This study addresses the pain points of traditional credit risk control and fraud detection, and constructs a big data-driven intelligent risk control and fraud detection model system. Experimental results demonstrate that the hybrid intelligent model based on XGBoost and BiLSTM achieves an accuracy of 92% and an AUC value of 0.93 on the test set, significantly outperforming traditional single models.
Although this study has achieved significant results, there are still some limitations, such as the use of a single commercial bank dataset, model architecture complexity, and time constraints on model inference due to real-time requirements. Based on the findings and limitations of this study, further research on related issues of data utilization and model optimization is expected.

Springer

4th International Conference on Computational Finance and Business Analytics

# A Hybrid XGBoost-BiLSTM Model for Credit Risk Control and Fraud Detection

CFBA-2609

Xinru Zhang and Yifan Zhang

Xuzhou Vocational Technology Academy of Finance & Economics

## Introduction or abstract

Traditional credit risk control and fraud detection systems face pain points such as single data dimensions, poor model timeliness, and lagging fraud identification, making it difficult to meet the risk control needs of massive credit business in the digital finance era. Based on a big data technology architecture, this paper integrates multi-source heterogeneous credit data to construct an intelligent risk control and fraud detection model system that combines machine learning and deep learning. First, it analyzes the core pain points of credit risk control and fraud detection and the empowerment path of big data technology; second, it designs a multi-source data preprocessing and feature engineering scheme to solve problems such as data noise and dimensionality disaster; finally, it constructs a hybrid intelligent model based on XGBoost-BiLSTM and verifies the model performance through real credit datasets.

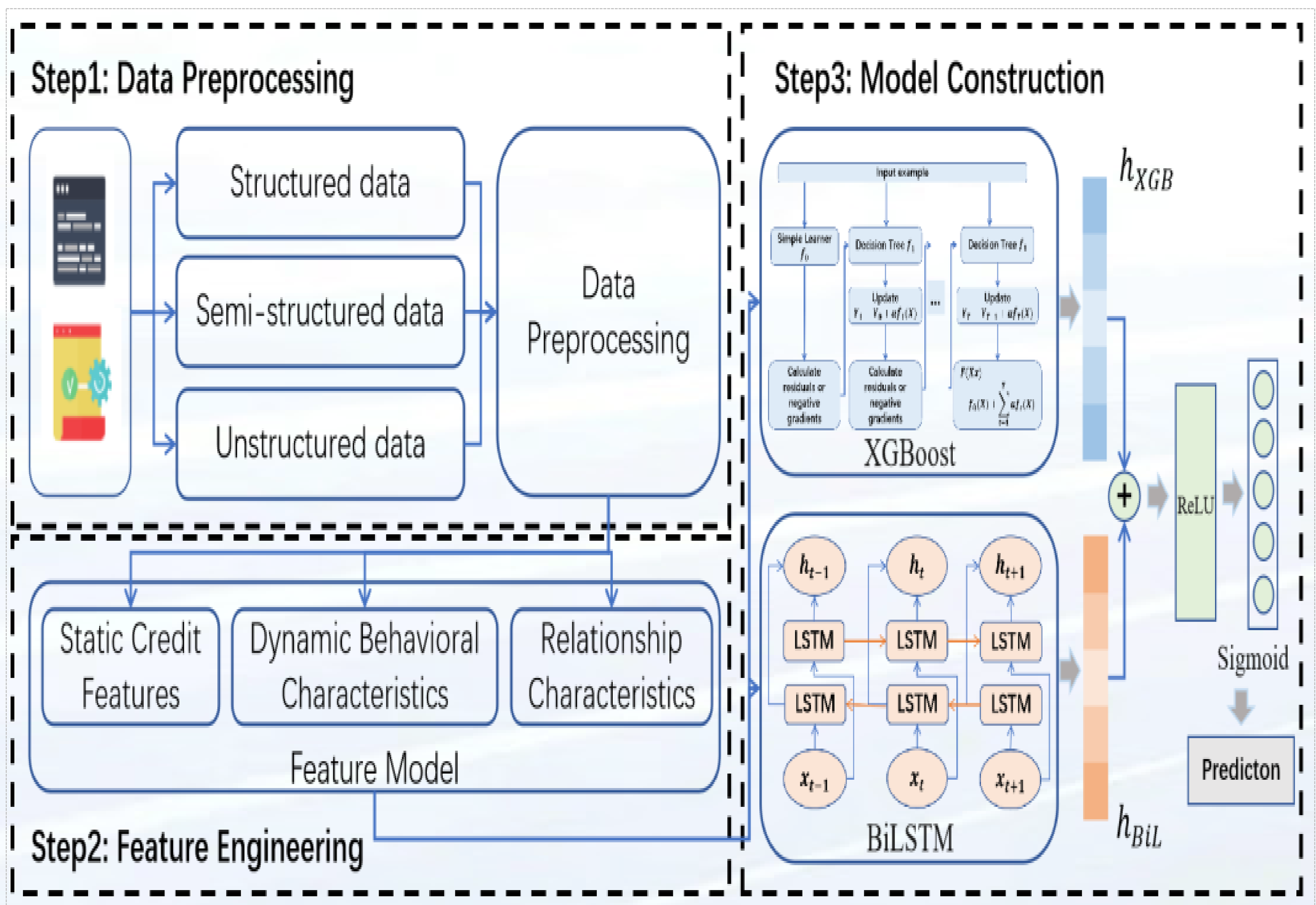
## Objectives

1.Propose a fusion preprocessing framework for multi-source heterogeneous data to solve the problem of traditional data singularity.

