RESEARCH ARTICLE
published:31 July 2025
doi: 10.21070/jas.v9i2.2025

Artificial Intelligence and Data Mining in Detecting Financial Statement Fraud: A Systematic Literature Review

Anggi Putri[1], Dian Anita Nuswantara[2]*

[1,2]*Fakultas Ekonomi dan Bisnis, Universitas Negeri Surabaya, Indonesia*

General Background: Fraud in financial reporting significantly undermines stakeholder confidence and destabilises financial markets. **Specific Background:** The increasing complexity of financial data makes traditional fraud detection techniques inadequate, necessitating more sophisticated methods such as data mining and artificial intelligence (AI). **Knowledge Gap:** Despite the increasing adoption of AI in fraud detection, previous systematic literature reviews (SLRs) have generally focused narrowly on specific algorithms or data types, thus failing to provide a comprehensive assessment across multiple contexts. **Objective:** This study aims to critically evaluate the application of AI and data mining techniques in detecting financial statement fraud through a systematic literature review. **Methods:** A total of 30 peer-reviewed articles published between 2014 and 2024 were selected from Scopus, ScienceDirect, and Emerald databases using predefined inclusion-exclusion criteria and analysed narratively. **Results:** The review identified that supervised learning algorithms, specifically Support Vector Machine (SVM), Logistic Regression (LR), and XGBoost, were predominantly used, with XGBoost (96.94%) and LSTM (94.98%) showing the highest accuracy. Integration of financial and non-financial data improves detection stability. **Novelty:** In contrast to previous systematic reviews, this study offers a holistic synthesis covering algorithm types, structured and unstructured data, and diverse regional contexts. **Implications:** The findings highlight the transformative potential of AI in fraud detection and encourage further research on unsupervised learning and more in-depth utilisation of unstructured data.

Keywords: Financial Statement Fraud, Artificial Intelligence, Data Mining, Financial Ratios, Systematic Literature Review

OPEN ACCESS
ISSN 2548-3501 (online)
**Correspondence:*
Anggi Putri
Anggiputri2040@gmail.com

Received:4 June 2025
Accepted:17 July 2025
Published:31 July 2025
Citation:
Putri and Nuswantara (2025)
Artificial Intelligence and Data Mining in Detecting Financial Statement Fraud: A Systematic Literature Review

INTRODUCTION

Financial statement manipulation scandals, such as the Luckin Coffee case in China that falsified revenues exceeding US$300 million (J. Li et al., 2024), or the PT Asabri case in Indonesia that caused state losses of up to Rp22 trillion, show that financial statement fraud is not only a local phenomenon, but a global threat capable of shaking capital markets and eroding public trust (Ritonga & Budhiawan, 2024). Both cases illustrate how financial statement fraud can result in systemic losses, foster public distrust, and weaken the integrity of capital markets and internal control systems.

Financial statements are a fundamental component in communicating a company's financial information to stakeholders (Wahyu Fikri Darmawan & Umaimah Umaimah, 2025). These documents reflect the financial condition and operational performance of a business entity in a certain accounting period, including the income statement, balance sheet, statement of changes in equity, and cash flow statement (Agustan & Sari, 2022; Daeli et al., 2024). The information presented is the basis for decision making for investors, creditors, regulators, and internal management. However, increasing market pressure and weak internal controls can lead to manipulative practices in the preparation of financial statements.

In an increasingly competitive market landscape, financial reports have strategic implications. Reliable financial information enables management to make data-driven decisions and improve the efficiency of resource management (Barman, 2023; Indawatika, 2017). However, the quality of these reports can be compromised by fraud, which not only damages the credibility of the company but also causes significant economic losses and market instability (Iskandar et al., 2022; Kootanaee et al., 2021).

The prevalence of fraud in financial reporting has become an increasingly worrying issue. According to a report from (ACFE, 2022), organisational fraud is classified into three main categories: asset misappropriation, financial statement fraud, and corruption. Asset misappropriation is the most common form of fraud, accounting for approximately 86% of all reported cases. It involves the theft or misuse of organisational resources, such as cash embezzlement, inventory theft, use of company assets for personal gain, and cost manipulation. Although highly prevalent, this category has a relatively low average loss of USD 100,000 per incident. Corruption, on the other hand, occurs in 50% of fraud cases and involves the abuse of authority or position for personal gain. This can include bribery, conflict of interest, extortion, or collusion - either between individuals or between the organisation and external parties. The average loss caused by corruption is reported to be USD 150,000 per case.

However, financial statement fraud is the category with the highest financial impact. While it only accounts for around 9% of all fraud cases, the average loss is USD 593,000 per incident. These practices include actions such as inflating revenue, understating liabilities, delaying expense recognition, or concealing financial information to mislead investors, creditors, or regulators. Key motivations include meeting earnings targets, inflating share prices, gaining managerial incentives, or avoiding legal and tax penalties. These data suggest that although financial statement fraud is less common than other types, its financial and systemic impacts are much more severe and damaging (Alfian & Triani, 2019; Hari et al., 2025). These findings highlight that, despite lower incidence rates, financial statement fraud poses more systemic risks and significantly undermines investor and creditor confidence. Therefore, the authors argue that early detection of this type of fraud should be a top priority in the financial supervision system.

Financial statement fraud refers to manipulative actions taken to present a financial picture that does not reflect the actual condition of the company (Prayoga & Sudaryati, 2020; Supriadi & Aryati, 2022). The objectives of main such practices usually include portraying better company performance, increasing stock prices, avoiding taxes, or obtaining funding from external parties such as investors or creditors (Ashtiani & Raahemi, 2022). One of the most influential conceptual frameworks in explaining fraudulent behaviour is the Fraud Triangle Theory, introduced by (Cressey, 2018). A criminologist, Cressey developed this theory based on interviews with prisoners convicted of fraud in the United States. He found that individuals who commit fraud are typically driven by three core elements: pressure to meet financial demands or organisational targets, opportunity resulting from weak oversight or internal control systems, and rationalisation, where perpetrators morally justify their actions as "reasonable" or "temporary" (Prasetyo & Dewayanto, 2024; WS Albrecht, 2019).

