ITM Web of Conferences **78**, 02023 (2025) https://doi.org/10.1051/itmconf/20257802023
CSEIT 2025

Credit Card Fraud Detection: Machine Learning and Deep Learning Advances, Challenges, and Future Directions

Chaobo Bian

School of Economics and Management, Xidian University, Xi'an, China

Abstract. Credit card fraud poses significant risks to financial institutions and cardholders, necessitating robust detection systems. This paper reviews advancements in machine learning and deep learning for credit card fraud detection, including traditional rule-based systems and advanced deep learning methods like ANNs, GNNs, LSTMs, and GANs. These techniques enhance detection accuracy but face challenges in interpretability, applicability, and data management. For instance, deep learning models often lack transparency due to their black-box nature, while class imbalance and dynamic fraud patterns complicate model applicability. Data issues, such as limited availability and privacy concerns, further hinder system development. To address these challenges, this study proposes theoretical solutions: integrating expert systems to improve model interpretability, employing hybrid techniques like GAN-SMOTE combinations and federated learning to enhance applicability, and leveraging advanced methods for data quality and privacy preservation. The findings highlight the need for future research to balance model complexity with transparency, ensuring scalability and addressing data limitations. These insights aim to guide the development of more efficient, accurate, and robust fraud detection frameworks to safeguard financial security and support the sustainable growth of the credit card market.

1 Introduction

Credit card fraud refers to activities aimed at obtaining illicit benefits without the cardholder's permission. Common types of credit card fraud include unauthorized transactions and cash-out schemes. With the rapid development of Internet technology, the convenience of credit cards has led to their widespread use in all aspects of social life, significantly expanding the cardholder population and providing strong support for economic growth. Despite robust legal frameworks in place to protect this domain, credit card fraud still occurs, causing direct economic losses to financial institutions and potentially damaging cardholders' credit records, thereby affecting their quality of life. Therefore, developing efficient and accurate fraud transaction detection systems and establishing corresponding emergency response mechanisms are of crucial importance for maintaining the healthy operation of the credit card market.

chaobo@vt.edu

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

In the field of credit card fraud detection, Stolfo et al. proposed a meta-learning-based multi-classification strategy aimed at identifying and preventing credit card fraud activities [1]. In the same year, Chan P.K. and WFan focused on the application of distributed data mining techniques in credit card fraud detection [2]. This process involves integrating and applying various data mining algorithms and models in a distributed computing environment to analyse large-scale credit card transaction datasets, with the goal of identifying any anomalies that may indicate fraudulent behaviour. The strategy employed by Chan et al. is designed to enhance the precision of credit card fraud detection while effectively controlling the false alarm rate.

In related literature, Brause et al. in 2023 explored the application of neural data mining techniques in identifying credit card transaction fraud [3]. They described how this method enables the real-time detection of fraudulent activities, thereby effectively preventing potential financial risks and safeguarding both corporate assets and cardholder interests. The construction of neural networks involves training with datasets that include records of both legitimate and fraudulent transactions, with the aim of distinguishing between the two types of transactions. Subsequently, the neural network applies the learned patterns to classify new transaction instances, identifying whether they are legitimate or fraudulent.

Given that machine learning algorithms can provide more accurate fraud detection capabilities and effectively reduce false positives, their application in this field is increasingly valued. Yao proposed an innovative strategy that integrates traditional logistic regression models with adaptive boosting algorithms to construct a superior ensemble classifier [4]. Compared to a single traditional logistic regression classifier, this ensemble model has demonstrated significantly enhanced performance in empirical evaluations. Similarly, Khan et al. integrated principal component analysis as a data mining technique with the fuzzy C-means clustering algorithm, successfully conducting a comprehensive assessment of various algorithm configurations to achieve higher classification accuracy [5].

As can be seen from the above, the field of credit card fraud detection has witnessed a series of novel advancements and methodologies, ranging from the application of traditional machine learning algorithms to the latest deep learning techniques. In light of this, there is a need to conduct a comprehensive literature review aimed at summarizing existing research findings, analysing technological bottlenecks, and indicating prospective pathways for future research.

The remainder of this paper is structured as follows. Initially, this paper will outline the typical methodology employed in this domain. Subsequently, the review will engage in a discussion of the findings, comparing the performance of different approaches and identifying challenges within the current research landscape. This will include an exploration of issues such as data-related constraints and the interpretability of results, along with potential solutions and future research avenues. Finally, the paper will conclude with a summary of the research outcomes, underscoring the implications and potential applications of the study, while also acknowledging its limitations and suggesting directions for further investigation.

