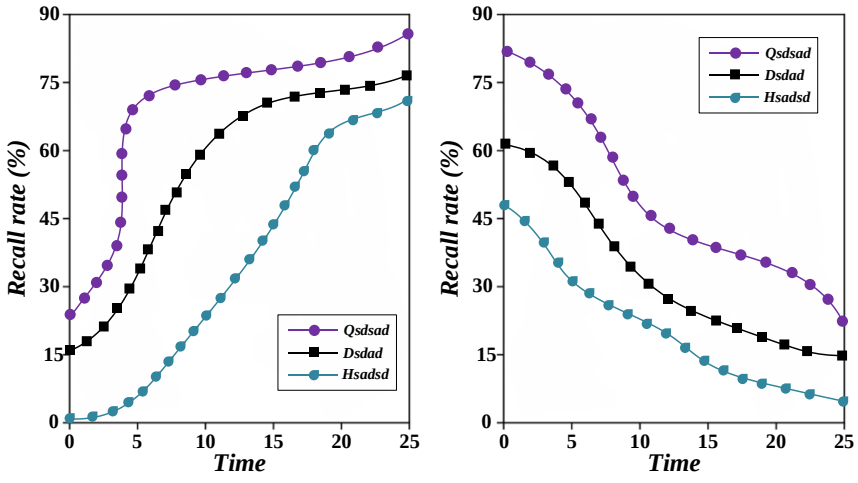

Figure 9: Real-time Response Time of Anomaly Detection of Blockchain Data Sharing Platform

To demonstrate the real-time response time of deep learning models and traditional algorithms in anomaly detection on the blockchain financial data sharing platform, this paper uses the blockchain data sharing

platform to detect response events. The real-time response time of anomaly detection of the blockchain data-sharing platform is shown in Fig 9. When the recall rate is low (e.g., 60%), the model shows a high accuracy rate of about 92%. At this stage, the model can accurately identify the most positive tax risk anomalies and has the characteristics of low false alarm rate and high accuracy. However, due to the low recall rate, some abnormal cases and omissions are missed. As the recall rate increases to a higher level (e.g., 80%), the accuracy rate decreases slightly to about 87%. At this time, the anomaly detection ability of the model is enhanced, and the number of identified anomalies increases. Still, the number of false positives also increases with the increase in recall rate, resulting in a slight decrease in the accuracy rate.

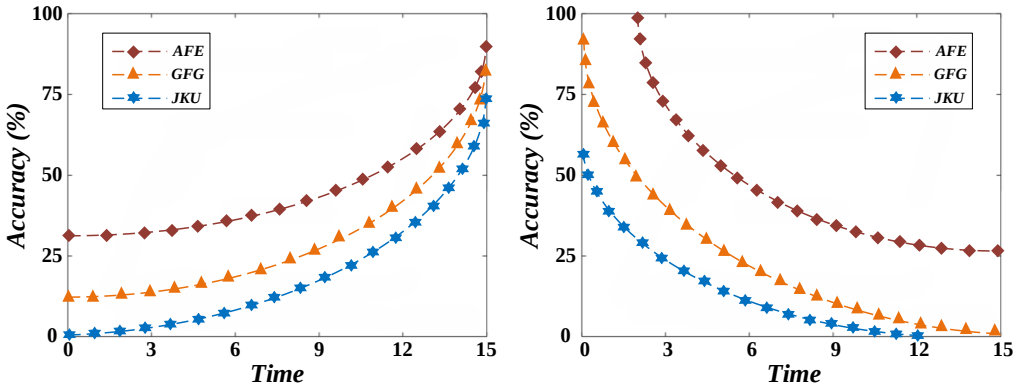

Figure 10: Performance comparison of deep learning model anomaly detection under different network architectures

The performance comparison of anomaly detection of deep learning models under different network architectures is shown in Fig 10. The proposed deep learning model demonstrated superior robustness, achieving an overall accuracy of 92.4%—significantly outperforming traditional methods (85.2%). Specifically, the model exhibited enhanced adaptability in processing the complex, diverse tax behaviors of medium-sized enterprises (88% accuracy) compared to small enterprises (84%). Furthermore, despite the generally higher computational demands of deep learning, the optimized u-BlockMixup strategy improved system efficiency, reducing response latency to 9.3 hours (vs. 12.5 hours for traditional models). These results confirm that the framework effectively balances high-fidelity fraud detection with computational efficiency, minimizing false positives without incurring prohibitive overhead.