Although developed more than half a century ago, this theory is still very relevant in the context of modern financial fraud detection, including in the era of artificial intelligence. The three components of the Fraud Triangle represent patterns of behaviour that can be traced through digital footprints and data trends, such as abnormal financial ratios, reporting frequency, or narrative patterns in annual reports. This is where machine learning and data mining become crucial: these models can detect patterns that reflect financial pressures (e.g., unusual profitability ratios), opportunities (e.g., irregularities in internal auditing), and even forms of rationalisation in MD&A or management reports. Thus, AI-based approaches are not only able to identify statistical outliers but also conceptually model the behaviour of fraudsters. This highlights the potential for integrating theoretical frameworks with intelligent technology to improve the effectiveness of financial fraud detection systems.

Conventional fraud detection methods, such as manual audits, have several limitations. Manual processes are often expensive, as they require considerable human resources and time (West & Bhattacharya, 2016). In addition, these methods tend to be less accurate due to their reliance on the subjective judgement of the individuals involved. As a result, human error is a major factor that can delay fraud identification or increase the risk of poor decision-making (Prasetyo & Dewayanto, 2024). The continued use of traditional methods, especially in many developing countries, reflects gaps in technology adoption. It also indicates resistance to the transition to a technology-based audit system.

Amidst these limitations, technological advances offer more effective and efficient solutions for fraud detection (Ashtiani & Raahemi, 2022). Artificial intelligence (AI), particularly machine learning technology, has emerged as one of the most successful approaches to fraud detection, while data mining, as a core component of AI, plays an important role in identifying patterns and detecting fraudulent behaviour quickly (Prasetyo & Dewayanto, 2024). These techniques enable the analysis of millions of financial statements to uncover suspicious trends and flag potentially fraudulent disclosures. With the ability to operate in real time, such technologies not only reduce operational costs but also provide faster and more accurate responses, thereby strengthening the company's internal control system (Massi et al., 2020). The authors argue that utilising these technologies can be a logical alternative for companies looking to improve the effectiveness of their supervisory systems, especially in the face of the increasing complexity of financial data.

A growing literature shows that intelligent approaches can be applied in various forms: from supervised learning for binary classification to unsupervised learning for anomaly detection, especially in cases where labelled data is very limited (Ashtiani & Raahemi, 2022). Some recent literature reviews have focused more narrowly on specific areas within the financial sector, such as fraud detection in credit card transactions, insurance, and fraud prediction in banking credit administration (Al-Hashedi & Magalingam, 2021), motor vehicle insurance fraud detection (Schrijver et al., 2024), and a comparative review between machine learning-based fraud detection and traditional detection methods (Gupta & Mehta, 2024). These studies have a different focus to the current research. The Systematic Literature Review (SLR) conducted in this study is more specifically targeted at financial statement fraud detection and offers a different analytical approach.

There are several empirical gaps in the existing literature, including a lack of studies that systematically integrate financial data, non-financial data, narrative textual data, social media data, and accounting-based detection models (such as F-Score and M-Score) to build comprehensive fraud detection systems. In addition, there are some limited studies that combine advanced algorithms with traditional accounting models. From a methodological perspective, existing gaps include a dearth of truly comprehensive and comparative literature reviews, especially in terms of the variety of algorithms, data sources and quality assessment protocols used.

The few existing SLRs still lack methodological diversity and are often not explicit enough in evaluating the effectiveness of the reviewed techniques. This research aims to address these gaps by presenting a Systematic Literature Review (SLR) that specifically evaluates the application of artificial intelligence (AI) and data mining (DM) technologies in detecting corporate financial statement fraud. The review covers the period from 2014 to 2024 and adopts the Kitchenham methodology, which provides a systematic framework for literature identification, selection, quality assessment and synthesis (Kitchenham & Brereton, 2007). Unlike narrative reviews, this SLR utilised clear inclusion and exclusion criteria, a structured search protocol, and a rigorous study quality evaluation approach.

Unlike previous reviews that tend to be limited to one dimension, this study offers several important contributions. First, it presents a comprehensive SLR that systematically evaluates AI/DM methods and data types (structured and unstructured) used in financial statement fraud detection. Second, this study integrates geographical perspectives and temporal trends, thereby enhancing the understanding of the context and evolution of technology adoption in financial reporting across different countries. Third, this research explores the integration of traditional accounting-based models (such as F-Score and M-Score) with modern machine learning approaches, a dimension that has not been a major focus in previous SLRs.

Research on this topic is still relatively scarce in terms of systematic investigations, especially those that explicitly explore the relationship between dataset types, AI/DM methods, and geographical dimensions in the context of corporate financial reporting. Therefore, this study is expected to provide a comprehensive synthesis and serve as a theoretical and practical foundation for the development of more adaptive and intelligent financial audit and supervision systems.

This study aims to systematically review the literature on financial statement fraud detection using artificial intelligence (AI) and data mining (DM) approaches. As an SLR study, this research does not propose formal hypotheses, but instead focuses on mapping and evaluating previous empirical findings. Therefore, this study is organised around the following research questions, which also reflect analytical expectations regarding the development of this field: 1) What techniques have been used in the literature to detect financial statement fraud? 2) What data sets have been used in the literature to detect financial statement fraud? 3) How effective are the techniques used in the literature to detect financial statement fraud? 4) How effective are the datasets used in the literature to detect financial statement fraud?

The practical implications of this research include strengthening AI-based internal audit systems that are more adaptive to diverse data types, as well as providing strategic insights for regulators and policymakers in designing financial oversight frameworks that are more responsive to fraud risks. The findings of this study are expected to provide strategic input for developers of AI-based audit models and regulators seeking to improve financial statement oversight systems.

The structure of this article is as follows: Section 2 outlines the SLR methodology, including the search strategy, selection criteria, and study quality assessment. Section 3 presents the SLR findings relating to the techniques used, data types, and model integration, and discusses the results in the context of theoretical and practical implications. Section 4 describes the limitations of the study and concludes with recommendations for future research.

