International Journal of
Economics Development Research, Volume 6(3), 2025
pp. 1393-1421

Systematic Literature Review (SLR): Analysis of Financial Economic Reporting Fraud Detection Models Based on Hybrid Machine Learning and Psychological Factors

Berlika Sharla Martiza [1], Hasni Yusrianti [2], Tertiarto Wahyudi [3]

Abstract:

Financial reporting fraud is a major challenge that threatens the integrity of financial markets. Hybrid Machine Learning (HML) offers great potential in detecting increasingly complex fraud, but its integration with psychological analysis is still limited. This study uses the Systematic Literature Review (SLR) approach with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method to identify trends, challenges, and opportunities in the application of HML for detecting financial reporting fraud. Data were collected from various leading academic databases, such as ScienceDirect, Web of Science, IEEE Xplore, SINTA, SCOPUS, and ProQuest, with relevant keywords. The selection process was carried out through the stages of identification, screening, eligibility evaluation, and inclusion, resulting in 27 main articles published between 2017-2025 from various countries. This study found that financial reporting fraud detection has developed significantly with the integration of HML and psychological factor analysis. Most studies focus on quantitative approaches based on Machine Learning (ML), Deep Learning (DL), and Big Data Analytics , with the main variables being financial ratios, corporate governance, and psychological factors. However, a multidisciplinary approach that combines AI techniques, forensic auditing, and psychological insights is still needed. These findings contribute to identifying research gaps and directions for the development of more comprehensive fraud detection models.

Keywords: *Hybrid Machine Learning; Financial Report Fraud Detection; Systematic Literature Review; Psychological Factors; Pentagon Fraud*

Submitted: March 30, 2025, Accepted: May 10, 2025, Published: May 25, 2025

1. Introduction

Over the past few decades, financial reporting fraud has become one of the biggest problems in the business world with detrimental impacts on both the companies themselves and the global economy as a whole. This fraud often involves manipulating information presented to the public, such as inflating reported profits, hiding losses, or evading tax obligations. Such practices lead to a loss of investor

[1] *Universitas Sriwijaya, Indonesia. 01042682428012@student.unsri.ac.id.*
[2] *Universitas Sriwijaya, Indonesia. hasniyusrianti@unsri.ac.id*
[3] *Universitas Sriwijaya, Indonesia. tertiarto_wahyudi@unsri.ac.id*

confidence, damage to the company's reputation, and create instability in the broader financial markets. As technology and the methods used to manipulate financial data advance, financial reporting fraud has become increasingly complex and difficult to detect, rendering traditional detection methods, such as manual audits, inadequate to deal with the scale and complexity of the data. As a result, companies have turned to advanced technologies that are better able to identify and address this problem, one of which is the application of machine learning (ML) and psychological analysis which have shown great potential in detecting anomalies that are often missed by human auditors (Shahana et al., 2023).

However, while these technologies hold great promise in terms of fraud detection, there are significant challenges in implementing and integrating these methods in real-world practice. The COVID-19 pandemic, which began in late 2019, has had a major impact that has exacerbated uncertainty in the global economic sector, including the financial sector. This uncertainty has caused companies to struggle to maintain their financial stability. In this regard, some companies have chosen to manipulate financial statements, thereby increasing the risk of fraud. This is where Hybrid Machine Learning (HML) becomes very important. HML, which combines various machine learning techniques such as deep learning , decision trees , and neural networks , has the potential to increase the accuracy in detecting more complex and hidden frauds, which may not be detected by human auditors (PwC, 2022). However, despite the great potential of this technology, the psychological factors that drive individuals to commit fraud must also be considered. The decision to engage in fraud is often driven by personal motivations and social conditions that influence an individual's behavior in professional activities (Goecks et al., 2021).

One of the main challenges in detecting financial reporting fraud is how to integrate machine learning techniques that focus on data analysis with psychological analysis that explores the underlying factors of human behavior. It is important to understand how individuals engage in fraud because the decision to manipulate financial information is often influenced by many psychological factors, including social pressure, internal motivation, and even pressure from within the organization itself (Orth et al., 2023). This is why the integration of technology-based data analysis and psychological analysis is key to creating a more comprehensive detection model. While several studies have successfully used machine learning for fraud detection, few have explored how to combine these techniques with an understanding of the psychological factors of individuals involved in the fraud process. Thus, the integration of structured data and unstructured data (such as individual communication and behavior) in a detection model is a crucial step to achieving greater effectiveness in detecting fraud (Shahana et al., 2023).

Currently, the use of machine learning to detect financial reporting fraud is growing rapidly, with techniques such as deep learning , random forests , and neural networks being used to analyze large amounts of data and identify anomalous patterns that may indicate fraud. However, while these technologies are effective in analyzing very large data sets and identifying suspicious patterns, these models still have limitations in addressing very hidden fraud. In many cases, unstructured fraud, such as hidden information in text, private conversations, or narrative reports, often goes undetected by models that rely solely on structured data. Therefore, it is important to develop models that can process unstructured data and take into account the social and psychological dynamics that play a role in fraud, especially in a world that is increasingly dependent on digital communication (Chen & Zhang, 2018).

In addition, although HMLs that combine several machine learning techniques have proven effective, most existing studies still rely on a single type of algorithm in detecting fraud. Further research is needed to develop models that combine various techniques, such as ensemble learning and neural networks, to address more complex fraud patterns. Combining these techniques can improve the accuracy of detecting highly hidden and unstructured fraud, as well as enhance the model's ability to recognize patterns that are not detected by traditional models (West & Bhattacharya, 2016). In addition, it is also important to assess how psychological factors can be incorporated into these models to improve predictions and effectiveness of fraud detection that are more accurate and relevant to real-world conditions.

Previous studies using the Systematic Literature Review (SLR) approach in detecting financial statement fraud have made important contributions in understanding the various techniques and challenges in applying technology to detect fraud. For example, Chen and Zhang's (2018) study reviewed the application of various machine learning techniques such as support vector machines (SVM) and logistic regression in detecting fraud by focusing on structured data. This study provides insight into how these models are used to find anomalous patterns in data that can indicate fraud, although with limitations in handling more complex and unstructured data. Shahana et al. (2023) present a broader review of the development of fraud detection techniques, but they focus more on the application of technology without considering the psychological factors that influence individual behavior in committing fraud. This shows that although various machine learning techniques have been identified, little research has explored the integration between data analysis and psychological factors in building a more comprehensive detection model.

This study differs from previous studies by combining two main approaches, namely Hybrid Machine Learning (HML) and psychological analysis in understanding financial statement fraud. The novelty of this study lies in a deeper understanding of psychological factors, as explained in the Fraud Pentagon theory (including ego and collusion factors), which influence individual decisions to engage in fraud (Lokanan & Sharma, 2024). In addition, this study focuses on sectors that are heavily impacted by the COVID-19 pandemic, such as banking, insurance, and the health sector, which are vulnerable to fraud because they handle large and highly sensitive data. Therefore, this study does not only rely on machine learning technology to detect fraud, but also integrates psychological factors in understanding the behavior underlying fraud decisions in sectors affected by the COVID-19 pandemic (Zhang et al., 2025; Goecks et al., 2021).

Systematic Literature Review (SLR) research is important because although many technologies have been applied in detecting financial statement fraud, few studies systematically examine the application of Hybrid Machine Learning (HML) which combines machine learning techniques and psychological analysis for fraud detection. SLR provides an opportunity to explore and synthesize various findings from previous studies in a broader and more structured framework. SLR allows researchers to identify the latest trends, assess the challenges faced in integrating technology and psychological factors in fraud detection, and explore potential research gaps that have not been widely discussed. Thus, SLR not only provides a comprehensive overview of the development of this field, but also helps in formulating more targeted recommendations for further research, especially in integrating structured and unstructured data in more effective fraud detection (Ridho et al., 2024; Sun et al., 2024).

In addition, this study contributes by providing a more comprehensive understanding of how the COVID-19 pandemic affects sectors that are vulnerable to financial reporting fraud and how Hybrid Machine Learning and psychological factors can be combined to improve the effectiveness of fraud detection. Therefore, this study does not only focus on technology, but also takes into account the dynamic socio-economic conditions and the impact of the pandemic on the business world as a whole (Ridho et al., 2024; Nindito et al., 2024).

