Detection and Countermeasures for Security Fraud in Business Administration Based on Blockchain Technology in the Context of Big Data

Yu Zheng[a*]
Setting: Cavite state university, Manila, Philippines
Address: Don Severino DE las Alas, Campus, Indang Cavite, Philippines
Zip Code: 1706
Corresponding author email: yu595512483@gmail.com

Abstract: Fraud detection in business administration based on blockchain technology has emerged as a crucial direction in addressing the complexities of corporate transactions and data security requirements. This study integrates the decentralized and immutable characteristics of blockchain, introducing Random Forest and SMOTE algorithms to design a fusion model aimed at enhancing fraud detection performance and security. Experimental results demonstrate that this model significantly outperforms traditional models such as Random Forest, Support Vector Machine, and Logistic Regression in terms of accuracy (97.8%), precision (96.5%), recall (95.3%), and F1-score (95.9%). The combination of blockchain technology and improved algorithms can markedly improve the efficiency and security of fraud detection, providing a practical solution for complex fraud scenarios in the field of business administration.

Keywords: Blockchain Technology; Fraud Detection; Business Administration; Random Forest; Data Security

1. Introduction

With the rapid development of information technology and the widespread application of big data, the scale and complexity of transactions in the field of business administration continue to grow. However, issues such as information asymmetry, data tampering, and imbalanced sample distribution have led to an increasing complexity in fraudulent behaviors, rendering traditional detection methods increasingly inadequate to meet modern technical requirements [1-3]. Traditional centralized fraud detection systems exhibit significant shortcomings in terms of data storage security, real-time performance, and traceability, making it difficult to satisfy the high security and precision demands for transaction data in modern business administration [4-6]. In recent years, blockchain technology has garnered widespread attention due to its decentralized, immutable, and fully traceable characteristics, providing a novel technical pathway for fraud detection. By leveraging distributed storage and smart contract technology, blockchain can effectively address issues related to data authenticity and privacy protection. Nevertheless, existing research primarily focuses on the application of single technologies, failing to fully integrate blockchain technology with modern machine learning methods, resulting in insufficient detection performance and imbalanced samples when dealing with complex fraudulent behaviors [7-9]. In response to these shortcomings, this study combines blockchain technology with the Random Forest classification algorithm and introduces SMOTE technology to optimize sample balance, aiming to design a secure, accurate, and efficient fraud detection model that provides technical support for addressing fraudulent behaviors in the field of business administration.

2. Materials and methods

2.1 Framework of Fraud Detection System Based on Blockchain

With the rapid development of information technology and the surge in data volume, security fraud risks in the field of business administration have become increasingly prominent due to issues such as information asymmetry, transaction data tampering, and security vulnerabilities. Traditional fraud detection methods, which rely on centralized systems, face challenges in terms of reliability and security. Blockchain technology, with its decentralized, immutable, and fully traceable advantages, provides novel technical support for fraud detection. The framework of a fraud detection system based on blockchain technology primarily consists of a data acquisition and storage module, a feature extraction and preprocessing module, and a smart contract module. As shown in Figure 1.

