

A Review of Financial Fraud Detection Methods Based on Machine Learning

Ziqi Yang*

University of Guizhou, Guizhou, China

**Corresponding author: Ziqi Yang.*

Abstract

Financial fraud seriously reduces capital market efficiency and harms investors' interests. Traditional auditing methods also have clear limits. Based on stages of technological development, this paper divides the use of machine learning in financial fraud identification into three stages, from statistical learning and classical machine learning to deep learning. It reviews how machine learning technologies have been applied in financial fraud detection and examines the main principles, strengths, and limits of different methods. The study finds that ensemble learning and graph neural networks are effective in identifying related-party transactions and abnormal patterns, but they still face problems such as data imbalance and weak interpretability. This paper offers a reference point for studying intelligent financial fraud identification from the perspective of technological change.

Keywords

financial fraud, machine learning, anomaly detection

1. Introduction

1.1 Definition, Hazards and Limitations of Traditional Detection Methods of Financial Fraud

Financial fraud refers to acts in which enterprise management deliberately makes false statements or hides important information in financial reports for specific purposes. According to the statistics of the Association of Certified Fraud Examiners (ACFE), the amount of losses caused by fraud in various organizations each year accounts for approximately 5% of their total revenue. Among them, the median loss caused by financial report fraud is the highest [1]. The harm caused by financial fraud mainly appears in three areas. It misleads investors and directs capital toward inefficient enterprises. Once the fraud is exposed, the company's stock price often falls sharply, financing costs increase, and in serious cases the firm may even go bankrupt. Frequent fraud cases also weaken public trust in the capital market and raise both transaction costs and regulatory costs for the whole market. Traditional fraud detection methods mainly fall into two categories. One is the rule-based expert system, which detects anomalies by setting preset thresholds for financial indicators. The problem is adaptability. This method is difficult to update when fraud tactics change. The other is manual sampling audit,

where auditors select some transactions for substantive testing based on the principle of importance. Because sampling is limited by nature, fraudulent activities outside the sample can easily be missed. These two methods do not deal well with complex transactions found in modern enterprises. For that reason, the use of machine learning methods has gradually become necessary. Machine learning models can learn fraud patterns from historical data, detect anomalies, and adjust to changing fraud strategies through retraining.

1.2 Research Objectives and Significance

This paper examines the research objectives at three levels. The first is to identify the common methods enterprises use to commit financial fraud, summarize the unusual features of fraud shown in financial data and annual report text, and build a basic theoretical framework for financial fraud detection as a basis for the later model analysis. The second is to examine how different machine learning algorithms work in fraud detection, compare their performance in practice, and assess the strengths and weaknesses of each model. The third is to review existing fraud detection technologies, point out where they do not match the practical needs of front-line auditors, and then propose optimization strategies that fit auditing scenarios and may offer a reference for later research.

2. Literature Review

2.1 Characteristics and Core Indicator System for Identifying Financial Fraud

Whether a machine learning model can be truly effective largely depends on whether the input feature variables have “soul”. Therefore, before applying various technical algorithms, it's essential to understand one key question: what quantifiable errors actually occur in accounting when financial fraud takes place? The following are several relatively critical dimensions that can be identified.

The first and most direct sign is abnormal movement in financial ratios. When companies manipulate financial statements, the adjustment usually appears in revenue, costs, or accrual items. The classic M-Score model, developed at the end of the last century, combines eight key ratios, such as accounts receivable index, gross margin index, asset quality index, and revenue growth rate, and this provided an early framework for later studies [2]. In practical screening, however, three signals appear more often. One is a sharp rise in accounts receivable that is faster than revenue growth. Another is inventory turnover moving in the opposite direction from gross margin. The third is a gap between operating cash flow and net profit.

Accrual quality and the sustainability of earnings are also central when identifying fraud risk. Accrual accounting can describe operating performance more fully, but it also leaves room for earnings manipulation. In actual financial statement analysis, a company may report high book profits for several periods and still fail to produce the matching operating cash inflows. That mismatch deserves attention.

Related-party transactions create a different problem because they are harder to trace from a single account or a single statement. In many fraud cases, the key issue is not one abnormal item but the path of fund transfers across multiple entities. That path can be hidden well. For example, some listed companies transfer resources to major shareholders through several shell companies that appear unrelated on the surface. Under this kind of “chain-like” arrangement, each entity's statements may look normal on their own, while the overall structure still allows profit transfers.