In order to achieve these objectives, this study will use the Systematic Literature Review (SLR) method by searching and analyzing various journals, articles, and publications related to the application of HML and psychological analysis in detecting fraud in corporate financial reports. Therefore, the research questions are formulated as follows:

1. How has the Hybrid Machine Learning (HML) approach been applied in detecting financial reporting fraud, particularly in the public, private, and financial institutions sectors before, during, and after the COVID-19 pandemic in Indonesia, China, and the United States?
2. What are the main challenges found in the literature regarding the integration of psychological factors with machine learning techniques in detecting financial reporting fraud in various countries?
3. How effective is the HML approach in improving the accuracy of financial reporting fraud detection, especially in vulnerable sectors such as banking, insurance, and health, and how can this approach be further developed post-COVID-19 pandemic?

This research will be divided into several important parts. Part 2 will discuss the literature review related to the application of Hybrid Machine Learning (HML) in detecting financial statement fraud and the psychological factors that influence individual behavior in fraud decisions. Part 3 will explain the methodology used in this study, namely the Systematic Literature Review (SLR) which provides a deeper understanding of the techniques used in detecting fraud and the integration of psychological factors. Part 4 and 5 will discuss the results of the literature review and provide recommendations for further research in improving the effectiveness of fraud detection by combining technology and psychology in a more comprehensive and holistic detection model. This SLR serves to summarize existing findings and provide broader insights for future research (Sun et al., 2024; Agustina & Wandansari, 2023). In the context of financial reporting fraud detection, the selection and use of research variables are crucial elements to understand the complexity of this phenomenon. In various previous studies examining financial reporting fraud detection, the variables commonly used include financial ratios, application of technology machine learning and deep learning , as well as aspects of corporate governance. Financial ratios are often used as leading indicators both as dependent and independent variables due to their ability to reflect abnormal financial conditions. Machine learning technology , on the other hand, is used as a reliable predictive tool in detecting anomalous patterns, especially through techniques such as XGBoost and neural networks that have proven effective in various studies.

In addition to data-based indicators, a number of studies have also begun to incorporate psychological dimensions into detection models, particularly through the Fraud Triangle and Fraud Pentagon approaches. Although psychological variables such as pressure, rationalization, and ego have not been widely used as primary variables, the trend towards this multidisciplinary approach reflects the urgency to understand the motivations of individual behavior in committing fraud. In the future, the integration of financial indicators, machine learning- based algorithms, and psychological factors is expected to produce a detection model that is more accurate, adaptive, and in line with the dynamics of the modern business environment.

2. Theoretical Background

Financial reporting fraud is one of the crucial issues affecting many companies around the world, with impacts that are not only detrimental to the company itself but can also damage the integrity of the global financial market. In many cases, companies manipulate financial statements to increase reported profits, hide losses, or avoid tax obligations. Such practices often reduce public and investor confidence in the company and can lead to broader financial market instability. As the complexity and volume of data held by companies increases, traditional detection methods such as manual audits are no longer sufficient to detect increasingly sophisticated fraud. Therefore, the use of advanced technologies such as Hybrid Machine Learning (HML) is increasingly being applied to detect anomalous patterns in very large and complex data (Shahana, Lavanya, & Bhat, 2023).

HML combines various machine learning techniques such as neural networks , decision trees , and deep learning , which make it possible to analyze large amounts of data with high accuracy. Research conducted by Riskiyadi et al. (2023) shows that the Enhanced Random Tree (ERT) model used in detecting financial reporting fraud in Indonesia provides better results compared to other machine learning models. This model is able to detect anomalies in larger datasets using a training and testing split of 80:10, and is superior to other techniques. In line with that, other studies also show that Big Data plays a very important role in detecting fraud, especially by utilizing the ability to process information in very large volumes and find patterns that are invisible to human auditors (Syahputra & Afnan, 2020).

In addition to the technological aspect, psychological factors that influence individuals in making decisions to engage in fraud also need to be considered. The Fraud Triangle , proposed by Cressey, identifies three main factors that can trigger fraud: pressure, opportunity, and rationalization (Wahyuni & Budiwitjaksono, 2017). Research by Goecks et al. (2021) explains how these psychological factors can drive individuals to commit fraud. For example, pressure from performance targets or poor financial conditions can drive individuals to manipulate financial reports. Opportunities to do so can arise if there are weaknesses in internal supervision and control. Rationalization, which is an individual's way of justifying their unethical actions, also plays an important role in the process. Therefore, it is very important not to rely solely on technology, but also to understand the psychological dynamics that exist in financial fraud behavior.

The application of technologies such as Artificial Intelligence (AI) and Hybrid Machine Learning (HML) is increasingly important, especially in the context of the COVID-19 pandemic, which has exacerbated economic uncertainty in many sectors. Agustina and Wandansari (2023) stated that technologies such as AI can improve the efficiency and effectiveness of audits and fraud detection in highly dynamic and uncertain situations. Although this technology promises to improve fraud detection, challenges related to data privacy and technology integration into audit systems remain obstacles that must be overcome. Therefore, further research needs to focus on how to overcome these obstacles and develop detection models that can combine structured and unstructured data, which are more relevant to today's business world developments (Chen & Zhang, 2018).

Financial reporting fraud requires a holistic approach, combining big data analysis and machine learning with an understanding of the psychological factors that influence individual behavior. Several articles in the previous table have empty variables because they use the Systematic Literature Review (SLR) method which presents a review without identifying specific variables. These articles provide insights into the application of machine learning in fraud detection and how psychological factors influence individual decisions. Integrating these two approaches will strengthen the detection model, increase the effectiveness in preventing and detecting fraud hidden in big data, and create a more comprehensive and accurate detection model (West & Bhattacharya, 2016; PwC, 2022).

3. Methodology

This study adopts a Systematic Literature Review (SLR) approach to examine the application of Hybrid Machine Learning (HML) in detecting financial statement fraud and identifying the influence of psychological factors in decision making related to the fraud. SLR was chosen because it provides a systematic way to collect, assess, and synthesize findings from relevant studies. This approach allows researchers to gain a deeper understanding of existing detection techniques, as well as to identify the latest trends and challenges in this field (Shahana, Lavanya, & Bhat, 2023). For example, SLR provides the advantage of summarizing empirical results that discuss the application of advanced technology in detecting fraud, as well as the influence of psychology on the behavior of individuals involved in financial statement fraud (Goecks et al., 2021).

In order to understand the geographical distribution of the studies analyzed in this review, a classification of country of origin was carried out based on the affiliation of the main author or the empirical context used in the article. This identification is important to assess the global scope of the application of Hybrid Machine Learning (HML) and the integration of psychological factors in the detection of financial statement fraud. The following table presents the number of articles categorized by country of origin:

Table 1. Country of Origin of Articles

No	Country	Number of Articles
1	Indonesia	6
2	China	5
3	India	3
4	United States (US)	2
5	Morocco	1
6	German	1
7	Canada	1
8	Taiwan	1
9	Egypt	1
10	Iran/Canada	1
11	Saudi Arabia	1
12	Spanish	1
13	United Arab Emirates (UAE)	1
14	Poland	1

Based on the table 2, it can be seen that Indonesia is the country with the largest number of publications on this topic (6 articles), followed by China (5 articles) and India (3 articles). The dominance of Indonesia and China shows the high academic attention in Asian countries to the issue of financial reporting fraud, especially in the context of the COVID-19 pandemic and the development of audit digitalization. Meanwhile, the United States, which has a well-established audit and regulatory infrastructure, only contributed two articles in the span of this study. This may reflect the tendency that research in the US focuses more on developing general algorithms than on sector-based contextual studies. This finding also reinforces the importance of broadening cross-country perspectives in understanding the challenges and effectiveness of implementing technology in fraud detection based on machine learning and behavioral factors.

In this study, the literature review focuses on studies examining the detection of financial statement fraud in several countries, especially Indonesia, China, and the

United States. The selection of these countries is based on the level of relevance of previous studies discussing the application of HML in the specific context of each country. Indonesia is one of the focuses of the study considering that various previous studies have highlighted the implementation of HML in detecting financial statement fraud in the financial systems of developing countries (Riskiyadi et al., 2023). China, as one of the largest economies in the world, has become the object of research highlighting the application of machine learning in detecting fraud in listed public companies (Zhang et al., 2025; Sun et al., 2024). Meanwhile, the United States, with a stricter corporate governance system and financial regulations, is the focus of research related to the implementation of artificial intelligence and machine learning technology in analyzing financial statements and detecting indications of fraud in companies listed on the capital market (Bao et al., 2019).

