Available online at
https://apmaj.uitm.edu.my/

APMAJ

e-ISSN: 2550-1631

Asia-Pacific Management Accounting Journal (21) 2026, 93-116

Asia-Pacific Management Accounting Journal

# Mapping the Evolution of Machine Learning-Based Financial Statement Fraud Detection: A Systematic Literature Review and Science Mapping Approach

Masumi Nakashima

*Bunkyo Gakuin University, Tokyo, Japan*

ARTICLE INFO

*Article history:*
Received 22 November 2025
Accepted 12 February 2026
Published 30 April 2026

*Keywords:*
Financial Statement Fraud
Machine Learning
Fraud Detection
Science Mapping
Text Mining
Systematic Literature Review

*DOI:*
10.24191/APMAJ.v21i1-06

ABSTRACT

This study provides a systematic analysis of machine learning-based financial statement fraud detection research using a systematic literature review, science mapping, and text-mining techniques. Based on 85 peer-reviewed articles published between 2009 and 2023, the study examines the intellectual structure, dominant themes, and evolution of the field. The findings reveal a shift from traditional statistical models to advanced machine learning and deep learning approaches, together with the increasing use of unstructured data such as textual disclosures. Major research themes include model development, feature selection, performance evaluation, and data integration. Despite methodological advances, the literature shows limited integration with fraud-related theories, including the fraud triangle and fraud diamond frameworks, reducing model interpretability and practical applicability. The study highlights challenges related to explainability, data imbalance, and ethical concerns, while emphasizing the need for stronger theoretical foundations. By synthesizing fragmented findings and linking data-driven methods with accounting theory, this study contributes to a deeper understanding of fraud detection research and offers implications for auditing practice, corporate governance, regulatory oversight, and future research.

## 1. INTRODUCTION

Financial statement fraud remains a critical issue in modern capital markets, as it distorts the allocation of resources and undermines the credibility of financial reporting (Akerlof, 1970; Stiglitz, 1989). Fraudulent financial reporting, defined as the intentional misstatement or omission of material information, poses

Corresponding author. *E-mail address* : mnakashima@bgu.ac.jp

significant challenges for auditors, regulators, and investors (Association of Certified Fraud Examiners, 1995). Traditional audit procedures, which relied heavily on sampling and manual judgment, often struggle to detect increasingly sophisticated fraudulent activities in complex business environments (Wells, 2001; Association of Certified Fraud Examiners, 2020). In recent years, advances in data analytics and machine learning have transformed the landscape of fraud detection. Machine learning techniques enable the identification of hidden patterns within large datasets and provide predictive capabilities that surpass traditional statistical approaches (Shalev-Shwartz, 2014). As a result, machine learning-based fraud detection has become a central topic in forensic accounting research, with a growing number of studies exploring various algorithms, including support vector machines, random forests, and deep learning models (Perols, 2011; Hajek & Henriques, 2017; Almaskati, 2022).

However, despite this rapid growth, the literature on machine learning-based financial statement fraud detection remains fragmented and lacks a coherent structure. First, the increasing number of publications has made it difficult to obtain a comprehensive understanding of methodological trends and research directions. Second, prior studies report inconsistent findings regarding the effectiveness of different machine learning models, largely due to variations in data, evaluation metrics, and research design. Third, the existing literature often lacks integration with established fraud-related theories, limiting the interpretability and practical implications of empirical results. Although several review and bibliometric studies have attempted to summarize this field (Ramírez-Alpízar et al., 2020; Soltani et al., 2023; Shahana et al., 2022), these studies primarily relied on descriptive analyses of publication trends and algorithm frequencies. As a result, they provided limited insights into the underlying intellectual structure and thematic evolution of the research domain. Therefore, a systematic literature reviewwas necessary to synthesize fragmented findings, identify patterns and inconsistencies, and provide a structured understanding of the development of this field. In particular, understanding how research themes were interconnected and how they evolved over time requires analytical approaches that go beyond conventional descriptive reviews.

To address these limitations, this study adopted a science mapping approach combined with text-mining techniques. Science mapping enabled the visualization and analysis of the intellectual structure of a research field by examining relationships among keywords, topics, and publications. By integrating science mapping with text-based analysis, this study provided a more comprehensive understanding of the development and transformation of machine learning-based fraud detection research. Furthermore, this study interpreted the findings in relation to established fraud-related theoretical frameworks, thereby strengthening the linkage between machine learning techniques and accounting research. This theoretical integration provided important implications for auditing practice, corporate governance, and regulatory oversight in the context of sustainable financial markets.

This study contributed to the literature in three main ways. First, it provided a systematic and reproducible synthesis of machine learning-based fraud detection research. Second, it revealed the intellectual structure and evolution of the field through science mapping and text-mining techniques. Third, it bridged the gap between technical methodologies and theoretical perspectives, highlighting key challenges such as model selection, data imbalance, explainability, and ethical considerations in AI-driven fraud detection

(Draschener et al., 2022). The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the data and methodology. Section 4 presents the results. Section 5 discusses the findings and implications. Section 6 concludes the study.

## 2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

### 2.1 Prior research on machine learning-based financial statement fraud detection

Financial statement fraud detection has been extensively examined using both statistical and machine learning approaches. Early studies primarily relied on traditional statistical models, such as logistic regression, to identify fraudulent financial reporting (Perols, 2011). While these models provided an initial analytical framework, their ability to capture complex and nonlinear relationships in high-dimensional data was limited. With the advancement of data mining and machine learning techniques, researchers increasingly adopted classification algorithms such as support vector machines (SVM), decision trees, and random forests. Empirical studies have demonstrated that these machine learning models generally outperform traditional statistical methods in detecting fraudulent financial statements (Hajek & Henriques, 2017). In addition, ensemble learning techniques, such as random forest and XGBoost, have shown strong predictive performance due to their robustness and ability to handle imbalanced datasets (Chen et al., 2023). More recently, deep learning approaches have been introduced to enhance fraud detection by incorporating unstructured data, including textual disclosures. Models such as long short-term memory (LSTM) and gated recurrent units (GRU) have been applied to financial reporting data, improving detection accuracy and enabling more comprehensive analysis (Xiuguo & Du Shengyo 2022; Li & Wang, 2023). These developments have indicated a shift from traditional structured-data models toward more complex and integrated analytical approaches.

In parallel, several review and bibliometric studies have attempted to synthesize the existing literature. For example, Ramírez-Alpízar et al. (2020) identified SVM as the most frequently used method, while Soltani et al. (2023) document the rapid growth of research in this field. Shahana et al. (2022) highlighted the dominance of supervised learning techniques in fraud detection models. However, these studies were largely descriptive and focus primarily on publication trends and algorithm frequencies. As a result, they provided limited insights into the intellectual structure, thematic relationships, and evolution of the research field. Furthermore, the integration of systematic review approaches with text-based analytical methods remained underdeveloped. This study addressed these limitations by combining science mapping with text-mining techniques to provide a deeper and more structured understanding of the field.

### 2.2 Theoretical framework

A theoretical foundation is essential for interpreting machine learning-based fraud detection beyond predictive performance. Fraud-related theories offered valuable insights into the motivations and mechanisms underlying fraudulent financial reporting. The Fraud Triangle Theory is one of the most widely used frameworks, suggesting that fraud occurs when three elements—pressure, opportunity, and rationalization—are present (Nakashima, 2021). This framework has been extensively applied in accounting research to explain fraudulent behavior. Building on this, the Fraud Diamond Theory introduced a fourth element, capability, emphasizing the role of individual skills and authority in enabling fraud

https://doi.org/10.24191/APMAJ.v21i1-06

(Tragouda et al., 2023). Additionally, the Gone Theory expanded the analysis by incorporating broader organizational and environmental factors influencing fraudulent activities (Xu et al., 2023).

