Deep Neural Networks and Hybrid Architectures for Credit Card Fraud Prevention

Jiayi Liu

School of Economics, Beijing Technology and Business University, Beijing, 100048, China
ljy2377819563@outlook.com

Abstract. Credit card fraud is becoming increasingly serious in recent years for people at home and abroad. The three representative ways in this paper are: Deep Neural Networks for multi-feature integration; a hybrid Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM) with attention mechanisms for spatiotemporal transaction modeling; and Generative Adversarial Networks for dealing with class imbalance through synthetic fraud data generation. Experimental comparison on benchmark datasets shows that hybrid models integrating CNNs, LSTMs and attention mechanisms have superior performance, with an accuracy exceeding 99% and a recall of 0.89 on standard benchmarks. Based on the above, although different parts of a single deep learning model can be used for fraud detection, more effective results are achieved by combining several such models in a hybrid system. The paper also points out the following problems: data imbalance, lack of model interpretability, requirements for real-time processing, and changes in fraudulent behaviour. The above results offer a system of reference for researchers and practitioners engaged in the construction of strong, high-performance credit card fraud detection systems in actual financial environments.

Keywords: CNNs, LSTM, GAN, Credit Card Fraud

1 Introduction

At present, with the development of the financial system, a serious problem of credit card fraud has occurred. With the all-around development of global digital transactions and the wider application of cashless payment systems, both the scale and severity of fraudulent behaviour have been increasing. By 2027, due to credit card fraud, the financial losses for all parties - consumers, merchants, and financial institutions - across the country will be more than US$40 billion [1]. Fraud losses across the entire country in the United States reached over $10 billion in 2023, and this trend of continuous increase highlights the deficiencies of the current safeguards [2]. Fraud will cause damage to the financial health of people's banks and increase operational expenses for banks and card issuers.

© The Author(s) 2026
L. C. Yoong (ed.), *Proceedings of the 2026 2nd International Conference on Financial Innovation and Marketing Management (FIMM 2026)*, Advances in Economics, Business and Management Research 408,
https://doi.org/10.2991/978-94-6239-801-6_152

The main types of traditional fraud detection systems are based on rules and, thus, can only identify problems in the transaction speed, location or amount. Although the above systems offer some protection, they are not entirely free of flaws. Fraudsters are constantly improving their methods to bypass the old rules. The other is that the rule-based method has a high false-positive rate and will therefore bother actual cardholders and increase the verification workload for banks. The first few models of machine learning included logistic regression, support vector machines and random forests, all of which learn patterns from the previous history of transaction data. Recently, some researchers have been studying the performance of these. Sangit and Sangam studied cutting-edge deep learning methods for credit card fraud detection, and several models based on artificial neural networks and hybrid deep learning were introduced [3]. Bian has investigated new studies in machine learning and deep learning for credit card fraud detection, including traditional rule-based systems as well as advanced ones such as GNNs (Graph Neural Networks), LSTMs, and GANs (Generative Adversarial Networks), and has also pointed out problems of interpretability and data management [4].

Although the performance of rule-based systems has gradually improved in recent years, they are no longer suitable for the high-demanding conditions of modern fraud detection. The most serious problem is severe class imbalance; fraudulent transactions account for less than 0.2% of all transactions, and thus the model tends to be biased towards the majority legitimate class. In addition, due to the change in the shape of fraud patterns, the model should be able to adapt to concept drift and has a low-latency requirement for real-time processing. Recently, with the development of deep learning technology, the above problems have been effectively addressed; it can learn features and representations automatically to improve pattern recognition performance.

This paper provides an all-round review of deep learning architectures for credit card fraud prevention and introduces three typical types: Deep Neural Networks for multi-feature fusion, hybrid CNN-LSTM with attention mechanisms for spatiotemporal modelling, and Generative Adversarial Networks to handle class imbalance. Systematically compare the above methods in terms of architectural design, performance characteristics and practical applications to identify effective strategies for real-world fraud detection deployments.

2 Methodology

2.1 Datasets

Several benchmark datasets have been released to assist in the progress of deep learning applications for credit card fraud detection. The most frequently used dataset is the Kaggle Credit Card Fraud Detection dataset, which contains 284,807 transactions of European cardholders in September 2013, and among them, 492 were fraudulent and accounted for 0.172% of the total [5]. This dataset has a serious class imbalance problem, and this is also

the case for actual fraud detection. An additional dataset is the IEEE-CIS Fraud Detection dataset, which contains approximately 590,000 transactions and 393 attributes; many synthetic datasets have also been generated by banks for research cooperation. Preprocessing of the above datasets generally includes feature scaling, handling missing values, and converting categorical variables through encoding.

2.2 Deep Learning Models for Fraud Detection

Deep Neural Networks (DNNs). Deep Neural Networks are the basic structures of deep learning-based fraud detection. A DNN has an input layer for transaction features, several hidden layers to learn hierarchical representations gradually, and an output layer for a binary classification decision. Backpropagation is used by the network, and ReLU is an activation function for non-linearity. DNNs in fraud detection combine various features, such as transaction amount, merchant category, time and customer behaviour, to create a joint-space representation. Deep neural networks (DNNs) can model the non-linear relationship between the features of transactions and the occurrence of fraud automatically, thus reducing the need for manual feature engineering.

