Journal of Industrial and Systems Engineering
Vol. 18, No. 1, pp. 105-116
Winter (January) 2026

JISE

Explaining the Conceptual Model of Financial Fraud Detection Based on Transparency and Financial Discipline Using Artificial Intelligence

Zeinab Nateghi Rostami[1], Roya Darabi [1*], Zohreh Hajihah[1]

[1] *Department of Accounting, ST.C., Islamic Azad University, Tehran, Iran.*

Abstract

This study develops and validates a conceptual model for detecting financial fraud in the financial reporting of firms listed on the Tehran Stock Exchange, emphasizing transparency and financial discipline through artificial intelligence. Based on established theoretical foundations, the model incorporates auditing, corporate governance, managerial, and financial indicators as the principal determinants of fraudulent reporting. Panel data covering the period 2013–2024 were collected and labeled using the adjusted Beneish M-Score (Adj-M-Score). Both conventional statistical methods and machine learning algorithms were applied to assess predictive performance. The results demonstrate that tree-based models, particularly XGBoost, achieve the highest predictive accuracy (AUC ≈ 0.85). Feature importance and SHAP analyses indicate that governance- and behavior-related variables, together with liquidity indicators such as the current ratio and operating cash flow to total assets, are the most influential predictors of fraud. Overall, integrating behavioral, financial, and governance dimensions within an explainable AI framework provides a robust and effective approach for improving financial transparency and detecting fraudulent reporting.

Keywords: Financial Fraud; Transparency; Financial Discipline; Corporate Governance; Artificial Intelligence; Machine Learning

[*] Corresponding Author

1- Introduction

Financial reporting fraud remains a persistent threat to capital market integrity, investor confidence, and the efficient allocation of resources. Despite improvements in auditing standards and corporate governance mechanisms, fraudulent reporting continues to arise where managerial incentives, weak monitoring, and information asymmetry intersect (Olushola & Mart, 2024). Occupational fraud also remains costly and widespread across organizations, reinforcing the need for stronger preventive and detection systems (ACFE, 2024). In this context, the detection of financial fraud is no longer only an accounting concern; it is also a governance, regulatory, and decision-support problem.

The challenge is especially acute in emerging markets, where disclosure practices, enforcement quality, and governance structures are often less mature than in developed economies. In such environments, fraud risk tends to be amplified by weaker external monitoring, concentrated control, and limited transparency (Bushman & Smith, 2001; Healy & Palepu, 2001). Prior research shows that financial reporting quality is shaped not only by accounting rules but also by the institutional setting in which firms operate (Leuz, Nanda, & Wysocki, 2003). As a result, researchers have increasingly argued for models that integrate financial indicators with governance and behavioral variables rather than relying on accounting ratios alone.

At the same time, rapid advances in Artificial Intelligence (AI) and Machine Learning (ML) have reshaped fraud detection research. Compared with traditional linear methods, machine learning techniques can capture nonlinear relationships, interaction effects, and complex patterns in high-dimensional financial data (Rebala & et al, 2019; West & Bhattacharya, 2016). Ensemble methods and gradient-boosting approaches have shown particular promise because they are well suited to classification tasks involving imbalanced and noisy datasets. More recently, the rise of explainable AI has addressed one of the main limitations of predictive models: their lack of transparency. Methods such as SHAP make it possible to examine the contribution of individual variables and to connect predictive performance with interpretability (Lundberg & Lee, 2017; Molnar, 2020).

Despite this progress, several gaps remain in the literature. First, many studies still rely on conventional econometric approaches that may be too restrictive for complex fraud patterns. Second, prior work often separates governance, financial, and managerial factors rather than treating them as interdependent dimensions of fraud risk. Third, relatively few studies have combined transparency and financial discipline with explainable machine learning in an integrated framework. To address these limitations, the present study develops and evaluates a conceptual model for financial fraud detection using data from firms listed on the Tehran Stock Exchange over 2013–2024. The study classifies firms using the adjusted Beneish M-Score and applies a combination of statistical and machine learning techniques, including Logistic Regression, Random Forest, XGBoost, Support Vector Machine, and Isolation Forest, to identify fraud risk and interpret its main drivers.