2 Method

2.1 Traditional approaches

2.1.1 Rule-based Systems

Vaquero has pointed out that rule-based systems have long been employed to detect potentially fraudulent transactions by applying predefined rules [6]. These rules can range from relatively simple ones, such as flagging transactions that exceed a certain amount, to more complex ones that involve identifying unusual transaction patterns. For instance, if a single card is used to make purchases in two different countries within a very short time frame, the system will flag this as suspicious. Essentially, these systems rely on a set of logic-based rules to classify data into either non-suspicious or suspicious activities. To illustrate, if a transaction surpasses a specific predefined limit, it will be identified as a potential candidate for fraudulent activity. Additionally, another rule might stipulate that if a series of transactions occur in quick succession, this pattern will also be flagged as suspicious.

2.1.2 Manual Verification

Manual verification is a prevalent method for detecting credit card fraud, which entails human analysis and verification of transactions. For instance, if a transaction deviates from a customer’s typical spending patterns, the credit card company may contact the customer directly to confirm the transaction. Similarly, transactions made in locations significantly distant from the customer’s usual whereabouts may be flagged for manual verification [6].

2.1.3. Authentication Measures

In the fight against credit card fraud, diverse authentication steps have been adopted to boost security. These steps generally call for extra details during transactions, like a Personal Identification Number (PIN) or a postal code. Certain systems also utilize challenge questions to confirm the cardholder's identity. This strategy is based on the concept of multifactor authentication, which involves verifying the cardholder's identity via multiple forms of proof. These can include knowledge-based factors (e.g., passwords), possession-based factors (e.g., physical cards), and biometric factors (e.g., fingerprints) [7].

2.1.4. Random Forest (RF)

Lin and Jiang stress the key part the probabilistic Random Forest (RF) plays in their AE-PRF method for credit card fraud detection [8]. As an ensemble learning technique, the RF model builds multiple decision trees through bootstrap aggregating (bagging) to improve robustness and reduce overfitting. Each tree is trained on a random selection of data and features, ensuring varied viewpoints on the dataset. The RF model yields the likelihood of a transaction being fraudulent, enabling flexible classification with adjustable detection thresholds. This probabilistic method is vital for balancing the detection of fraudulent activities and the minimization of false alarms. The RF's natural capability to manage imbalanced data makes it especially effective for credit card fraud detection, where normal transactions far exceed fraudulent ones.

2.2 Deep Learning-based approaches

2.2.1 ANN

Asha and Suresh investigated the utilization of Artificial Neural Networks (ANN) in detecting credit card fraud and made a comparison with other machine learning algorithms [9]. When contrasted with other machine learning algorithms (like Support Vector Machine and k-Nearest Neighbour), ANN exhibits greater accuracy and resilience when dealing with extensive datasets. The experimental outcomes indicate that ANN accomplished a detection accuracy of 99.92% in credit card fraud cases, markedly surpassing other approaches. In addition, the deep learning attributes of ANN allow it to automatically identify significant features within data, lessening the need for manual feature engineering and thereby enhancing the model's generalization capacity.

2.2.2 GNN

Liu et al. present an innovative Graph Neural Network (GNN)-centred method for credit card fraud detection, employing a weighted multiple graph known as the Transaction Graph (TG) alongside an enhanced Graph SAGE algorithm [10]. The TG offers a comprehensive representation of transaction relationships and characteristics, where nodes symbolize transactions and edges are assigned weights based on logical statements that denote their significance. The upgraded Graph SAGE introduces a novel sampling strategy for choosing informative neighbours and an attention mechanism to adjust neighbour importance according to time intervals, thereby bolstering the model's robustness. Upon evaluation with a large real-world dataset, this approach surpasses current state-of-the-art models, attaining a recall rate of 91.7%, an F1-score of 87.8%, and an Area Under Curve (AUC) of 81.6%. Future research will concentrate on building heterogeneous graphs and devising interpretable models to further enhance fraud detection capabilities.

2.2.3 LSTM

Benchaji et al. present an advanced credit card fraud detection framework that incorporates the attention mechanism and Long Short-Term Memory (LSTM) deep recurrent neural networks [11]. This model capitalizes on the sequential nature of transaction data to pinpoint crucial transactions within input sequences, thereby enhancing the accuracy of fraud prediction. More specifically, it amalgamates three distinct sub-methods: Uniform Manifold Approximation and Projection (UMAP) for the purpose of feature selection, LSTM for capturing transaction sequences, and an attention mechanism to boost LSTM's performance. The model undergoes optimization through feature selection and dimensionality reduction, employs Synthetic Minority Oversampling Technique (SMOTE) to tackle imbalanced datasets, and utilizes LSTM to construct contexts of consumer spending behaviour. Experiments conducted on two datasets reveal that the model demonstrates remarkable efficiency and effectiveness, achieving results that are on par with existing LSTM-based methods. Future endeavours will delve into models that exclusively rely on attention mechanisms and transformer architectures for sequence processing in the realm of fraud detection.