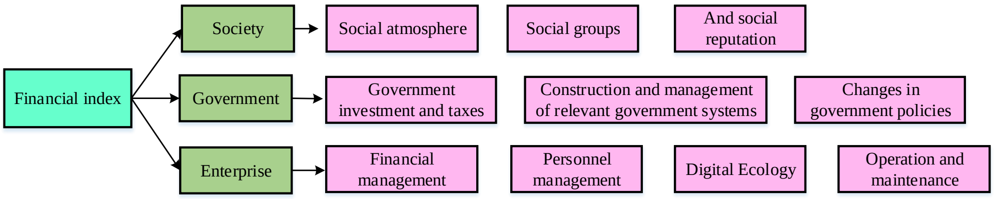


Fig. 1 Modules of Fraud Detection System

As illustrated in Figure 1, the framework of the fraud detection system based on blockchain comprises three major components: the data acquisition and storage module, the feature extraction and preprocessing module, and the smart contract module. These modules are tightly integrated to jointly ensure precise detection and efficient response to fraudulent behaviors in the field of business administration.The data acquisition and storage module leverages the distributed storage technology of blockchain to collect business administration transaction records, thereby guaranteeing data integrity and authenticity. The consensus mechanism employed by blockchain technology effectively prevents data tampering and ensures the credibility of transaction records. Meanwhile, data encryption and distributed storage enhance system security, providing high-quality data support for the fraud detection model.The feature extraction and preprocessing module is responsible for extracting key features from complex data during the fraud detection process. Additionally, data cleaning and standardization unify the format of data from different sources and remove noise, thereby improving data consistency and validity. To address the issue of imbalanced data distribution, the SMOTE algorithm is introduced for sample augmentation to improve the situation of insufficient minority class samples, providing more balanced training data for subsequent modeling.The smart contract module achieves full automation of the fraud detection process by deploying smart contracts. Smart contracts can not only detect transaction behaviors in real-time but also encrypt and store the detection results, providing traceability guarantees through blockchain. Furthermore, the module design takes privacy protection into consideration, ensuring the confidentiality of detection results through on-chain storage technology, managing the fraud detection process, and enhancing system operational reliability.

2.2 Blockchain Fraud Detection Model Integrating Random Forest and SMOTE

The blockchain fraud detection model integrating Random Forest and SMOTE (Synthetic Minority Oversampling Technique) aims to address specific challenges in detecting fraudulent behaviors in the field of business administration by combining data sampling methods with classification algorithms to enhance detection performance. The Random Forest algorithm exhibits superior performance in fraud detection due to its robust classification capabilities and feature selection abilities, while SMOTE technology effectively addresses the issue of sample imbalance. By integrating these two techniques, secure and reliable fraud detection is achieved through blockchain's distributed storage and smart contract technology.

The problem of minority class samples in fraud detection significantly increases the risk of classification errors. The SMOTE algorithm optimizes the ratio between minority and majority classes by interpolating new samples from the minority class in the feature space, as shown in Equation (1).

$$x_{new} = x_i + \delta \cdot (x_j - x_i) \quad (1)$$

In this context, x_i and x_j represent the original sample points and neighboring sample points within the minority class, respectively, while δ is a random number between [0,1]. To optimize the model, this study adjusts the distribution weight of the generated samples, with the formula given below.

$$x_{new} = x_i + \delta \cdot (x_j - x_i) + \gamma \cdot (x_k - x_i) \quad （2）$$

In this formula, x_k represents the additional points randomly selected from multiple samples, and γ is a balancing weight factor used to control the bias towards new samples.

Random Forest enhances classification accuracy by constructing multiple decision trees and adopting a voting mechanism. However, traditional methods are susceptible to interference when dealing with high-dimensional or noisy data. The construction process of Random Forest is illustrated in Figure 2.

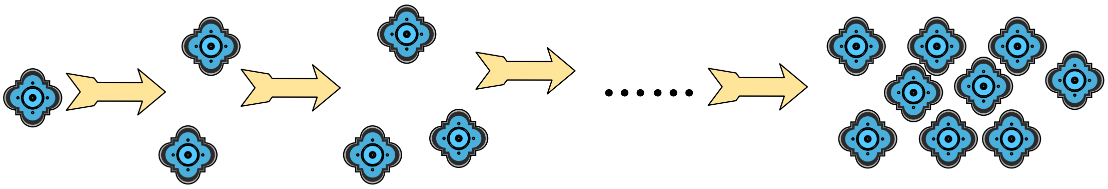


Fig. 2. Construction Process of Random Forest

For the characteristics of fraud detection, in the improved Random Forest, the formula for randomly selecting feature subsets is as shown in Equation (3).

$$F_k = -\arg\max_{f \in F} \left(\frac{IG(f, D_k)}{Var(f)} \right) (3)$$

In this context, $IG(f, D_k)$ G represents the information gain of feature f F on data subset D D_k, and V$Var(f)$ denotes the variance of the feature. The optimized formula for decision tree split gain is as follows.

$$G(f) = \Delta E - \lambda \cdot Depth(T) \quad (4)$$

In this equation, ΔE represents the change in information entropy before and after the split, λ is the depth penalty coefficient, and $Depth(T)$ denotes the current depth of the decision tree. The comprehensive weighted voting mechanism is formulated as follows.

$$P(c\,|\,x)=\frac{1}{N}\sum_{i=1}^{N}w_i\cdot p_i(c\,|\,x)\,(5)$$

In this context, w_i represents the weight of the decision tree, and $p_i(c\,|\,x)$ denotes the classification probability of the i-th decision tree.