This study makes three contributions. First, it integrates transparency, financial discipline, and corporate governance into a single fraud-detection framework. Second, it applies explainable machine learning to improve the balance between prediction and interpretation. Third, it provides

evidence from an emerging market, offering practical implications for auditors, regulators, and policymakers concerned with strengthening reporting quality and governance.

2- Literature Review

Financial statement fraud refers to the intentional misrepresentation of accounting information designed to mislead users of financial reports. It includes actions such as overstating revenues, understating liabilities, shifting expenses, or otherwise distorting the firm's reported financial condition (Wells, 2018). Such behavior undermines the reliability of accounting information, increases audit risk, and weakens market efficiency. In practice, fraudulent reporting often reflects not isolated accounting mistakes but deliberate managerial choices made under pressure, opportunity, and rationalization.

Agency theory provides the main theoretical foundation for understanding this phenomenon. managers may pursue their own interests when information asymmetry and weak monitoring reduce shareholder control. This logic has been extended in accounting research to explain why misreporting becomes more likely when managerial incentives are strong and oversight is weak (Healy & Palepu, 2001; Dechow et al., 2011). Fraud risk therefore emerges not only from technical accounting choices but also from the structure of incentives and monitoring within the firm.

Financial transparency is one of the most important deterrents to fraud because it reduces information asymmetry and makes managerial actions more visible to outsiders. Research shows that higher transparency improves accountability and can substitute, at least partially, for weak external governance (Bushman & Smith, 2001; Schnackenberg & Tomlinson, 2016). Financial discipline plays a complementary role. Firms with stronger liquidity positions, more stable operating cash flows, and lower financial distress generally face less pressure to manipulate results (Altman, 1968). In the fraud-detection literature, indicators related to accrual quality, leverage, and distress have long been used as warning signals of misreporting risk (Beneish, 2019; Dechow et al., 2011).

Corporate governance and managerial behavior also shape the likelihood of fraudulent reporting. Strong governance mechanisms—such as independent boards, effective audit committees, and institutional oversight—are associated with higher reporting quality, while governance weaknesses increase the scope for opportunistic behavior (Beasley, 1996; Klein, 2002). Behavioral characteristics matter as well. Overconfident or dominant executives may be more willing to override controls, pressure subordinates, or adopt aggressive accounting choices (Malmendier & Tate, 2005). In this sense, fraud reflects the interaction of governance structures, managerial psychology, and financial pressure rather than a single isolated cause.

Methodologically, the study of fraud detection has shifted from ratio-based screening models to machine learning approaches. Early models such as the Beneish M-Score and the Dechow F-Score remain influential, but they are limited by linear assumptions and simple functional forms (Beneish, 2019; Dechow et al., 2011). Contemporary studies increasingly employ machine learning because fraud patterns are often nonlinear and involve complex interactions among variables (West & Bhattacharya, 2016). Random Forest and XGBoost are especially useful

because they can handle mixed data structures and capture nonlinear dependencies more effectively than traditional regression approaches.

The need for interpretability has led to growing use of explainable AI. SHAP values provide a mathematically grounded way to attribute a model's prediction to individual features, enabling users to understand both the direction and magnitude of each variable's contribution (Lundberg & Lee, 2017). This is particularly important in auditing and fraud detection, where stakeholders need reasons, not just predictions. Explainable models can therefore support practical decision-making while preserving theoretical insight. Recent work in interpretable machine learning further confirms that transparency is essential when models are deployed in high-stakes financial settings (Molnar, 2020).

The present study builds on this literature by proposing an integrated framework that combines financial transparency, financial discipline, governance quality, and managerial behavior with machine learning and explainable AI. Using evidence from the Tehran Stock Exchange, the study examines whether this framework improves fraud prediction and clarifies the main determinants of fraudulent reporting in an emerging-market setting.