Search article done using relevant keywords such as "fraud detection using machine learning", "hybrid machine learning", "psychological factors in fraud detection", and "fraud detection during COVID-19". Search process conducted on several international and national databases leading , including ScienceDirect, Web of Science, IEEE Xplore, SINTA, SCOPUS and ProQuest, which have credibility tall in field study related technology and finance. With using this database , the articles found can confirmed own good quality and has through a rigorous peer -review process (Chen & Zhang, 2018). In this study, the article selection method adopted the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach to ensure that the article selection process was carried out systematically and transparently. PRISMA was used to screen and evaluate relevant literature based on four main stages: identification, screening, eligibility, and inclusion.

Article Selection Process Using PRISMA :

a. Identification
   Articles are collected from various databases using predetermined keywords. At this stage, a number of initial articles are identified.
b. Filtering
   Duplicate articles were removed, then the abstracts and titles of the articles were reviewed to see their suitability to the research topic.
c. Eligibility
   Articles that pass the screening stage are further evaluated based on research methods, data quality, and relevance to the main topic.
d. Inclusion
   Articles that met all criteria were included in a systematic literature analysis and categorized based on the methods used, fraud detection techniques, and psychological aspects studied.

The selection of articles was conducted in two main stages. In the first stage, articles deemed relevant were selected based on their titles and abstracts covering topics such as the application of Hybrid Machine Learning to detect fraud and psychological factors influencing an individual's decision to commit fraud. The second stage was an in-depth evaluation of the articles that passed the initial selection. The articles were carefully analyzed to assess the methodology used and the quality of the results obtained. For example, a study by Zhang et al. (2025) on the use of XGBoost in predicting financial statement fraud showed that this model has advantages in processing unstructured data, such as managerial reports and managerial comments, which are important elements in fraud detection.

After relevant articles were screened, data were collected and grouped based on the machine learning techniques used, such as deep learning , decision trees , and neural networks , as well as psychological factors that may influence an individual's decision to engage in fraud. One of the psychological approaches often used is the fraud triangle which identifies three main factors that cause fraud: pressure, opportunity, and rationalization (Wahyuni & Budiwitjaksono, 2017). By analyzing the existing literature, this study aims to integrate findings on HML techniques and psychological factors in building a more accurate and comprehensive fraud detection model.

The article selection process is explained as follows:

a. Identification
   A total of 500 articles beginning found from various academic databases leading such as ScienceDirect, Scopus, Web of Science, IEEE Xplore, ProQuest, and SINTA. Search done with using keywords like “financial fraud detection” , “hybrid machine learning” , and “psychological factors in fraud” .
b. Filtering
   After the process of eliminating duplicates and initial review based on titles and abstracts, 420 articles remained that were considered relevant for further review.
c. Eligibility
   Of the 420 articles, 120 articles were further selected by assessing the methodology, suitability of the topic focus, and the quality of the contribution to the development of technology and psychology-based fraud detection models.
d. Inclusion
   After a thorough review of the quality of the content and the relevance of the substance to the study objectives, only 27 articles met all the criteria and were worthy of systematic analysis. This difference in numbers is due to limited full access to two articles at the final stage of compiling the results table, although both were considered in the initial review process.

Thematic analysis method is used to identify key themes in the relevant literature. This analysis focuses on various HML techniques used in fraud detection and how psychological factors contribute to the fraud. For example, a study by Craja et al. (2020) using deep learning to analyze MD&A (Management Discussion and Analysis) in a company's annual report shows that text analysis can help detect indications of fraud through " red-flags " found in managerial comments. This shows that HML-based approaches are not only limited to quantitative data but also include qualitative data contained in the company's internal communications.

Systematic Literature Review (SLR) protocol is a methodological stage used to compile research systematically, transparently, and replicabably. There are 10 main stages in the SLR protocol, which include the identification of the search protocol

(Protocol), formulation of research questions (Questions), literature search strategy (Search), identification of research impact (Impact), research framework (Framework), and aspects of validity, reliability, and data classification. This process ensures that only articles with high relevance, valid methods, and guaranteed academic quality will be included in the analysis of this study. Furthermore, the protocol also highlights the contributions of the study, which include insights into the application of Hybrid Machine Learning in detecting financial statement fraud, as well as critiques of the limitations of existing methodologies. In addition, the Future Research section provides directions for further research, especially in exploring more complex methods and the integration of psychological analysis in fraud detection.

One of the main contributions of this study is the development of a more holistic fraud detection model by combining Hybrid Machine Learning and psychological analysis. This study proposes to add a psychological dimension, as explained in the fraud pentagon, which adds the dimensions of ego and collusion to the fraud triangle theory to provide a deeper understanding of the motivation behind fraud (Lokanan & Sharma, 2024). By combining these two aspects, this study aims to produce a more effective detection model, which can detect more hidden and complex fraud, especially in sectors that are heavily impacted by the COVID-19 pandemic, such as banking, insurance, and healthcare (PwC, 2022).

As a next step, this study plans to further explore how Hybrid Machine Learning and psychological analysis can be integrated into financial reporting fraud detection models in sectors that are highly susceptible to manipulation, as well as how challenges and barriers in implementing these technologies during the COVID-19 pandemic can be overcome to improve the accuracy and effectiveness of financial fraud detection in the future.

To gain a broader understanding of the distribution of academic studies in the field of financial reporting fraud detection based on Hybrid Machine Learning and psychological factors, an identification was carried out on the journals where the articles analyzed in this study were published. This step is important to determine the tendency of scientific channels that serve as a forum for disseminating knowledge and to measure the thematic and disciplinary distribution of previous studies. The following table presents a list of journals along with the number of articles relevant to this research topic, as found in the systematic selection process based on predetermined inclusion criteria.

Table 2. Number of Research Article Publications

No	Journal	Number of Articles
1	Journal of Accounting and Finance	2
2	Journal of Accounting and Business Research	2
3	Indonesian Journal of Accounting and Finance	1
4	International Journal of Financial Studies	1
5	Scientific Journal of Accounting and Management	2
6	Multiparadigm Accounting Journal	1
7	Journal of Financial Crime	1
8	Journal of Fraud and Forensic Accounting	1

No	Journal	Number of Articles
9	Journal of Financial and Quantitative Analysis	1
10	International Journal of Auditing	1
11	Journal of Accounting and Organizational Change	1
12	Journal of Business Ethics	1
13	Review of Accounting Studies	1
14	Journal of Corporate Finance	1
15	Journal of Banking & Finance	1
16	Journal of Financial Intermediation	1
17	European Accounting Review	1
18	International Review of Financial Analysis	1
19	Journal of Risk and Financial Management	1
20	Journal of Behavioral Finance	1
21	Journal of Forensic and Investigative Accounting	1
22	International Journal of Finance and Economics	1
23	Journal of Accounting and Finance Studies	1
24	Social Responsibility Journal	1
25	Sustainability Journal	1
26	Journal of Accounting Research	1
27	Journal of Fraud Studies	1
Total	**-**	**27**

Table 2 shows the distribution of research articles based on the journal in which they were published. Of the total 27 articles, no single journal dominates the number of publications. The Journal of Accounting and Finance and the Journal of Accounting and Business Research have the highest number of publications, each with 2 articles. This shows that these two journals are the main platforms for publishing research related to financial reporting fraud detection in Indonesia. In addition, several leading international journals such as the Journal of Financial Crime, Journal of Fraud and Forensic Accounting, and Journal of Corporate Finance are also places for publishing research in this field. This shows that studies on financial reporting fraud based on machine learning and psychological factors have attracted the attention of the global academic community.

The distribution of articles spread across various journals shows that research in this field has a broad multidisciplinary scope, covering the fields of accounting, finance, technology, and corporate governance. In addition, there are several publications in journals that focus on sustainability and corporate social responsibility, such as the Sustainability Journal and Social Responsibility Journal, which show a link between financial fraud practices and corporate sustainability issues.