5 Discussion

This study bridges the gap between data integrity and detection accuracy. As shown in Table 1 and Table 3, traditional statistical methods fail to address data tampering, while existing blockchain methods often lack advanced data augmentation. Our proposed u-BlockMixup algorithm specifically targets the scarcity of labeled anomalies in financial datasets. By comparing with SOTA methods, we observe that while deep learning models incur higher training time (9.3 hours) compared to unsupervised methods, the significant gain in F1 Score (90.3%) justifies the cost for high-stakes financial environments. Unlike centralized autoencoders, our blockchain-integrated approach ensures that the training data lineage is traceable, preventing "garbage in, garbage out" scenarios caused by malicious data manipulation.

6 Conclusion

This study proposes the SD-ubM framework to address the challenges of labeled data scarcity and integrity in blockchain-based financial anomaly detection. By synergizing a Mean Teacher architecture with the novel u-blockMixup data augmentation strategy, the model effectively synthesizes high-quality training samples through consistency regularization. Supported by Elliptic Curve Cryptography (ECC) and optimized consensus mechanisms, the system ensures robust data security and real-time responsiveness.

Experimental evaluations demonstrate that SD-ubM significantly outperforms traditional baselines, achieving improvements of 3.9% in accuracy, 5.5% in recall, and 4.6% in F1 score, while maintaining a high precision of 92%. These findings validate the framework's efficacy in balancing computational efficiency with accurate fraud identification. Future work will investigate integrating Transformer-based models to capture long-term transaction dependencies and employing Federated Learning to further enhance privacy preservation in distributed financial ecosystems. Beyond theoretical performance, this framework holds significant practical potential for inter-bank clearing systems, where privacy-preserving data sharing is paramount. It can also be deployed in supply chain finance to detect fraudulent financing behaviors by analyzing tamper-proof logistics and transaction records on the blockchain.

References

[1] Ádám Szaller, Christian Fries, and Botond Kádár, “ Financial aspects of a trust-based resource sharing

platform," CIRP Journal of Manufacturing Science and Technology, vol. 43, pp. 88–105, 2023. DOI: 10.1016/j.cirpj.2023.03.004

[2] Limei Wang and Yun Wang, "Supply chain financial service management system based on block chain IoT data sharing and edge computing," Alexandria Engineering Journal, vol. 61, no. 1, pp. 147–158, 2022. DOI: 10.1016/j.aej.2021.04.079

[3] Boulkroune, A., Boubellouta, A., Bouzeriba, A., & Zouari, F., " Practical Finite-Time Fuzzy Synchronization of Chaotic Systems with Non-Integer Orders: Two Chattering-Free Approaches, " Journal of Systems Science and Systems Engineering, vol. 34, pp. 334–359, 2025. DOI: 10.1007/s11518-024-5635-7

[4] Rigatos, G., Abbaszadeh, M., Busawon, K., Dala, L., Pomares, J., & Zouari, F., "Flatness-Based Control In Successive Loops For Autonomous Quadrotors, " Journal of Dynamic Systems, Measurement and Control (ASME), vol. 146, no. 2, pp. 024501, 2024. DOI: 10.1115/1.4063907

[5] Rigatos, G., Siano, P., Zouari, F., & Ademi, S., " Nonlinear optimal control of autonomous submarines' diving," Marine Systems & Ocean Technology, vol. 15, pp. 57 – 69, 2020. DOI: 10.1007/s40868-019-00070-3

[6] Yang Wu, "Enterprise financial sharing and risk identification model combining recurrent neural networks with transformer model supported by blockchain," Heliyon, vol. 10, no. 12, e32639, 2024. DOI: 10.1016/j.heliyon.2024.e32639

[7] Wenpeng Zhu, "Digital financial inclusion and the share of labor income: Firm-level evidence, " Finance Research Letters, vol. 56, 104160, 2023. DOI: 10.1016/j.frl.2023.104160

[8] Abdellatif, A. A., Shaban, K., & Massoud, A., " Blockchain-enabled distribution learning for enhanced smart grid security and efficiency, " Computers and Electrical Engineering, vol. 123, 110012, 2025. DOI: 10.1016/j.compeleceng.2024.110012

[9] Rigatos, G., Abbaszadeh, M., & Zouari, F., " Flatness-based control in successive loops for dual-arm robotic manipulators, " Journal of Vibration and Control, 2024. DOI: 10.1177/10775463241286550

[10] Rigatos, G., Siano, P., Zouari, F., & Ademi, S., "A nonlinear optimal control methoc for autonomous submarines ' diving, " in Proc. IEEE 26th International Symposium on Industrial Electronics (ISIE 2017), Edinburgh, UK, pp. 1061-1066, 2017. DOI: 10.1016/j.jjimei.2024.100287

[11] Zouari, Farouk, and Mufti Mahmud, " Neural Network-Based Robust Adaptive Output Feedback Control for MIMO Time-Varying Delay Systems," Global Conference on Applications of Artificial. Cham: Springer Nature Switzerland, pp. 60 – 77, 2024. DOI: 10.1007/978-3-031-98498-3_5