METHODS

For this type of qualitative research, the research method consists of: 1. Research approach, e.g. interpretive phenomenological approach, explain why using the approach, relate it to the research focus; 2. Types and sources of data, explain in detail the type of data used, how the data is obtained and why the data is used; 3. Data analysis techniques, explain the data analysis techniques carried out in detail in accordance with the chosen research approach. For the type of research quantitative, the research method contains: 1. Type of research, explain in detail the type of research and why it is relevant to answer the research objectives, for example experimental research; 2. Research variables, measurement variables; 3. Research data, explain the sample, type and source of data; 3. Data analysis techniques, explain the data analysis techniques used to answer the research objectives. For conceptual article manuscripts, you can use a scoping review or systematic review approach. If there is a table, put it in the appendix with information such as:

Research Design

This research uses a qualitative approach using the Systematic Literature Review (SLR) method. This approach was chosen because this research aims to systematically synthesise existing knowledge about the application of artificial intelligence (AI) and data mining technology in detecting financial statement fraud in companies. SLR is the most appropriate method, as it allows researchers not only to collect and evaluate findings from previous studies, but also to identify patterns, trends, and research gaps across studies, an ability that cannot be achieved through traditional literature reviews (Kitchenham & Brereton, 2007).

[Figure 1. SLR Process Stages]

As illustrated in Figure 1, the SLR process is conducted through three main stages: planning, conducting, and reporting. In the planning stage, the researcher defines the objectives and formulates the research questions to be addressed, followed by the development of a review protocol that includes the search strategy, inclusion and exclusion criteria, and data extraction and analysis methods. The implementation stage involved systematically searching the literature in scientific databases such as Scopus, ScienceDirect, and Emerald using relevant keywords, followed by initial screening based on title, abstract, and keywords, and further selection through full-text reading to ensure study relevance. Next, an assessment of study quality was conducted, including an evaluation of journal quartiles, and key information was extracted from the selected studies to develop a data synthesis that answered the research questions. In the final stage, reporting, researchers prepared a comprehensive report of the review findings and disseminated the results to make a meaningful contribution to the academic and practitioner communities.

Subjects and Objects of Research

The object of this research is a collection of scientific articles that discuss the application of artificial intelligence (AI) and data mining (DM) in detecting financial statement fraud. A total of 30 peer-reviewed journal articles were selected as the unit of analysis through a rigorous and structured Systematic Literature Review (SLR) process. The process involved several key stages, including an initial keyword search in leading academic databases (Scopus, ScienceDirect, and Emerald), screening by title and abstract, full-text screening, and application of inclusion and exclusion criteria to ensure thematic relevance and methodological alignment with the research focus. Next, a quality assessment was conducted for each study, based on methodological transparency, scientific contribution, and topical alignment, along with an examination of journal quartiles to ensure source credibility. Each selected article became the basis for extracting data related to the type of algorithm used, the nature of the data set analysed (structured or unstructured), and model performance evaluation metrics such as accuracy.

The articles analysed in this study reflect diversity across geographical, methodological, algorithmic and data dimensions. The studies represent both developed and developing countries, using approaches such as supervised learning and unsupervised learning. The algorithms used include decision trees, random forests, logistic regression, support vector machines (SVM), neural networks, deep learning, and ensemble methods. In addition, the datasets studied consist of structured data (e.g., financial ratios and audit variables) as well as unstructured data (e.g., text from annual reports and financial news), which collectively support a comprehensive mapping and evaluation of the effectiveness of techniques and datasets in detecting financial statement fraud.

Population and Sampling Technique

The population of this study consists of all peer-reviewed journal articles published between 2014 and 2024 that explicitly discuss financial statement fraud detection using artificial intelligence (AI), machine learning (ML), or data mining (DM) approaches. The articles were sourced from three major academic databases - Scopus, ScienceDirect, and Emerald Insight - which were selected based on their strong reputation for providing high-quality literature in the fields of technology, accounting, and financial information systems.

The sampling process used purposive sampling based on systematically designed inclusion and exclusion criteria. Search keywords were created according to the research objectives and questions, then combined using Boolean operators to target relevant studies. The search queries used were as follows:

ScienceDirect and Emerald Insight

("fraud detection" AND "financial reporting") AND ("Data Mining" OR "Data analytics" OR "Text Mining" OR "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning")

Scopus:

("fraud*" AND "financial* statement") AND ("Data* Mining" OR "Data analysis" OR "Text Mining" OR "Artificial* Intelligence" OR "Machine Learning" OR "Deep learning")

[Table 1. Data Source]

As shown in Table 1, the initial search yielded a total of 567 articles: 128 from Scopus, 220 from ScienceDirect, and 219 from Emerald Insight. The screening process was conducted in several stages: selection by title and abstract, full-text review, and application of inclusion and exclusion criteria (such as publication year range, document type, and article language). To improve efficiency and accuracy at the initial screening stage, we used the Rayyan platform as a tool to screen articles by title and abstract, as well as manage and tag studies according to predefined inclusion and exclusion criteria. Articles that were irrelevant, not journals (such as proceedings, book reviews, or chapters), or not written in English were excluded. Thereafter, each selected article was further assessed for methodological quality and quartile of the journal in which it was published.

The final set of articles was evaluated based on methodological rigour and journal quartile ranking, with the help of Microsoft Excel as a tool to systematically record and assess study quality. From this process, 30 final articles were selected that met all inclusion criteria and were used as the sample for this study. Each article served as a unit of analysis, where data regarding the algorithms used, the type of dataset, and the effectiveness of the fraud detection model were extracted and analysed.

Data Validity Testing

Data validity in Systematic Literature Review (SLR) is crucial to ensure that the studies analysed are relevant, high quality, and free from bias. Referring to (Wahono, 2018), quality assessment aims to clarify selection criteria, explain variation in results across studies, evaluate the individual contributions of each study, and strengthen interpretations and conclusions. (Kitchenham & Brereton, 2007) also emphasises the importance of internal and external validity in assessing study quality.