Hybrid CNN-LSTM with Attention Mechanism. A hybrid CNN-LSTM network with an attention mechanism has been developed to address the limitations of individual DNNs in extracting both spatial and temporal features from transaction data. Convolutional Neural Networks are feature extractors that learn to recognize local patterns in the data, such as clusters of unusual expenditures or abnormal merchant categories. Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network that can remember information over extended periods to understand how people have been spending money in the past and detect any abnormal behaviour in their spending. Dynamically weight the importance of different features through an attention mechanism, focus on essential indicators of fraudulent behaviour, and ignore irrelevant information in routine transactions. According to the above study, the accuracy is as high as 99.93% and the recall is 0.89 for the Credit Card Fraud dataset. The combined effect of the above modules is to obtain spatiotemporal feature representations and identify abnormal behaviour over time in fraud data.

Generative Adversarial Networks (GANs). GANs help to solve the problem of class imbalance in the training data for fraud detection. The two competing neural networks in the GAN architecture are a generator to create fake fraudulent transaction samples and a discriminator to determine whether these samples are real or not. Adversarial training is used to make the generator create realistic fake transaction data and increase the number of these synthetic samples to balance the classes. The high-end version has been added to

provide more realistic synthetic data for sophisticated fraudulent behaviours via a gated attention mechanism. Experimental results show that GAN-based augmentation can improve the detection performance of oversampling compared to traditional oversampling methods such as SMOTE, and the GA-GAN framework has outperformed the best-in-class method for class-imbalanced credit card datasets.

2.3 Comparative Analysis Framework

To conduct a systematic comparison of the different deep learning models, the same set of evaluation indicators will be employed. The first is the general proportion of correct classifications; the second is the ratio of predicted fraud cases that are actually fraudulent. Recall is also known as the rate of true positives per all actual fraud occurrences. The F1 score is a combination of precision and recall; therefore, both are required. AUC measures how well a model performs at different thresholds in a binary classification problem. Practically speaking, the latency of inference, training time, interpretability and adaptability to concept drift are also considered to determine whether it can be used in practice.

3 Results

3.1 Experimental Configuration and Datasets

The papers listed in this paper used many public datasets for model evaluation. Table 1 shows the features of the first set of data from some typical deep learning-based fraud detection studies.

Table 1. Summary of Benchmark Datasets for Credit Card Fraud Detection

Dataset	Source	Transactions	Fraud Cases	Fraud Ratio	Year
Kaggle CCF	European cardholders	284,807	492	0.172%	2013
IEEE-CIS	Vesta Corporation	590,000	—	0.151%	2019
Synthetic	Various institutions	555,719+	—	≤0.200%	2024

3.2 Performance Comparison

Table 2 shows the results of a comparison of several deep learning models for credit card fraud detection in recent studies and their corresponding performance indices.

Table 2. Performance comparison of Deep Learning Models for Credit Card Fraud Detection

Model	Accuracy	Precision	Recall	F1 Score	AUC
DNN	95.2%–98.5%	0.82–0.91	0.78–0.88	0.80–0.89	0.96–0.98
CNN-LSTM-Attention	99.93%	0.98	0.89	—	—
GA-GAN	98.3%–99.1%	0.94–0.97	0.91–0.95	0.92–0.96	0.97–0.99
LSTM with SHAP	—	—	—	0.76–0.84	—
SV2LSTM	99.99%–100%	99.99%–100%	1.00	—	—
Limitations	No temporal modeling, high FPR	High computational cost	Training instability, mode collapse	Requires user interaction	Complex architecture, overfitting risk

The experiment shows that hybrid architectures have been better than single-model approaches in terms of many indices. A CNN-LSTM-Attention model has achieved a high accuracy of 99.93% and significantly surpassed that of a DNN alone. SV2LSTM is a model that uses the soft voting ensemble learning method to combine an Autoencoder-LSTM sequential model, and it has shown good accuracy and precision across multiple datasets. GA-GAN is a class-imbalance-aware model that achieved F1 scores of 0.92 to 0.96, significantly exceeding those of traditional SMOTE-based augmentation. Among them, the recall metric has shown more difficulty in all the models; it is difficult to detect rare instances of fraud without generating many false alarms.

4 Discussion

4.1 Key Findings and Interpretations

Comparison and analysis of how deep learning has been used for credit card fraud detection. First, hybrid architectures that combine several learning paradigms consistently outperform single-paradigm models. The CNN-LSTM-Attention model has achieved the best results so far by extracting spatial features and temporal sequences simultaneously and focusing on key elements in the data dynamically. Based on the above analysis, the next stage of work should be to detect anomalous spending behaviour through sequential analysis of transactions. Second, Class imbalance is also one of the reasons for poor detection performance. GAN-based methods have shown that generating realistic synthetic fraud samples can significantly increase the recall rate without decreasing the precision, and GA-GAN is a typical case study that has achieved a higher F1 score than traditional oversampling methods. Third, the models need to be more understandable in order to be

applied in practice. SHAP can be used to provide interpretability and explainability for LSTM models in the context of a continuous-learning user-feedback loop.