Despite their importance, these theoretical perspectives were often underutilized in machine learning-based studies, which tended to focus primarily on algorithm performance and predictive accuracy. This lack of theoretical integration limited the interpretability and practical relevance of empirical findings.

Integrating fraud-related theories with machine learning approaches can enhance both model design and interpretation. For instance, theory-driven feature selection can improve model performance by incorporating variables related to governance, internal control, and managerial incentives. Moreover, theoretical frameworks help explained why certain variables and models are effective in detecting fraud. By incorporating these theoretical perspectives, this study bridged the gap between technical methodologies and accounting theory, providing a more comprehensive understanding of financial statement fraud detection using machine learning.

### 2.3 Hypothesis development

Based on prior literature and theoretical considerations, this study proposed that the development of machine learning-based fraud detection research was characterized by a transition from traditional statistical methods to more advanced and data-intensive approaches.

**H1**: Research on financial statement fraud detection has shifted from traditional statistical models toward machine learning and deep learning techniques over time.

**H2**: Recent studies have increasingly incorporate unstructured data, such as textual disclosures, alongside financial data.

**H3**: The application of machine learning in fraud detection remained weakly integrated with established fraud-related theoretical frameworks.

## 3. RESEARCH DESIGN

### 3.1 Research approach

This study adopted a systematic literature review combined with a science mapping approach to analyze the development of machine learning-based financial statement fraud detection research. Given the rapid expansion and fragmentation of this research domain, a structured and reproducible methodology was essential for synthesizing existing knowledge. Unlike traditional narrative reviews, which were often subject to researcher bias, the systematic literature review ensured transparency and replicability through clearly defined procedures for article selection, screening, and analysis. In addition, science mapping enabled the visualization of the intellectual structure of the field by examining relationships among keywords, topics, and publications. By integrating systematic review techniques with quantitative text-mining and science mapping methods, this study provided a comprehensive and objective analysis of research trends, thematic structures, and methodological evolution in the field.

https://doi.org/10.24191/APMAJ.v21i1-06 

### 3.2 Data collection and sample selection

Relevant studies were identified using two major academic databases, Scopus and Web of Science. The search was conducted using combinations of keywords such as “financial statement fraud” or “fraudulent financial reporting” in conjunction with “machine learning.” The search period covered publications from 2009 to 2023, reflecting the emergence and growth of machine learning applications in fraud detection.

The initial search was conducted using the keywords “financial statement fraud” or “fraudulent financial reporting” AND “machine learning” in the Scopus and Web of Science databases. This process yielded 44 articles from Scopus and 61 articles from Web of Science, resulting in a total of 105 records. After removing 13 duplicate records, 92 articles remained. Subsequently, a screening process based on titles and abstracts was conducted to ensure relevance to financial statement fraud detection. In this stage, 7 studies that were not directly related to financial statement fraud were excluded. As a result, the final sample consisted of 85 peer-reviewed journal articles. It should be noted that while the main analyses in this study were based on the full sample of 85 articles, certain sub-analyses, such as country-level analysis, were conducted using a subset of 56 articles due to data availability constraints, particularly regarding country-specific information. To enhance transparency and reproducibility, the sample selection process is summarized in Figure 2 as a PRISMA (Page et al. 2021)-style flow diagram, while the overall analytical workflow is presented in Figure 1. In addition, Table 1 provides the list of 56 articles used for country-level analysis, ensuring transparency regarding the subset selection. The list of articles used for country-level analysis is provided in Appendix A.

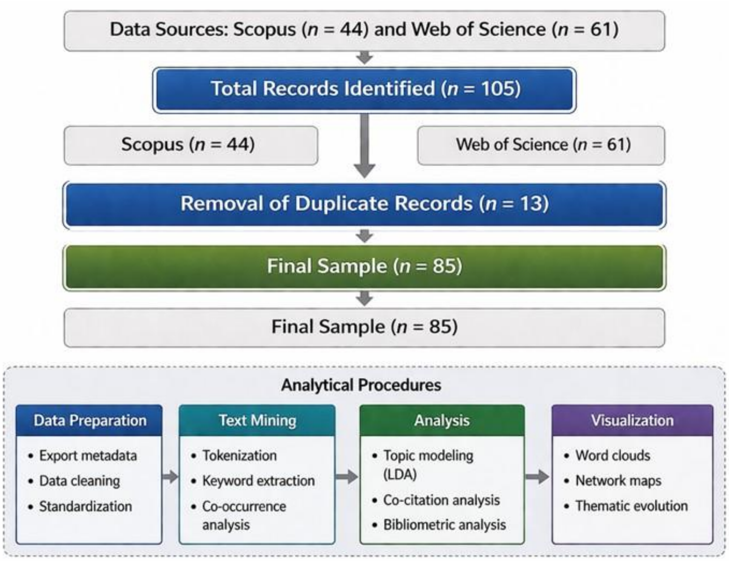



Fig. 1. Study selection and analytical workflow

Source: Author (2026)

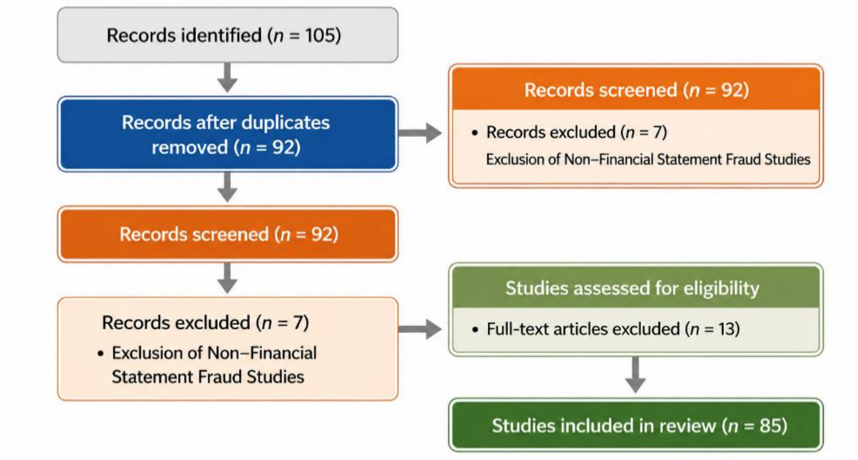



Fig. 2. PRISMA flow diagram of study selection.

Source: Author (2026)

Note: Fig 1 and 2 were created by the author based on original research data. AI tools were used only for language refinement and visual design assistance; all analytical processes and interpretations were conducted by the author.

