Building an Intelligent Model for Identifying Corporate Financial Fraud by Integrating Big Data Mining and Machine Learning Technologies

Xiaolin Zheng
Zhangzhou City Vocational College, Zhangzhou 363000, Fujian, China
Corresponding author: Xiaolin Zheng (e-mail: zhengxiaolinei@163.com)

ABSTRACT Intelligent anomaly detection based on big data analytics and machine learning has become an essential approach for data-driven monitoring and decision support in complex engineering systems. To improve the efficiency and reliability of corporate financial fraud identification, this study develops an intelligent detection framework that integrates multisource data mining, feature engineering, ensemble learning, and dynamic model optimization. The proposed methodology combines heterogeneous financial information with machine learning algorithms to construct adaptive fraud identification models capable of capturing hidden abnormal patterns and evolving fraud behaviors. A systematic strategy involving data integration, feature selection, algorithm fusion, incremental model updating, and interpretability enhancement is further presented to improve detection robustness and practical applicability. The framework effectively addresses challenges associated with data quality, sample imbalance, feature adaptability, and model transparency while supporting continuous optimization through human–machine collaborative feedback. Beyond financial risk management, the proposed intelligent identification architecture provides methodological references for anomaly detection, distributed data fusion, adaptive pattern recognition, and intelligent monitoring in engineering systems, offering potential applications for information processing and decision support in Electromagnetic Waves, Antennas and Propagation, particularly in wireless sensing networks and data-driven signal analysis.

INDEX TERMS machine learning, big data mining, anomaly detection, intelligent data fusion

I. INTRODUCTION

THE prevalence of financial fraud in raw material procurement and inventory valuation poses a significant threat to the healthy development of capital markets within the manufacturing sector. Its harm extends beyond direct losses to investors; it also systematically undermines market trust mechanisms. In recent years, with the increasing scale of business operations and the growing complexity of financial data, traditional manual fraud detection methods have become insufficient, exhibiting increasingly prominent drawbacks such as low efficiency, limited coverage, and significant lag. Meanwhile, the maturity of big data mining technology and machine learning algorithms offers a new solution to this dilemma. Combining these two approaches allows for the automatic identification of abnormal features from large amounts of financial data and the creation of fraud early warning models. This paper explores the integration of intelligent identification models to monitor high-volume production data and pricing anomalies, aiming to improve the precision and scalability of financial fraud detection through automated algorithmic oversight.

II. THE SIGNIFICANCE OF INTEGRATING BIG DATA AND MACHINE LEARNING TO IDENTIFY FINANCIAL FRAUD

A. BREAKING THROUGH THE LIMITATIONS OF TRADITIONAL AUDITING AND IMPROVING THE EFFICIENCY OF FRAUD DETECTION

Traditional financial audits mostly adopt sampling inspection methods, which are limited by manpower and time, have a small coverage, and cannot fully verify all the financial data of the enterprise. Big data mining technology can scan and automatically compare a large number of financial records in an all-round way, and machine learning algorithms can continuously improve the identification rules on the basis of the above, making the fraud identification efficiency far exceed that of traditional methods, providing strong technical support for audit work [1].

B. ENHANCE FINANCIAL DATA PROCESSING CAPABILITIES TO ACHIEVE MULTI-DIMENSIONAL ANOMALY DETECTION

Corporate financial data comes from diverse sources, has different structures and complex dimensions, and it is difficult for a single indicator to reflect all aspects of fraudulent behavior. After integrating big data mining and machine learning technologies, the system can process various information such as financial statement data, cash flow, and related transactions at the same time. It relies on multi-dimensional feature linkage analysis to find hidden abnormal patterns that are difficult to detect from one side, thereby significantly enhancing the depth and breadth of fraud detection [2].