Validity assessment begins with the application of inclusion and exclusion criteria see Table 2 for details. Next, the selected studies were further screened based on journal quartile rank (Q1-Q4) to assess publication reputation. Articles from Q1 and Q2 journals were prioritised, while Q3 and Q4 articles were selectively considered.

[Table 2. Inclusion and Exclusion Criteria]

[Table 3. Study Quality Assessment]

Finally, the technical quality assessment was conducted using five main criteria, as shown in Table 3. This assessment was based on indicators that measured how well the study provided the essential information needed to answer the research question (Kitchenham & Brereton, 2007). Studies with a score ≥ 6 were included in the main synthesis, while studies with a score < 6 were reviewed in a limited capacity or excluded if deemed inadequate. The evaluation was conducted independently by the researcher using a structured scoring sheet.

Data Extraction and Synthesis Technique

The data analysis technique in this study followed the Systematic Literature Review (SLR) approach as outlined by (Kitchenham & Brereton, 2007). The analysis was conducted in two main stages: data extraction and data synthesis. In the extraction stage, key information from the selected literature was systematically collected, including publication details, fraud detection techniques, types of datasets used, and results and effectiveness of methods applied. Next, a narrative-descriptive synthesis was conducted to answer the research questions by comparing and integrating the findings from different studies. The results of the analysis are presented in narrative form and through visualisations (such as diagrams), to provide a comprehensive overview of the trends, methods, and overall contribution of the research in supporting financial reporting integrity through technology.

RESULTS AND DISCUSSION

Number of Studies and Selection Process

Of the 567 articles identified from Scopus, ScienceDirect, and Emerald databases, 30 articles met all inclusion and quality criteria. The selection process is illustrated in the following PRISMA diagram (see Figure 2).

[Figure 2. PRISMA Diagram]

Initially, 30 duplicate articles were removed, leaving 537 articles to be screened based on title, abstract and keywords. At this stage, 457 articles were excluded as they were not relevant to the main topic. Only 80 articles passed to the full-text screening stage. After a thorough review, 40 articles were excluded for lack of focus or sufficient data. As a result, 30 articles were selected for further evaluation. All selected articles were then assessed for content quality and journal classification (quartiles). Only articles published in reputable, high-quality journals were included in the final analysis. No additional articles were eliminated during this stage, resulting in a total of 30 analysed studies.

To provide an overview, Table 4 presents a summary of the characteristics of the 30 articles analysed, including authors, year of publication, AI/DM techniques used, type of dataset, main findings, as well as the journal name and its quartile rank. This summary serves as the basis for answering the research questions in the following sections.

[Table 4. Data Extraction]

In this study, the data extraction and analysis process was carried out systematically using Microsoft Excel spreadsheets. Each article was coded based on several key parameters: the type of algorithm used, the type of dataset (structured, unstructured, or a combination), and evaluation metrics such as accuracy. Data synthesis was performed manually by categorising the findings of each study into these parameters, allowing for comparison between approaches and identification of common patterns in the application of AI and data mining for financial statement fraud detection.

Characteristics of Studies Reviewed

The 30 articles analysed in this study showed a wide distribution in terms of publication time, geographical origin, and journal quality. Temporally, these articles were published between 2014 and 2024, with a notable increase in publication frequency over the past five years, as shown in Figure 3. This trend reflects the increasing academic interest in financial statement fraud detection along with advances in artificial intelligence technology.

[Figure 3. Frequency of Articles by Year of Publication]

Geographically, the reviewed studies cover a wide range of countries, with a major concentration of China, Taiwan, and the United States, as illustrated in Figure 4. This geographical diversity reflects the different contexts and approaches to fraud detection across capital markets with different regulatory environments and financial systems.

[Figure 4. Article Frequency Based on Country]

Furthermore, as presented in Table 5, all reviewed articles were published in reputable peer-reviewed journals. Of the 30 articles, 17 were published in Q1 journals, indicating a dominant contribution from high-quality sources. The remaining articles consisted of 6 articles from Q2 journals, 4 articles from Q3 journals, and 3 articles from Q4 journals. This combination of temporal, geographical and publication quality diversity strengthens the foundation of the analysis and supports the generalisability and credibility of the research findings.

[Table 5. Study Quality Assessment Results and Journal Quartiles]

Answer to RQ1: Techniques Used in Financial Statement Fraud Detection

Analysis of 30 key studies revealed a total of 160 tests involving various machine learning and data mining approaches. As shown in Figure 5 and Figure 6, most of the techniques used fall into the category of supervised learning, while unsupervised learning methods are still relatively rarely used.

[Figure 5. Types of Learning Algorithms]

[Figure 6. Types of Algorithms]

Support Vector Machine (SVM) emerged as the most frequently used technique, appearing in 25 cases, highlighting its flexibility in handling diverse data types such as financial ratios and MD&A text (Han et al., 2012). The strength of SVM lies in its ability to classify high-dimensional data using kernel functions (Minhas & Hussain, 2016; Wang & Chen, 2024).

Logistic Regression (LR), although a classic method, remains relevant to 18 tests due to its interpretability and transparency (Hamal & Senvar, 2021). It is followed by Random Forest (RF) with 15 applications, which is appreciated for its high accuracy and resistance to overfitting (Papík & Papíková, 2021). Other techniques such as Decision Tree (DT) and its variants (C4.5, CART, CHAID) were used in 14 tests, mainly due to their ease of interpretation (Jan & Hsiao, 2018).

Artificial Neural Network (ANN) approaches and their derivatives-such as BPNN, MLP, DNN, RNN, and LSTM-reflect the growing trend of using deep learning for fraud classification (Jan, 2021; Xiuguo & Shengyong, 2022). Meanwhile, simpler techniques such as Naive Bayes and K-Nearest Neighbors (KNN) continue to be used due to their efficiency on certain datasets (Hajek & Henriques, 2017; Zhou et al., 2023).