Table 1. Number of Articles by Country

| Country | Number of Articles |
|---|---|
| Canada | 0.5 |
| Vietnam | 0.5 |
| China | 19 |
| Thailand | 0.5 |
| Czech Republic | 1 |
| Germany | 1 |
| Greece | 3 |
| India | 3 |
| Iran | 1 |
| Indonesia | 2 |
| Korea | 1 |

https://doi.org/10.24191/APMAJ.v21i1-06 

| Country | Number of Articles |
|---|---|
| Lebanon | 1 |
| Netherlands | 1 |
| Malaysia | 1 |
| New Zealand | 1 |
| Taiwan | 3 |
| South Africa | 3 |
| Sweden | 1 |
| UK | 1.5 |
| UAE | 2 |
| Ukraine | 1 |
| US | 7 |
| Spain | 1 |
| Total | 56 |

### 3.3 Analytical methods

To analyse the selected studies, this research employed a combination of science mapping and text-mining techniques. First, word frequency analysis was conducted to identify the most commonly used terms and key research topics in the literature. Second, co-occurrence network analysis was applied to examine relationships among keywords and to identify major research themes, enabling the visualization of how concepts are interconnected within the field. Third, correspondence analysis was used to explore the evolution of research topics over time and across different contexts, such as geographical regions and time periods. These complementary methods allowed for a systematic and data-driven examination of the intellectual structure and thematic development of machine learning-based fraud detection research.

### 3.4 Research validity and limitations

This study ensured methodological rigor in several ways. The use of multiple databases enhanced the coverage and reliability of the dataset, while clearly defined inclusion and exclusion criteria improve transparency and replicability. Furthermore, the integration of science mapping and text-mining techniques provided robust and objective analytical insights, thereby reducing subjective bias. However, several limitations should be acknowledged. First, the analysis was limited to articles indexed in selected databases,

https://doi.org/10.24191/APMAJ.v21i1-06 

which may exclude relevant studies from other sources. Second, only English-language publications were included, potentially introducing language bias. Third, the results of text-mining analysis may have been influenced by preprocessing decisions and parameter settings. Finally, as this study adopts a literature-based approach, it did not provide empirical validation of model performance. Despite these limitations, the research design provided a systematic, transparent, and reproducible framework that meets the methodological expectations of international academic journals.

## 4. RESULTS

### 4.1 Publication trends

The analysis revealed a significant increase in research on machine learning-based financial statement fraud detection, particularly after 2015, with a sharp rise observed from 2019 onward. This trend reflected the growing importance of data-driven approaches in accounting and auditing, as well as the increasing availability of large-scale financial and textual data. Geographically, the majority of publications originated from China and the United States, indicating that these countries played a leading role in advancing research in this field. This concentration may be attributed to the availability of large financial datasets and strong research infrastructure. Figure 3 illustrates the distribution of the selected studies across subject areas. The results showed that research was concentrated in accounting, finance, and data-related disciplines, indicating the interdisciplinary nature of financial fraud detection studies. Figure 4 presents the distribution of studies based on data resources. The results indicated that most studies relied on financial statement data, while a smaller but growing number of studies incorporated textual and hybrid data sources.

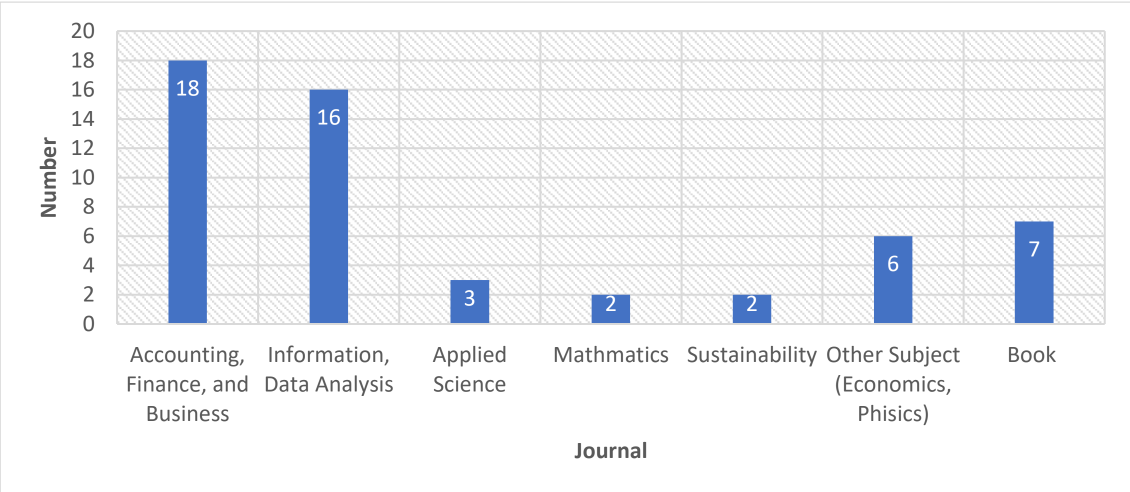



Fig.3. Distribution of Publications across Subject Areas

Note: Figure 3 shows the distribution of the selected studies across subject areas based on the final sample (n = 85). Some studies are classified into multiple subject areas.

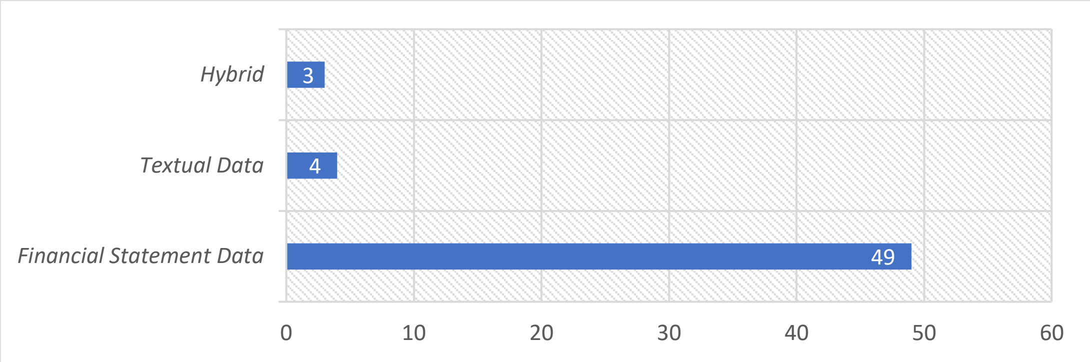



Fig. 4. Distribution of Accounting Fraud Detection Studies by Data Resource

Note: Figure 4 illustrates the distribution of the selected studies by data resource based on the final sample (n = 85).

### 4.2 Dominant research themes

Word frequency analysis showed that the most frequently occurring terms include “fraud,” “machine learning,” “financial statement,” and “fraud detection,” indicating a strong focus on classification-based detection models. In addition, terms such as “algorithm,” “classification,” and “feature selection” highlight the technical orientation of the field. Co-occurrence network analysis identified several major research themes, including (1) machine learning model development, (2) feature selection and data preprocessing, (3) performance evaluation and comparison, and (4) emerging use of textual data. These themes demonstrated that the literature is structured around methodological development rather than theoretical exploration. Figure 5 shows the science mapping of accounting fraud detection research based on titles and authors’ keywords. The network structure revealed the major research themes and their interrelationships within the field.

