Machine Learning-Based Detection of Accounting Irregularities in Corporate Financial Reports

Grant Kimmer
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
gzimmerman@binghamton.edu

Jereme Meran
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
moran1997@uab.edu

Rishi C. Gokhale
Department of Computer Science, University of North Texas, Denton, TX, USA.
gokhalerishi@unt.edu

Jakub C. Hoffman
Department of Computer Science, University of Houston, Houston, TX, USA.
jakubchoffman@uh.edu

Abstract

The detection of accounting irregularities in corporate financial reports has become a complex systems problem that extends beyond traditional ratio analysis and accrual modeling. Machine learning methods are increasingly deployed to identify anomalous reporting behavior in large populations of public firms, yet their performance depends on the design of data infrastructures, the selection of textual and numeric features, the management of imbalanced data, and the integration of model outputs into audit and regulatory workflows. This paper presents a system-level examination of machine learning-based detection of accounting irregularities. It analyzes the structural trade-offs between predictive accuracy, interpretability, false alert burden, and operational cost. It further considers how narrative disclosures, particularly risk factor sections and management discussion, can be transformed into semantic features that complement structured financial variables. The discussion addresses governance requirements, including data lineage, model refresh cycles, fairness, robustness, and the explainability expectations of auditors, boards, and regulators. Rather than treating detection as a purely algorithmic task, the paper frames it as a socio-technical infrastructure in which machine learning components interact with organizational expertise, regulatory standards, and market incentives. It concludes by identifying forward-looking research and policy directions for building accountable, sustainable, and operationally credible machine learning systems in financial reporting oversight.

Keywords

accounting irregularities, financial reporting, machine learning, natural language processing, governance, explainability, audit automation, anomaly detection.

1. Introduction

The detection of accounting irregularities in corporate financial reports has long occupied a central position in accounting research, securities regulation, and capital market oversight [1]. As reporting complexity and the volume of unstructured disclosure have increased, conventional screening tools have struggled to keep pace with the diversity of earnings manipulation strategies. Foundational work on earnings manipulation emphasized the decomposition of reported earnings into expected and unexpected components, with unexpected accruals treated as a signal of managerial discretion [2]. This conceptual framework was grounded in the recognition that accruals, estimates, and classification decisions provide a tractable but imperfect window into reporting integrity. Subsequent scholarship broadened the view from a narrow focus on accruals toward a more comprehensive understanding of reporting quality, including revenue recognition, expense capitalization, reserve estimation, segment reporting, and disclosure completeness [3].

Machine learning introduces a different research and operational paradigm. Rather than relying on a small set of handcrafted indicators, supervised and unsupervised models can learn high-dimensional representations from structured financial variables and unstructured narrative texts. This shift creates new opportunities for detecting subtle and evolving patterns that are difficult to capture with static analytical procedures. At the same time, it creates substantial system-level challenges related to data quality, model opacity, imbalanced training data, regulatory acceptance, and the allocation of scarce investigative resources. This paper examines the structural, infrastructural, governance, and policy dimensions of machine learning-based detection of accounting irregularities. It emphasizes trade-offs that arise when algorithmic systems are embedded in audit firms, regulatory agencies, and corporate control environments. The goal is not to evaluate a single algorithm but to analyze the architecture of detection systems as institutional and technical arrangements.

2. The Accounting Irregularity Detection Landscape

The empirical literature on machine learning for accounting fraud detection reveals a progression from logistic regression and decision trees to ensemble classifiers and deep representation learning [4]. Early contributions compared multiple predictive algorithms using historical fraud cases and matched non-fraud firms, often reporting improvements in sensitivity but comparatively high false positive rates when models were evaluated on realistic populations [5]. Data mining approaches applied to financial statement data demonstrated that algorithmic screening can complement traditional audit procedures, yet their practical value depends heavily on the prevalence of fraud, the cost of erroneous alerts, and the tolerance of investigators for low precision [6]. In parallel, textual analysis of regulatory filings showed that managerial language in annual reports contains information about risk, uncertainty, and obfuscation that is not fully captured by numerical ratios alone [7].

The landscape is marked by a persistent tension between detection power and operational feasibility. Fraudulent reporting is rare, which means that even models with strong relative performance can generate many false positives when deployed across thousands of firms. The ground truth for fraud is also imperfect because regulatory enforcement and restatement data capture only detected cases. This partial observability creates a selection problem: the labeled examples available for training may not represent the full distribution of irregular reporting. Consequently, model evaluation must account for label noise, discovery lag, and the possibility that some non-prosecuted cases still contain material misstatements. These features distinguish accounting irregularity detection from many other anomaly detection problems

and shape the design of data pipelines, model selection criteria, and post-deployment monitoring.