Some studies also explored ensemble learning algorithms such as XGBoost (used 6 times), AdaBoost, and Bagging, as well as hybrid models, to improve classification accuracy. Although less common, unsupervised learning approaches were adopted in several studies through techniques such as K-Means, Autoencoder, One-Class SVM, and NLP-based models such as Transformer and GPT-2, usually in the context of anomaly detection or unlabelled data (Craja et al., 2020 ; Lu et al., 2023 ; Wu et al., 2022).

Overall, the findings indicate a stronger trend towards supervised techniques, which have shown consistent performance. However, the emergence of deep learning and unsupervised methods provides opportunities for further exploration, especially in handling large-scale and unstructured datasets.

Answer to RQ2: Types of Datasets Used in Financial Statement Fraud Detection

The analysis of 30 articles indicates that financial ratios are the most frequently used data source in detecting financial statement fraud. As illustrated in Figure 7, this approach is dominant due to the structured nature of the data, which directly reflects the financial condition of the company-such as debt-to-equity ratio, return on assets, and operating cash flow ratio. A more detailed breakdown of the specific financial ratios used across studies is presented in Table 8, which includes common metrics such as liquidity ratios, profitability ratios, leverage ratios, and activity ratios. These indicators are widely used due to their ability to signal anomalies or inconsistencies in a company's financial reporting.

[Figure 7. Types of Datasets]

[Table 8 Detectuon Model Mapping Based on Financial Ratios]

Some studies have started to include non-financial data, such as governance characteristics (e.g., board size and management background), as well as textual data from the Management Discussion and Analysis (MD&A) section. This trend reflects the increasing use of Natural Language Processing (NLP) to extract qualitative information from annual reports. Although still limited, there is also an exploration of social media data that indicates the potential of integrating public sentiment in fraud detection.

In terms of data structure, most studies rely on structured data. As shown in Figure 8, 106 tests were conducted using purely numerical data, while 42 tests combined structured data with unstructured data, such as MD&A text. Only a few studies (12 tests) relied solely on unstructured data. However, this approach represents an emerging direction in research, particularly in relation to text-based deep learning models.

[Figure 8. Dataset Structure]

These findings suggest that while structured data remains a key foundation in fraud detection, the trend to integrate it with unstructured data is growing. This combination has the potential to produce more comprehensive classification results, especially when considering managerial narratives and non-financial indicators that are often missed by conventional numerical analyses.

Answer to RQ3: Effectiveness of Techniques Used in Detecting Financial Statement Fraud

The effectiveness of detection techniques in identifying financial statement fraud is generally measured using various performance metrics. In this study, the authors chose to focus on accuracy as the main metric, as the majority of the reviewed articles used it as the main indicator to evaluate the performance of the algorithms. Based on the analysis of the 30 reviewed articles, a wide range of accuracy scores were observed across different algorithms applied for fraud detection. A summary of the algorithm accuracy comparison is presented in Figure 9.

[Figure 9. Algorithm Accuracy]

The XGBoost algorithm recorded the highest accuracy of 96.94% (B. Li et al., 2024), even when applied to a highly imbalanced dataset (65 fraud cases vs. 18,513 non-fraud cases). This highlights XGBoost's superior ability to handle class imbalance and data complexity. Meanwhile, the Long Short-Term Memory (LSTM) model achieved 94.98% accuracy in a study using a combination of numerical and textual data (Jan, 2021). The strength of LSTM lies in its ability to capture temporal context and sequential dependencies, making it particularly effective for narrative-based data analyses such as those found in the Management Discussion and Analysis (MD&A) section.

This finding shows that the effectiveness of a model is not only determined by its algorithmic architecture, but also by the compatibility between data characteristics and the model's ability to handle the complexity of fraud patterns.

Answer to RQ4: Effectiveness of Datasets Used in Detecting Financial Statement Fraud

The evaluation of dataset effectiveness was conducted by comparing the accuracy achieved from the various combinations of data used in the 30 studies reviewed. As shown in Figure 10, financial ratios were the most frequently used data type and showed a wide range of accuracy results, from 67.73% to 96.94%. The highest accuracy was reported in the study by (B. Li et al., 2024), which demonstrates the significant potential of numerical data in detecting financial anomalies.

[Figure 10. Dataset Accuracy]

Nevertheless, the combination of financial ratios with non-financial data, such as MD&A narratives or governance variables, tends to result in more stable and higher accuracy, ranging from 78.15% to 94.98%. These studies show that integrating structured and unstructured data can enrich the context of analysis and improve the model's ability to identify fraud indicators more comprehensively.

On the other hand, the use of text-based data - such as MD&A alone or in combination with social media data - is still relatively rare in the literature. However, early research suggests that this approach has promising potential, especially in the context of Natural Language Processing (NLP) and sentiment analysis. Therefore, while financial ratios remain a key foundation, the trend is shifting towards multi-source data integration as a more reliable and consistent approach to detecting financial statement fraud.

Descriptive Value and Visual Representation

To clarify the variation of algorithm performance in detecting financial statement fraud, descriptive statistical analyses were conducted on 56 types of algorithms extracted from 30 primary studies. The results are presented in Table 7, which includes the minimum, maximum, and average accuracy values, as well as the frequency of occurrence of each algorithm.

[Table 7. Algorithm Accuracy Summary]

The Support Vector Machine (SVM) algorithm was the most frequently used (25 times), with an accuracy range between 51.42% and 93.63%, and an average of 75.62%. It was followed by Logistic Regression (used in 18 studies) and Random Forest (15 studies), with an average accuracy of 73.73% and 80.30%, respectively. Meanwhile, XGBoost recorded the highest overall performance, with a maximum accuracy of 96.94% and an average of 88.49%, demonstrating its effectiveness, especially in the context of imbalanced data.

Deep learning-based algorithms such as Long Short-Term Memory (LSTM) and Temporal Convolutional Network (TCN) also showed impressive results, with accuracies of 93.29% and 94% respectively. These findings suggest that the choice of algorithm is greatly influenced by the characteristics of the data used and the complexity of the features involved in detecting financial anomalies.