https://doi.org/10.24191/APMAJ.v21i1-06 

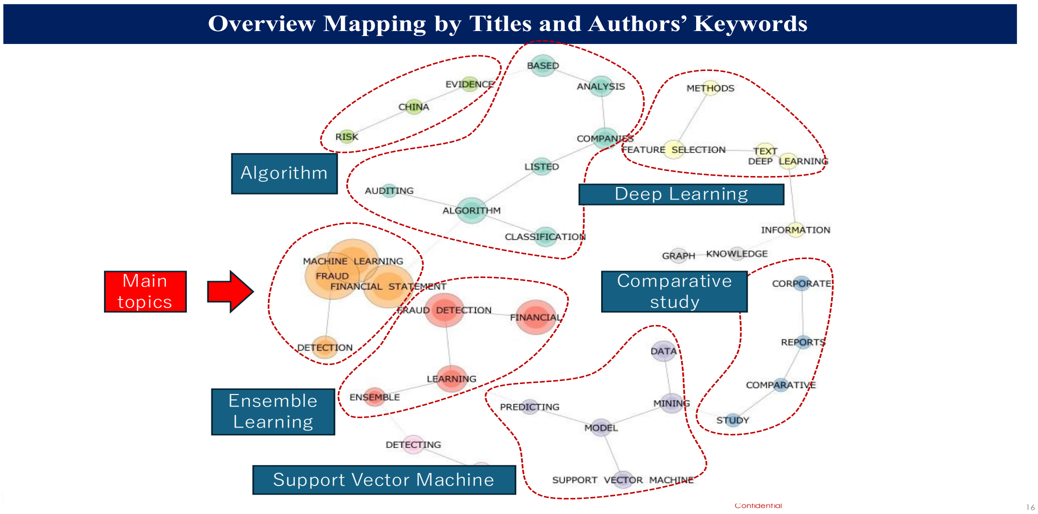



Fig. 5. Science Mapping of Accounting Fraud Detection Research Based on Titles and Authors' Keywords

Table 2 summarizes the most frequently occurring words extracted from the selected studies. The results confirmed the dominance of machine learning and fraud-related terms in the literature.

Table 2. The Number of Occurrences of Words

| Index | Extracted Words | Number of Occurrences |
|---|---|---|
| 1 | FRAUD | 54 |
| 2 | MACHINE LEARNING | 48 |
| 3 | FINANCIAL STATEMENT | 46 |
| 4 | FINANCIAL | 29 |
| 5 | FRAUD DETECTION | 28 |
| 6 | LEARNING | 17 |
| 7 | ALGORITHM | 16 |
| 8 | DETECTION | 12 |
| 9 | BASED | 11 |
| 10 | COMPANIES | 11 |
| 11 | ANALYSIS | 10 |

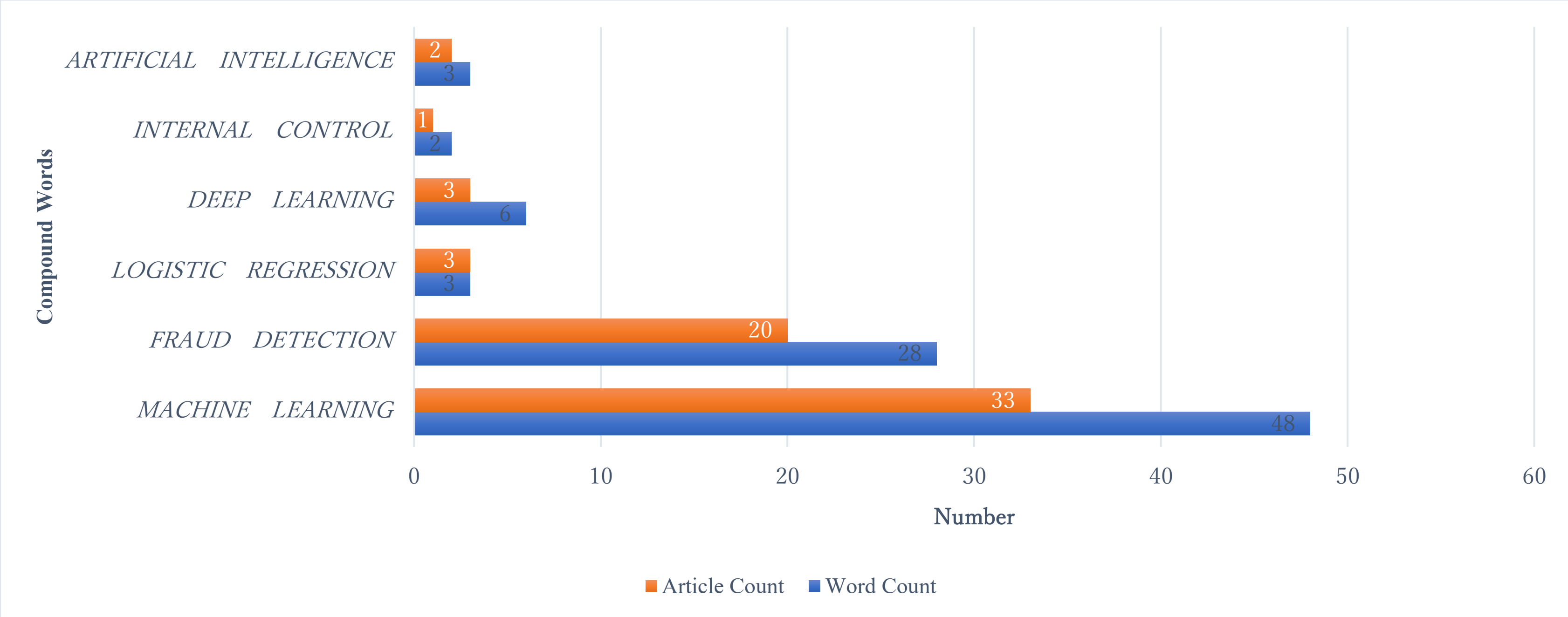
| Index | Extracted Words | Number of Occurrences |
|---|---|---|
| 12 | DATA | 10 |
| 13 | CLASSIFICATION | 9 |
| 14 | MINING | 9 |
| 15 | APPROACH | 8 |
| 16 | DETECTING | 8 |
| 17 | ENSEMBLE | 8 |
| 18 | FEATURE SELECTION | 8 |
| 19 | PREDICTION | 8 |
| 20 | ACCOUNTING | 7 |
| 21 | LISTED | 7 |
| 22 | MODEL | 7 |
| 23 | SUPPORT VECTOR MACHINE | 7 |
| 24 | DEEP LEARNING | 6 |
| 25 | NEURAL NETWORK | 6 |
| 26 | TEXT | 6 |

Note: Table 2 shows the most frequently occurring words extracted from the titles and authors' keywords of the selected studies (n = 85). The number of occurrences represents the frequency of each term. Multi-word expressions (e.g., "machine learning," "financial statement," and "support vector machine") are treated as single terms. The extraction and counting of words were conducted as part of the text mining process.

## 4.3 Evolution of research topics

Correspondence analysis indicated a clear temporal shift in research focus. Early studies primarily relied on statistical models, such as logistic regression, while more recent studies increasingly adopted machine learning and deep learning techniques. In particular, the emergence of deep learning and textual data analysis after 2019 suggests a transition toward more complex and data-intensive approaches. This shift reflected the growing recognition that financial fraud detection requires the integration of both numerical and unstructured data. Figure 6 illustrates the temporal evolution of key research topics based on both word frequency and article counts. Figure 6 highlights a clear shift toward machine learning and deep learning approaches over time.

