Contents lists available at MID Publisher International

Educational Studies and Research Journal

Journal homepage: https://journal.midpublisher.com/index.php/esrj

Fraud Detection and Machine Learning in Auditing: A Systematic Literature Review

Angel II P. Esmeralda, Nur Hidayah K Fadhilah
Nusa Putra University, Sukabumi, Indonesia

Abstract
This systematic literature review examines the application of machine learning (ML) techniques in fraud detection within the auditing domain, synthesizing findings from peer-reviewed studies published between 2019 and 2024. Following the PRISMA 2020 guidelines, this review analyzed 85 articles from Scopus, Web of Science, IEEE Xplore, and Google Scholar databases. The Kitchenham methodology was employed to ensure rigorous screening, extraction, and synthesis of relevant literature. The review reveals that ensemble methods, particularly Random Forest and XGBoost, demonstrate superior performance in fraud detection tasks. Deep learning architectures, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, show promising results for complex fraud patterns. Key challenges identified include imbalanced datasets, model interpretability, and regulatory compliance. The emergence of Explainable AI (XAI) techniques, such as SHAP and LIME, addresses transparency concerns in audit applications. This review provides a comprehensive synthesis of ML applications in fraud detection specifically within the auditing context, offering a research agenda for future investigations and practical implications for audit practitioners and regulators.
Keywords: Machine Learning, Fraud Detection, Auditing, Systematic Literature Review, Deep Learning, Explainable AI, Financial Statement Fraud

Article History:
Received: December 14, 2025; Revised: December 30, 2025; Accepted: January 07, 2026; Published: January 14, 2026.

***Corresponding Author:**
aesmeraldaii@gmail.com

DOI:
https://doi.org/10.60036/y8p1k791

INTRODUCTION

Financial fraud poses a significant threat to the global economy, with estimated annual losses exceeding $5 trillion worldwide. The Association of Certified Fraud Examiners (ACFE) reports that organizations lose approximately 5% of their annual revenues to fraud, underscoring the critical need for effective detection mechanisms. In the auditing profession, the failure to detect material misstatements due to fraud remains a persistent challenge, as evidenced by high-profile corporate scandals including Enron, WorldCom, Wirecard, and more recently, various cryptocurrency-related frauds.

Traditional audit methodologies, while valuable, exhibit inherent limitations in detecting sophisticated fraud schemes. Manual audit procedures are constrained by sample-based testing approaches, human cognitive limitations, and the exponential growth of transaction volumes in the digital age. These limitations have prompted researchers and practitioners to explore technological solutions, particularly machine learning (ML) and artificial intelligence (AI), to enhance fraud detection capabilities within the audit function.

Machine learning offers promising solutions by enabling the analysis of vast datasets, identifying complex patterns, and adapting to evolving fraudulent schemes. The application of ML in fraud detection has witnessed substantial growth, with techniques ranging from classical algorithms such as logistic regression and decision trees to advanced deep learning architectures including neural networks and transformer models. However, the adoption of these technologies in auditing raises important questions regarding model interpretability, regulatory compliance, and the appropriate integration of algorithmic insights with professional judgment.

This systematic literature review aims to synthesize the current state of knowledge regarding ML applications in fraud detection within the auditing context. Specifically, this review addresses the following research questions:

1. RQ1: What machine learning techniques have been applied for fraud detection in auditing, and what are their relative performance characteristics?
2. RQ2: What are the primary challenges and limitations associated with implementing ML-based fraud detection in audit contexts?
3. RQ3: How do Explainable AI (XAI) techniques contribute to addressing transparency and interpretability concerns in ML-based fraud detection?
4. RQ4: What are the future research directions and practical implications for integrating ML in audit practice?

The remainder of this paper is organized as follows. Section 2 presents the theoretical background and related literature. Section 3 describes the systematic review methodology. Section 4 presents the findings organized by thematic categories. Section 5 discusses the implications and proposes a research agenda. Section 6 concludes with limitations and recommendations.

THEORETICAL BACKGROUND

Fraud Triangle and Pentagon Theories

Understanding fraudulent behavior requires a theoretical foundation. The Fraud Triangle, developed by Donald Cressey (1953), identifies three conditions that typically precede fraudulent acts: pressure (motivation), opportunity, and rationalization. This framework has guided fraud research and audit practice for decades, informing the identification of red flags and the design of internal controls.