3. Machine Learning Architectures for Financial Report Screening

Recent architectures move beyond static numeric inputs toward hybrid models that jointly encode financial statement variables and narrative disclosures. Structural deviation in risk factor narratives has been modeled as a semantic anomaly detection task, connecting textual coherence and latent meaning to the stability of reporting choices [8]. Large-sample machine learning estimates using public company data suggest that models trained on accounting ratios, audit variables, and market information can improve fraud detection relative to traditional screening heuristics [9]. Reviews of classification frameworks have highlighted that feature construction, sampling strategy, and cost-sensitive evaluation often matter more than the choice of a single base learner [10]. The broader fraud detection literature similarly frames model deployment as a system design problem in which data availability, alert review capacity, and regulatory constraints shape the operational meaning of accuracy [11].

The choice of architecture is therefore not solely a technical decision. Supervised models require labeled historical cases, which may be sparse and subject to regulatory delay. Semi-supervised and unsupervised methods can be trained on large unlabeled populations but produce alerts that are harder to interpret. Ensemble methods reduce variance and improve stability across reporting regimes, but they introduce additional complexity in model refresh and explanation. Deep learning models can capture nonlinear interactions and semantic structure in disclosure text, yet they may amplify biases in training data and require more intensive governance. In practice, organizations often deploy layered architectures in which broad screening models produce a risk-ranked set of candidates, followed by focused models that incorporate textual, governance, and market signals for a smaller set of high-priority firms.

4. Textual and Semantic Features in Corporate Disclosures

Textual features derived from management discussion, footnotes, and risk factor sections have become central to modern detection systems. Meta-learning frameworks suggest that combining multiple base classifiers across heterogeneous feature sets can improve fraud detection and reduce overfitting to a specific reporting regime [12]. Linguistic predictors, including word usage patterns, syntactic complexity, and deceptive language markers, have been shown to discriminate between fraudulent and non-fraudulent filings in controlled settings [13]. Credibility analysis has further explored how deceptive narratives exhibit measurably different structure, affect, and specificity when compared with truthful disclosures [14]. Topic modeling and other unsupervised semantic methods have revealed that misreporting firms systematically alter the thematic content of their disclosures, providing an additional signal beyond accounting quality [15].

However, the transformation of unstructured text into model-ready features introduces important conceptual problems. Dictionary-based methods depend on word lists that may not generalize across industries, regulatory regimes, or time periods. Embedding-based methods provide dense semantic representations but may lose interpretability for auditors who need to explain a high-risk classification. Moreover, corporate disclosures are strategic documents. Managers may adjust language in anticipation of regulatory scrutiny, and the meaning of risk factor disclosures changes as disclosure norms evolve. A machine learning model that treats historical text as a stable signal may therefore degrade when disclosure practices shift after

enforcement actions or new accounting standards. The effective use of textual features requires continuous retraining, domain-aware preprocessing, and careful alignment with structured financial variables.

5. Data Infrastructure, Governance, and Deployment Economics

Deploying machine learning detection systems at scale requires integrated data pipelines that extract, align, and validate data from heterogeneous sources. Evidence from IPO prospectus analysis demonstrates that narrative documents carry information beyond accounting variables, but the extraction of that information creates processing challenges related to document structure, comparability, and temporal consistency [16]. A broader survey of textual analysis research in accounting emphasizes that disclosure language is not a homogeneous stream; it is shaped by regulation, disclosure strategy, and industry norms [17]. The development of natural language processing in accounting and finance has similarly highlighted the need for domain-specific tokenization, phrase detection, and semantic disambiguation rather than generic language models alone [18].

These technical requirements are inseparable from governance and deployment economics. Data lineage must be documented so that model inputs can be traced to source filings and corrected when errors propagate. Model refresh cycles must balance stability with responsiveness to changes in reporting standards and disclosure tone. The cost of false alerts is not limited to investigator time; it also includes the reputational burden on firms that are erroneously flagged and the potential dampening of legitimate risk disclosure. Organizations therefore need governance structures that define thresholds, escalation procedures, and human review responsibilities. The machine learning model is one component of a larger control infrastructure, and its value depends on the quality of the surrounding data engineering and organizational workflows.

6. Robustness, Fairness, and Explainability

Fraud detection datasets are highly imbalanced, and misspecification of training objectives can produce models that appear accurate but fail on rare positive cases [19]. The imbalance problem is not merely statistical; it interacts with class overlap, small disjuncts, and temporal drift in corporate reporting behavior [20]. If a model optimizes overall accuracy, it may simply ignore fraud cases because they are scarce. Cost-sensitive learning, resampling, and ensemble methods can mitigate this problem, but each approach introduces its own biases and operational consequences. Robustness also requires stress testing under different economic conditions, because the features of fraudulent reporting may differ across expansionary and downturn periods.