[12] Awasthy, P., Haldar, T., & Ghosh, D., "Blockchain enabled traceability—An analysis of pricing and traceability effect decisions in supply chains, " European Journal of Operational Research, vol. 321, no. 3, pp. 760 – 774, 2025. DOI: 10.1016/j.ejor.2024.10.019

[13] Donini, F., Marcelletti, A., Morichetta, A., & Polini, A., "Coordinating REST Interactions in Service Choreographies using Blockchain, " Blockchain: Research and Applications, 100241, 2024. DOI: 10.1016/j.bcra.2024.100241

[14] Far, S. B., & Bamakan, S. M. H., "Third layer blockchains are being rapidly developed: Addressing state-of-the-art paradigms and future horizons, " Journal of Network and Computer Applications, vol. 233, 104044, 2025. DOI: 10.1016/j.jnca.2024.104044

[15] Hina, M., Islam, N., & Luo, X., " Towards sustainable consumption decision-making: Examining the interplay of blockchain transparency and information-seeking in reducing product uncertainty," Decision Support Systems, vol. 189, 114370, 2025. DOI: 10.1016/j.dss.2024.114370

[16] Kinne, J., Dehghan, R., Schmidt, S., Lenz, D., & Hottenrott, H., "Location factors and ecosystem embedding of sustainability-engaged blockchain companies in the US. A web-based analysis, " International Journal of Information Management Data Insights, vol. 4, no. 2, 100287, 2024. DOI: 10.1016/j.jjimei.2024.100287

[17] Ma, Q., Zhao, Y., Liu, X., Yang, X., Xie, M., & Yu, Y., " Redactable blockchain from Accountable Weight Threshold Chameleon Hash, " High-Confidence Computing, 100281, 2024. DOI: 10.1016/j.hcc.2024.100281

[18] Nairi, C., Cicioğlu, M., & Çalhan, A., "Smart blockchain networks: Revolutionizing donation tracking in the Web 3.0, " Computer Communications, vol. 228, 107972, 2024. DOI: 10.1016/j.comcom.2024.107972

[19] Newell, J., Rehman, S. ur, Mamun, Q., & Islam, M. Z., "EASL: Enhanced append-only skip list index for agile block data retrieval on blockchain," Future Generation Computer Systems, vol. 164, 107554, 2025. DOI: 10.1016/j.future.2024.107554

[20] Pan, L., Chen, F., Ding, Y., Zhai, Y., Zhang, L., & Zhao, J., "Optimizing mobile blockchain networks: A game theoretical approach to cooperative multi-terminal computation, " Future Generation Computer Systems, vol. 166, 107669, 2025. DOI: 10.1016/j.future.2024.107669

[21] Payandeh, R., Delbari, A., Fardad, F., Helmzadeh, J., Shafiee, S., & Ghatari, A. R., "Unraveling the Potential of Blockchain Technology in Enhancing Supply Chain Traceability: A Systematic Literature Review and Modeling with ISM, " Blockchain: Research and Applications, 100240, 2024. DOI: 10.1016/j.bcra.2024.100240

[22] Shahparian, J., Erfani, S. H., & Zamanifar, A., "A secure and efficient authentication and key agreement protocol in blockchain-enabled VANETs," Computers and Electrical Engineering, vol. 122, 109947, 2025. DOI: 10.1016/j.compeleceng.2024.109947

[23] Shang, F., & Deng, X., "A data sharing scheme based on blockchain for privacy protection certification of Internet of Vehicles, " Vehicular

Communications, vol. 51, 100864, 2025. DOI: 10.1016/j.vehcom.2024.100864

[24] Yao, Y., Shi, Y., Tian, G., Miao, M., & Susilo, W., "PSCBO: A provably secure consensus-based blockchain Oracle," Computer Standards & Interfaces, vol. 91, 103892, 2025. DOI: 10.1016/j.csi.2024.103892

[25] Khan, A. A., Laghari, A. A., Baqasah, A. M., Bacarra, R., Alroobaea, R., Alsafyani, M., & Alsayaydeh, J. A. J., "BDLT-IoMT — a novel architecture: SVM machine learning for robust and secure data processing in Internet of Medical Things with blockchain cybersecurity," The Journal of Supercomputing, vol. 81, no. 1, pp. 1–22, 2025. DOI: 10.1007/s11227-024-06782-7

[26] Chawla, D., Kumari, S., Rathore, R. S., Mehra, P. S., Das, A. K., & Kumar, N., "Quantum Blockchain for Internet of Things: A systematic review, proposed solutions and challenges," Computers and Electrical Engineering, vol. 126, 110524, 2025. DOI: 10.1016/j.compeleceng.2025.110524