The feature system above provides the financial theory needed to choose input variables for machine learning models. The following sections explain how different machine learning methods build fraud identification frameworks from these features.

2.2 Statistical Learning and Classical Machine Learning Methods

Statistical learning and classical machine learning methods were among the earliest techniques used to identify financial fraud. These mainly include logistic regression, decision trees, random forests, and support vector machines. Their basic logic is straightforward. Using historical data, the model learns differences in financial characteristics between fraud samples and non-fraud samples, and then uses those patterns to classify new cases. Logistic regression is the most basic method at this stage. It uses multiple financial indicators as input variables and outputs a probability value between 0 and 1 to represent the likelihood of fraud. For

example, the model might output “The probability of fraud for a certain enterprise is 0.78”, and the auditor can treat that enterprise as a key concern. One reason this method is widely used is that its coefficients show both the direction and size of each indicator's effect on fraud probability, so the auditor can explain why an enterprise is marked as high-risk. For example, “for every 1% increase in the growth rate of accounts receivable, the probability of fraud increases by 0.3%” [3].

A decision tree generates a set of decision rules that resembles a flowchart. It does this by repeatedly selecting the financial indicators that best separate fraudulent samples from non-fraudulent ones. For instance, a decision tree might produce the following rule: “If the growth rate of accounts receivable is greater than 30%, the gross profit margin decreases by more than 10%, and the inventory turnover rate increases abnormally, then it is determined to be at a high risk of fraud”. This rule-based form is close to the way auditors make judgments when preparing audit working papers, so it is easier to understand and communicate in audit practice.

Random forest builds on the decision tree by introducing ensemble learning. It creates multiple decision trees, lets each tree make its own judgment, and then uses voting to decide whether an enterprise is at high risk of fraud. It could reduce the risk of error that can occur when a single decision tree gives an unstable result. From an accounting perspective, random forest can automatically output “which financial indicators are the most important for fraud identification”—for example, the model may show that “the growth rate of accounts receivable, inventory turnover rate, and the matching degree of operating cash flow and net profit” are the three most crucial indicators, which provides directional guidance for auditors to design substantive procedures [3].

The core idea of support vector machines is to identify a classification boundary that can best distinguish between fraudulent enterprises and non-fraudulent ones. It is particularly suitable for situations where the sample size is small but there are many financial indicators. For example, there are only a few dozen confirmed fraudulent enterprises, but each enterprise has dozens of financial indicators. Due to the scarcity of real fraudulent samples, the feature of support vector machines makes it have certain application value.

The common advantage of the above methods lies in their strong interpretability. Auditors need to explain the basis of their judgments to clients, record the reasoning process in audit records, and provide reasonable explanations in regulatory inquiries—these requirements are in line with the characteristics of statistical learning and classical machine learning methods. However, the limitations of these methods are also obvious: they highly rely on structured data in financial statements (profits, assets, liabilities, cash flows, etc.), and enterprises that commit fraud will precisely manipulate these figures, which creates a difficulty—the basis used to identify fraud, in some cases, has been modified [4].

2.3 Deep Learning Methods

Deep learning methods are an extension of statistical learning and classical machine learning. Their core feature is the ability to automatically learn hierarchical features in data through multi-layer network structures. In the field of financial fraud detection, the application of deep learning mainly focuses on two directions: anomaly detection based on autoencoders and related party transaction identification based on graph neural networks.

The typical application of autoencoders in the identification of financial fraud is “unsupervised anomaly detection”. First, a large amount of financial data from normal enterprises is used to train the autoencoder, allowing it to learn the characteristics of normal enterprises. Then, the financial data of the enterprise to be tested is input into the trained model. If the model can accurately reproduce the data of that enterprise, it indicates that the financial model of that enterprise is consistent with that of normal enterprises; if the model's reconstruction effect is poor, it suggests that the financial situation of that enterprise may be suspicious of fraud. This method has special value in accounting practice. Traditional fraud identification methods usually require prior knowledge of which enterprises have committed fraud, but in reality, a large number of fraud cases have not been discovered. Autoencoders do not rely on pre-labeled fraud samples and can help auditors initially screen out those “seemingly abnormal” enterprises from a large number of enterprises, narrowing the scope of further audits [5].

The application of graph neural networks in the identification of financial fraud mainly focuses on related-party transaction fraud. Related-party transactions are a high-incidence area of financial fraud, but traditional

financial indicator analysis methods are difficult to effectively identify it – because fraud often does not occur independently but involves complex transactions among multiple related companies.