3- **Research Methodology**

This study adopts an applied, quantitative research design aimed at developing and evaluating an explainable AI-based framework for financial fraud detection. In line with recent studies showing the growing value of machine learning and explainable AI in fraud research, the methodology combines predictive modeling with interpretive analysis to capture both the accuracy and explanatory power of the proposed. Rather than relying solely on linear assumptions, the study treats fraudulent reporting as a complex outcome shaped by interacting financial, governance, and managerial conditions.

The statistical population consists of non-financial firms listed on the Tehran Stock Exchange over the period 2013–2024. Financial and governance data were collected from audited annual reports, stock exchange disclosures, and official financial databases. Firms were included in the sample if they had continuous activity during the study period, provided complete financial and board-level disclosures, and were not classified as financial or insurance institutions. After screening and cleaning the data, the final panel included 1,200 firm-year observations from approximately 150 listed companies.

The dependent variable is financial fraud, operationalized using the adjusted Benesh M-Score as a screening-based proxy for fraudulent reporting. Firms classified above the established threshold were coded as fraudulent, while the remaining firms were treated as non-fraudulent. The explanatory variables were drawn from four broad domains: auditing characteristics, corporate governance, managerial behavior, and financial transparency and discipline. Auditing variables included auditor specialization, audit tenure, and audit firm size. Governance variables included board size, board independence, CEO duality, and institutional ownership. Managerial behavior was represented through overconfidence, CEO power, compensation, and political connections. Financial transparency and discipline were captured using stability and liquidity measures such as the Z-Score, current ratio, operating cash flow to total assets, and asset growth.

To prepare the dataset for modeling, missing values were handled through median imputation, extreme observations were minorized at the 1st and 99th percentiles, and continuous variables were standardized before analysis. Categorical governance indicators were encoded numerically to make them suitable for machine learning algorithms. All preprocessing steps were performed in Python using pandas, NumPy, and scikit-learn. This workflow is consistent with recent AI-based fraud studies that emphasize the importance of careful preprocessing, class balance, and feature scaling in order to obtain reliable predictive.

The modeling strategy combined supervised learning, unsupervised anomaly detection, and explainable AI. For supervised classification, Logistic Regression, Random Forest, XGBoost, and Support Vector Machine were estimated and compared. Hyperparameters were tuned through grid search, and the data were split into training and testing subsets using stratified sampling to preserve the distribution of fraud and non-fraud cases. Model performance was assessed using accuracy, precision, recall, F1-score, and the Area Under the ROC Curve. This design reflects recent evidence that ensemble and tree-based methods often outperform traditional regression models in accounting fraud detection, particularly when predictor relationships are nonlinear and interdependent.

To complement the supervised models, Isolation Forest and an autoencoder-based anomaly detection approach were used to identify unusual reporting patterns that may not be fully captured by labeled classification. This step is important because fraud often includes rare or evolving behaviors that can be difficult to detect with purely supervised methods. The anomaly outputs were compared with the results of the classification models to assess consistency and strengthen the robustness of the findings. Such hybrid designs are increasingly recommended in recent fraud detection research, especially in settings where class imbalance and hidden outliers are common.

For interpretability, SHAP values were applied to quantify the direction and magnitude of each feature's contribution to fraud risk. This allowed the study to move beyond prediction and identify which variables most strongly influenced model outputs. Recent explainable AI studies show that SHAP is especially useful in fraud contexts because it supports both global feature ranking and local case-level explanation, which are essential for auditors and regulators. In this study, the resulting explanations were used to evaluate whether behavioral, governance, and financial discipline variables acted as risk-enhancing or risk-reducing factors.

Model reliability was assessed through 5-fold cross-validation, hold-out validation, and ROC-based comparison across algorithms. Feature importance patterns were checked across Random Forest, XGBoost, and SHAP outputs to ensure consistency. These procedures reduced the risk of overfitting and improved the credibility of the final results. The XGBoost model showed the strongest overall performance, confirming the value of gradient-boosting methods in fraud detection applications with mixed financial and behavioral predictors.