2.Design a hierarchical feature engineering scheme oriented towards credit risk control to enhance the representation ability of features for risk factors.

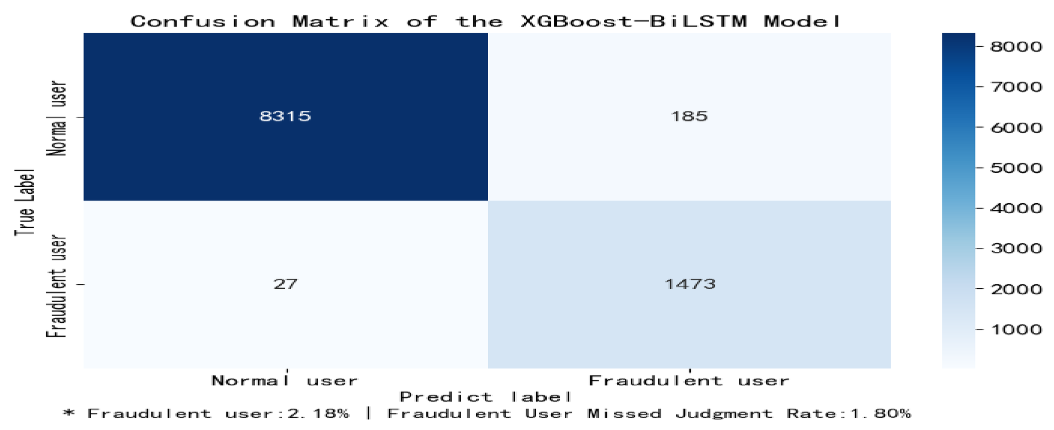
3.Construct an hybrid intelligent model based on XGBoost-BiLSTM. The model integrates static credit assessment and dynamic behavior sequence analysis to achieve collaborative decision-making in risk control and fraud detection.

## Materials & Methods




First, analyze the core pain points of credit risk control and fraud detection, as well as the empowerment path of big data technology; second, design a multi-source data preprocessing and feature engineering scheme to address issues such as data noise and dimensionality disaster; finally, construct a hybrid intelligent model based on XGBoost-BiLSTM and verify the model performance using real credit datasets.

## Results




Experimental results demonstrate that the proposed method improves model performance by more than 10% compared to models using a single data source and achieves an accuracy of 92% and an AUC value of 0.93 on the test set, significantly outperforming traditional single models.

## Conlcusion

This study addresses the pain points of traditional credit risk control and fraud detection, and constructs a big data-driven intelligent risk control and fraud detection model system. Experimental results demonstrate that the hybrid intelligent model based on XGBoost and BiLSTM achieves an accuracy of 92% and an AUC value of 0.93 on the test set, significantly outperforming traditional single models.
Although this study has achieved significant results, there are still some limitations, such as the use of a single commercial bank dataset, model architecture complexity, and time constraints on model inference due to real-time requirements. Based on the findings and limitations of this study, further research on related issues of data utilization and model optimization is expected.

## References

1. Wei L, Dong J, Yu H. NERHF: a hybrid machine learning-driven efficient credit risk control framework. Sci Rep. Dec 1(2025)
2. Dongmei L, Liping L, Research on Efficiency in Credit Risk Prediction Using Logistic-SBM Model, Wireless Communications and Mobile Computing, 5986295 (2022).
3. Amarnadh, V., Moparthi, N.R. Range control-based class imbalance and optimized granular elastic net regression feature selection for credit risk assessment. Knowl Inf Syst 66, 5281–5310 (2024).
4. Zhang Y , Ye S , Liu J ,et al: Can digital transformation of commercial banks reduce green credit risks?. International Review of Financial Analysis 99(2025).
5. Hosseini, Mohammad R, Godini, et al: Comparing classic regression with credit scorecard model for predicting sick building syndrome risk: A machine learning perspective in environmental assessment. Building and Environment, vol.253,0360-1323(2024).
6. Lu H, Xiuying W, et al: Integrated multivariate segmentation tree for heterogeneous credit data analysis in small- and medium-sized enterprises. Expert Systems with Applications vol.300, 0967-4174(2026).
7. Chang, V., Xu, Q.A., Akinloye, S.H. et al.: Prediction of bank credit worthiness through credit risk analysis: an explainable machine learning study. Ann Oper Res 354, 247–271 (2025).
8. Song, Y., An, X., Zhang, W., Sun, K., & Zhang, A.: Catalyzing Financial Risk Control Excellence: A Novel Fusion Model - PSO-Boost-Trans. Journal of Organizational and End User Computing, 36(1), 1-29(2024).
9. RezvaniNejad, Masoud and Ali Sabzali Yameqani.: WDAE-GAN: A hybrid dual autoencoder and generative adversarial framework with wavelet denoising for credit card fraud detection. Expert Syst. Appl., vol.299 (2025).
10. Dubey, Parul et al.: Innovative IoT-enabled mask detection system: A hybrid deep learning approach for public health applications. MethodsX, vol. 14, 103291.(2025).