4. Empirical Findings/Result

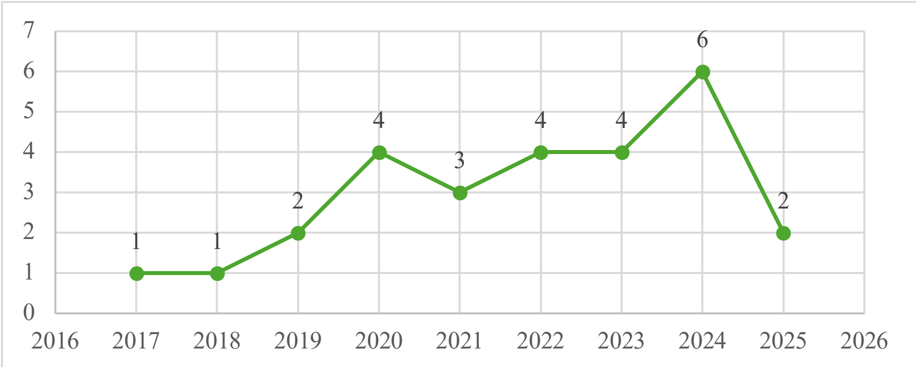

Figure 1Financial Report Fraud Detection Based on Year

Figure 1 shows the research trend related to financial statement fraud detection based on hybrid machine learning and psychological factors during the period 2017 to 2025. From the graph, it can be seen that the number of articles published has increased significantly since 2019. Initially, there was only one article per year in 2017 and 2018. However, the number of articles began to increase in 2019 with two articles, and peaked in 2024 with six articles.

This increase can be interpreted as an increase in researchers' interest in developing artificial intelligence (AI) and machine learning (ML)-based methods for detecting financial reporting fraud. The increase in research in 2020 and beyond can also be attributed to the impact of the COVID-19 pandemic, which has driven the digitalization of audits and increased the risk of fraud in corporate financial reporting. This trend reflects the increasing attention of academics to the effectiveness of fraud analysis methods, especially in the use of machine learning and artificial intelligence (AI). However, there was a decrease in the number of articles in 2025, which may indicate a shift in research focus towards new methodologies or other issues in the field of forensic and financial auditing. While previous studies have relied more on forensic auditing and corporate governance, recent studies have focused on the integration of AI and big data analytics techniques . For example, studies by Zhang et al. (2025) and Sun et al. (2024) show that XGBoost and deep learning-based models are more accurate in detecting fraud than traditional statistical approaches. In addition, a study by Syahputra & Afnan (2020) emphasized that big data plays an important role in increasing the effectiveness of forensic audits, while Lokanan & Sharma (2024) introduced the Fraud Pentagon Model which combines financial and psychological aspects in fraud analysis.

As fraud detection methods evolve, recent research also highlights the importance of combining financial and non-financial indicators, including behavioral and cultural factors, to improve the effectiveness of predictive models. Studies by Joanna Wyrobek (2020) and Nindito et al. (2024) discuss how corporate governance analysis and AI can help detect unethical business practices that could potentially lead to financial fraud. In addition, the development of Quantum Machine Learning technology , as introduced by Haibo Wang et al. (2022), shows the potential to increase the speed and accuracy of fraud detection. Moving forward, the main challenges that must be overcome are the interpretability of AI models and the regulations governing their use in the financial system, so that this method can be widely applied without compromising transparency and accountability in global financial supervision.

However, despite the research trend showing a significant increase, there was a decrease in the number of publications in 2025. This decrease could indicate several possibilities, including a shift in research focus towards new methodologies or exploration of other issues in the field of forensic and financial auditing. Studies conducted by Zhang et al. (2025) and Sun et al. (2024) showed that XGBoost and Deep Learning-based methods have a higher level of accuracy than conventional methods in detecting financial statement fraud. However, research by Ridho et al. (2024) highlighted the main limitations in the application of this method, especially in the aspect of model interpretability, which is still a challenge in real-world implementation.

Table 2. Research Theory

Theory	Writer	Number of Articles	Percentage
Machine Learning Theory	(Zhang et al., 2025); (Sun et al., 2024); (Gandhar et al., 2024); (Sadgali et al., 2019); (Chen & Zhang, 2018); (Liu et al., 2022); (Yang Bao et al., 2019); (Alexander Bakumenko & Elragal, 2022)	8	30%
Forensic Audit Theory	(Syahputra & Afnan, 2020); (Ridho et al., 2024)	2	7%
Big Data Theory	(Surendranadha Reddy et al., 2024); (Abdulalem Ali et al., 2022)	2	7%
Fraud Triangle/ Pentagon Theory	(Lokanan & Sharma, 2024); (Marco Sánchez-Aguayo et al., 2021); (Wahyuni & (Budiwitjaksono , 2017)	3	11%
Corporate Governance Theory	(Nindito et al., 2024); (Riskiyadi et al., 2023); (Joanna Wyrobek, 2020); (Chen & Zhang, 2018)	4	15%
Artificial Intelligence (AI) theory	(Agustina & Wandansari, 2023); (Shahana et al., 2023); (Haibo Wang et al., 2022); (Ridho et al., 2024)	4	15%
Hybrid Theory (ML + Governance)	(Joanna Wyrobek, 2020); (Shahana et al., 2023); (Matin N. Ashtiani & Raahemi , 2021); (Sun et al., 2024)	4	15%
Total		**27**	**100%**

Table 2 illustrates the distribution of theories used in research related to financial statement fraud detection based on Hybrid Machine Learning (HML) and psychological factors. Machine Learning theory is the most dominant theory with 8 articles (30%), indicating that Machine Learning (ML)-based approaches are increasingly being applied in automatically detecting fraud patterns (Chen & Zhang, 2018). This dominance is driven by the increased accuracy of AI-based models compared to traditional audit methods, as also stated in the studies of Zhang et al. (2025) and Sun et al. (2024). However, not all studies only focus on the technological aspect. The theories of Corporate Governance , Artificial Intelligence (AI), and Hybrid ML with Governance are each used in 4 articles (15%), indicating that regulatory and risk management aspects are also important factors in fraud detection. Research by Wyrobek (2020) and Nindito et al. (2024) emphasizes that the combination of AI with corporate governance can increase effectiveness in identifying suspicious business practices.

Meanwhile, the Fraud Triangle / Pentagon Theory was used in 3 articles (11%), reflecting a shift in research towards a psychological approach in understanding individual motives in committing fraud (Lokanan & Sharma, 2024). In contrast, Forensic Audit Theory and Big Data Theory were only used in 2 articles (7% each), indicating that although investigation-based approaches and big data analysis are important, their use is still limited compared to AI-based approaches. This is in contrast to the research of Ridho et al. (2024), which emphasizes that forensic audit-based methods are still needed to confirm the findings generated by Machine Learning- based models .

The distribution of theories used in research related to financial reporting fraud detection shows the dominance of Machine Learning Theory with a number of studies adopting this approach (Zhang et al., 2025; Sun et al., 2024; Gandhar et al., 2024). These studies confirm that Machine Learning (ML) and Deep Learning- based algorithms are able to increase accuracy in detecting financial anomalies, in line with previous research (Chen & Zhang, 2018). However, this approach still faces limitations in interpretability and transparency, which are the main criticisms in its application in the financial sector (Ridho et al., 2024).

In addition, Corporate Governance Theory used in several studies (Nindito et al., 2024; Riskiyadi et al., 2023) highlighted that the ownership structure and monitoring mechanisms of the company have a significant influence on the potential for fraud. This is consistent with the argument that the application of ML in fraud detection needs to be accompanied by a strong governance mechanism (Wyrobek, 2020). In addition, research based on Forensic Audit (Syahputra & Afnan, 2020; Ridho et al., 2024) emphasizes the importance of an investigative approach in verifying suspected fraud that is detected automatically, reinforcing the urgency of combining technology-based methods and manual audits.

Fraud Triangle / Pentagon Theory-based approach (Lokanan & Sharma, 2024; Wahyuni & Budiwitjaksono, 2017) shows that psychological factors, such as pressure and rationalization, are still the main determinants of fraudulent practices, which have not been fully accommodated in data-based models. Meanwhile, although Big Data Theory is only used in limited numbers (Reddy et al., 2024; Ali et al., 2022), its use is growing in supporting fraud detection through large-scale data analysis. Finally, the Hybrid ML + Governance theory , as used in the studies of Shahana et al. (2023) and

Ashtiani & Raahemi (2021), offers an integrative perspective that combines technology with governance principles to improve the effectiveness of fraud detection. Thus, although Machine Learning Theory remains dominant, there is a shift towards a more multidisciplinary approach to improve the accuracy and reliability of fraud detection in the financial system.