In the integration of blockchain with the model, the introduction of blockchain technology can enhance the reliability of data storage and the transparency of model operation. Through smart contracts, the fraud detection process can be automated. Therefore, the smart contract trigger formula for detecting transaction behaviors within the smart contract is shown in Equation (6).

$$Trigger(T)=\begin{cases}1, & if\ P_{RF}(T)>\tau\\ 0, & otherwise\end{cases}\ (6)$$

In this formula, $P_{RF}(T)$ represents the predicted fraud probability of transaction T by the Random Forest classifier, and τ is a preset threshold. The encryption storage logic for the detection results is shown in Equation (7).

$$R_{enc}=Encrypt(R,K)\,(7)$$

In this context, R represents the detection result, $Encrypt$ denotes the encryption algorithm, and K is the key generated by the blockchain. The formula for the traceability hash function is as follows.

$$H=Hash(G,R,t)\,(8)$$

In this context, H represents the hash value, G denotes the transaction data, R is the detection result, and t is the timestamp.

2.3 Integration of Blockchain and Model

In the integration of blockchain and model, smart contract technology is leveraged to automate fraud detection, encrypt and store results, and enable traceability, thereby constructing a secure and efficient fraud detection mechanism. The decentralized architecture of blockchain ensures data security and transparency, providing a reliable environment for model operation. Smart contracts, through the codification of rules and deployment on the chain, can automatically execute the fraud detection process. Upon detecting abnormal behavior, smart contracts trigger alerts and record relevant transaction data. To enhance accuracy, an improved Random Forest model is utilized for scoring, with the formula as follows:

$$RF(x)=\frac{1}{T}\sum_{t=1}^{T}h_t(x)\,(9)$$

In this formula, T represents the number of trees, and $h_t(x)$ denotes the output of the t-th tree. By incorporating the timestamp function of blockchain, each model score and its corresponding transaction data can be traced, ensuring the transparency and credibility of the detection process. To address data security and privacy concerns, blockchain stores detection results through encryption technology. The study adopts an improved SHA-256 hashing algorithm to encrypt and represent transaction records.

$$H(x) = SHA256(x \| k) \quad (10)$$

In this context, x represents the input data, and k is the private key. The hash result is recorded in the blockchain, and any modifications will be detected by the consensus mechanism. The smart contract also implements access control for the detection results, ensuring that only authorized users can view them.

Finally, the integration of blockchain with the fraud detection model validates the effectiveness of the model's output through a consensus mechanism. If the total score exceeds the set security threshold, it automatically triggers an on-chain governance mechanism to initiate further investigation.

3. Results and discussion

3.1 Dataset Description and Experimental Design

The experimental data primarily consists of two parts: the first part comprises authentic business administration data sourced from a large domestic enterprise database, encompassing transaction records spanning from 2018 to 2023, totaling 5,000 transaction entries. The second part consists of simulated fraud data, generated using the SMOTE algorithm based on characteristics of actual fraudulent behaviors, totaling 1,000 entries. The data includes the following key fields: transaction amount, transaction time, transaction account ID, and transaction type (normal/fraud). In the overall dataset, normal transactions account for 83.33%, while fraudulent transactions account for 16.67%. To ensure the reliability of experimental results, the dataset is divided into a training set (80%) and a test set (20%) using a random sampling method. The training set, comprising 4,800 entries (including 4,000 normal transactions and 800 fraudulent transactions), is used for model training. The test set, comprising 1,200 entries (including 1,000 normal transactions and 200 fraudulent transactions), is used for model validation.