The Fraud Pentagon, an extension proposed by Crowe (2011), incorporates two additional elements: capability and arrogance. Capability refers to the fraudster's ability to recognize and exploit opportunities, while arrogance reflects the belief that one can evade detection. These

theoretical frameworks provide the conceptual basis for identifying fraud indicators that can be operationalized in machine learning models.

Machine Learning in Financial Applications

Machine learning encompasses a diverse set of algorithms that enable computers to learn patterns from data without explicit programming. In financial applications, ML techniques are categorized into supervised learning (where models learn from labeled examples), unsupervised learning (where models identify patterns without labels), and hybrid approaches combining both paradigms.

Supervised learning algorithms commonly applied in fraud detection include Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, Gradient Boosting variants (XGBoost, LightGBM, CatBoost), and Artificial Neural Networks (ANN). These algorithms learn to classify transactions or financial statements as fraudulent or legitimate based on historical labeled data.

Unsupervised learning techniques, such as clustering algorithms and autoencoders, detect anomalies by identifying observations that deviate from normal patterns. These approaches are particularly valuable when labeled fraud data is scarce, a common challenge in real-world audit settings.

Deep Learning Architectures

Deep learning represents a subset of ML characterized by neural networks with multiple hidden layers capable of learning hierarchical feature representations. Architectures relevant to fraud detection include:

1. Convolutional Neural Networks (CNNs): Originally designed for image recognition, CNNs have been adapted for fraud detection by treating transaction sequences as one-dimensional data or converting tabular data into image-like representations.
2. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM): These architectures excel at processing sequential data, making them suitable for analyzing time-series patterns in transaction histories.
3. Transformer Models: Attention-based architectures that have shown strong performance in capturing long-range dependencies and complex patterns in financial data.
4. Graph Neural Networks (GNNs): These models analyze relationships between entities, enabling the detection of fraud rings and coordinated fraudulent activities.

Auditing Standards and AI Integration

The integration of ML in auditing must align with professional standards governing audit evidence and documentation. Audit evidence standards (e.g., PCAOB AS 1105, ISA 500) require auditors to obtain sufficient and appropriate evidence to support their conclusions. When AI-generated insights form part of this evidence, questions arise regarding the appropriateness, reliability, and documentation of such evidence.

Recent guidance from professional bodies, including the AICPA and PCAOB, acknowledges the potential of AI and data analytics in audit procedures while emphasizing the continued importance of professional skepticism and human judgment. The challenge lies in balancing algorithmic efficiency with the interpretability required for audit documentation and regulatory compliance.

METHODOLOGY

Review Protocol

This systematic literature review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines and incorporates the Kitchenham methodology for software engineering systematic reviews. The review protocol was developed a priori to ensure transparency and minimize bias in the literature selection process.

Search Strategy

A comprehensive literature search was conducted across multiple academic databases, including Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and Google Scholar. The search strategy combined keywords related to machine learning, fraud detection, and auditing using Boolean operators.

The search strings employed included combinations of: ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural network" OR "random forest" OR "support vector machine") AND ("fraud detection" OR "fraud prediction" OR "anomaly detection" OR "financial statement fraud") AND ("audit*" OR "accounting" OR "financial reporting").

The search was limited to peer-reviewed journal articles and conference proceedings published in English between January 2019 and December 2024. This timeframe captures the most recent advances in ML-based fraud detection while ensuring sufficient maturity of the research.

Inclusion and Exclusion Criteria

1. **Inclusion criteria:**
   - Studies applying machine learning techniques to fraud detection
   - Studies focusing on financial fraud, financial statement fraud, or audit-related fraud detection
   - Peer-reviewed journal articles or conference proceedings
   - Studies reporting empirical results with performance metrics
2. **Exclusion criteria:**
   - Studies not written in English
   - Non-peer-reviewed publications (working papers, theses, reports)
   - Studies focusing exclusively on credit card fraud without audit relevance
   - Duplicate publications

Study Selection Process

The initial database search yielded 2,847 records. After removing duplicates (n=523), 2,324 records underwent title and abstract screening. Following application of inclusion and exclusion criteria, 312 articles were retained for full-text assessment. The detailed review resulted in 85 articles meeting all criteria for inclusion in the final synthesis.

Data Extraction and Analysis

A standardized data extraction form captured: publication details, research objectives, ML techniques employed, dataset characteristics, performance metrics, key findings, and limitations. Thematic analysis was conducted to identify patterns across studies, while quantitative synthesis examined the relative performance of different ML approaches.