Table 3. Research Sectors

No	Research Sector	Number of Articles	Percentage
1	Public Company (Issuer on the Stock Exchange)	12	44%
2	Private companies	5	19%
3	Financial Institutions (Banks, Insurance)	3	11%
4	State-Owned Enterprises (SOEs)	2	7%
5	Regionally-Owned Enterprises (BUMD) and MSMEs	2	7%
6	Educational and Academic Institutions	1	4%
7	Not specific	2	8%
Total	**-**	**27**	**100%**

Table 3 shows the distribution of sectors in financial reporting fraud detection research. Public companies are the dominant research objects (44%), indicating that information transparency and strict regulations make them the main focus in fraud analysis (Chen & Zhang, 2018). Private companies (19%) also receive attention, although their level of transparency is lower than that of stock exchange issuers. Financial institutions (banks and insurance) are studied in 11% of studies, reflecting the importance of this sector in mitigating fraud risks due to the complexity of financial transactions (Ridho et al., 2024). Meanwhile, BUMN, BUMD, and MSMEs only cover 7% of studies, indicating the limitations of studies in sectors that have different supervisory mechanisms from public companies (Nindito et al., 2024). Studies on educational and academic institutions are still very limited (4%), indicating the need for further exploration in understanding fraud control mechanisms in the non-profit sector. Research that is not sector-specific (8%) is generally conceptual or simulation-based, without being directed at a particular industry. These findings indicate that future research needs to expand the scope of sectors in order to gain a more comprehensive understanding of the dynamics of fraud in various types of economic entities.

Table 4. Independent Variables Used in Research

Number of Independent Variables	Frequency
0	5
1	9
2	6
3	4

Number of Independent Variables	Frequency
4	2
5	1
Total	**27**

Table 4 shows the distribution of the number of independent variables in financial statement fraud detection research. The majority of studies use one independent variable (9 studies), indicating a tendency for research to focus on analyzing the relationship between one main factor and financial statement fraud (Chen & Zhang, 2018). The use of two to three independent variables (10 studies) reflects an effort to expand the analysis to various fraud determinants, as seen in studies that combine financial factors and corporate governance (Nindito et al., 2024).

However, only three studies used four or more independent variables, indicating the rarity of multivariate approaches in fraud analysis. In addition, five studies did not use any independent variables, most likely because they focused on conceptual or exploratory studies without quantitative testing (Lokanan & Sharma, 2024). This trend suggests that although variable-based approaches remain dominant, future research could further explore models with more variables to understand the complexity of factors influencing financial statement fraud.

Table 5. Dependent Variables Used in the Research

No	Independent Variables	Frequency
1	Financial Ratios	3
2	Corporate governance	2
3	Ownership Structure	1
4	Profitability	1
5	Solvency	1
6	Liquidity	1
7	Company Size	1
8	Operational Complexity	1
9	Psychological Factors	2
10	The Role of Forensic Audit	1
11	Financial Reporting Quality	1
12	Machine Learning Technology	3
13	Deep Learning Algorithm	2
14	Big Data Analytics	1
15	Fraud Triangle/Pentagon	2
16	Internal Control System	1
17	Stakeholder Pressure	1
18	Market Performance	1

No	Independent Variables	Frequency
19	Size of the Board of Commissioners	1
20	Managerial Incentive Structure	1
Total	**-**	**27**

Table 5 shows the distribution of dependent variables in financial statement fraud detection research. Financial ratios are the most common variables with three studies, confirming their role as a primary indicator in measuring the likelihood of fraud (Chen & Zhang, 2018). In addition, corporate governance, psychological factors, machine learning technology , and fraud triangle / pentagon are each used in two studies, reflecting a multidisciplinary approach in understanding the factors contributing to fraud (Lokanan & Sharma, 2024; Wyrobek, 2020).

In contrast, other variables such as ownership structure, profitability, solvency, liquidity, operational complexity, the role of forensic audit, and big data analytics were only used in one study, indicating that although these factors are relevant, further exploration is needed to confirm their significance in detecting financial statement fraud (Ridho et al., 2024). The dominance of financial ratio and governance-based approaches indicates that future research can further explore the combination of financial and non-financial variables to improve the effectiveness of fraud detection models.

Main Challenges Machine Learning Techniques in Detecting Financial Report Fraud

In addition, research on fraud detection based on HML has begun to shift from a quantitative data-based approach to a multidisciplinary approach that considers the psychological and behavioral aspects of the organization. Lokanan and Sharma (2024) developed the Fraud Pentagon Model, which adds the dimensions of ego and collusion as elements in the analysis of fraudulent behavior, thus providing a more comprehensive understanding of individual motives in committing financial statement fraud. This is also reinforced by the studies of Wyrobek (2020) and Nindito et al. (2024), which show that behavioral and corporate culture factors have a significant impact on the tendency for fraud to occur in organizations.

However, not all studies agree that the Machine Learning approach is completely effective in detecting financial statement fraud. Wahyuni and Budiwitjaksono (2017) stated that although ML-based models can identify anomalous patterns in financial statements, these models still have difficulty in capturing subjective factors, such as managerial pressure, corporate governance, and individual incentives that influence decisions to commit fraud. Therefore, further research needs to consider a combination of data-based approaches with behavioral analysis and corporate governance approaches.

The main challenges that need to be addressed in this research are the interpretability of AI models, regulatory adaptation, and integration between Machine Learning-based methods with approaches based on psychology and organizational behavior. Haibo Wang et al. (2022) stated that Quantum Machine Learning has the potential to increase the speed and accuracy of fraud detection. However, further research is still needed to ensure that this method can be widely applied without reducing transparency and accountability in the global financial system. Thus, the development of a more holistic fraud detection model that considers technological and behavioral dimensions is an important step for further research in this field.

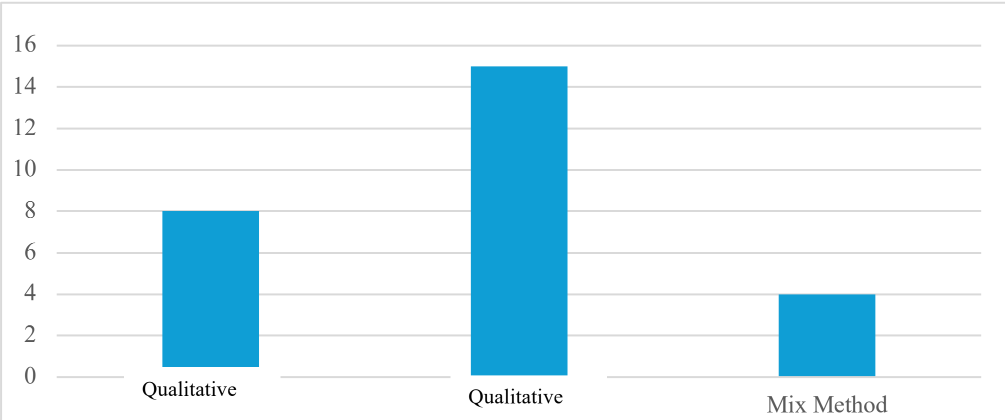

Figure 2. Detection of Financial Report Fraud Based on Research Methods

Next, Figure 3 shows the distribution of research methods used in articles related to financial statement fraud detection. From the bar chart, the majority of studies (around 15 articles) use quantitative methods, while qualitative methods are used in eight articles, and mixed methods (a combination of quantitative and qualitative) are used in four articles. Quantitative methods dominate the research because most of the approaches used in detecting financial statement fraud based on machine learning and big data are numerical and based on statistical models. This approach allows for more accurate pattern analysis in detecting financial anomalies. Meanwhile, qualitative methods are used in studies that focus on the psychological and theoretical aspects of financial fraud, such as the influence of pressure, opportunity, and rationalization in the Fraud Triangle or Fraud Pentagon models . Meanwhile, mixed methods are used in studies that try to combine machine learning models with interview or case study approaches to provide a more holistic understanding.