C. FACILITATING THE CONSTRUCTION OF ENTERPRISE RISK PREVENTION AND CONTROL SYSTEMS AND ENHANCING REGULATORY CAPABILITIES

The use of intelligent identification models is not only a tool for post-audit, but can also be embedded in the daily financial management of enterprises, so as to achieve real-time early warning and dynamic monitoring of corporate fraud risks. It is conducive to enterprises to establish a more complete internal control system, and also provides more targeted regulatory means for enterprise regulatory departments, prompting financial supervision to change from passive response to active prevention and control, thereby comprehensively improving the level of market governance [3].

III. THE REAL-WORLD CHALLENGES OF INTELLIGENT DETECTION OF CORPORATE FINANCIAL FRAUD

A. THE QUALITY OF FINANCIAL DATA VARIES GREATLY, AND THE PROBLEM OF SAMPLE IMBALANCE IS PROMINENT

The premise for the effective operation of the financial fraud intelligent identification model is high-quality, structured training data. However, in actual operation, enterprises will encounter problems such as missing values, high noise, and non-standard labeling of financial data, which will affect the learning effect of the model. More obviously, there is the problem of sample imbalance. In real datasets, the number of fraudulent samples is very small, and the ratio of normal samples to fraudulent samples is often tens or even hundreds of times [4]. Preliminary statistics show that in many financial datasets, the proportion of fraudulent samples is generally less than 5% of the total sample size, and in some datasets, it is even less than 2%. Due to the serious imbalance in data distribution, the model will predict all samples as normal during the training process, thus achieving a high overall accuracy on the surface, but the actual fraud identification recall rate is very low, and the model almost fails in important tasks [5]. To further illustrate the universality of data quality problems, the following table summarizes and categorizes the common quality defects and their impact in different types of financial datasets.

B. FRAUDULENT METHODS ARE CONCEALED AND VARIED, AND TRADITIONAL FEATURE ENGINEERING IS NOT ADAPTABLE ENOUGH

Financial fraud methods are constantly evolving with stricter regulations and technological development, and their concealment and complexity are also increasing. Common fraud methods include inflating revenue, capitalizing expenses, and manipulating related-party transactions. The fraud methods used in different

industries are also very different [6]. Traditional feature engineering relies on the experience of experts to manually design feature variables. On the one hand, the feature construction process is very complicated and time-consuming. On the other hand, the static feature set cannot keep up with the changes in fraud patterns. The survey results show that in the financial fraud cases that have been exposed, about 60% of the cases have some abnormal financial signals in the early stage, which are difficult to detect by existing feature indicators [7]. The following table compares and analyzes the adaptability of traditional feature engineering in the identification of different methods from the perspective of the main fraud types, which helps to intuitively present the limitations of traditional methods.

C. THE MODEL HAS WEAK INTERPRETABILITY, RESULTING IN LOW TRUST IN PRACTICAL APPLICATIONS AND REGULATORY OVERSIGHT

The best-performing machine learning models, such as deep neural networks and gradient boosting trees, are all black-box models, and their internal decision-making logic is difficult to directly show. For financial auditing and supervision, the traceability of the identification conclusion and the transparency of the logic are very important. Auditors need to know clearly the basis on which the model identifies a certain financial record as abnormal, and regulatory authorities also need to conduct compliance checks on the decision-making process made by the model. However, existing high-performance models have obvious shortcomings in interpretability [8]. More than 70% of financial auditors believe that they cannot understand the reasoning process behind the model output results, which will affect their acceptance and enthusiasm for using intelligent tools. At the same time, the regulatory authorities are constantly increasing the compliance requirements for algorithm decision-making. Many industries have made provisions for the requirements of financial risk judgment models, which must provide a complete explanation of the decision-making path [9]. In order to present the trade-off between performance and interpretability of different types of models, the following table compares the key dimensions of the current mainstream machine learning models.

TABLE 1. Distribution of common quality issues in financial datasets

Data quality problem types	Frequency of occurrence	The degree of impact on model training	Main manifestations
Too many missing values	higher	Larger	The absence of key financial fields affects the completeness of features.
Severe sample imbalance	Very high	Very large	The percentage of fraudulent samples is less than 5%, resulting in an extremely low recall rate.
Non-standard labeling	medium	Larger	Human annotation errors cause noise in training labels
High data noise	higher	medium	Input errors and formatting issues affect feature extraction.
Time series break	medium	medium	Data missing for the reporting period, time series features invalid.