2.2.4 Generative Adversarial Networks (GANs)

Fiore et al. propose a method to enhance credit card fraud detection using Generative Adversarial Networks (GANs) to tackle class imbalance [12]. By training a GAN to generate synthetic examples of fraudulent transactions, they augment the training dataset, improving classifier sensitivity. The GAN’s generator creates realistic synthetic data, while the discriminator evaluates its authenticity. Merging synthetic and real data enhances the classifier’s ability to detect fraud without significantly increasing false positives. Experiments on a public dataset show improved sensitivity compared to classifiers trained on original data alone. This approach highlights GANs’ potential for generating realistic synthetic data to rebalance datasets, offering a promising solution for fraud detection.

3 Discussion

3.1 Limitations and challenges

Despite the notable progress in credit card fraud detection via the utilization of machine learning and deep learning approaches, a number of limitations and challenges persist, especially concerning interpretability, applicability, and data-related issues.

3.1.1 Interpretability

One of the most critical challenges in the field of credit card fraud detection is the interpretability of complex models, especially those based on deep learning techniques such as GANs. While these models offer superior performance in detecting fraudulent activities, their black-box nature often raises concerns regarding transparency. This lack of interpretability is particularly problematic in highly regulated industries like finance, where regulatory scrutiny requires financial institutions to explain their automated fraud detection processes. For instance, GANs, which consist of a generator and a discriminator, are highly effective in generating synthetic data to address class imbalance but are difficult to interpret due to their complex architecture. Similarly, neural networks, including ANN and LSTM, also face interpretability issues, as they automatically extract features from data without providing clear explanations for their decisions. This opacity can hinder the adoption of these models in real-world applications, as stakeholders may be reluctant to rely on systems they cannot fully understand or explain.

3.1.2 Applicability

The applicability of credit card fraud detection methods confronts notable constraints when considering data distribution. First, class imbalance poses a substantial challenge. Fraudulent transactions are typically very infrequent relative to legitimate ones, producing a skewed dataset. Most machine learning algorithms are inclined to favor the majority class, which results in subpar detection rates for fraud instances. Traditional oversampling methods like SMOTE may fail to capture the intricate patterns of fraud in high-dimensional data, while under sampling risks losing valuable information. Secondly, the dynamic nature of fraud patterns poses another limitation. Fraudsters constantly adapt their tactics, causing the data distribution to shift over time. Models trained on historical data may become less effective in detecting new types of fraud. Although online learning algorithms can update models in real-time to some extent, their performance stability and accuracy remain challenges. Additionally, different financial institutions or regions may have varying credit card fraud data distributions due to differences in business models, customer demographics, and risk management strategies. Models trained on data from one institution or region may not generalize well to others, complicating the development of unified fraud detection models and increasing the complexity of cross-institutional and cross-regional fraud detection collaboration.

3.1.3 Data-related Issues

Data-related issues continue to pose a significant hurdle in creating and implementing effective fraud detection systems. A key challenge is the class imbalance found in credit card transaction datasets, where fraudulent transactions are far less numerous than legitimate ones. This disparity can result in models that are skewed toward the majority class, thereby reducing the detection rate of fraudulent activities. While techniques such as SMOTE and GANs have been proposed to address this issue, they come with their own limitations. For example, SMOTE may introduce noise into the dataset, while GANs require careful tuning to generate realistic synthetic data. Another data-related challenge is the quality and availability of data. High-quality, labeled datasets are essential for training accurate models, but obtaining such datasets can be difficult and expensive. Additionally, the privacy and security of data must be ensured, as sensitive financial information is involved. This requires robust data protection measures and compliance with regulatory requirements, adding another layer of complexity to the development of fraud detection systems.

3.2 Future prospects

3.2.1. Enhancing Interpretability with Expert System

To address the interpretability issues of complex models in credit card fraud detection, an expert system approach can be theoretically employed. An expert system could integrate domain-specific knowledge and rules with the black-box models. It would utilize a knowledge base encompassing expert insight on fraud patterns and a reasoning engine to simulate expert decision-making. By extracting and translating the complex model's features and decision logic into understandable rules or explanations, the expert system can provide transparent reasoning for fraud detection outcomes, thereby enhancing the interpretability for stakeholders and meeting regulatory requirements.