Graph neural networks process by treating each enterprise or individual as a “node”, and the equity relationships, transactions, and the same registered addresses between enterprises are regarded as “edges” to construct a relationship network graph. By passing information layer by layer in the graph neural network, the model can identify transaction chains such as “Company A → Shell Company B → Audited Company C” that span multiple entities. For example, a listed company may transfer benefits to its affiliated parties through a seemingly unrelated shell company. If the financial statements of the listed company are examined alone, no abnormalities may be detected; however, in the relationship network constructed by the graph neural network, the hidden relationship between the company and the shell company will be identified [6].

The advantage of deep learning methods lies in their ability to handle complex scenarios that traditional methods struggle to deal with. However, the promotion of deep learning methods in auditing practice faces a fundamental obstacle: insufficient interpretability. When a model labels a certain enterprise as high-risk, auditors find it difficult to explain to the client “why”—because the internal decision-making process of deep learning models is a “black box”. This issue will remain the main constraint for the application of deep learning methods in the auditing field in the current and future periods.

2.4 Imbalanced Learning and Ensemble Methods

The identification of financial fraud faces a particular practical challenge: In real enterprises, the proportion of those with significant financial fraud is extremely low, typically less than 1% [4]. This means that the number of fraud samples and non-fraud samples is severely imbalanced. If a classification model is directly trained, the model will tend to classify all enterprises as “non-fraud”, because doing so can achieve an accuracy rate of over 99%, but the proportion of successfully identifying fraudulent enterprises is almost zero.

The imbalance learning method is specifically designed to address this issue. The most representative one is the SMOTE method. Unlike simply replicating the existing fraudulent samples, SMOTE generates some synthetic samples between the existing fraudulent samples, which are located between two real fraudulent samples. For example, if a fraudulent enterprise A and a fraudulent enterprise B have distinct characteristics in the “accounts receivable growth rate” and “inventory turnover rate” indicators, SMOTE will generate some “synthetic fraudulent samples” that possess some features of both, to balance the training data [5].

The ensemble learning method enhances the stability and accuracy of recognition by combining multiple models. The core of bagging-based methods, such as Random Forest, is to train multiple models, each focusing on different subsets of samples and different combinations of financial indicators. Finally, the judgment results of all models are integrated. Boosting-based methods first train a basic model, identify the samples where it makes mistakes, and then in the next round of training, let the new model focus on these erroneous samples. Through multiple iterations, the ability to recognize difficult samples gradually improves. One significant advantage of this method is that it can effectively reduce the false positive rate.

Unbalanced learning and ensemble methods do not alter the fundamental logic of fraud detection, but rather solve the practical problem of data imbalance. They can be combined with classical machine learning and deep learning.

3. Key Finding

Looking across the three technical approaches discussed above, and consistent with data-mining classification frameworks in the literature [7], machine learning progress in financial fraud detection is not a simple story of “new methods replacing old ones”. The shift is better understood as a move from “explicitly encoded rules” to “implicit pattern learning”. In the statistical learning stage, such as logistic regression and decision trees, the main task is to convert human judgment rules, often drawn from auditor experience, into measurable mathematical form. Deep learning aims at something different. It lets the model learn patterns on its own from raw data, including patterns that human analysts may not recognize directly. That does not mean these two approaches compete in every situation. They fit different auditing needs. Statistical learning is still necessary when the judgment criteria must be stated clearly, especially in regulatory reporting where the reason for a decision matter as much as the decision itself. By contrast, graph neural networks have shown strengths

that traditional methods usually do not have, especially when the task is to detect hidden related-party transaction networks.

A more basic problem appears in the data itself. Human manipulation of financial information conflicts directly with the objectivity that algorithms are supposed to rely on, and this is the main criticism developed in this paper. Most discussions of machine learning performance assume that the input data reflect the firm's actual operations with reasonable accuracy. In financial fraud cases, that assumption may fail from the start, because the perpetrators can intentionally change or fabricate the very indicators the model uses. The result is that the model analyzes data that may already be distorted before training or prediction begins. This creates a “data paradox”. The algorithm may be objective, but the data entering it may be shaped by subjective and intentional falsification. Existing studies have not paid enough attention to this contradiction, yet it affects the validity of the whole approach. For that reason, improving model accuracy alone is unlikely to solve the core problem. If the inputs are distorted, more complex computations do not necessarily produce better judgments. A more workable direction for future research may be to bring in external evidence that management cannot easily control, such as tax records, customs data, or on-site transaction details, and use that material to cross-check model outputs.