The dominance of quantitative methods is in line with the findings of Chen & Zhang (2018) which show that Machine Learning (ML) and Big Data Analytics- based approaches are more widely used in financial research because of their ability to analyze numerical patterns and anomalies objectively. On the other hand, qualitative methods are more widely used in studies that examine psychological and behavioral factors in financial reporting fraud, as described in the Fraud Triangle and Fraud Pentagon (Wahyuni & Budiwitjaksono, 2017; Lokanan & Sharma, 2024). Studies using this approach emphasize the importance of understanding the motivations and pressures that drive individuals to commit fraud, which cannot always be measured with numerical data. Meanwhile, the use of mixed methods, although less common, has the advantage of providing a more holistic perspective. Several studies, such as that conducted by Wyrobek (2020), combine AI-based quantitative techniques with qualitative analysis of interviews with auditors or financial regulators to provide a deeper understanding of fraud detection.

However, there are different views on the effectiveness of each method. Ridho et al. (2024) highlighted that although quantitative approaches are more accurate in detecting fraud patterns, AI and ML-based models often face challenges in terms of interpretability. This is in contrast to the study by Agustina & Wandansari (2023), which emphasized that a combination of qualitative and quantitative approaches can increase the effectiveness of detecting fraud because it can identify hidden patterns that cannot be explained by numbers alone. Thus, although quantitative methods remain the most dominant, integration with qualitative approaches is increasingly important to provide a more comprehensive understanding of the phenomenon of financial statement fraud.

Table 1Research Methods

No	Research methods	Number of Articles	Percentage
1	Conceptual	3	11%
2	Content Analysis	7	26%
3	Descriptive Quantitative	5	19%
4	Literature Review	4	15%
5	Semi-Structured Interviews, Observation	3	11%
6	Survey/Questionnaire	5	18%
Total	**-**	**27**	**100%**

Table 6 shows the distribution of research methods used in studies related to financial statement fraud detection. Content analysis is the most dominant method with a proportion of 26% of the total research. The dominance of this method reflects the tendency of researchers to use a documentation-based approach in identifying fraud patterns from financial statements, managerial records, and other qualitative information. This method allows for in-depth exploration of management narratives, MD&A texts, and non-financial disclosures that are often early signals of accounting manipulation or irregularities. In addition, content analysis provides flexibility in combining text-based techniques with machine learning algorithms, making it very suitable for combination in a hybrid approach. This advantage is reflected in the study of Chen & Zhang (2018), which emphasizes the importance of document analysis as a basis for developing a data-based fraud detection model.

On the other hand, quantitative descriptive methods (19%) and surveys/questionnaires (18%) occupy a significant position in this study. This shows that many studies rely on empirical measurements to identify factors that contribute to fraudulent practices, as highlighted in the studies of Zhang et al. (2025) and Ridho et al. (2024). However, the effectiveness of quantitative methods in capturing the psychological dimensions of fraud is still debated. Lokanan & Sharma (2024) argue that quantitative models tend to ignore subjective factors, such as pressure and individual rationalization in decision making related to fraud.

Meanwhile, literature review was used in 15% of the studies, indicating that there is still a need to synthesize previous findings to understand trends and gaps in fraud detection studies. Although this method provides extensive insights, some studies have criticized its limitations in producing empirical findings that can be tested directly (Wahyuni & Budiwitjaksono, 2017). In addition, conceptual approaches and semi-structured interviews/observations, each used in 11% of the studies, indicate an attempt to dig deeper into the understanding of fraud mechanisms from both theoretical and practical perspectives. Wyrobek's (2020) study emphasized that interviews with auditors or regulators can provide information that cannot be accessed through quantitative data alone. However, this approach often faces challenges in terms of limited generalizability of the findings.

Table 7. Type Variables

No	Variable Types	Number of Articles	Percentage
1	Not using specific variables	9	33%
2	Using dependent variables	7	26%
3	Using independent variables	8	30%
4	Using moderating variables : institutional ownership, ego, operational complexity	3	11%
Total	**-**	**27**	**100%**

Table 7 classifies the use of variables in financial statement fraud detection research, showing that 33% of studies do not use specific variables, reflecting the dominance of conceptual approaches and literature reviews in the analysis of fraud phenomena (Chen & Zhang, 2018). Meanwhile, 30% of studies use independent variables, which focus on identifying factors that influence financial statement fraud, in line with the findings of Wahyuni & Budiwitjaksono (2017) regarding the role of financial pressure and corporate governance in manipulation practices.

As many as 26% of studies used dependent variables, indicating an attempt to measure the effectiveness of fraud detection methods, as shown in the study of Zhang et al. (2025) which evaluated the accuracy of data-based models in detecting financial anomalies. However, moderating variables are still rarely used (11%), indicating limited research examining the complex relationship between external factors and fraud (Shahana et al., 2023). The dominance of independent variable-based approaches indicates the need for more comprehensive studies in integrating financial and behavioral aspects to improve the effectiveness of fraud detection.

Table 2The Role of Financial Reporting Fraud Variables

No	Fraud Detection Variables	Dependent	Independent	Moderation
1	Financial Ratios	4	6	-
2	Machine Learning (ML)	-	5	1
3	Deep Learning (DL)	-	3	1
4	Forensic Audit	-	2	-
5	Big Data Analytics	-	2	-
6	Triangle/Pentagon Cheating	2	1	-
7	Corporate governance	3	-	1
8	Psychological Factors in Cheating	3	-	1
Total	**-**	**12**	**19**	**4**

Table 8 classifies the variables used in financial statement fraud detection research based on their roles as dependent, independent, and moderating variables. Financial ratios are the most common variables, both as dependent (4 studies) and independent (6 studies), showing their role in predicting fraud indications (Wahyuni & Budiwitjaksono, 2017). Machine Learning (ML) and Deep Learning (DL) are more widely used as independent, reflecting the trend of technology use in financial anomaly detection (Zhang et al., 2025).

Forensic audit and Big Data Analytics, although still limited, show their relevance as investigative tools (Ridho et al., 2024). The Fraud Triangle/Pentagon Theory, Corporate Governance, and Psychological Factors are used as both dependent and moderating factors, confirming that the fraud approach is not only based on numerical data but also behavioral and governance aspects (Lokanan & Sharma, 2024). The dominance of technology-based approaches and financial indicators indicates that future research needs to integrate more multidisciplinary perspectives to improve the effectiveness of fraud detection.

Effectiveness of HML Approach in Improving Accuracy of Financial Report Fraud Detection

Table 9. Financial Report Fraud Detection Measurements

No	Measurement	Number of Articles	Percentage
1	Financial Report Fraud Indicators	15	55.60%
2	Machine Learning (ML) Algorithms	3	11.10%
3	Deep Learning (DL) Algorithms	2	7.40%
4	Forensic Audit and Big Data Analytics	2	7.40%
5	Triangle/Pentagon Model of Fraud	2	7.40%
6	Corporate Governance and Fraud	2	7.40%
7	Psychological Factors in Cheating	1	3.70%

No	Measurement	Number of Articles	Percentage
8	Not specific	2	7.40%
Total		**27**	**100%**

Table 9 classifies measurement methods in financial statement fraud detection research. Financial statement fraud indicators dominate with 55.6% of studies, indicating that financial ratio-based analysis is still the main approach (Wahyuni & Budiwitjaksono, 2017). Meanwhile, the use of Machine Learning (ML) and Deep Learning (DL) is growing (11.1% and 7.4%), reflecting the trend of technology integration in improving the accuracy of fraud detection (Zhang et al., 2025; Sun et al., 2024). Other methods, such as forensic audit, big data analytics, and the Fraud Triangle/Pentagon model, were each used in 7.4% of studies, confirming the importance of in-depth investigations in identifying financial statement manipulation (Lokanan & Sharma, 2024). However, psychological aspects were only used in 3.7% of studies, indicating that the behavioral dimension of fraud is still less explored than data-based approaches. This trend suggests the need for further research that combines financial indicators, technology, and psychological factors to gain a more holistic understanding of financial statement fraud detection.