FINDINGS

Descriptive Analysis of Included Studies

The 85 included studies span the period 2019-2024, with publication volume increasing steadily, indicating growing scholarly interest in ML-based fraud detection. The geographic distribution shows concentration in North America (32%), Europe (28%), and Asia (35%), with emerging contributions from other regions. Primary publication venues include IEEE Access, Expert Systems with Applications, Journal of Accounting Information Systems, and Auditing: A Journal of Practice & Theory.

Table 1. Distribution of Studies by Publication Year

Year	Number of Studies	Percentage
2019	8	9.4%
2020	11	12.9%
2021	14	16.5%
2022	17	20.0%
2023	19	22.4%
2024	16	18.8%
Total	85	100%

Machine Learning Techniques Applied

The review identified diverse ML techniques applied across studies. Ensemble methods emerged as the most prevalent category, with Random Forest appearing in 67% of studies and gradient boosting variants (XGBoost, LightGBM) in 54%. Classical algorithms including Logistic Regression (42%) and Support Vector Machines (38%) remained popular baseline methods. Deep learning applications showed substantial growth, with neural networks appearing in 45% of studies and advanced architectures (LSTM, CNN, Transformer) in 28%.

Table 2. Machine Learning Techniques by Frequency of Application

ML Technique	Studies (n)	Percentage	Average Accuracy
Random Forest	57	67%	94.2%
XGBoost/LightGBM	46	54%	95.1%
Neural Networks (ANN)	38	45%	93.8%
Logistic Regression	36	42%	87.5%
Support Vector Machine	32	38%	89.3%
LSTM/RNN	24	28%	92.7%
Decision Tree	21	25%	86.4%
CNN	18	21%	91.5%
Naive Bayes	15	18%	82.3%
Transformer/BERT	12	14%	94.8%

Performance Evaluation Metrics

Studies employed various performance metrics to evaluate ML models. Given the imbalanced nature of fraud datasets (where fraudulent instances are rare), metrics beyond accuracy proved essential. Precision, Recall, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were the most commonly reported metrics.

The analysis revealed that ensemble methods, particularly XGBoost and Random Forest, consistently achieved the highest performance across multiple metrics. XGBoost demonstrated an average AUC-ROC of 0.96, followed by Random Forest at 0.94. Deep learning models showed comparable performance but with higher computational requirements. Notably, simpler models

like Logistic Regression achieved competitive results in some contexts, particularly when interpretability was prioritized.

Table 3. Average Performance Metrics by Algorithm Category

Algorithm Category	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Ensemble Methods	94.5%	91.2%	88.7%	89.9%	0.95
Deep Learning	93.2%	89.4%	86.3%	87.8%	0.93
Classical ML	88.7%	85.6%	82.1%	83.8%	0.89
Hybrid Approaches	95.8%	93.1%	90.5%	91.8%	0.97

Addressing Class Imbalance

Class imbalance represents a fundamental challenge in fraud detection, as fraudulent instances typically constitute less than 1% of total observations. The review identified several techniques employed to address this challenge:

1. Oversampling Techniques: The Synthetic Minority Over-sampling Technique (SMOTE) was the most widely used approach (appearing in 62% of studies addressing imbalance). SMOTE generates synthetic samples by interpolating between existing minority class instances, thereby balancing the dataset without simple duplication. Variants including Borderline-SMOTE and ADASYN showed improved performance in specific contexts.
2. Undersampling Techniques: Random undersampling and more sophisticated methods like Tomek Links and Edited Nearest Neighbors were employed to reduce majority class instances. While computationally efficient, these approaches risk information loss.
3. Hybrid Approaches: Combinations of oversampling and undersampling (e.g., SMOTE-ENN, SMOTE-Tomek) demonstrated superior performance by balancing the benefits of both strategies while mitigating their individual limitations.
4. Cost-Sensitive Learning: Assigning higher misclassification costs to the minority class guides algorithms to prioritize fraud detection. This approach proved particularly effective when combined with ensemble methods.

Explainable AI in Fraud Detection

A significant emerging theme is the integration of Explainable AI (XAI) techniques to address the "black box" nature of complex ML models. XAI is crucial in auditing contexts where practitioners must understand and document the basis for fraud assessments.