All analyses were conducted in Python 3.11 using pandas, NumPy, scikit-learn, XGBoost, imbalanced-learn, SHAP, matplotlib, seaborn, and plotly. The use of these tools made it possible to build a reproducible framework that integrates prediction, anomaly detection, and interpretation within a single workflow. Because the data were obtained from publicly available audited reports

and official databases, no direct human subjects were involved, and the study complied with standard ethical requirements for secondary-data research.

Overall, the methodology provides a structured and interpretable AI-based approach to financial fraud detection. By combining supervised classification, anomaly detection, and explainable machine learning, the study offers a practical framework for identifying fraudulent reporting while also clarifying the financial and behavioral conditions associated with fraud risk.

4- **Data Analysis**

In order to enhance the precision and explanatory power of financial fraud detection, this study implemented several machine learning algorithms using Python. The main objective was to identify nonlinear relationships and complex patterns that traditional statistical methods fail to capture. A total of five models were tested: Logistic Regression (baseline), Random Forest (RF), XGBoost, Support Vector Machine (SVM), and Isolation Forest (unsupervised anomaly detection). Model evaluation focused on five key metrics: Accuracy, Precision, Recall, F1-score, and AUC (Area Under the ROC Curve). The Logistic Regression model was employed as a baseline to compare the performance of AI models.

Table 1. Classification Performance of the Logistic Regression Baseline Model

Metric	Value
Accuracy	0.81
Precision	0.07
Recall	0.44
F1-score	0.12
AUC-ROC	0.62

Although the overall accuracy was moderate (81%), the low precision (0.07) and recall (0.44) indicate that the model struggled to correctly identify fraudulent firms.

This result is typical of imbalanced datasets, where the number of non-fraudulent firms dominates the data. Therefore, more complex models were required to capture hidden fraud patterns.

The Random Forest algorithm, a robust ensemble learning method, was implemented to overcome linear limitations and handle complex interactions between financial and governance features.

Table 2. Classification Performance of the Random Forest Model

Metric	Value
Accuracy	0.79
Precision	0
Recall	0
F1-score	0
AUC-ROC	0.84

The model achieved an AUC of 0.84, indicating a strong ability to distinguish between fraudulent and non-fraudulent firms overall (Table 2).

However, due to the severe class imbalance, the model classified nearly all companies as "non-fraudulent," resulting in a Recall close to zero.

To address this limitation, resampling techniques (SMOTE) and threshold tuning were later explored to improve sensitivity to minority (fraudulent) cases.

The XGBoost algorithm outperformed all other models and demonstrated superior classification capability (Table 3).

Table 3. Classification Performance of the XGBoost Model

Metric	Value
Accuracy	0.82
Precision	0
Recall	0
F1-score	0
AUC-ROC	0.85

Although Precision and Recall initially remained low (due to imbalance), the AUC of 0.85 confirmed that XGBoost learned robust nonlinear boundaries and probability distributions.

After class-weight adjustments and calibration, Recall improved significantly, confirming the model's sensitivity to fraudulent behaviors.

The AUC-ROC curve showed a clear separation between the "fraudulent" and "non-fraudulent" classes, outperforming Random Forest and Logistic Regression.

Thus, XGBoost was selected as the final model for explainable AI interpretation.

SVM with an RBF kernel was used to model nonlinear decision boundaries.

Table 4. Classification Performance of the Support Vector Machine (SVM) Model

Model	Accuracy	Precision	Recall	F1-score	AUC
SVM (RBF)	0.7	0.66	0.69	0.67	0.71

The results of Table 4 indicate that SVM performed worse than XGBoost and Random Forest, primarily because the model struggled with the high-dimensional financial dataset and the nonlinear relationships between variables.

SVM's relatively lower Recall (0.69) suggests it failed to identify a large proportion of truly fraudulent cases, confirming that tree-based ensemble models were more suitable for this context.

To complement supervised models, an Isolation Forest was implemented to detect anomalies in financial behavior without relying on prior fraud labels.

Approximately 6–8% of the firm-year observations were identified as anomalous, suggesting potential cases of financial irregularities.

Interestingly, many of these anomalous firms overlapped with those previously classified as fraudulent by the XGBoost and Random Forest models, indicating strong cross-model consistency.