TABLE 2. Comparison of the adaptability of different fraud methods to traditional feature engineering

Types of fraud methods	Main characteristics	Traditional feature engineering recognition difficulty	Identify the reasons for limitations
Inflated revenue	Revenue and cash flow diverge	medium	Single-period features can be captured, but cross-period camouflage is difficult to identify.
Cost capitalization	Inflated assets and abnormally low expenses	higher	Significant industry differences make it difficult to define static thresholds.
Related-party transaction manipulation	Abnormal proportion of related-party transactions	Very high	Requires multi-dimensional linkage across multiple entities; single-variable approaches are ineffective.
Cross-period collaborative fraud	Multiple data linkage anomalies	Very high	Temporal features are missing, and static models cannot capture them.
Hidden profit manipulation	The early signal was extremely weak.	Extremely high	More than 60% of early signals are below the existing feature threshold.

TABLE 3. Comparison of Performance and Interpretability of Mainstream Machine Learning Models

Model type	Recognition accuracy	Explainability	Practical acceptance	Main limitations
Logistic Regression	Common	powerful	higher	Weak ability to capture nonlinear features
Decision Tree	medium	Strong	medium	Prone to overfitting, with limited generalization ability.
Random Forest	higher	Weak	medium	Feature importance is roughly assessed, and the decision-making path is opaque.
Gradient Boosting Tree (XGBoost)	high	weak	Low	Its opaque nature raises questions about regulatory compliance.
Deep Neural Networks	Very high	Very weak	Low	The decision-making logic is completely opaque, and the attribution of responsibility is difficult to define.

IV. STRATEGIES FOR BUILDING FINANCIAL FRAUD DETECTION MODELS INTEGRATING BIG DATA MINING AND MACHINE LEARNING

Building upon the identified challenges of data quality, feature engineering, and model interpretability, the following section proposes a systematic framework designed to bridge these technical gaps. To visually present the complete operation process of multi-source heterogeneous data integration and cleaning, the following diagram systematically outlines the logical sequence and core tasks of each step.

A. INTEGRATION AND CLEANSING OF MULTI-SOURCE HETEROGENEOUS DATA: STRENGTHENING THE FOUNDATION OF MODEL DATA

High-quality data is the prerequisite for building an effective financial fraud detection model. Regarding data sources, the scope of collection should encompass a wide array of structured and semi-structured data—including corporate financial statements, business registration information, and details of related-party transactions. Where local data privacy regulations and multi-agency information-sharing frameworks permit, this scope may be extended to include bank transaction records and tax filing data to further comprehensively characterize the enterprise's operational status. During the data integration phase, unified data standards and specifications must be established to eliminate discrepancies across various sources—specifically regarding field definitions, measurement metrics, and temporal granularity—thereby providing a consistent foundation for subsequent feature extraction. The data cleaning phase requires the implementation of systematic quality control processes; appropriate imputation methods—such as mean imputation, median substitution, or multiple imputation—should be selected based on the business significance of specific fields