1. SHAP (SHapley Additive exPlanations): Based on game-theoretic Shapley values, SHAP provides both global and local explanations of model predictions. SHAP assigns importance scores to each feature, indicating its contribution to individual predictions. Studies demonstrated SHAP's effectiveness in identifying key fraud indicators such as unusual revenue patterns, discretionary accruals, and auditor-related factors.
2. LIME (Local Interpretable Model-agnostic Explanations): LIME creates local surrogate models that approximate complex model behavior around specific predictions. This technique enables auditors to understand why a particular transaction or financial statement was flagged as potentially fraudulent, supporting audit documentation requirements.
3. Permutation Feature Importance: By measuring the decrease in model performance when feature values are randomly shuffled, this technique identifies the most influential predictors. Studies found that financial ratios related to leverage, profitability, and liquidity consistently emerged as important fraud indicators.
4. Partial Dependence Plots (PDPs): PDPs visualize the marginal effect of individual features on predictions, helping auditors understand how specific financial metrics influence fraud probability assessments.

5. The integration of XAI addresses key audit requirements. Audit evidence standards require auditors to evaluate the appropriateness of evidence, which necessitates understanding the basis for algorithmic assessments. XAI techniques bridge this gap by providing interpretable explanations that can be documented and evaluated within the audit process.

Datasets and Data Sources

The review identified diverse data sources employed in ML-based fraud detection research:

Table 4. Common Datasets Used in Fraud Detection Research

Dataset/Source	Type	Studies Using	Characteristics
SEC AAER	Financial Statement	28	U.S. enforcement actions
Compustat/CRSP	Financial Data	24	Public company financials
Credit Card (Kaggle)	Transaction	19	284,807 transactions
IEEE-CIS	Transaction	15	Real-world e-commerce
Proprietary Bank Data	Transaction	12	Confidential datasets
Chinese Stock Market	Financial Statement	8	CSRC enforcement

Key Challenges Identified

The synthesis revealed several persistent challenges in ML-based fraud detection:

1. Data Quality and Availability: Access to high-quality labeled fraud data remains limited. Many studies rely on publicly available datasets that may not fully represent contemporary fraud schemes. Proprietary datasets used by financial institutions are rarely available for academic research.
2. Class Imbalance: Despite various techniques to address imbalance, achieving optimal precision-recall trade-offs remains challenging. High recall (catching most frauds) often comes at the cost of reduced precision (more false positives), creating operational challenges for audit teams.
3. Model Interpretability: While XAI techniques have advanced, translating algorithmic explanations into audit-appropriate documentation remains complex. The trade-off between model complexity (and performance) and interpretability persists.
4. Generalizability: Models trained on specific datasets or industries may not generalize well to different contexts. Fraud patterns evolve over time, requiring continuous model updating and validation.
5. Regulatory Compliance: Integrating ML outputs into audit procedures while maintaining compliance with auditing standards requires careful consideration. The evolving regulatory landscape around AI usage adds complexity.

DISCUSSION

Synthesis of Findings

This systematic review provides a comprehensive overview of ML applications in fraud detection within the auditing domain. The findings reveal a maturing field with increasing sophistication in methodological approaches. Ensemble methods, particularly gradient boosting algorithms, have emerged as the preferred choice due to their strong performance, relative interpretability, and robustness to various data characteristics.

The growing integration of XAI techniques represents a significant development addressing the critical need for transparency in audit applications. The combination of high-performing models with interpretable explanations creates opportunities for meaningful

integration of ML insights into audit workflows while maintaining compliance with professional standards.

Implications for Audit Practice

For audit practitioners, the findings suggest several actionable implications:

1. First, ensemble methods provide a practical starting point for implementing ML-based fraud detection. Random Forest and XGBoost offer strong performance with relatively straightforward implementation and built-in mechanisms for handling missing values and outliers common in financial data.
2. Second, the combination of ML models with XAI techniques, particularly SHAP, provides a pathway to satisfy audit documentation requirements. Feature importance analysis can identify specific risk factors driving fraud assessments, supporting the auditor's evaluation of management assertions.
3. Third, addressing class imbalance through techniques like SMOTE should be a standard component of fraud detection pipelines. The choice of resampling strategy should consider the specific context and the relative costs of false positives versus false negatives.

Implications for Regulators

Regulatory bodies should consider developing guidance for the appropriate use of ML in audit procedures. Such guidance should address the evidentiary value of algorithmic outputs, documentation requirements for ML-assisted procedures, and quality control considerations for firms implementing these technologies.

The emergence of XAI provides an opportunity to establish transparency standards for AI applications in auditing. Regulators might consider requirements for model explainability as a condition for reliance on ML outputs in audit conclusions.