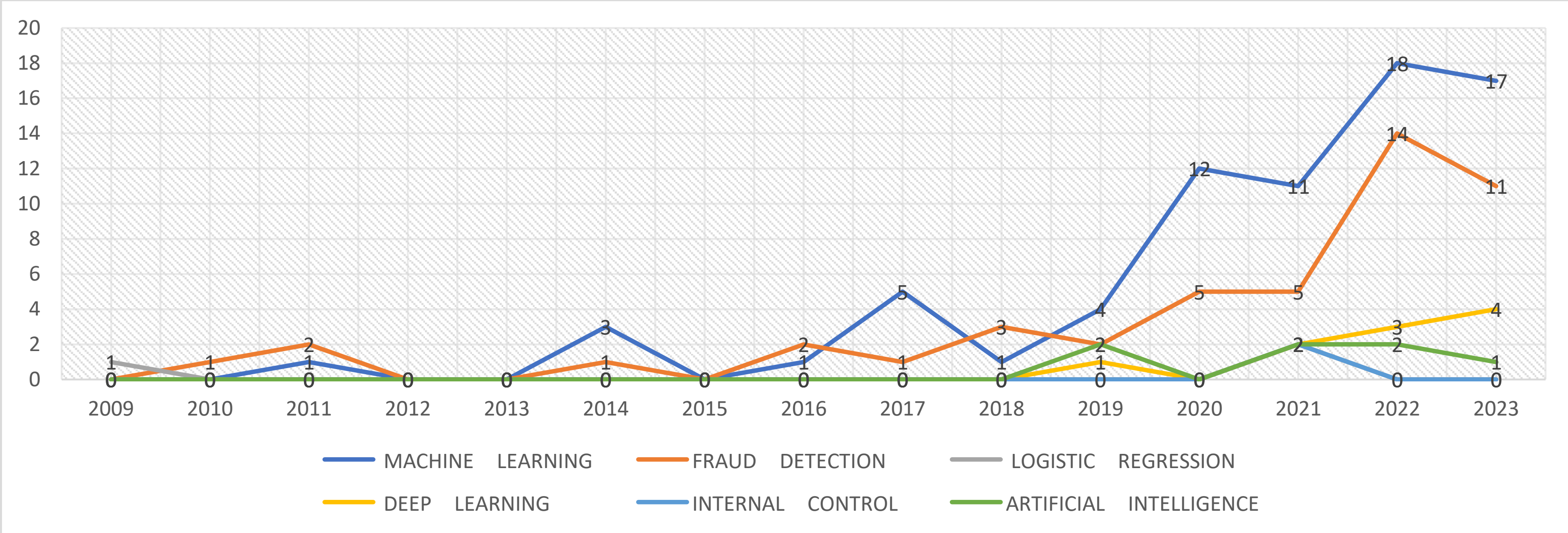
https://doi.org/10.24191/APMAJ.v21i1-06 




Panel A: Total Words Count and Articles Count.

This panel compares the total word counts and article counts for the selected compound terms extracted from the titles and authors' keywords of the reviewed studies.

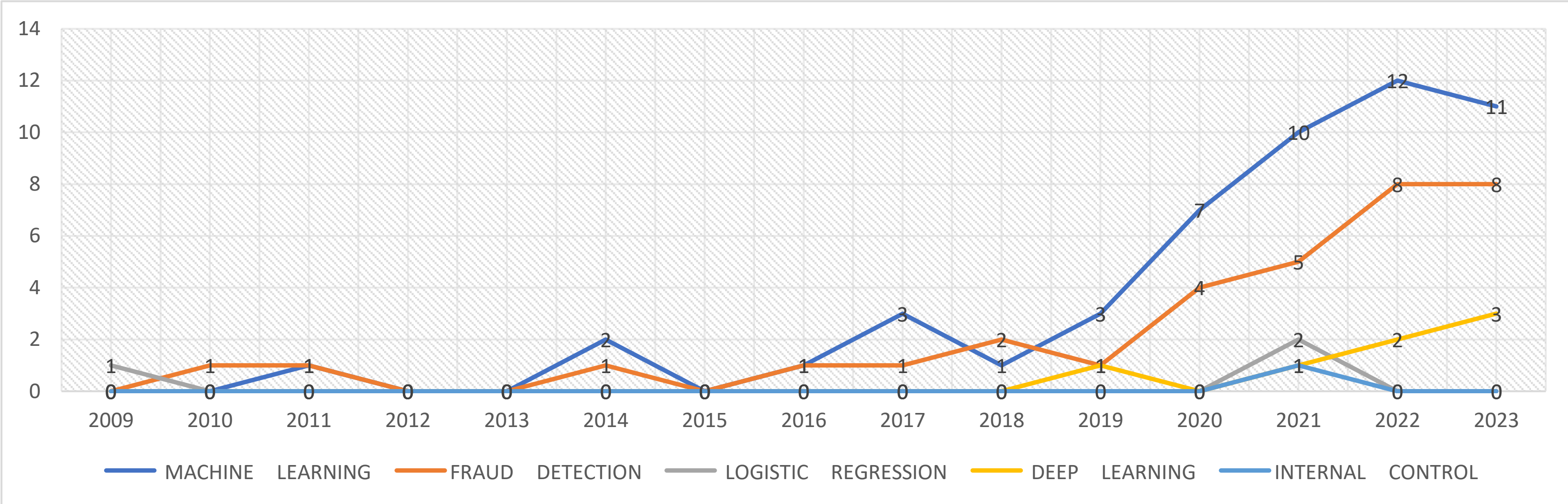
 




Panel B: Time-Series Analysis of Word Counts for Key Research Topics

This panel illustrates the annual changes in the frequency of selected compound terms, highlighting the evolution of major research topics over the study period.




Panel C: Time-Series Analysis of Articles Counts for Key Research Topics

This panel presents the annual number of articles containing each selected compound term, illustrating changes in research attention over time.

Fig. 6. Comparison of Word Frequency and Article Count for Selected Compound Terms

**Note**: Figure 6 presents a comprehensive comparison of word frequency and article count for selected compound terms in accounting fraud detection research. Panel A shows the total word counts and article counts, while Panel B and Panel C illustrate the time-series analysis of word counts and article counts, respectively, for key research topics. Word count represents the total frequency of keyword occurrences, whereas article count indicates the number of studies in which each keyword appears. The analysis is based on the selected studies (n = 85) and highlights both the relative importance and the temporal evolution of major research topics.

With time period as a variable, co-occurrence network analysis and correspondence analysis were conducted. In the correspondence analysis, terms located near the origin between the "before COVID-19" and "during COVID-19" categories represented general terms that are commonly used across periods. In contrast, terms positioned farther from the origin are more distinctive and characterize specific time periods. Furthermore, co-occurrence analysis was conducted using both time period and country as variables. The results indicated that "deep learning" was strongly associated with China, while "neural networks" are more prominent in the United States and other regions. In addition, "support vector machine" was frequently associated with studies from other countries. Figure 7 presents the co-occurrence network and correspondence analysis of research topics by time period and country. Figure 7 provides a visual comparison of how research themes differed before and during the COVID-19 period, as well as across countries.