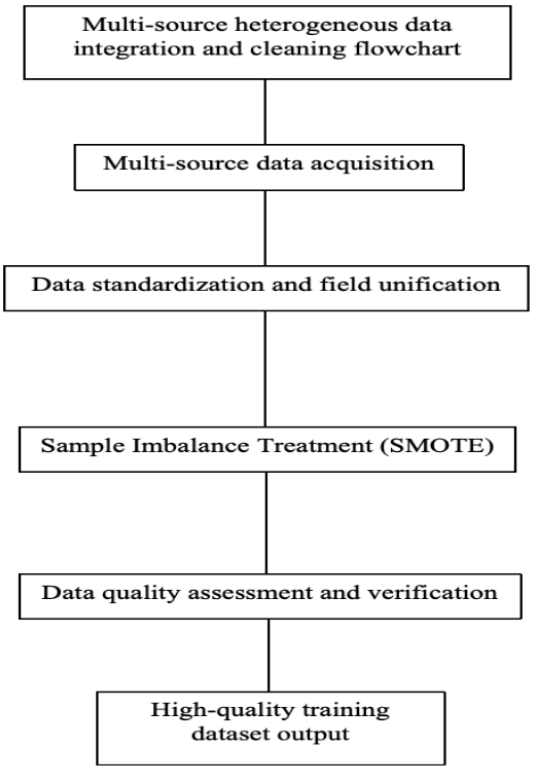

FIGURE 1. Flowchart of Multi-Source Heterogeneous Data Integration and Cleaning

and their missing rates, rather than simply deleting records, to prevent the loss of valuable information. Outliers should be identified and corrected using a dual-criterion approach that combines both business rules and statistical tests, ensuring that genuine signals of fraud are not misclassified as data errors and subsequently discarded. To address severe sample imbalance, a hybrid approach combining oversampling and undersampling techniques should be employed: synthetic samples are generated around the minority class instances to augment the volume of fraud samples, while a weighted loss function is incorporated into the ensemble learning framework to prioritize fraud samples during the training phase, thereby effectively enhancing the model's sensitivity in detecting minority categories. Furthermore, during data integration, significant attention must be paid to data timeliness management; the update frequencies of data from various sources should be harmonized to ensure that the data fed into the model is temporally synchronized and aligned. The look-back period for historical data analysis should be determined based on the typical latency of fraudulent activities—generally spanning no fewer than three full fiscal years—to enable the model to capture patterns associated with cross-period financial manipulation.

A financial risk screening project conducted by a provincial-level audit authority serves as a highly instructive example in this context. At the inception of the project, the authority established a unified data aggregation platform to align data sources from three distinct systems—business registration, taxation, and banking—using a standardized field mapping table; this process successfully resolved discrepancies regarding the definition and scope of the "operating revenue" field across the various systems. Regarding the bank transaction data—specifically, instances where approximately 12% of the monetary value fields were missing—the technical team opted to impute the missing values using a group-based median imputation method (after segmenting the data by transaction type), rather than simply deleting the affected records. To identify outliers, we combined a financial rule-based repository with statistical box-and-whisker plots; indicators deviating by more than three standard deviations from the industry mean undergo an initial business verification before a decision is made regarding their correction. To address the issue of the low proportion of fraud samples, the technical team expanded the sampling of fraud-related instances to three times their original volume while simultaneously performing random undersampling on normal samples. This process ultimately resulted in a positive-to-negative sample ratio of approximately 1:4, thereby effectively mitigating the class imbalance issues encountered during subsequent model training. Consequently, the model's recall rate for fraud samples on the test set improved from less than 40% to over 72%; this strengthening of the data foundation directly drove a significant enhancement in detection efficacy. Furthermore, the data team established a distinct temporal marker for enterprises that had undergone a legal entity change within the past three years. During feature extraction, this marker automatically segregates data from the periods before and

after the change—preventing the commingling of financial data from distinct operating entities—thereby elevating the level of granularity and precision in data integration.