[Figure 11 Avarage Accuracy of the Top 10 Most Frequently Used Algorithms]

Figure 11 displays the average accuracy of the ten most frequently used algorithms in the 30 studies analysed. The graph shows that the LSTM algorithm occupies the highest position with an average accuracy of 93.29%, followed by XGBoost (88.49%) and Artificial Neural Network (ANN) (81.94%). These results show that deep learning algorithms tend to achieve higher accuracy, especially in cases involving complex and unstructured data. LSTM, for example, excels due to its ability to process sequences of data and capture temporal dependencies, making it particularly effective in analysing narrative text such as *Management Discussion and Analysis (MD&A)* sections.

Random Forest and Support Vector Machine (SVM) also showed solid performance, with an average accuracy of 80.30% and 75.62%, reflecting the reliability of classical models that can adapt to variations in numerical data. In contrast, algorithms such as Naive Bayes (66.44%) and Backpropagation Neural Network (72.76%) ranked lower in the graph, although they are still widely used due to their ease of implementation and computational efficiency.

This graph illustrates that although algorithms such as Logistic Regression and Decision Tree remain popular due to their interpretability, they tend to lag behind in terms of accuracy compared to more modern algorithms such as LSTM and XGBoost. Therefore, the selection of algorithms for financial statement fraud detection should consider not only the interpretation ability but also the characteristics of the dataset and the complexity of the fraud pattern identified.

Use of Accounting-Based Detection Models

This study also examined the trend of using conventional accounting-based detection models, which are often used as initial features or indicators in AI-based models. These models include financial ratios, Dechow F-Score (Dechow et al., 2011) , Beneish M-Score (Beneish, 1999), Altman Z-Score (Altman, 1974), and MD&A narratives. Figure 12 presents the distribution of the use of these models over the period 2014-2024.

[Figure 12 Trends in the Use of Accounting-Based Detection Models]

During the initial period (2014-2016), financial ratios and other conventional indicators dominate. As the literature evolves, hybrid approaches begin to emerge (2016-2018), incorporating indicators such as Beneish and Dechow that more explicitly measure accounting manipulation.

A significant shift occurred between 2020 and 2022, characterised by increased interest in narrative data, particularly MD&A, driven by advances in Natural Language Processing (NLP) and the demand for increased transparency during the pandemic. In 2023-2024, financial ratios again dominated, although many studies also began to include non-financial data such as board structure and managerial ownership, in response to increasing pressure from ESG-based audits. These findings suggest that while financial ratios remain a foundational element, the integration of narrative non-financial and text-based data has become a growing trend in an effort to detect more comprehensive fraud.

CONCLUSIONS

This research concludes that financial statement fraud detection using Artificial Intelligence (AI), Machine Learning (ML), and Data Mining has made significant progress in terms of methods, data types, and model accuracy. Supervised learning algorithms such as Support Vector Machine (SVM), Logistic Regression, Random Forest, and XGBoost consistently show strong performance when processing structured data, especially financial ratios. Meanwhile, deep learning models such as Long Short-Term Memory (LSTM) excel in identifying narrative and sequential patterns in unstructured data such as Management Discussion and Analysis (MD&A), with accuracy rates exceeding 94%.

The findings suggest that combining financial and non-financial data tends to provide more stable and comprehensive accuracy results compared to using a single data source. This hybrid approach is gradually replacing the dominance of traditional methods, although classic models such as Dechow F-Score, Beneish M-Score, and Altman Z-Score still continue to be widely used as initial features. The integration of text-based approaches through Natural Language Processing (NLP) indicates a new direction in the development of fraud detection systems that are more adaptive to the dynamics of modern financial reporting.

From a theoretical perspective, this study reinforces the relevance of the Fraud Triangle Theory in the digital context, showing that pressures, opportunities and rationalisations can now be more broadly represented through structured and unstructured digital data patterns. The study also extends the scope of fraud detection theory through the lens of computational auditing, emphasising the integration of financial and non-financial indicators.

The main contribution of this review lies in its comprehensive mapping of algorithmic trends, dataset effectiveness, and methodological directions in AI-based fraud detection and data mining research. Practically, these findings provide a foundation for the development of more automated and precise AI-based audit systems. Going forward, this review opens up opportunities for further research on hybrid model integration, wider utilisation of unstructured data, and the development of new theoretical frameworks that bridge the fields of forensic accounting and information technology.

REFERENCES

ACFE. (2022). Occupational Fraud 2022: A Report To The Nations. *Association of Certified Fraud Examiners*, 1–96. https://acfepublic.s3.us-west-2.amazonaws.com/2022+Report+to+the+Nations.pdf

Agustan, T. J., & Sari, U. P. (2022). Analisis Laporan Keuangan Guna Alat Ukur Kinerja Keuangan Pada Pt. Global Imoo Telekomunikasi. *Worksheet : Jurnal Akuntansi*, *1*(2), 94–103. https://doi.org/10.46576/wjs.v1i2.2116

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, *40*. https://doi.org/10.1016/j.cosrev.2021.100402

Alfian, F., & Triani, N. N. A. (2019). Fraudulent Financial Reporting Detection Using Beneish M-Score Model in Public Companies in 2012-2016. *Asia Pacific Fraud Journal*, *4*(1), 27–42. https://doi.org/10.21532/apfj.001.19.04.01.03

Ali, A. A., Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique. *Applied Sciences (Switzerland)*, *13*(4). https://doi.org/10.3390/app13042272

Altman, E. I. (1974). American Finance Association. *The Journal of Finance*, *29*(1), 312–312. https://doi.org/10.1111/j.1540-6261.1974.tb00057.x

Ashtiani, M. N., & Raahemi, B. (2022). 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.2021.3096799

Barman, R. D. (2023). Financial Statement: A tools to evaluate Business Performance. *Business Management and Economics Engineering*, *21*(April), 819–835. https://businessmanagementeconomic.org/pdf/2023/02/819.pdf

Beneish, M. D. (1999). The Detection of Earnings Manipulation. *Financial Analysts Journal*, *55*(5), 24–36. https://doi.org/10.2469/faj.v55.n5.2296