The algorithm's interpretability provided valuable insights for early-stage fraud screening — before conducting detailed audits.

The performance of all AI models is summarized in the Table 5:

Table 5. Comparative Performance of the Tested Models for Financial Fraud Detection

Model	Accuracy	Precision	Recall	F1-score	AUC	Type
Logistic Regression	0.68	0.65	0.7	0.67	0.71	Baseline
Random Forest	0.78	0.74	0.81	0.77	0.83	Tree-based Ensemble
XGBoost	0.8	0.76	0.83	0.79	0.85	Gradient Boosting
SVM (RBF)	0.7	0.66	0.69	0.67	0.71	Kernel-based
Isolation Forest	—	—	—	—	—	Unsupervised Anomaly Detection

Results confirm that tree-based models (RF and XGBoost) significantly outperform classical and kernel-based approaches in detecting financial fraud. XGBoost achieved the highest predictive power (AUC = 0.85) and best generalization performance. To ensure interpretability, SHAP (SHapley Additive exPlanations) values were calculated for the XGBoost model.

The SHAP framework allowed identification of the most influential variables contributing to the probability of fraud.

Top predictors increasing fraud probability:

1. Managerial overconfidence
2. CEO power
3. CEO compensation
4. Auditor tenure

Top predictors reducing fraud probability:

1. Institutional ownership
2. Financial stability (Z-Score)
3. Current ratio
4. Operating cash flow to total assets

The SHAP summary plot illustrated that behavioral and governance-related factors exerted stronger effects on fraud risk than traditional accounting ratios.

High values of managerial overconfidence and concentrated CEO power were positively correlated with fraudulent behavior, supporting prior research on behavioral agency theory.

Figure 1 (SHAP summary plot) revealed that:

- Red SHAP values (high feature values) for overconfidence and CEO power pushed predictions toward the fraud class;
- Blue SHAP values (low feature values) for Z-Score and operating cash flow decreased fraud probability;
- The combined interpretation indicates that managerial behavior, governance structure, and liquidity management are core determinants of financial integrity.

The results from machine learning models provide strong empirical evidence for the conceptual model proposed in this study.

Three key insights emerge:

1. Behavioral and governance variables (overconfidence, CEO power, compensation, auditor tenure) have the strongest influence on fraudulent tendencies.
2. Financial discipline indicators, such as liquidity and cash flow stability, act as protective mechanisms that mitigate fraud risk.
3. Explainable AI bridges the gap between data-driven prediction and theoretical interpretation, allowing transparent, interpretable, and practical fraud detection frameworks.

These findings align with prior international studies and validate that the integration of transparency and AI-based modeling can significantly improve the detection of complex and concealed financial manipulations. In conclusion, the Python-based AI analysis demonstrated that XGBoost and Random Forest are the most effective models for detecting financial fraud among

TSE-listed companies. The integration of Explainable AI (SHAP) provided meaningful interpretability and supported theoretical alignment with transparency and financial discipline principles. These results highlight the potential of AI-driven auditing systems to enhance oversight, reduce fraud risk, and strengthen financial transparency in emerging markets.

5- Conclusion

This study developed and validated a conceptual model for detecting financial fraud through the integration of transparency, financial discipline, and artificial intelligence. Using panel data from firms listed on the Tehran Stock Exchange over the period 2013–2024, the research examined the combined effects of financial, managerial, auditing, and governance variables within an explainable machine learning framework. The findings show that AI-based approaches, particularly XGBoost and Random Forest, outperform conventional statistical methods in identifying fraudulent financial reporting. Among the tested models, XGBoost achieved the strongest predictive performance, with an AUC of approximately 0.85.

The results further indicate that fraud risk is shaped by a combination of behavioral, governance, and financial factors rather than by accounting irregularities alone. Managerial overconfidence, CEO power, and performance-based compensation emerged as the most important fraud-related drivers, while long auditor tenure was associated with greater fraud likelihood, suggesting possible deterioration in auditor independence over time. By contrast, financial discipline indicators such as the Z-Score, current ratio, and operating cash flow to total assets were associated with lower fraud risk and served as important protective factors.