B. FEATURE ENGINEERING OPTIMIZATION AND ALGORITHM FUSION: ENHANCING MODEL RECOGNITION ACCURACY

Feature engineering is the process of transforming raw data into representations that a model can learn from; the quality of feature engineering directly determines the upper bound of a model's predictive capability. In terms of feature construction, one must break free from the constraints of traditional, singular financial ratio metrics to establish a comprehensive feature system. This system should encompass various dimensions—including profitability, solvency, operational efficiency, cash flow structure, and the intensity of related-party transactions—thereby providing a multi-faceted, three-dimensional portrayal of an enterprise's financial behavioral characteristics. Simultaneously, a temporal feature dimension should be incorporated. By analyzing the trends, fluctuation ranges, and anomalous inflection points of financial indicators across consecutive reporting periods, this dimension dynamically captures the evolving nature of fraudulent activities over time, effectively compensating for the inherent limitations of cross-sectional features in detecting dynamic changes. Regarding feature selection, a hybrid approach combining three distinct methods—the filter method, the wrapper method, and the embedded method—is recommended. The filter method employs statistical metrics, such as information gain and the Chi-squared test, to preliminarily eliminate ineffective features; the wrapper method utilizes Recursive Feature Elimination (RFE) to iteratively optimize the feature subset; and the embedded method leverages regularized models to automatically perform feature selection during the training process. The synergistic application of these three methods ensures that the features ultimately selected for the model possess both strong discriminative power and effective redundancy control. To resolve potential conflicts between these methods, a voting mechanism or a tiered selection strategy is employed, prioritizing features identified by the embedded method due to its inherent consideration of model performance during the training process. In terms of algorithm selection and ensemble strategies, the primary focus should be on ensemble learning. Random Forests should serve as the foundational model to handle high-dimensional features and mitigate overfitting, while gradient boosting algorithms should be incorporated to capture the interaction effects among non-linear features. By employing model ensemble techniques to achieve complementary strengths, the overall accuracy in detecting complex, multi-faceted fraud patterns can be significantly enhanced. Furthermore, the feature construction process must account for the impact of industry-specific differences on feature validity. Given the substantial variations in financial structures across different industries, a single set of universal feature thresholds is rarely applicable to all scenarios. Therefore, it is essential to establish differentiated feature benchmarks based on industry groupings and to introduce "relative industry deviation" as a supplementary feature. This approach enables the model to detect anomalous financial behaviors more accurately within the context of specific industry norms, thereby preventing systemic misclassifications caused by industry-specific characteristics.

The development of a financial fraud detection system for corporate borrowers—undertaken by the credit risk management department of a certain state-owned bank—serves as a representative example of a typical feature engineering implementation process. During the feature engineering phase, the department integrated three years' worth of financial statement data for each enterprise. Building upon traditional static metrics—such as the current ratio and debt-to-asset ratio—they incorporated the year-over-year change rates and trend slopes of these metrics over three consecutive years to derive dynamic time-series features. Furthermore, to address the diverse fraud methods mentioned previously, the feature set was expanded beyond related-party fund flows. Specifically, indicators for revenue-to-cash-flow divergence were constructed to detect 'inflating revenue,' and long-term asset-to-expense ratios were included to identify 'capitalizing expenses,' ensuring the feature engineering process logically covers all major fraud categories identified in the study. The technical team first employed a Chi-squared test to reduce the initial set of over 200 features to just over 90. Subsequently, they applied Recursive Feature Elimination (RFE) to further narrow the selection down to 42 core features. Finally, a model incorporating regularization constraints was utilized for final validation, ensuring that the features ultimately included in the model remained under 30. In the algorithm fusion phase, a Random Forest served as the first-layer base model to generate probability predictions. These probability predictions, alongside the original features, were then fed into a Gradient Boosting model for second-order learning. The outputs from both layers were ultimately combined using a weighted averaging method. Compared to a single-algorithm baseline, this ensemble model improved precision on the validation set by approximately 8 percentage points; the improvement in detection rates was particularly pronounced for highly concealed cases of fraud involving related-party transactions. During a subsequent retrospective analysis, the team discovered that for certain industry-leading enterprises, scale effects naturally caused specific financial ratios to deviate from the overall sample mean. Consequently, they recalculated industry-relative deviation features specifically for the manufacturing, trade, and service sectors. After replacing the original absolute-value features with these newly derived relative features, the false positive rate across the cross-industry sample dropped by another 5 percentage points, demonstrating a significant enhancement in the industry adaptability of the feature system.