Auditing also has a professional logic that differs from standard model evaluation. In audit work, the priority is not only “accuracy rate” but also “traceable reasonable suspicion”. Academic studies usually assess performance with metrics such as AUC and F1 scores, but audit practice uses a different standard when conclusions need to be defended. Auditors are expected to provide “reasonable assurance”. Their judgments must stand up to review by regulators, questions from clients, and in some cases judicial review. Because of that, it is not enough to flag a person or enterprise as potentially problematic. Auditors also need to explain why the suspicion arose and how the conclusion follows from the available evidence. Here the “black box” problem matters. Deep learning models often do not provide explanation paths that fit this professional requirement. This paper therefore argues that, until explainable artificial intelligence (XAI) develops further, deep learning is unlikely to be widely used in key audit judgments. Even so, in supporting tasks such as preliminary risk assessment and anomaly detection, its high sensitivity still appears useful.

Intelligent enhancements in financial fraud detection should not be treated as technological progress alone, but evaluated against the benchmarks of “algorithmic capability”, “data reliability”, and “audit professional logic”.

4. Discussion

4.1 Current Main Challenges

Although machine learning methods have made significant progress in the field of financial fraud detection, they still face many challenges. The first is the scarcity of fraud samples. The proportion of genuine fraud enterprises is extremely low. Even with the use of imbalance learning methods such as SMOTE, the synthetic samples are not real samples, and boosting-based ensemble methods like XGBoost, despite their strong performance on balanced datasets, still face challenges when applied to highly imbalanced fraud detection scenarios [8]. The second is the continuous evolution of fraud methods. Fraud models are usually built based on historical cases, but fraud methods evolve with technological upgrades. A model that performs well on historical data may completely fail when faced with new fraud methods. This confrontation is a unique challenge in the field of fraud detection. In addition, the black box problem of the model is also a major challenge. Deep learning methods provide higher prediction accuracy while reducing interpretability. However, for auditors, they not only need to identify the problem but also need to be able to explain the source of the problem.

4.2 Future Trends

In response to the aforementioned challenges, future research can focus on the following directions. The first is the analysis of language and financial texts. Besides financial data, the annual reports, management discussions and analyses, performance presentations, and responses to inquiries of enterprises also contain abundant clues of fraud. Large language models can extract semantic features from these texts that traditional

methods often miss, such as management deliberately avoiding certain topics or using overly optimistic or ambiguous language [6].

Another direction is the use of interpretable artificial intelligence in auditing. Researchers are trying to improve model interpretability without losing too much prediction accuracy. For example, techniques such as SHAP values analysis can generate explanation reports for each enterprise marked as high-risk and show which financial indicators contributed to that result. This could make the output more usable in auditing practice, where the decision itself and the reasoning behind it both matter.

The combination of graph neural networks and knowledge graphs is also very important. In the future, external industry knowledge, regulatory rules, and historical cases of fraud can be utilized to enrich the semantic information of the enterprise relationship network, thereby further improving the ability to identify related party transaction fraud [6].

5. Conclusion

This article systematically reviews the application of machine learning in financial fraud detection by dividing it into three stages of technological development: statistical learning and classical machine learning, deep learning, and imbalanced learning and ensemble methods. Machine learning methods have significantly improved the efficiency of financial fraud detection, surpassing the limits that traditional manual sampling could achieve. Ensemble learning and graph neural networks are currently the most active research directions in this field. Despite continuous technological advancements, challenges such as data imbalance, dynamic evolution of fraud tactics, and insufficient model interpretability still exist and will persist in the future. Therefore, future research should focus on modeling dynamic fraud behaviors to enable models to adapt to evolving fraud tactics; further, research on model interpretability to narrow the gap between the academic research frontier and the needs of auditing practice.

References

- [1] Association of Certified Fraud Examiners. (2024). Occupational Fraud 2024: A Report to the Nations. ACFE.
- [2] Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 55(5), 24--36.
- [3] Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research*, 58(1), 199–235.
- [4] Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, 30(2), 19–50.
- [5] Zhou, Z. H. (2021). *Machine Learning*. Tsinghua University Press. (Chapter 11: Anomaly Detection)
- [6] Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. *Expert Systems with Applications*, 240, Article 122156.
- [7] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559–569.
- [8] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

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.