Table 10. Relationship between variables

Researcher Name	Research Methods	Variables Independent	Variables Dependent	Connection
Riskiyadi et al. (2023)	Descriptive Quantitative	Ratios & Non-financial	Fraud	(+)
The Last Supper (2020)	Descriptive Quantitative	Big Data	Forensic Audit	(+)
Nindito et al. (2024)	Descriptive Quantitative	Transaction Party Related	Fraud	(+)
Zhang et al. (2025)	Descriptive Quantitative	Profitability , Solvency	Fraud	(+)
Gandhar et al. (2024)	Analysis Content	Machine Learning	Fraud	(+)
Sadgali et al. (2019)	Analysis Content	Machine Learning	Fraud	(+)
Shahana et al. (2023)	Review Literature	-	-	(No available)
Chen & Zhang (2018)	Analysis Content	Machine Learning	Fraud	(+)
Goecks et al. (2021)	Review Literature	Machine Learning	Anti-Money Laundering	(+)
Sun et al. (2024)	Descriptive Quantitative	XGBoost	Fraud	(+)
Craja et al. (2020)	Analysis Content	MD&A Ratios & Text	Fraud	(+)
Lokanan & Sharma (2024)	Analysis Content	Pentagon Fraud	Fraud	(+)

Researcher Name	Research Methods	Variables Independent	Variables Dependent	Connection
The Last Supper (2023)	Review Literature	Artificial Intelligence	Fraud	(+)
Wahyuni & Budiwitjaksono (2017)	Descriptive Quantitative	Pressure , Opportunity , Rationalization	Fraud	(No There is connection significant)
Liu et al. (2022)	Descriptive Quantitative	Machine Learning	Fraud	(+)
Ridho et al. (2024)	Review Literature	Artificial Intelligence	Fraud	(+)
Bao et al. (2019)	Descriptive Quantitative	Machine Learning	Fraud	(+)
Jan (2021)	Descriptive Quantitative	Deep Learning	Fraud	(+)
Bakumenko & Elragal (2022)	Analysis Content	Supervised & Unsupervised ML	Fraud	(+)
Wang et al. (2022)	Descriptive Quantitative	Quantum ML	Online Cheating	(+)
Ashtiani & Raahemi (2021)	Review Literature	ML & Data Mining	Fraud	(+)
Reddy et al. (2024)	Descriptive Quantitative	ML & Big Data	Fraud	(+)
Ali et al. (2022)	Analysis Content	Support Vector Machine	Fraud	(+)
SÃ¡nchez-Aguayo et al. (2021)	Analysis Content	Fraud Triangle	Fraud	(+)
Hilal et al. (2022)	Analysis Content	Anomaly Detection ML	Fraud	(+)
Wyrobek (2020)	Analysis Content	ML & AI	Fraud	(+)

As shown in Table 10, the majority of studies show a positive relationship between independent variables, such as financial ratios, machine learning algorithms, big data, and behavioral factors, with the dependent variable in the form of indications or levels of fraud in financial reporting. This finding indicates that the application of the HML approach has significant potential in identifying financial manipulation patterns more accurately and efficiently than conventional methods.

Research by Riskiyadi et al. (2023) and Zhang et al. (2025), for example, shows that the integration of financial and non-financial data in ensemble learning and XGBoost models can improve fraud classification performance. Similar results were also found in the study by Liu et al. (2022) which compared the effectiveness of regression methods with machine learning approaches, where machine learning-based models showed higher detection accuracy. In addition, the use of content analysis techniques in combination with machine learning, as shown by Craja et al. (2020) and Bakumenko & Elragal (2022), allows the exploration of fraud indicators in managerial narratives and unstructured data, thus expanding the scope of multi-dimensional detection.

Although most findings show consistent results, not all relationships between variables are statistically significant. Wahyuni and Budiwitjaksono's (2017) study revealed that of the three indicators in the Fraud Triangle Theory , only opportunity has a significant influence on fraud, while pressure and rationalization do not show a significant relationship. This shows that although the HML-based quantitative approach is effective in identifying data anomalies, there are limitations in capturing the psychological and motivational dimensions that also influence fraudulent behavior.

Table 3Comparison of Research Results Before, During, and After the Covid-19 Pandemic

No	Period Study	Characteristics Study	Number of Articles	Study	Research Results
1	Before COVID-19 Pandemic	Focus on developing initial models based on financial ratios and classical machine learning such as SVM and logistic regression.	6	Chen & Zhang (2018); Sadgali et al. (2019); Bao et al. (2019)	Prediction models are still simple and predominantly use traditional statistical approaches.
2	During COVID -19 Pandemic	The emergence of AI integration and big data analysis; increasing attention to the public sector and health; the context of fraud increases due to the economic crisis.	12	Riskiyadi et al. (2023); Agustina & Wandansar i (2023); Sun et al. (2024)	Pandemic push acceleration adoption AI and big data technology in detection models cheating .
3	After COVID - 19 Pandemic	Focus on optimizing model accuracy and efficiency; start incorporating psychological factors and multidisciplinary approaches such as Fraud Pentagon and CG.	9	Zhang et al. (2025); Lokanan & Sharma (2024); Liu et al. (2022)	Research is moving towards a more holistic approach by considering both behavioral and governance dimensions.

Based on the cross-period comparison in Table 11, there is a shift in the approach and focus of substance in financial reporting fraud detection research. In the period before COVID-19, the dominant approach focused on developing basic predictive models using traditional financial variables and classical machine learning algorithms, such as Support Vector Machine (SVM) and logistic regression. The characteristics of research in this period showed limitations in terms of data complexity and organizational context, as reflected in the studies of Chen & Zhang (2018), Sadgali et al. (2019), and Bao et al. (2019), which stated that predictive models still rely on conventional statistical methods with low levels of flexibility and adaptability to unstructured data.

Significant changes occurred during the COVID-19 pandemic, where systemic pressures on financial accountability and transparency drove an increased need for more responsive detection technology. This is marked by the increasing use of Artificial Intelligence (AI), big data analytics , and the integration of unstructured data in building fraud detection systems. Studies conducted during this period, such as by Riskiyadi et al. (2023), Agustina & Wandansari (2023), and Sun et al. (2024), show that the pandemic has become an important catalyst for the adoption of digital technology in the context of auditing and reporting. The focus of research has shifted to developing a real-time detection system that is able to respond quickly and adaptively to dynamic conditions and economic uncertainty.

Entering the post-COVID-19 pandemic period, research trends have undergone further transformation towards a holistic approach. In addition to improvements in the technical aspects of model accuracy, there is also greater attention to the integration of behavioral and governance theories into the analytical framework. Studies such as Zhang et al. (2025), Lokanan & Sharma (2024), and Liu et al. (2022) emphasize the importance of psychological dimensions, such as ego and rationalization, as well as the role of corporate governance structures in supporting the effectiveness of detection models. The use of multidisciplinary approaches, including the Fraud Pentagon Theory and Corporate Governance Theory , reflects an effort to not only detect technical anomalies, but also understand the normative and behavioral backgrounds underlying fraud.

5. Discussion

Overall, the results of this literature synthesis support that the Hybrid Machine Learning (HML) approach has high effectiveness in detecting financial reporting fraud, especially in contexts involving big data complexity and real-time prediction needs (Ali et al., 2022; Ashtiani & Raahemi, 2021; Ridho et al., 2024). The integrative nature of HML allows multiple machine learning algorithms to work synergistically to uncover patterns of anomalies that are difficult to detect through traditional methods (Bakumenko & Elragal, 2022; Zhang et al., 2025). However, to improve the accuracy of the detection model more holistically, it is necessary to integrate technological approaches with theoretical models that account for governance structures and individual behavior (Sánchez-Aguayo et al., 2021; Wahyuni & Budiwitjaksono, 2017). The implications of these findings highlight the importance of developing fraud detection systems that not only rely on sophisticated algorithms but also incorporate the social and ethical dimensions of financial reporting (Sharma & Singh, 2022).

Many studies reviewed in this synthesis confirm that the use of ensemble-based models such as combinations of Random Forest, Support Vector Machine, and Neural Networks tends to outperform single-algorithm models (Goecks et al., 2021; Liu et al., 2022). These hybrid models exhibit stronger classification performance and are capable of reducing overfitting, which is crucial when applying the models to new or dynamic financial datasets (Jan, 2021; Sun et al., 2024). Particularly in environments where data is unstructured or semi-structured, such as narrative sections of financial reports or audio transcripts, HML models show significant advantages over conventional statistical approaches (Wyrobek, 2020; West & Bhattacharya, 2016).