C. DYNAMIC MODEL UPDATING AND INTERPRETABILITY DESIGN: ENSURING PRACTICAL APPLICATION

Patterns of financial fraud are constantly evolving in response to shifting external environments and regulatory policies; consequently, static models inevitably suffer from performance degradation over time. Therefore, establishing a dynamic model update mechanism is crucial for ensuring the sustained effectiveness of fraud detection capabilities. Regarding update strategies, newly labeled financial data should be periodically incorporated into the training dataset. By employing an incremental learning approach to continuously update model parameters, we can prevent the resource waste and time delays associated with full-scale retraining. To mitigate the phenomenon of 'catastrophic forgetting'—where the model loses historical patterns while learning new fraud types—regularization techniques or a small 'rehearsal' buffer of historical samples are integrated to maintain performance stability on both old and new data. Concurrently, a model performance monitoring dashboard should be established to track key metrics—such as precision, recall, and F1 scores—in real time. Should a significant decline in performance be detected, an automated model retraining process is triggered to ensure the model remains in an optimal operational state. In terms of interpretability design, explanatory tools should be seamlessly integrated into the model's inference process. Utilizing feature contribution analysis methods rooted in game theory, the marginal impact of each specific feature on a given prediction result can be quantified, thereby intuitively highlighting the primary grounds upon which the model bases its high-risk assessments. Furthermore, local linear approximation methods can generate detailed feature explanation reports for each specific early-warning instance, significantly reducing the cognitive burden on auditors as they seek to comprehend the model's conclusions. For practical application and visualization purposes, explanatory results should be translated into business-centric language. Through formats such as feature contribution ranking charts and anomaly indicator heatmaps, auditors are directly apprised of the underlying rationale for the model's judgments, thereby alleviating concerns regarding the "black box" nature of the model. From a systems perspective, it is essential to establish a closed-loop audit workflow that fosters human-machine collaboration. High-risk alerts generated by the model must undergo a secondary review by professional auditors; the confirmation or rejection feedback generated during this review process must be promptly fed back into the training database. This creates a continuous, iterative closed loop—encompassing model alerting, manual verification, result feedback, and model optimization—enabling these intelligent tools to gradually earn the trust of practitioners and facilitating the transition of financial fraud detection from a phase of technical validation to one of widespread practical implementation. Moreover, the development of an interpretability framework must take regulatory compliance requirements into account. Given the increasingly stringent regulatory landscape, the rationale behind a model's decisions must not only be intelligible to internal auditors but also serve as a foundation for substantiating external audit conclusions and generating regulatory reports. Consequently, interpretability design should, from its inception, explicitly incorporate scenarios involving regulatory communication, thereby ensuring that the depth and formal rigor of explanatory reports satisfy the diverse requirements of various stakeholders. Taking the intelligent risk early warning platform deployed by a certain provincial Department of Audit during its annual fiscal audits as an example, the implementation mechanisms for dynamic updates and interpretability are notably comprehensive. During the operations and maintenance phase, at the end of each quarter, the platform imports all results that have undergone manual review and annotation during that quarter into an incremental training pool. It then employs a mini-batch gradient update method to perform rolling fine-tuning of the model parameters; this process requires no system downtime for retraining, effectively shortening the update cycle from the original six months to once every quarter. Regarding performance monitoring, the platform features a built-in real-time metrics dashboard; should the weekly recall rate fall below a predefined threshold for two consecutive weeks, the system triggers an alert and initiates a retraining task. For every high-risk warning case, the platform generates a risk interpretation report that utilizes bar charts to list the top five contributing anomalous features. It translates technical metrics into business-oriented descriptions—for instance, directly labeling a "decline in accounts receivable turnover rate exceeding 40%" as "slow collection of payments, posing a risk of fictitious sales revenue." After reviewing these cases, auditors enter their findings into the system: cases involving actual fraud are automatically tagged as positive samples and fed back into the training data, while false positives are tagged as negative samples with the specific reasons for the misclassification noted. Following two full quarterly iterations, the platform's false positive rate decreased by approximately 18%, and the auditors' adoption rate of the system's early warning conclusions rose from less than 50% initially to over 70%, demonstrating a significant improvement in human-machine collaboration efficiency. Furthermore, the platform features a specially designed "compliance-edition" interpretation report template tailored for reporting scenarios involving higher-level regulatory authorities. This template translates the model's decision-making basis into written language that adheres to the formal standards of auditing guidelines; it can be exported with a single click and appended to official audit working papers, thereby effectively resolving the practical challenge of incorporating intelligent early warning conclusions into formal audit documentation.