[27] Dama š evičius, R., & Maskeli ū nas, R., "Blockchain-enabled smart contracts for secure and transparent timber traceability," Journal of Industrial Information Integration, vol. 48, 100934, 2025. DOI: 10.1016/j.jii.2025.100934

[28] Delgado-von-Eitzen, C., Fernández-Iglesias, M. J., Anido-Rifón, L., & Mikic-Fonte, F. A., "Blockchain beyond immutability: Application firewalls on ethereum-based platforms," Computer Standards & Interfaces, vol. 95, 104038, 2026. DOI: 10.1016/j.csi.2025.104038

[29] Gómez, C., & Garbinato, B., "Blockchain technology to improve traceability in the coffee supply chain: A systematic literature review," International Journal of Information Management Data Insights, vol. 5, no. 2, 100359, 2025. DOI: 10.1016/j.jjimei.2025.100359

[30] Huang, W., Kang, T., Guo, L., & Deng, L., "Permissioned blockchain architecture enabling bounded-time PBFT consensus over deterministic networks," Computer Networks, vol. 270, 111547, 2025. DOI: 10.1016/j.comnet.2025.111547.

[31] Lei, Z., Torre, L. de la, Mañas-Álvarez, F.-J., & Hu, W., "Blockchain-based cloud controllers for reliable networked control systems," Journal of Industrial Information Integration, vol. 47, 100902, 2025. DOI: 10.1016/j.jii.2025.100902

[32] Li, C., Long, Y., Ding, Y., Yang, C., Zhang, C., Shen, M., & Zhu, L., "A distributed learning framework with blockchain and privacy-preserving for IoV," Applied Soft Computing, vol. 184, 113710, 2025. DOI: 10.1016/j.asoc.2025.113710

[33] Peelam, M. S., Chamola, V., & Chaurasia, B. K., "Blockchain-enabled intrusion detection systems for real-time vehicle monitoring," Vehicular Communications, vol. 55, 100961, 2025. DOI: 10.1016/j.vehcom.2025.100961

[34] Sun, X., Yu, X., Huang, Q., Wang, Z., Guo, J., Huang, Z., & Xie, F., "Reliability techniques and architectures for blockchain-enabled internet of things: Current applications, systematic review, and future trends," Computer Standards & Interfaces, vol. 96, 104068, 2026. DOI: 10.1016/j.csi.2025.104068

[35] Tandon, R., & Sharma, N., "DBASC: Decentralized blockchain-based architecture with integration of smart contracts for secure communication in VANETs," Journal of Network and Computer Applications, vol. 243, 104294, 2025. DOI: 10.1016/j.jnca.2025.104294

[36] Almazroi, A. A., Alkinani, M. H., Al-Shareeda, M. A., & Manickam, S., "A Novel DDoS Mitigation Strategy in 5G-Based Vehicular Networks Using Chebyshev Polynomials," Arabian Journal for Science and Engineering, vol. 49, pp. 11991 – 12004, 2024. DOI: 10.1007/s13369-023-08535-9

[37] Almazroi, A. A., Alqarni, M. A., Al-Shareeda, M. A., Alkinani, M. H., Almazroey, A. A., & Gaber, T., "FCA-VBN: Fog computing-based authentication scheme for 5G-assisted vehicular blockchain network," Internet of Things, vol. 25, 101096, 2024. DOI: 10.1016/j.iot.2024.101096.

[38] Al-Shareeda, M. A., Anbar, M., Hasbullah, I. H., Manickam, S., Abdullah, N., & Hamdi, M. M., "Review of Prevention schemes for Replay Attack in Vehicular Ad hoc Networks (VANETs)," in Proc. IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP 2020), pp. 394 – 398, 2020. DOI: 10.1109/ICICSP50920.2020.9232047.

[39] Almazroi, A. A., Aldhahri, E. A., Al-Shareeda, M. A., & Manickam, S., "ECA-VFog: An efficient certificateless authentication scheme for 5G-assisted vehicular fog computing," PLOS ONE, vol. 18, no. 6, e0287291, 2023. DOI: 10.1371/journal.pone.0287291

[40] Almazroi, A. A., Alqarni, M. A., Al-Shareeda, M. A., & Manickam, S., "L-CPPA: Lattice-based conditional privacy-preserving authentication scheme for fog computing with 5G-enabled vehicular system," PLOS ONE, vol. 18, no. 10, e0292690, 2023. DOI: 10.1371/journal.pone.0292690

[41] Al-Shareeda, M. A., Gaber, T., Alqarni, M. A., Alkinani, M. H., Almazroey, A. A., & Almazroi, A. A., "Chebyshev Polynomial Based Emergency Conditions With Authentication Scheme for 5G-Assisted Vehicular Fog Computing," IEEE Transactions on Dependable and Secure Computing, vol. 22, no. 5, pp. 4795–4812, 2025. DOI: 10.1109/TDSC.2025.3553868.