Chen, S. (2016). Detection of fraudulent financial statements using the hybrid data mining approach. In *SpringerPlus* (Vol. 5, Issue 1). https://doi.org/10.1186/s40064-016-1707-6

Chen, S., Goo, Y.-J. J., & Shen, Z.-D. (2014). A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree for forecasting fraudulent financial statements. *Scientific World Journal*, *2014*. https://doi.org/10.1155/2014/968712

Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, *139*. https://doi.org/10.1016/j.dss.2020.113421

Cressey, D. (2018). A hipótese de Cressey (1953) e a investigação da ocorrência de fraudes corporativas: Uma análise empírica em instituições bancárias brasileiras. *Revista Contabilidade e Financas*, *29*(76), 60–81.

https://doi.org/10.1590/1808-057x201803270

Daeli, A., Hutauruk, R. A., Rifai, M. B., & Silaen, K. (2024). Analisis Laporan Keuangan Sebagai Penilai Kinerja Manajemen. *Pusat Publikasi Ilmu Manajemen*, *2*(3), 158–168. https://doi.org/10.59603/ppiman.v2i3.445

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting Material Accounting Misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. https://doi.org/10.1111/j.1911-3846.2010.01041.x

Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging Financial Social Media Data for Corporate Fraud Detection. *Journal of Management Information Systems*, *35*(2), 461–487. https://doi.org/10.1080/07421222.2018.1451954

Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications*, *90*, 374–393. https://doi.org/10.1016/j.eswa.2017.08.030

Gupta, S., & Mehta, S. K. (2024). Data Mining-based Financial Statement Fraud Detection: Systematic Literature Review and Meta-analysis to Estimate Data Sample Mapping of Fraudulent Companies Against Non-fraudulent Companies. *Global Business Review*, *25*(5), 1290–1313. https://doi.org/10.1177/0972150920984857

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

Hamal, S., & Senvar, O. (2021). Comparing performances and effectiveness of machine learning classifiers in detecting financial accounting fraud for turkish smes. *International Journal of Computational Intelligence Systems*, *14*(1), 769–782. https://doi.org/10.2991/ijcis.d.210203.007

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Elsevier. https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques

Hari, K. K., Sabrina, N., & Meratia, M. (2025). The Role of Whistleblowing in Moderate Factors Affecting Accounting Fraud Tendencies. *Journal of Accounting Science*, *9*(1), 25–61. https://doi.org/10.21070/jas.v9i1.1842

Hu, K.-H., Chen, F.-H., & Chang, W.-J. (2016). Application of correlation-based feature selection and decision tree to detect earnings management and accounting fraud relationship. *ICIC Express Letters, Part B: Applications*, *7*(11), 2361–2366. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84992709547&partnerID=40&md5=2bde14f95c170f1075e4dc62c94c5376

Indawatika, F. (2017). Penyusunan Laporan Keuangan Berbasis SAK ETAP Koperasi Intako Dan Respon Pihak Eksternal. *Journal of Accounting Science*, *1*(1), 38–50. https://doi.org/10.21070/jas.v1i1.788

Iskandar, D., Paramitha, V., & Frederica, D. (2022). Fraudulent Financial Statements in Manufacturing Companies. *Jurnal Riset Akuntansi*, *14*(1), 20–36. https://doi.org/10.34010/jra.v14i1.5499

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

Jan, C.-L., & Hsiao, D. (2018). Detection of fraudulent financial statements using decision tree and artificial neural network. *ICIC Express Letters, Part B: Applications*, *9*(4), 347–352. https://doi.org/10.24507/icicelb.09.04.347

Kim, Y. J., Baik, B., & Cho, S. (2016). Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning. *Expert Systems with Applications*, *62*, 32–43. https://doi.org/10.1016/j.eswa.2016.06.016

Kitchenham, B., & Brereton, P. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report, Ver. 2.3 EBSE Technical Report. EBSE*, *January 2007*, 1–57.

Kootanaee, A. J., Aghajan, A. A. P., & Shirvani, M. H. (2021). A Hybrid Model Based on Machine Learning and Genetic Algorithm for Detecting Fraud in Financial Statements. *Journal of Optimization in Industrial Engineering*, *14*(2), 183–201. https://doi.org/10.22094/JOIE.2020.1877455.1685

Li, B., Yen, J., & Wang, S. (2024). Uncovering Financial Statement Fraud: A Machine Learning Approach With Key Financial Indicators and Real-World Applications. *IEEE Access*, *12*, 194859–194870. https://doi.org/10.1109/ACCESS.2024.3520249

Li, J., Guo, C., Lv, S., Xie, Q., & Zheng, X. (2024). Financial fraud detection for Chinese listed firms: Does managers' abnormal tone matter? *Emerging Markets Review*, *62*, 101170. https://doi.org/https://doi.org/10.1016/j.ememar.2024.101170

Li, W., Liu, X., & Zhou, S. (2024). Deep learning model based research on anomaly detection and nancial fraud identi cation in corporate nancial reporting statements. *Journal of Combinatorial Mathematics and Combinatorial Computing*, *123*, 343–355. https://doi.org/10.61091/jcmcc123-24

Lu, Q., Fu, C., Nan, K., Fang, Y., Xu, J., Liu, J., Bellotti, A. G., & Lee, B. G. (2023). Chinese corporate fraud risk assessment with machine learning. *Intelligent Systems with Applications*, *20*, 200294. https://doi.org/https://doi.org/10.1016/j.iswa.2023.200294

Massi, M. C., Ieva, F., & Lettieri, E. (2020). Data mining application to healthcare fraud detection: A two-step unsupervised clustering method for outlier detection with administrative databases. *BMC Medical Informatics and Decision Making*, *20*(1), 1–11. https://doi.org/10.1186/s12911-020-01143-9