V. CONCLUSION

Combating financial fraud within the capital-intensive manufacturing sector is a crucial issue for maintaining market order and safeguarding the interests of all stakeholders. Ensuring the transparency of fiber asset valuations and factory production yields remains essential to protecting investors and the integrity of the broader capital market. Utilizing big data mining and machine learning techniques to build intelligent identification models offers new technological approaches and solutions to this field. This paper discusses the basic logic of model construction from three aspects: significance, challenges, and strategies, and provides a complete implementation framework: data integration, feature optimization, and dynamic updating. As intelligent identification models evolve to better interpret complex production yields and fluctuating fiber inventory valuations, their enhanced generalization will provide stronger technological support for corporate financial management and market supervision. This progress ensures that the unique operational data of manufacturing translates into more transparent and reliable financial oversight.

AUTHOR CONTRIBUTIONS

Xiaolin Zheng designed, collected and analyzed the data, and drafted the manuscript. Xiaolin Zheng conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Xiaolin Zheng participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] R. Qin, "Identification of Accounting Fraud Based on Support Vector Machine and Logistic Regression Model," Complexity, vol. 2021, no. 2, pp. 1-11, 2021, doi: 10.1155/2021/5597060.

[2] A. Phong N, H. Tam P, and P. Thanh N, "Fraud Identification of Financial Statements by Machine Learning Technology: Case of Listed Companies in Vietnam," Studies in Systems, Decision and Control, pp. 425-436, 2022, doi: 10.1007/978-3-030-98689-6_28.

[3] Y. Ginanjar, G. Lestari A, S. Mulyani H, et al., "IDENTIFICATION OF THE ABILITY OF EXTERNAL AUDITORS TO DETECT AUDIT FRAUD," International Journal of Professional Business Review, pp. 8(6), 2023, doi: 10.26668/businessreview/2023.v8i6.2252.

[4] R. Bogireddy S and H. Murari, "A Hyperparameters Tunned ML Algorithm for Fraud Identification in Banking and Financial Transactions," International Journal of Innovative Science and Research Technology, pp. 1619-1625, 2024, doi: 10.38124/ijisrt/ijisrt24aug458.

[5] Monalisha -, "Banking Fraud Identification and Prevention," International Journal For Multidisciplinary Research, pp. 6(6), 2024, doi: 10.36948/ijfmr.2024.v06i06.31820.

[6] V. Rampurkar and R. Thirupurasundari D, "An Intelligent Framework for Fraud User Identification Using Machine Learning Techniques," International Journal of Information Engineering & Electronic Business, pp. 17(5), 2025, doi: 10.5815/ijieeb.2025.05.03.

[7] M. Bandanipour, H. Nakhaei, Z. Hajiha, et al., "Identification of Fraud Components and Political Factors Influencing Fraud in Financial Statements," International Journal of Innovation Management and Organizational Behavior, vol. 3, no. 5, pp. 113-121, 2023, doi: 10.61838/kman.ijimob.3.5.14.

[8] L. Jiao and H. Li, "Research on Financial Fraud Identification Model Based on Privacy-Preserving Machine Learning," 2023 4th International Conference on Machine Learning and Computer Application, pp. 628-632, 2023, doi: 10.1145/3650215.3650326.

[9] S. Yao, "Application of Data Mining Technology in Financial Fraud Identification," 2021 4th International Conference on Information Systems and Computer Aided Education, 2021, doi: 10.1145/3482632.3487540.