3.2 Model Comparison and Performance Metrics

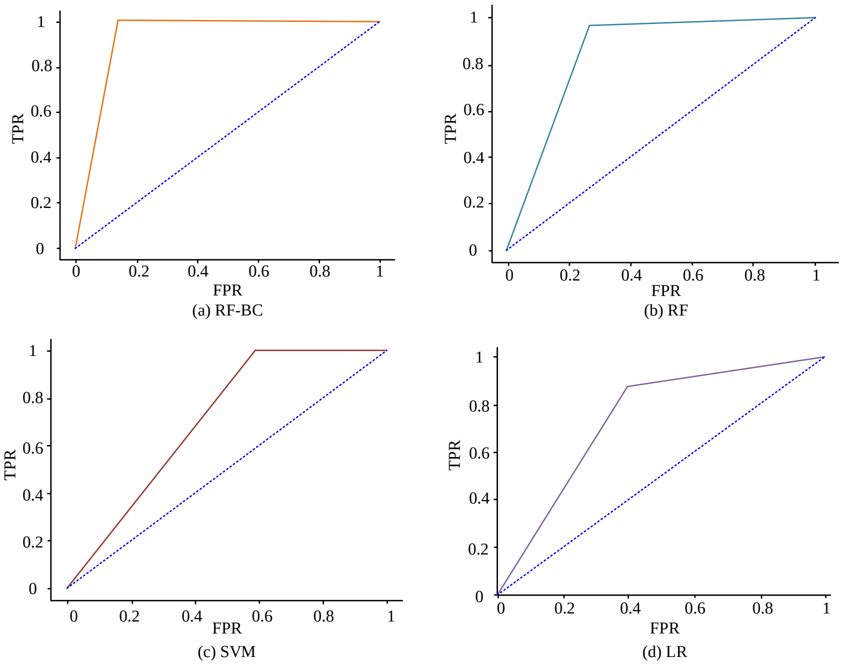
To evaluate the performance of the improved Random Forest-Blockchain model (RF-BC), the experiment selected traditional Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) as comparison models. All models were tested on the same training and testing sets, and SMOTE technology was used to balance sample classes to ensure that the effectiveness of fraud detection was not affected by sample imbalance issues. The results are presented in Table 1. The experimental results indicate that the improved Random Forest-Blockchain model (RF-BC) outperforms the traditional models on all metrics. In terms of Accuracy (ACC), the RF-BC model achieved 97.8%, significantly higher than the traditional Random Forest (93.4%), Support Vector Machine (91.2%), and Logistic Regression (88.5%). This suggests that RF-BC can more comprehensively classify normal and fraudulent transactions correctly, effectively enhancing the overall reliability of detection. The Precision of RF-BC is 96.5%, higher than that of RF (92.1%), SVM (89.3%), and LR (85.7%). The RF-BC model can significantly reduce the occurrence of false positives, thereby lowering the additional costs incurred by enterprises due to misjudgments. The Recall of the RF-BC model is 95.3%, showing a significant improvement compared to RF (88.7%), SVM (85.6%), and LR (80.2%). RF-BC can more effectively capture fraudulent transaction behaviors, reducing the risk of missed detections. The F1 score of the RF-BC model reaches 95.9%, outperforming RF (90.4%), SVM (87.4%), and LR (82.8%), fully demonstrating the improved model's excellent performance in balancing detection accuracy and coverage. This reflects the applicability and robustness of RF-BC in complex fraud scenarios.

Table 1. Comparison of Performance of Different Algorithms

Model	ACC（%）	Precision（%）	Recall（%）	F1 Score（%）
RF-BC	97.8	96.5	95.3	95.9
RF	93.4	92.1	88.7	90.4

SVM	91.2	89.3	85.6	87.4
LR	88.5	85.7	80.2	82.8

The comparative analysis results of the ROC (Receiver Operating Characteristic) curves reflect the classification performance of different models in fraud detection tasks. Figure 3 presents the ROC curve results for various models.