Explainability has become a prerequisite for audit and regulatory acceptance. Unified model interpretation methods provide post hoc attributions that allow reviewers to inspect which inputs drive a high-risk classification [21]. Local explanation approaches can produce instance-level explanations, but their reliability depends on the underlying model and the choice of perturbation neighborhood [22]. In high-stakes settings, interpretable model families may be preferable when their transparency can be preserved without unacceptable loss of detection power [23]. Fairness is another pressing concern. Models trained on historical enforcement data may reproduce existing regulatory attention patterns, focusing on industries or firm profiles that were scrutinized in the past. This can lead to uneven monitoring burdens and may fail to detect emerging fraud types in less examined sectors. A fair and credible

detection system must therefore include audits of model performance across firm size, industry, and disclosure complexity.

7. Policy, Regulatory, and Organizational Implications

From a policy perspective, machine learning detection systems must be evaluated according to their total cost of ownership, including data engineering, model refresh, and alert investigation capacity. The fraud detection literature underscores that data engineering choices, such as timestamp alignment and feature leakage prevention, can determine operational success more than algorithm selection [24]. Reviews of data mining applications in accounting show that adoption has been uneven, with governance and interpretability concerns often limiting deployment inside audit firms and regulatory agencies [25].

Regulators face the challenge of setting expectations for model risk management without stifling innovation. Audit standards have historically emphasized professional skepticism and evidence-based judgment. Machine learning outputs do not replace these responsibilities but rather reframe them: auditors must decide how to evaluate model-generated risk indicators and how to document the limitations of algorithmic support. Cross-jurisdictional differences in disclosure requirements and enforcement practices further complicate the design of global detection infrastructures. A system that performs well in one regulatory environment may be less effective in another because the meaning and frequency of restatements, enforcement actions, and disclosure language differ. Policy discussions should therefore focus on transparency, auditability, and the appropriate allocation of human expertise alongside machine learning systems. The long-term goal is not complete automation but the construction of governance arrangements that make detection more consistent, explainable, and responsive to emerging risks.

8. Conclusion

Machine learning-based detection of accounting irregularities represents a significant evolution in the monitoring of corporate financial reporting. It offers the potential to integrate structured financial signals with unstructured narrative disclosures, to identify subtle reporting anomalies, and to support more efficient allocation of audit and regulatory resources. Yet the transition from algorithmic research to institutional deployment is not straightforward. Detection systems must contend with rare and partially observed fraud labels, imbalanced data, evolving disclosure practices, regulatory heterogeneity, and the operational cost of false alerts. Moreover, the interpretability and fairness of machine learning outputs shape their legitimacy among auditors, boards, and enforcement agencies. A sustainable detection infrastructure requires careful attention to data governance, model refresh cycles, and the integration of algorithmic signals with professional judgment. It also requires policy frameworks that recognize both the potential and the limits of machine learning in high-stakes financial oversight. Future research should continue to examine hybrid architectures, temporal robustness, and the institutional conditions under which machine learning can improve the detection of accounting irregularities without introducing new forms of bias or opacity.

References

1. Beneish, M. D. (1999). The detection of earnings manipulation. Financial Analysts Journal, 55(5), 24-36.

2. Dechow, P. M., Ge, W., & Schrand, C. M. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. Journal of Accounting and Economics, 50(2-3), 344-401.

3. Jones, M. J. (1991). Earnings management during import relief investigations. Journal of Accounting Research, 29(2), 193-228.

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

5. Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. Auditing: A Journal of Practice & Theory, 30(2), 19-56.

6. 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.

7. Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35-65.

8. Sun, F., He, S., Wang, R., Ke, L., Shen, H., & Liao, Q. (2026). Modeling Structural Deviation in 10-K Risk Factors: A Semantic Anomaly Detection and Explainable AI Approach. Risks, 14(4), 87.

9. Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in public companies: A machine learning approach. Journal of Accounting Research, 58(2), 493-541.

10. 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.

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

12. Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A meta-learning framework for detecting financial fraud. MIS Quarterly, 36(4), 1293-1327.

13. Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings? Journal of Emerging Technologies in Accounting, 7(1), 25-46.

14. Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. Decision Support Systems, 50(3), 585-594.

15. Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. Journal of Accounting Research, 58(1), 237-291.

16. Hanley, K. W., & Hoberg, G. (2010). The information content of IPO prospectuses. Review of Financial Studies, 23(7), 2821-2864.

17. Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. Journal of Accounting Literature, 29, 143-165.

18. Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. Intelligent Systems in Accounting, Finance and Management, 23(3), 157-214.

19. Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. Progress in Artificial Intelligence, 5(4), 221-232.

20. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284.

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

22. 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.

23. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

24. Baesens, B., Höppner, S., & Verdonck, T. (2021). Data engineering for fraud detection. Decision Support Systems, 150, 113492.

25. Amani, F. A., & Fadlalla, A. M. (2017). Data mining applications in accounting: A review of the literature and organizing framework. International Journal of Accounting Information Systems, 24, 32-58.