4.2 Challenges and Limitations

There are still some problems now. All detection systems have been affected by the problem of data imbalance; as a result, very few examples of fraudulent behaviour are available for model training. As Bian pointed out, deep learning models are generally less interpretable because of their "black-box" nature, and problems such as class imbalance and fluctuating fraud patterns have reduced the model's performance; furthermore, data deficiencies, such as limited availability and privacy issues [6], have been further impeding system construction. Ge and others believe that due to the problem of class imbalance in credit card transaction data, it is difficult to train suitable models, so they have developed GA-GAN [7]. Real-time processing is also an engineering problem; otherwise, the complex deep learning model for transaction authorisation would have too high a latency. Fraud is adversarial by nature; thus, there is an ongoing arms race and model adaptation to new attack methods. Cheng and others have shown that the traditional rule-based system and classical machine learning method are unable to handle the complicated fraudulent behaviour in a large volume of financial data [8]. Deep learning models are also relatively computationally expensive and not very suitable for resource-limited banks.

4.3 Future Directions

The five future research directions are as follows. First, develop more explainable AI frameworks that can meet the regulatory requirements for transparency and help build trust with users. Second, due to privacy issues of the data, federated learning can be used to train a model jointly across multiple banks without sharing sensitive transaction data. Thirdly, graph neural networks are emerging from research, and since transaction networks have an inherent graph structure, GNNs can be applied to learn relational patterns and identify coordinated fraud rings. Fourthly, a learning mechanism will be employed to address concept drift and continuously update the model in light of modifications to the pattern of fraud. Finally, to ensure that the laboratory results can be applied in practice, stringent tests need to be conducted in a real-world environment, such as clinical trials and A/B tests in the production area [9, 10].

5 Conclusion

Systematically review the deep learning methods for credit card fraud prevention in this paper, and focus on the three representative architectures: Deep Neural Networks for multi-feature integration, hybrid CNN-LSTM with attention mechanisms for spatiotemporal

modeling, and Generative Adversarial Networks for handling class imbalance. Based on the above comparison, hybrid architectures that combine several learning paradigms have shown excellent performance, and the CNN-LSTM-Attention model achieved a remarkable accuracy of 99.93% and the SV2LSTM model showed almost perfect detection in all the test sets. GAN-based augmentation generates synthetic fraud samples to address the problem of severe class imbalance in transaction data and improves the recall and F1 score. There are still some defects: the model is difficult to interpret, computationally expensive, not suitable for real-time applications, and changes in the pattern of new fraud require frequent model updates.

Explainable AI frameworks will be favoured in the future to offer transparent and trustworthy predictions for regulatory compliance in research. Federated Learning can be used to address the problem of data privacy and conduct collaborative model training among multiple organisations. Graph neural networks have shown that the previous architecture cannot effectively capture the relational structure of the transaction network. The results of this review provide a system for reference for other scholars and practitioners to build high-performance, cost-effective, and practical credit card fraud detection systems in actual financial institutions. As fraudsters continuously develop more sophisticated attack methods, the development of deep learning architecture will continue to be used in the future to protect the safety and stability of the world's payment systems.

References

1. The Nilson Report: Nilson Report Issue 1236: Global Card Fraud Losses. Oxnard, CA (2023)
2. Federal Trade Commission (FTC): Consumer Sentinel Network Data Book 2023. Washington, D.C. (2024)
3. P. N. Sangit and S. Sangam, "The role of deep learning in credit card fraud detection: A state-of-the-art review," Cureus J. Comput. Sci., vol. 2, 2025, Art. no. es44389-025-04298-6, doi: 10.7759/s44389-025-04298-6.
4. C. Bian, "Advances in machine learning and deep learning for credit card fraud detection: A review," ITM Web Conf., vol. 78, 2025, Art. no. 02023.
5. "Hybrid CNN-LSTM with attention mechanism for robust credit card fraud detection," IEEE Access, vol. 13, pp. 114056–114068, Jun. 25, 2025, doi: 10.1109/ACCESS.2025.3583253.
6. Bian, C.: Credit card fraud detection: Machine learning and deep learning advances, challenges, and future directions. ITM Web of Conferences 78, 02023 (2025)
7. Ge, J., Yin, L., Zhang, S., Zhao, X.: Gated attention based generative adversarial networks for imbalanced credit card fraud detection. PeerJ Computer Science 11, e2972 (2025)

8. Cheng, D., Zou, Y., Xiang, S., Jiang, C.: Graph neural networks for financial fraud detection: A review. Frontiers of Computer Science 19(9) (2025)
9. Awoyemi, J.O., Adentumbi, A.O., Oluwadare, S.A.: Credit card fraud detection using machine learning techniques: A comparative analysis. In: ICCNI 2017, pp. 1–9 (2017)
10. Mohbey, K.K., Khan, M.Z., Indian, A.: Credit card fraud prediction using XGBoost: An ensemble learning approach. International Journal of Information Retrieval Research 12(2), 1–17 (2022)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