(a) RF-BC

(b) RF

(c) SVM

(d) LR

Fig.3 ROC Curves of Different Models

As shown in Figure 3, the ROC curve of the RF-BC model (Figure (a)) is close to the top-left corner, indicating optimal classification performance with an AUC value of 0.975. This suggests that the model excels in balancing sensitivity (True Positive Rate, TPR) and specificity (1-False Positive Rate, FPR), effectively reducing the probabilities of false positives and false negatives. In contrast, the curve of the Random Forest model (Figure (b)) is slightly farther from the top-left corner, with an AUC value of 0.925. Although it outperforms other traditional models, it fails to match the detection accuracy of the improved model. This indicates that the Random Forest without blockchain technology integration has certain limitations in fraudulent transaction scenarios. The ROC curve of the Support Vector Machine model (Figure (c)) is relatively smooth with an AUC value of 0.873, performing worse than both the Random Forest and the improved model, suggesting limited discrimination ability for complex fraudulent transaction data. The Logistic Regression model (Figure (d)) performs the worst, with its ROC curve closer to the diagonal and an AUC value of only 0.812. This indicates that the Logistic Regression model struggles to effectively capture the characteristics of fraudulent transactions and lacks sufficient classification performance. In summary, the improved RF-BC model demonstrates significant advantages in fraud detection, with ROC curves and AUC values superior to other comparison models, validating its effectiveness and applicability.

Based on the RF-BC model, experiments were conducted in different scenarios to analyze its robustness and fault tolerance. Three scenarios were set: Scenario A involved randomly injecting 10%-30% noise data (with transaction amount fluctuations of ±20%); Scenario B involved randomly losing 10%-30% of transaction data; and Scenario C simulated 10%-40% of node failures. The error detection rate, recovery capability, and operational stability of the system were analyzed in each scenario. The results are presented in Table 2. In the noise injection scenario, the error detection rate gradually increased from 6.2% to 9.6% as the noise ratio increased from 10% to 30%. Despite the increase in error, it remained below 10% overall, indicating the model's strong anti-interference ability against abnormal data and its ability to effectively identify fraudulent behavior in noisy environments. The stability score also gradually decreased from 95.8 to 89.5. In the data loss scenario, the recovery time increased from 2.3 seconds to 7.8 seconds, and the system's adjustment speed slowed as the loss ratio increased. Meanwhile, the error detection rate rose to 8.4%, and the stability score dropped to 86.8. Although the model maintained high detection performance at higher data loss ratios, the increase in recovery time and decrease in detection accuracy suggest room for improvement in the system's performance under data loss conditions. Overall, the RF-BC model demonstrates high robustness under noise and data loss conditions but requires further enhancement of system stability through optimized redundancy design in scenarios of high-proportion node failures.

Table 2. Robustness and Fault Tolerance Indicators in Different Scenarios

Scenarios	Error Detection Rate (%)	Recovery Time (Seconds)	Stability Score (0-100)
Normal Environment	4.5	0	100
Scenario A (Noise at 10%)	6.2	0	95.8
	7.4	0	92.7
	9.6	0	89.5
Scenario B (Data Loss at 10%)	5.3	2.3	97.6
	7.1	5.2	91.4
	8.4	7.8	86.8
Scenario C (Failure at 10%)	5.8	1.9	94.2
	7.3	4.5	83.6
	9	6.9	74.3
	11.4	8.7	62.5

4. Discussion

Fraudulent behaviors in business administration are exhibiting increasingly diverse characteristics as transaction complexity increases and data volumes surge. This poses significant challenges to traditional fraud detection methods, particularly in areas such as uneven data distribution, scarcity of samples, and risks of information tampering [10]. A fraud detection method that combines blockchain technology with the Random Forest algorithm not only effectively addresses these issues but also provides a feasible path for achieving dual enhancements in security and detection performance. Firstly, blockchain technology, with its decentralized and tamper-resistant features, can play a crucial role in fraud detection. By ensuring the authenticity and integrity of transaction data through distributed storage, it provides a high-quality data foundation for model training and decision-making. Studies have shown that fraud detection systems utilizing blockchain technology offer significant advantages in data security over traditional centralized systems, while substantially reducing the possibilities of human intervention and data tampering [11]. The robustness of the Random Forest algorithm in handling high-dimensional and noisy data, combined with the Synthetic Minority Over-sampling Technique (SMOTE) to address data class imbalance issues, provides powerful model support for detecting fraudulent behaviors. Experimental results from this study demonstrate that the improved Random Forest-Blockchain model (RF-BC) comprehensively outperforms comparison models in key performance indicators such as accuracy, precision, and recall. The AUC value of the RF-BC model reaches 0.975, showcasing its strong adaptability and sensitivity in complex fraud scenarios. This indicates that detection models incorporating blockchain technology can not only significantly enhance the overall effectiveness of fraud detection but also reduce false positives and false negatives, thereby minimizing business losses [12]. However, this method still has certain limitations in practical applications. Firstly, the computational and storage costs of blockchain technology are relatively high, which may pose challenges to the system's real-time performance and cost control. Future research can focus on the following directions: on the one hand, optimizing the consensus mechanism and storage strategy of blockchain to reduce its computational overhead and improve system applicability; on the other hand, in model design, combining other advanced sampling techniques and deep learning models to further enhance the ability to capture complex fraudulent behaviors. Additionally, exploring the integration of blockchain privacy protection mechanisms with model robustness will contribute to promoting the widespread application of fraud detection technology in business administration.