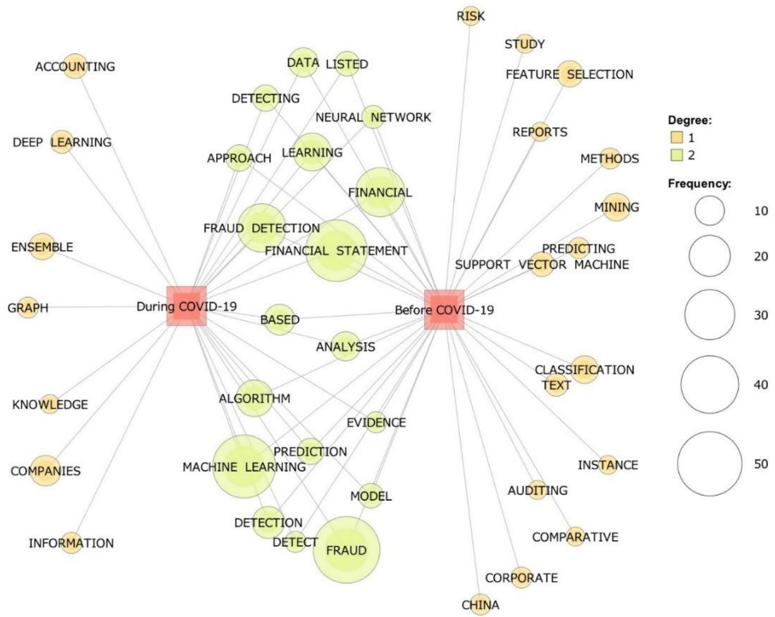



Fig. 7. Co-occurrence Network and Correspondence Analysis of Research Topics by Time Period and Country

Note: This figure presents a co-occurrence network and correspondence analysis of research topics based on titles and authors' keywords of the selected studies (n = 85). The network visualizes keyword co-occurrence relationships, where node size represents the frequency of keyword occurrence and links indicate co-occurrence between terms. The correspondence analysis maps the relationships between keywords and categorical variables, including time period (before COVID-19 and during COVID-19) and country. Terms located near the origin represent general topics common across categories, whereas terms positioned farther from the origin indicate more distinctive and category-specific topics. The highlighted structures illustrate the evolution of research themes over time and across regions.

Figure 8 illustrates the co-occurrence network of research topics by country. The results revealed notable differences in research focus across regions.

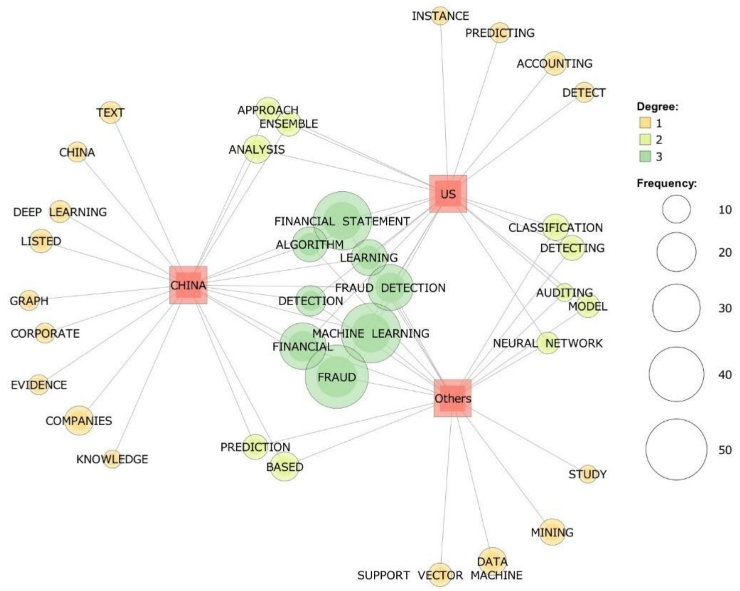



Fig. 8. Co-occurrence Network of Research Topics by Country

Note: Figure 8 presents a co-occurrence network of research topics based on titles and authors' keywords of the selected studies (n = 85), categorized by country (China, United States, and others). Node size represents the frequency of keyword occurrence, and links indicate co-occurrence relationships between terms. Colors reflect the strength of connections (degree) within the network. The spatial distribution of keywords illustrates differences in research focus across countries. For example, "deep learning" is more closely associated with China, while "neural networks" are more prominent in studies from the United States and other regions, and "support vector machine" appears more frequently in studies from other countries.

Figure 9 presents the hierarchical clustering of research topics. The dendrogram identified distinct clusters, which formed the basis for the classification of major research themes in this study.

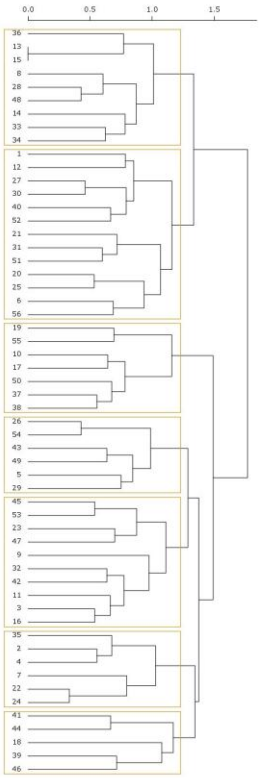



Fig. 9. Hierarchical Clustering of Research Topics in Accounting Fraud Detection

Note: Figure 9 presents a dendrogram resulting from hierarchical clustering analysis conducted on the extracted keywords from the selected studies (n = 85). The clustering is based on similarity measures derived from text mining techniques. The horizontal axis represents the distance (or dissimilarity) between clusters, while the vertical axis lists the individual observations (e.g., documents or keyword groups). The highlighted boxes indicate clusters corresponding to major research themes identified in this study.

## 4.4 Machine learning techniques and evaluation methods

The analysis showed that support vector machines (SVM) were the most frequently used method, followed by random forests and XGBoost. These models were widely adopted due to their strong predictive performance and ability to handle nonlinear relationships and imbalanced datasets. In terms of evaluation metrics, accuracy, precision, recall, and F1-score are commonly used, while more recent studies increasingly adopted ROC curves and AUC as more robust performance measures. This trend indicated a growing awareness of the limitations of simple accuracy measures in imbalanced classification problems.

https://doi.org/10.24191/APMAJ.v21i1-06

### 4.5 Interpretation of findings

Overall, the findings suggested that machine learning-based fraud detection research was characterized by rapid methodological advancement but limited theoretical integration. The dominance of algorithm-focused studies indicated that the field was still in a technically driven stage of development. Furthermore, the increasing use of complex models and diverse data sources highlighted the need for greater attention to model interpretability and ethical considerations. From a theoretical perspective, the limited integration of fraud-related theories suggested an opportunity for future research to bridge the gap between data-driven approaches and accounting theory. These findings contributed to a deeper understanding of the structure and evolution of machine learning-based fraud detection research and provide important implications for both academic research and practical applications.

## 5. DISCUSSION AND IMPLICATIONS

### 5.1 Discussion of findings

This study provided a structured analysis of machine learning-based financial statement fraud detection research and revealed several important insights. First, the findings indicated that the field has experienced rapid methodological advancement, particularly with the increasing adoption of machine learning and deep learning techniques. This shift reflected the growing need to process large and complex datasets in financial reporting and auditing contexts. However, despite these advancements, the literature remained heavily focused on algorithm performance and model comparison, with limited attention to theoretical interpretation. Second, the results highlighted a clear transition from traditional statistical models to more advanced machine learning approaches, as well as a growing integration of textual data. This development suggested that fraud detection was increasingly recognized as a multidimensional problem that requires both quantitative and qualitative analysis. The use of textual data, such as management disclosures, provided additional information that cannot be captured by numerical data alone. Third, the analysis revealed a lack of integration between machine learning techniques and established fraud-related theories. Although some studies reference frameworks such as the fraud triangle or fraud diamond, these theories were not systematically incorporated into model design or interpretation. As a result, many models function as “black boxes,” limiting their interpretability and practical applicability.