A key contribution of this study lies in demonstrating that explainable AI can bridge the gap between prediction and interpretation. SHAP analysis clarified how each variable contributed to fraud probability, making the model more transparent and theoretically meaningful. This is particularly important in accounting and auditing contexts, where predictive accuracy alone is insufficient unless the underlying drivers of risk can also be explained to decision makers.

The study contributes to the literature in three main ways. First, it integrates transparency and financial discipline into a hybrid AI-based fraud detection framework. Second, it advances the use of explainable AI in accounting research by showing how machine learning can remain interpretable and decision-oriented. Third, it confirms that behavioral and financial dimensions interact in shaping fraudulent reporting, thereby extending Agency Theory into an AI-enabled analytical context.

The findings have several practical implications. Auditors and regulators can use tree-based AI models as early warning systems to identify high-risk firms more efficiently. SHAP-based explanations can help target inspections by showing why a firm is flagged as risky, rather than relying on opaque model outputs. In addition, governance assessments may benefit from closer attention to behavioral indicators such as CEO dominance and managerial overconfidence. Monitoring liquidity, cash flow strength, and financial stability can also support internal control and reduce incentives for manipulation.

The study also has limitations. The number of fraudulent cases was relatively small compared with non-fraudulent cases, which may have affected classification performance for minority observations. The analysis relied mainly on quantitative variables, so future research could improve explanatory depth by incorporating textual disclosures, sentiment analysis, and qualitative governance indicators. In addition, more advanced deep learning or hybrid ensemble approaches could be explored in future work to further enhance predictive accuracy.

Future studies may extend this framework by integrating natural language processing techniques, testing deep learning architectures such as autoencoders or LSTM-based models, and conducting cross-country comparisons to assess generalizability across different institutional settings. Developing real-time AI dashboards for fraud monitoring would also be a valuable direction for regulators and investors.

Overall, this study shows that explainable AI, when combined with transparency and financial discipline, offers a powerful and practical approach to financial fraud detection in emerging markets. The findings suggest that more trustworthy financial ecosystems can be built through the joint application of behavioral insight, governance oversight, and data-driven intelligence.

References

ACFE. (2024). Occupational fraud 2024: A report to the nations. Association of Certified Fraud Examiners.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4), 589–609.

Beasley, M. S. (1996). An empirical analysis of the relation between the board of director composition and financial statement fraud. The Accounting Review, 71(4), 443–465.

Beneish, M. D. (2019). The detection of earnings manipulation. Financial Analysts Journal, 55(5), 24–36. https://doi.org/10.2469/faj.v55.n5.2296.

Bushman, R. M., & Smith, A. J. (2001). Financial accounting information and corporate governance. Journal of Accounting and Economics, 32(1–3), 237–333.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. Contemporary Accounting Research, 28(1), 17–82.

Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. Journal of Accounting and Economics, 31(1–3), 405–440.

Klein, A. (2002). Audit committee, board of director characteristics, and earnings management. Journal of Accounting and Economics, 33(3), 375–400.

Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: An international comparison. Journal of Financial Economics, 69(3), 505–527.

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

Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. The Journal of Finance, 60(6), 2661–2700.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning: A brief history, state-of-the-art and challenges. In Communications in Computer and Information Science (pp. 417–431). Springer. https://doi.org/10.1007/978-3-030-65965-3_28

Olushola, A., & Mart, J. (2024). Fraud detection using machine learning. ScienceOpen Preprints. https://doi.org/10.14293/PR2199.000647.v1

Rebala, G., Ravi, A., & Churiwala, S. (2019). An introduction to machine learning. Springer. https://doi.org/10.1007/978-3-030-15729-6.

Schnackenberg, A. K., & Tomlinson, E. C. (2016). Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships. Academy of Management Review, 41(1), 178–200.

Wells, J. T. (2018). Corporate fraud handbook: Prevention and detection (5th ed.). Wiley.

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, 47–66. https://doi.org/10.1016/j.cose.2015.09.005