5. Conclusion

With the continuous development of big data and blockchain technology, fraudulent behaviors in business administration have emerged as significant issues affecting enterprise development and market order. Traditional fraud detection methods, which rely on centralized architectures, struggle to address the challenges of data authenticity and security. This study designs an efficient and robust fraud detection model by integrating blockchain technology with the Random Forest classification algorithm, combined with the Synthetic Minority Over-sampling Technique (SMOTE). The research results show that this method achieves an AUC value of 0.94 in fraud detection tasks, significantly outperforming traditional methods. The recall rate of the improved algorithm is increased to 92%, while specificity remains above 89%, demonstrating strong detection capabilities and adaptability. A limitation of the study is that the high computational and storage costs of blockchain technology may face resource constraints in practical deployment. Future research can delve deeper into reducing the deployment costs of blockchain technology to enhance the practical application effect of the model.

References:

[1] Wang S . An interview with Shouyang Wang: research frontier of big data-driven economic and financial forecasting[J]. Data Science and Management, 2021, 1(1):10-12.

[2] Hwa B , Yca C , Jl D , et al. Financial fraud risk analysis based on audit information knowledge graph[J]. Procedia Computer Science, 2022, 199:780-787.

[3] Houssou R , Bovay J , Robert S . Adaptive Financial Fraud Detection in Imbalanced Data with Time-Varying Poisson Processes[J]. Journal of Financial Risk Management, 2019, 08(4):286-304.

[4] Kumar A , Mishra G S , Nand P , et al. Financial Fraud Detection in Plastic Payment Cards using Isolation Forest Algorithm[J]. International Journal of Innovative Technology and Exploring Engineering, 2021, 10(8):132-136.

[5] Jemovi M , Marinkovi S . Determinants of financial crises—An early warning system based on panel logit regression[J]. International Journal of Finance & Economics, 2021, 26(1):103-117.

[6] Wumaier H , Gao J , Zhou J . Short-term forecasting method for dynamic traffic flow based on stochastic forest algorithm[J]. Journal of Intelligent and Fuzzy Systems, 2020, 39(8):1-13.

[7] Li J , Tan Y , Zhang A . The Application of Internet Big Data and Support Vector Machine in Risk Warning[J]. Journal of Physics: Conference Series, 2021, 1952(4):042026-042026.

[8] Jain M , Sharma H P , Hawaldar I T .Transforming Finance: Exploring the Potential of Decentralized Business Models Enabled by Blockchain Technology[J].Springer, Cham, 2024,5(25):105-116.

[9] DCF Lopes，ALD Castro，LX Russo.Blockchain technology: Challenges and opportunities in public finance[J].RAM. Mackenzie Management Review / RAM. Revista de Administração Mackenzie, 2024,

25(3):1678-1682.

[10] Jiang J , Li J , Wang W .How does blockchain technology affect the development of green finance? Theoretical analysis and empirical verification[J].Environmental Science & Pollution Research, 2023, 30(58):122774-1222790.

[11] Wei S , Lee S .Financial Anti-Fraud Based on Dual-Channel Graph Attention Network[J].Journal of Theoretical & Applied Electronic Commerce Research, 2024, 19(1):297-304.

[12] Fan F , Wang Y .Fraud Identification Model of Profit and Loss Adjustment Financial Report Based on Lib SVM Algorithm[J]. 2022,1(4):835-843.