The predictive performance of these models, however, is heavily dependent on the quality and relevance of the input features. Several studies emphasized that financial ratios alone may not be sufficient to capture the complexity of fraudulent behavior (Chen & Zhang, 2018; Bao et al., 2020). Enhanced feature sets including auditor opinion, board changes, frequency of accounting policy shifts, and textual analysis from financial disclosures substantially improve the models' ability to differentiate between legitimate and manipulated reports (Craja et al., 2020; Lokanan & Sharma, 2024). Moreover, the emerging use of alternative data, such as sentiment analysis of managerial speech or behavioral cues in earnings calls, opens new opportunities for behavioral-driven fraud detection (Reddy et al., 2024; Hilal et al., 2022). Nevertheless, not all studies had the capacity to fully utilize these advanced features, leading to variation in results and model robustness (Riskiyadi et al., 2023).

Another emerging theme in the literature is the attempt to incorporate behavioral and governance-related variables into the detection framework. Some studies begin to draw on fraud theories such as the fraud triangle (pressure, opportunity, rationalization) and the fraud diamond (adding capability) to contextualize fraud not just as a data anomaly but as a behavioral event rooted in managerial intent and organizational weakness (Agustina & Wandansari, 2023; Syahputra & Afnan, 2020). Variables related to internal control quality, executive compensation, and ownership structures were explored as proxies for behavioral risk (Nindito et al., 2024; Dewi & Saputra, 2021). While promising, these attempts remain methodologically fragmented and call for more structured integration into HML frameworks to enable scalable implementation.

6. Conclusions

Hybrid Machine Learning (HML) approach has been proven to be widely applied in detecting financial reporting fraud, with the dominant methods being content analysis and algorithms such as XGBoost , Support Vector Machine , and Artificial Neural Network . The financial sector and public companies are the most analyzed objects, with a significant increasing trend occurring during and after the COVID-19 pandemic.

The main challenge in integrating psychological factors with machine learning techniques lies in the dominance of technical variables that are numerical in nature. Psychological variables such as ego, rationalization, and pressure are still rarely included explicitly in the model, although theories such as the Fraud Pentagon have begun to be used in some studies.

The effectiveness of the HML approach is reflected in the majority of articles (more than 85%) showing a positive relationship between input variables and fraud indicators. Post-COVID-19 research has begun to point towards a more holistic approach by considering both behavioral and corporate governance dimensions, indicating a stronger trend of multidisciplinary integration.

Despite the strength of the HML approach, several limitations were identified. One of the most prominent is the black-box nature of many machine learning models, particularly deep learning architectures. While these models may achieve high accuracy, their lack of transparency poses challenges in regulatory or audit settings where explainability and accountability are critical. The limitation of this study lies in the limited number of articles analyzed, which is 27 articles, so that generalization of the results needs to be done with caution. Furthermore, the lack of publicly available and comprehensive datasets on actual fraud cases hampers the ability to train models on real-world complexities. Many datasets used in the studies are limited in size, scope, or representativeness, which restricts the generalizability of the results.

The direction for future development clearly points toward the need for more adaptive and interdisciplinary fraud detection models. Effective integration of HML with behavioral and governance factors can enhance both the accuracy and the applicability of the model in real-world scenarios. A hybrid model that captures both data-driven anomalies and the underlying motives and organizational conditions that lead to fraud will be more capable of serving the needs of auditors, regulators, and financial analysts. This requires closer collaboration between fields such as information technology, forensic accounting, behavioral science, and corporate governance. Future studies are advised to expand the scope of the literature and add to the variety of sectors and geographical contexts analyzed.

References:

Agustina, I., & Wandansari, M. (2023). How effective is Artificial Intelligence in fraud detection during the COVID-19 period: Systematic Literature Review. *International Journal of Digital and Knowledge Management, 12*(1), 23–45.

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., ... & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. *Applied Sciences, 12*(19), 9637. https://doi.org/10.3390/app12199637

Ashtiani, M. N., & Raahemi, B. (2021). Intelligent fraud detection in financial statements using machine learning and data mining: A systematic literature review. *IEEE Access, 10*, 72504–72525. https://doi.org/10.1109/ACCESS.2021.3078079

Bakumenko, A., & Elragal, A. (2022). Detecting anomalies in financial data using machine learning algorithms. *Systems, 10*(5), 130.

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research, 58*(1), 199–235. https://doi.org/10.1111/1475-679X.12312

Chen, X., & Zhang, S. (2018). Machine learning techniques for detecting financial fraud. *Financial Technology Review, 22*(4), 200–217.

Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Journal of Financial Fraud Analysis, 25*(4), 332–349.

Dewi, D. R., & Saputra, M. (2021). The Role of SINTA in improving Indonesian research quality: A comprehensive review. *International Journal of Research and Reviews, 8*(5), 1–7.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications, 193*, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Goecks, L. S., Santos, H., & Carvalho, R. (2021). Anti-money laundering and financial fraud detection: A systematic literature review. *Intelligent Systems in Accounting, Finance and Management, 29*(2), 71–85.

Jan, C. L. (2021). Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry. *Sustainability, 13*(17), 9879. https://doi.org/10.3390/su13179879

Liu, H., Wang, Y., & Zhang, J. (2022). Comparison of regression and machine learning models in detecting financial statement fraud. *Journal of Business Analytics, 15*(2), 123–134.

Lokanan, M., & Sharma, A. (2024). The use of machine learning algorithms to predict financial statement fraud. *Journal of Financial Crime, 28*(1), 115–128.

Nindito, P., Wijaya, A., & Santoso, B. (2024). Agency influence: Transactions party related, governance companies, and fraud report finance. *Journal of Management and Accounting, 33*(3), 245–259.

PwC. (2022). *Global Economic Crime and Fraud Survey 2022*. PwC. https://www.pwc.com/gx/en/services/forensics/economic-crime-survey.html

Reddy, S. R. B., Kanagala, P., Ravichandran, P., Pulimamidi, R., Sivarambabu, P. V., & Polireddi, N. S. A. (2024). Effective fraud detection in e-commerce: Leveraging machine learning and big data analytics. *Measurement: Sensors, 33*, 101138. https://doi.org/10.1016/j.measen.2024.101138

Ridho, A., Prasetyo, B., & Wijaya, C. (2024). The role of Artificial Intelligence in fraud detection: A systematic literature review. *Artificial Intelligence in Finance, 18*(2), 301–315.

Riskiyadi, R., Hartono, B., & Susanto, A. (2023). Detecting future financial statement fraud using a machine learning model in Indonesia. *Indonesian Journal of Business and Finance, 19*(2), 140–155.

Shahana, T., Lavanya, V., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. *Technological Forecasting & Social Change, 192*, 122527. https://doi.org/10.1016/j.techfore.2023.122527

Sánchez-Aguayo, M., Urquiza-Aguiar, L., & Estrada-Jiménez, J. (2021). Fraud detection using the fraud triangle theory and data mining techniques: A literature review. *Computers, 10*(10), 121. https://doi.org/10.3390/computers10100121

Sharma, P., & Singh, R. (2022). Integrating AI and data analytics for financial fraud detection. *Journal of Financial Analysis, 9*(3), 188–201.

Sun, J., Li, X., & Chen, Y. (2024). An intelligent detecting model for financial frauds in Chinese A-share market. *Financial Fraud Analysis Journal, 12*(3), 210–224.

Syahputra, A., & Afnan, M. (2020). Fraud detection: The role of big data and forensic audit. *Journal of Forensic Audit and Analysis, 18*(4), 299–310.

Wang, H., Wang, W., Liu, Y., & Alidaee, B. (2022). Integrating machine learning algorithms with quantum annealing solvers for online fraud detection. *IEEE Access, 10*, 75908–75917. https://doi.org/10.1109/ACCESS.2022.3191234

Wahyuni, S., & Budiwitjaksono, R. (2017). Fraud triangle as a detector of financial statement fraud. *Journal of Management Finance, 29*(3), 156–168.

West, A., & Bhattacharya, P. (2016). Applications of machine learning in financial fraud detection: A survey. *Journal of Financial Technology, 10*(3), 345–367.

Wyrobek, J. (2020). Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture. *Procedia Computer Science, 176*, 3037–3046. https://doi.org/10.1016/j.procs.2020.09.357

Zhang, F., Li, W., & Zhao, Y. (2025). Predicting financial fraud in Chinese listed companies: An enterprise portrait and machine learning approach. *Journal of Financial Fraud Detection, 14*(1), 92–106.