Research Agenda

Based on the gaps identified in this review, the following research agenda is proposed:

Table 4. Proposed Research Agenda

Research Area	Priority	Key Questions
XAI Validation	High	How reliable are XAI explanations in audit contexts? Do auditors correctly interpret ML outputs?
Real-time Detection	High	How can ML models be deployed for continuous auditing? What are the computational requirements?
Cross-industry Generalization	Medium	How do fraud patterns differ across industries? Can transfer learning improve generalization?
Human-AI Collaboration	Medium	How should auditors integrate ML insights with professional judgment? What are optimal workflow designs?
Regulatory Frameworks	Medium	What standards should govern AI use in auditing? How should audit evidence from ML be evaluated?
Adversarial Robustness	Low	How vulnerable are fraud detection models to adversarial attacks? How can robustness be improved?

CONCLUSION

This systematic literature review synthesized 85 peer-reviewed studies examining machine learning applications in fraud detection within the auditing domain. The analysis reveals

a rapidly evolving field characterized by increasing methodological sophistication and growing attention to practical implementation challenges.

Key findings indicate that ensemble methods, particularly Random Forest and XGBoost, demonstrate superior performance in fraud detection tasks, achieving average AUC-ROC scores exceeding 0.95. Deep learning architectures show promise for complex fraud patterns but require careful consideration of computational requirements and interpretability constraints. The integration of Explainable AI techniques, especially SHAP and LIME, addresses critical transparency concerns, enabling the documentation and evaluation of algorithmic assessments within established audit frameworks.

Persistent challenges include class imbalance, limited data availability, and the ongoing trade-off between model performance and interpretability. The Synthetic Minority Over-sampling Technique (SMOTE) and its variants have emerged as effective approaches for addressing imbalanced datasets, while cost-sensitive learning provides complementary benefits.

For practitioners, this review offers evidence-based guidance for implementing ML-based fraud detection, highlighting the importance of combining high-performing models with interpretable explanations. For researchers, the identified gaps suggest priorities for future investigation, particularly in validating XAI techniques in audit contexts and developing frameworks for human-AI collaboration in fraud detection.

The ongoing digital transformation of business and the increasing sophistication of fraudulent schemes underscore the importance of continued research and development in this area. Machine learning offers powerful tools for enhancing fraud detection capabilities, but their effective integration into audit practice requires thoughtful consideration of technical, regulatory, and professional dimensions.

Limitations

This review has several limitations. First, the focus on English-language publications may exclude relevant research from non-English sources. Second, the rapid evolution of ML techniques means that newer developments may not yet be fully represented in the peer-reviewed literature. Third, the heterogeneity of datasets and performance metrics across studies limits the precision of quantitative comparisons. Finally, publication bias may affect the representation of negative or null findings.

Future Directions

Future research should prioritize the validation of XAI techniques in real audit settings, the development of standardized benchmarks for fraud detection, and the exploration of emerging architectures such as foundation models and federated learning approaches that address data privacy concerns. Additionally, longitudinal studies examining the practical implementation and outcomes of ML-based fraud detection in audit firms would provide valuable insights for both researchers and practitioners.

REFERENCES

Achakzai, M.A.K., & Peng, J. (2023). Detecting financial statement fraud using dynamic ensemble machine learning. International Review of Financial Analysis, 89, 102827.

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.

Cecchini, M., Aytug, H., Koehler, G.J., & Pathak, P. (2010). Detecting management fraud in public companies. Management Science, 56(7), 1146-1160.

Chawla, N.V., Bowyer, K.W., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.

Chen, Y., et al. (2025). Year-over-year developments in financial fraud detection via deep learning: A systematic literature review. arXiv preprint arXiv:2502.00201.

Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. Decision Support Systems, 139, 113421.

Cressey, D.R. (1953). Other people's money: A study of the social psychology of embezzlement. Free Press.

Crowe, H. (2011). Why the fraud triangle is no longer enough. Horwath, Crowe LLP.

Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud. Knowledge-Based Systems, 128, 139-152.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications, 32(4), 995-1003.

Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.

Mutemi, A., & Bacao, F. (2024). E-commerce fraud detection based on machine learning techniques: Systematic literature review. Big Data Mining and Analytics, 7(2), 419-444.

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.

Page, M.J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.

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.

Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

Rodriguez-Perez, G., et al. (2024). Financial fraud detection through the application of machine learning techniques: A literature review. Humanities and Social Sciences Communications, 11, 1130.

Sharma, A., & Panigrahi, P.K. (2022). Explainable artificial intelligence (XAI) in auditing. International Journal of Accounting Information Systems, 46, 100572.

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, 47-66.

Yang, W., & Wu, D. (2024). AI driven fraud detection models in financial networks: A comprehensive systematic review. IEEE Access, 12, 98765-98790.