3.2.2. Enhancing Applicability via Hybrid Techniques

To address the applicability limitations of credit card fraud detection methods, advanced techniques can be theoretically applied. For class imbalance, combine GANs with SMOTE to generate synthetic fraud data, preserving fraud patterns and augmenting the minority class. For dynamic fraud patterns, use online learning algorithms with adaptive feature selection to update models in real-time. For heterogeneous data across institutions, apply federated learning with domain adaptation to collaboratively train models without sharing raw data. These methods can improve the applicability of fraud detection models.

3.2.3. Addressing Data Challenges

To solve data-related issues in credit card fraud detection, theoretically, advanced techniques can be applied. For class imbalance, combine SMOTE with GANs to generate synthetic fraud data [13], preserving fraud patterns and augmenting the minority class. For data quality and availability, use feature selection methods like FID-SOM to reduce dimensionality and extract relevant features from limited datasets. To ensure data privacy and security, employ federated learning, allowing multiple institutions to collaboratively train models without sharing raw data, thus protecting sensitive financial information.

4 Conclusion

This paper has reviewed the current landscape of credit card fraud detection, highlighting advancements in machine learning and deep learning techniques. The findings indicate that while models like ANN, GNN, LSTM, and GANs significantly enhance detection accuracy, they face limitations in interpretability, applicability, and data-related issues. These challenges hinder the widespread adoption of advanced techniques in real-world scenarios.

To address these challenges, this paper has proposed theoretical solutions, including enhancing interpretability through expert systems, improving applicability via hybrid techniques, and tackling data challenges with advanced methods. These approaches aim to guide future research and development in the field. The analysis and proposed solutions presented in this paper underscore the critical need for more efficient, accurate, and robust fraud detection systems. Such systems are essential not only for safeguarding financial security but also for maintaining the healthy operation and sustainable development of the credit card market.

In summary, ongoing research is crucial to address current limitations and adapt to the dynamic nature of financial fraud. Future work should focus on balancing model complexity with transparency, ensuring scalability, and addressing data quality and privacy concerns. This will help maintain the robustness of the credit card market and protect cardholders from fraudulent activities.

References

1. Chan, P., Stolfo, S.: 'Toward scalable learning with non-uniform class and cost distributions', Proc. Int. Conf. Knowledge Discovery & Data Mining, 2008
2. Chan, P.K., Fan, W., Prodromidis, A.L., Stolfo, S.J.: 'Distributed data mining in credit card fraud detection', IEEE Intell. Syst., 2015, 14 (6), pp. 67–74
3. Brause, R., Langsdorf, T., Hepp, M.: 'Neural data mining for credit card fraud detection', Proc. IEEE Int. Conf. Tools with Artificial Intelligence, 2016
4. Yao, Z.: 'Application of ensemble classifiers in credit card fraud detection', PhD thesis, Chongqing University, 2019
5. Khan, M.Z., Shaikh, S.A., Shaikh, M.A., et al.: 'The performance analysis of machine learning algorithms for credit card fraud detection', Int. J. Online Biomed. Eng., 2023, 19 (3), pp. 82–98
6. Vaquero, P.R.: 'Literature review of credit card fraud detection with machine learning', 2023
7. Nguyen, T.T., Tahir, H., Abdelrazek, M., Babar, A.: 'Deep learning methods for credit card fraud detection', 2020
8. Lin, T.H., Jiang, J.R.: 'Credit card fraud detection with autoencoder and probabilistic random forest', Mathematics, 2021, 9 (21), 2683
9. Asha, R.B., KR, S.K.: 'Credit card fraud detection using artificial neural network', Glob. Trans. Proc., 2021, 2 (1), pp. 35–41
10. Liu, G., Tang, J., Tian, Y., Wang, J.: 'Graph neural network for credit card fraud detection', Proc. 2021 Int. Conf. Cyber-Phys. Soc. Intell. (ICCSI), 2021, pp. 1–6
11. Benchaji, I., et al.: 'Enhanced credit card fraud detection based on attention mechanism and LSTM deep model', J. Big Data, 2021, 8, pp. 1–21
12. Fiore, U., De Santis, A., Perla, F., Zanetti, P., Palmieri, F.: 'Using generative adversarial networks for improving classification effectiveness in credit card fraud detection', Inf. Sci., 2019, 479, pp. 448–455
13. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: 'SMOTE: synthetic minority over-sampling technique', J. Artif. Intell. Res., 2002, 16, pp. 321–357