### 5.2 Theoretical Implications

The findings have important implications for theory development in accounting and fraud research. The limited integration of fraud-related theories suggested a gap between data-driven methodologies and theoretical frameworks. Integrating theories such as the Fraud Triangle, Fraud Diamond, and Gone Theory (Bologna et al. 1993) into machine learning models can enhance both explanatory power and interpretability. For example, theory-driven feature selection can improve model relevance by incorporating variables related to pressure, opportunity, rationalization, and capability. Furthermore, theoretical frameworks can help explain why certain variables and models are effective in detecting fraud.

https://doi.org/10.24191/APMAJ.v21i1-06 

This study contributed to bridging this gap by emphasizing the importance of aligning machine learning techniques with established theoretical perspectives, thereby advancing a more theory-informed approach to fraud detection research. This study therefore contributed to the literature by providing a structured framework for integrating machine learning techniques with fraud-related theoretical constructs.

### 5.3 Practical Implications

From a practical perspective, the findings provided valuable insights for auditors, regulators, and practitioners. The increasing use of machine learning models highlighted their potential to improve fraud detection accuracy and efficiency in auditing processes. However, the growing complexity of models also raised concerns regarding interpretability and transparency. In auditing and regulatory contexts, the ability to explain model outputs is critical for decision-making and accountability. Therefore, practitioners should carefully consider the trade-off between model performance and interpretability. In addition, the integration of textual data into fraud detection models have suggested new opportunities for leveraging unstructured information, such as annual reports and management discussions. This development may enhance the effectiveness of fraud detection systems in real-world applications.

### 5.4 Implications for Future Research

This study also identified several directions for future research. First, future studies should focus on integrating theoretical frameworks with machine learning models to improve interpretability and explanatory power. Second, there is a need for research on explainable artificial intelligence (XAI) to address the black-box nature of complex models. Third, future research should explore ethical considerations, including data bias, fairness, and transparency in AI-based fraud detection. Finally, expanding data sources and incorporating cross-country analysis may provide a more comprehensive understanding of fraud detection in different institutional contexts. Future research should also examine the robustness of machine learning models across different institutional and regulatory environments. Overall, this study demonstrated that while machine learning has significantly advanced financial statement fraud detection, there remains a critical need to integrate theoretical perspectives and address practical challenges related to interpretability and ethics. By highlighting these issues, this study contributes to both academic research and practical applications in the field.

## 6. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

This study provides a systematic and structured analysis of machine learning-based financial statement fraud detection research by integrating a science mapping approach with text-mining techniques. By examining studies published between 2009 and 2023, this research identified key trends, dominant methodologies, and the evolving intellectual structure of the field. The findings revealed that the research domain has experienced rapid growth, particularly after 2015, with a strong shift toward advanced machine learning and deep learning techniques. In addition, the increasing use of textual data indicated a move toward more comprehensive and data-intensive approaches to fraud detection. Despite these methodological advancements, the study highlighted a critical gap in the integration of machine learning techniques with established fraud-related theoretical frameworks. This lack of theoretical grounding limited the interpretability and practical applicability of many existing models. Overall, this study contributes to the literature by providing a comprehensive synthesis of existing research, uncovering the structural development of the field, and emphasizing the need for greater integration between data-driven methods and theoretical perspectives. This study provides a foundation for future research that seeks to integrate data-driven approaches with theoretically grounded perspectives in accounting.

This study has several limitations. First, the analysis was based on articles retrieved from selected databases, which may not have captured all relevant studies in the field. Second, the study focused exclusively on English-language publications, which may have introduced language bias. Third, the results of text-mining analysis may be influenced by preprocessing decisions and methodological choices. In addition, as this study adopted a literature-based approach, it did not provide empirical validation of the findings. Therefore, caution should be exercised when generalizing the results.

Based on the findings, several directions for future research are suggested. First, future studies should aim to integrate fraud-related theoretical frameworks more explicitly into machine learning models to enhance interpretability and explanatory power. Second, research on explainable artificial intelligence (XAI) is needed to address the black-box nature of complex models. Third, future research should consider ethical issues, such as data bias, fairness, and transparency in AI-driven fraud detection. Fourth, expanding data sources, including cross-country datasets and alternative data types, may provide a more comprehensive understanding of fraud detection. Finally, empirical studies that test and validate the effectiveness of theory-driven machine learning models are encouraged to further advance this field.

## 7. ACKNOWLEDGEMENTS/FUNDING

This paper is based on the keynote speech delivered at the 2024 Webinar of the Bangladesh Chapter of the Asia-Pacific Management Accounting Association (APMAA). The author would like to express sincere gratitude to Dr. Prof. Farhana Begum for the kind invitation and the opportunity to deliver the keynote address. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 22K01813. The author gratefully acknowledges this financial support.

## 8. CONFLICT OF INTEREST STATEMENT

The author declares that there are no conflicts of interest regarding the publication of this paper. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funding agency, the Japan Society for the Promotion of Science (JSPS), had no role in the design of the study, the collection, analysis, or interpretation of data, the writing of the manuscript, or the decision to publish the results.

https://doi.org/10.24191/APMAJ.v21i1-06 

## 9. AUTHORS' CONTRIBUTIONS

The author solely contributed to all aspects of this study, including conceptualization, methodology, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, and project administration. The author has read and approved the final version of the manuscript.

## 10. DECLARATION OF GENERATIVE AI IN THE WRITING PROCESS

During the preparation of this work, the author used ChatGPT (OpenAI) to improve the readability, language quality, and overall clarity of the manuscript. The author reviewed and edited all AI-assisted content as necessary and takes full responsibility for the accuracy, originality, and integrity of the final manuscript.

## 11. DATA AVAILABILITY/SUPPLEMENTARY MATERIALS

The data supporting the findings of this study are included in this published article. Additional information, including the dataset used for the literature review, is available from the corresponding author upon reasonable request.

## 12. ETHICS STATEMENT

This study did not involve human participants, animals, or the collection of personal data. Therefore, ethical approval and informed consent were not required.

## 13. REFERENCES

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. The Quarterly Journal of Economics, 84(3), 488–500. https://doi.org/10.2307/1879431

Almaskati, N. (2022). Machine learning in finance: Major applications, issues, metrics, and future trends. International Journal of Financial Engineering, 9(3), 1–32.

Association of Certified Fraud Examiners. (1995). Cooking the books: What every accountant should know about fraud. Austin, TX.

Association of Certified Fraud Examiners. (2020). Report to the Nations,2020 Global Study Occupational Fraud and Abuse, Government Edition.

Ayinla, K., & Rasheedat, A. A. (2024). Exploring a Novel Approach of K-mean Gradient Boosting Algorithm with PCA for Drought Prediction. American Journal of Data Mining and Knowledge Discovery, 9(1), 1-19. https://doi.org/10.11648/j.ajdmkd.20240901.11

Bologna, G. J., Lindquist, R. J., & Wells, J. T. (1993). The accountant's handbook of fraud and commercial crime. Wiley.

https://doi.org/10.24191/APMAJ.v21i1-06 

Chen, X., & Zhai, C. (2023). Bagging or boosting? Empirical evidence from financial statement fraud detection. Accounting and Finance, 63(5), 5093–5142. https://doi.org/10.1111/acfi.13103

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

Draschener, C. F., Jabeen, H., & Lehmann, J. (2022). Ethical and sustainability considerations for knowledge graph-based machine learning. Proceedings of the IEEE International Conference on Artificial Intelligence and Knowledge Engineering.

Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud: A comparative study of machine learning methods. Knowledge-Based Systems, 128, 139–152. https://doi.org/10.1016/j.knosys.2017.05.001

Li, J., &Wang, C. (2023). A deep learning approach of financial distress recognition combining text. Electronic Research Archive, 31(8), 4683–4707.

Nakashima, M. (2021). Can the fraud triangle explain fraudulent financial statements? Evidence from Japan. Journal of Forensic and Investigative Accounting, 13(1), 198–232.

Page, M. J., Joanne E McKenzie, J. E., Patrick M Bossuyt, P. M., Isabelle Boutron, I., Tamara C Hoffmann, T. C., Cynthia D Mulrow, C. D., … David Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.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. https://doi.org/10.2308/ajpt-50009

Ramírez-Alpízar, A., Jenkins, M., Martínez, A., & Quesada-López, C. (2020). Data mining and machine learning techniques applied to financial fraud detection: A systematic mapping study. RISTI, 97–109.

Shahana, T., Vilvanathan, L., & Bhat, A. R. (2022). State of the art in financial statement fraud detection: A systematic review. Technological Forecasting and Social Change, 192, 1–24.

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.

Soltani, M., Kythreotis, A., & Cyprus, N. (2023). Two decades of financial statement fraud detection literature: A bibliometric analysis. Journal of Financial Crime, 30(5), 1367–1388. https://doi.org/10.1108/JFC-08-2022-0183

Stiglitz, J. E. (1989). Markets, market failures, and development. American Economic Review, 79(2), 197–203.

Tragouda, M., Doumpos, M., & Zopounidis, C. (2023). Corporate governance and financial fraud detection. In Governance and Financial Performance (pp. 171–192).

Wang Xiuguo, X., & Du Shengyong, S. (2022). Financial statement fraud detection using deep learning. IEEE Access, 10, 22516–22532. https://doi.org/10.1109/ACCESS.2022.3154238

Wells, J. T. (2001). Occupational fraud and abuse. Obsidian Publishing.

Wolfe, D. T., & Hermanson, D. R. (2004). The Fraud Diamond: Considering the four elements of fraud. The CPA Journal, 74(12), 38–42.

Xu, X., Xiong, F., & An, Z. (2023). Using machine learning to predict corporate fraud: Evidence based on the gone framework. Journal of Business Ethics, 186, 137–158. https://doi.org/10.1007/s10551-022-05208-7

https://doi.org/10.24191/APMAJ.v21i1-06 

https://doi.org/10.24191/APMAJ.v21i1-06 

# APPENDIX A: LIST OF ARTICLES USED FOR COUNTRY-LEVEL ANALYSIS (N = 56)

TABLE

| No. | Author | Year |
|---|---|---|
| 1 | McKee, T.E. | 2009 |
| 2 | Kotsiantis, S; Kanellopoulos, D; Tampakas, V | 2010 |
| 3 | Perols J. | 2011 |
| 4 | Tang L.; Peng P.; Luo C. | 2016 |
| 5 | Minhas S.; Hussain A. | 2016 |
| 6 | Hajek P.; Henriques R. | 2017 |
| 7 | Mohanty L.; Thakur K.; Manju G. | 2019 |
| 8 | Lokanan M.; Tran V.; Vuong N.H. | 2019 |
| 9 | Sheu G.Y. | 2019 |
| 10 | Gupta S.; Mehta S.K. | 2020 |
| 11 | Ramírez-Alpízar A.; Jenkins M.; Martínez A.; Quesada-López C. | 2020 |
| 12 | Patel H.; Parikh A. | 2020 |
| 13 | Walker S. | 2021 |
| 14 | Chyan-Long Jan | 2021 |
| 15 | Bao Y.; Ke B.; Li B.; Yu Y.J.; Zhang J. | 2021 |
| 16 | Kootanaee A.J.; Aghajan A.A.P.; Shirvani M.H. | 2021 |
| 17 | Liu X. | 2021 |
| 18 | Mongwe W.T.; Mbuvha R.; Marwala T. | 2021 |
| 19 | Xiuguo W.; Shengyong D. | 2022 |
| 20 | Papík M.; Papíková L. | 2022 |
| 21 | Zhang Y.; Hu A.; Wang J.; Zhang Y. | 2022 |
| 22 | Huang L.; Abrahams A.; Ractham P. | 2022 |
| 23 | Bakumenko A.; Elragal A. | 2022 |
| 24 | Tragouda M.; Doumpos M.; Zopounidis C. | 2023 |
| 25 | Riskiyadi M. | 2023 |
| 26 | Li J.; Chang Y.; Wang Y.; Zhu X. | 2023 |
| 27 | Ali A.A.; Khedr A.M.; El-Bannany M.; Kanakkayil S. | 2023 |
| 28 | Achakzai M.A.K.; Peng J. | 2023 |
| 29 | Akimova O.; Ivankov V.; Nykyforak I.; Andrushko R.; Rak R. | 2023 |
| 30 | Chen X.; Zhai C. | 2023 |
| 31 | Song, X.P. ; Hu, ZH ; Du, J.G. ; Sheng, Z.H. | 2014 |
| 32 | Moepya, .SO. ; Nelwamondo, F.V. ; Van der Walt, C. | 2014 |
| 33 | Kim, Y.J .; Baik, B.; Cho, S. | 2015 |
| 34 | Moepya, S.O ; Nelwamondo, F.V.; Twala, B. | 2017 |
| 35 | Stamatis Karlos,S.; Kostopoulos, G.;Kotsiantis, S. and Tampakas, V. | 2017 |
| 36 | Dbouk B. and Zaarour, I. | 2017 |
| 37 | Jianrong Yao; Jie Zhang; Lu Wang | 2018 |
| 38 | S.Hidayattullah, I. Surjandari, E.Laoh | 2020 |
| 39 | Angenent, M.N.; Barata, A.P.and Takes, F.W. | 2020 |
| 40 | Bao, Y.; Ke, B.;Juliayu, Y. and Zhang, J. | 2020 |
| 41 | Brown, N.C., Crowley, R.M.and W.B.Elliott. | 2020 |
| 42 | El-Bannany, M; Dehghan, AH; Khedr, AM | 2021 |
| 43 | Shen, Y.M. ; Guo, CC ; Li, H ; Chen, JJ; Guo, YCA; Qiu, XY | 2021 |
| 44 | Nasir, M ; Simsek, S ; Cornelsen, E ; Ragothaman, S; Dag, A | 2021 |
| 45 | Zhao, ZH ; Bai, T.Y. | 2022 |
| 46 | Almaskati, N. | 2022 |
| 47 | Fukas, P. ; Rebstadt, J ; Menzel, L.; Thomas, O. | 2022 |
| 48 | Hwang, T.K. ; Chen, W..C ; Chiang, WC; Li, Y.M. | 2022 |
| 49 | Wen, S.G. ; Li, J.P.; Zhu, X.Q. ; Liu, M.X. | 2022 |
| 50 | Zeng, S; Li, YR ; Li, Y.Q. | 2022 |
| 51 | Xu, Xiong, and An | 2023 |
| 52 | Rahman, M.J.; Zhu, H.T. | 2023 |
| 53 | Chen, Y.S.; Wu, Z.J. | 2023 |
| 54 | Li, J.P.; Chang, Y.P. ; Wang, YH ; Zhu, X.Q. | 2023 |
| 55 | Li, J.W. ; Wang, C.R. | 2023 |
| 56 | Qiu, S. ; Luo, Y.S. | 2023 |

https://doi.org/10.24191/APMAJ.v21i1-06