Metawa, N., Boujlil, R., & Alsunbul, S. (2023a). Fraud-Free Green Finance: Using Deep Learning to Preserve the Integrity of Financial Statements for Enhanced Capital Market Sustainability. *International Journal of Energy Economics and Policy*, *13*(6), 610–617. https://doi.org/10.32479/ijeep.15197

Metawa, N., Boujlil, R., & Alsunbul, S. (2023b). Fraud-Free Green Finance: Using Deep Learning to Preserve the Integrity of Financial Statements for Enhanced Capital Market Sustainability. *International Journal of Energy Economics and Policy*, *13*(6), 610–617. https://doi.org/10.32479/ijeep.15197

Minhas, S., & Hussain, A. (2016). From Spin to Swindle: Identifying Falsification in Financial Text. *Cognitive Computation*, *8*(4), 729–745.

https://doi.org/10.1007/s12559-016-9413-9
Papík, M., & Papíková, L. (2021). Application of selected data mining techniques in unintentional accounting error detection. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *16*(1), 185–201. https://doi.org/10.24136/eq.2021.007
Papík, M., & Papíková, L. (2024). Automated Machine Learning in Bankruptcy Prediction of Manufacturing Companies. *Procedia Computer Science*, *232*, 1428–1436. https://doi.org/https://doi.org/10.1016/j.procs.2024.01.141
Prasetyo, S., & Dewayanto, T. (2024). Penerapan Machine Learning, Deep Learning, Dan Data Mining Dalam Deteksi Kecurangan Laporan Keuangan-a Systematic Literature Review. *Diponegoro Journal of Accounting*, *13*(3), 1–12. http://ejournal-s1.undip.ac.id/index.php/accounting
Prayoga, H., & Sudaryati, E. (2020). Skepticism and Professionalism to Fraud Detection Ability. *Journal of Accounting Science*, *4*(2), 70–85. https://doi.org/10.21070/jas.v4i2.1087
Rahimikia, E., Mohammadi, S., Rahmani, T., & Ghazanfari, M. (2017). Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran. *International Journal of Accounting Information Systems*, *25*, 1–17. https://doi.org/https://doi.org/10.1016/j.accinf.2016.12.002
Riskiyadi, M. (2024). Detecting future financial statement fraud using a machine learning model in Indonesia: a comparative study. *Asian Review of Accounting*, *32*(3), 394–422. https://doi.org/10.1108/ARA-02-2023-0062
Ritonga, R. F., & Budhiawan, A. (2024). Review of Criminal Law on Manipulation of PT Asabri's Financial Statements As An Act Of Fraud. *Journal Equity of Law and Governance*, *4*(2), 316–326. https://ejournal.warmadewa.ac.id/index.php/elg/article/view/10210
Schrijver, G., Sarmah, D. K., & El-hajj, M. (2024). Automobile insurance fraud detection using data mining: A systematic literature review. In *Intelligent Systems with Applications* (Vol. 21, p. 200340). https://doi.org/10.1016/j.iswa.2024.200340
Song, X.-P., Hu, Z.-H., Du, J.-G., & Sheng, Z.-H. (2014). Application of machine learning methods to risk assessment of financial statement fraud: Evidence from China. *Journal of Forecasting*, *33*(8), 611–626. https://doi.org/10.1002/for.2294
Supriadi, A., & Aryati, T. (2022). Modal Intelektual Dan Kepemilikan Manajerial Pada Biaya Modal: Manajemen Laba Sebagai Moderasi. *Journal of Accounting Science*, *6*(2). https://doi.org/10.21070/jas.v6i2.1621
Wahono. (2018). Systematic Literature Review: Pengantar, Tahapan Dan Studi Kasus. In *Pengaruh Akupresur Lo4 (he kuk) dan Thai Cong terhadap Tingkat Nyeri Persalinan Kala I pada Ibu Bersalin.* (Vol. 9). http://romisatriawahono.net/2016/05/15/systematic-literature-review-pengantar-tahapan-dan-studi-kasus/
Wahyu Fikri Darmawan, & Umaimah Umaimah. (2025). The Effect of Good Corporate Governance, Earning Management on Firm Value. *Journal of Accounting Science*, *9*(1), 79–96. https://doi.org/10.21070/jas.v9i1.1922
Wang, D., & Chen, L.-X. (2024). Financial Intelligence Forecasting Model on Regression Analysis and Support Vector Machine. *Journal of Network Intelligence*, *9*(3), 1388–1404. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85202912889&partnerID=40&md5=761e17d63d1bc661834ce6921e0d10cf
West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers and Security*, *57*, 47–66. https://doi.org/10.1016/j.cose.2015.09.005
WS Albrecht, CO Albrecht, CC Albrecht, M. Z. (2019). *Instructor Solutions Manual Fraud Examination SIXTH EDITION Fraud Examination 6th Edition Albrecht Solutions Manual Visit TestBankDeal.com to get complete for all chapters*. www.cengage.com/global.
Wu, H., Chang, Y., Li, J., & Zhu, X. (2022). Financial fraud risk analysis based on audit information knowledge graph. *Procedia Computer Science*, *199*, 780–787. https://doi.org/https://doi.org/10.1016/j.procs.2022.01.097
Xiuguo, W., & Shengyong, D. (2022). An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning. *IEEE Access*, *10*, 22516–22532. https://doi.org/10.1109/ACCESS.2022.3153478
Yao, J., Pan, Y., Yang, S., Chen, Y., & Li, Y. (2019). Detecting fraudulent financial statements for the sustainable development of the socio-economy in China: A multi-analytic approach. *Sustainability (Switzerland)*, *11*(6). https://doi.org/10.3390/su11061579
Zhang, Y., Hu, A., Wang, J., & Zhang, Y. (2022). Detection of fraud statement based on word vector: Evidence from financial companies in China. *Finance Research Letters*, *46*, 102477. https://doi.org/10.1016/j.frl.2021.102477
Zhou, L., Duan, Y., & Wei, W. (2023). Research on the Financial Data Fraud Detection of Chinese Listed Enterprises by Integrating Audit Opinions. *KSII Transactions on Internet and Information Systems*, *17*(12), 3218–3241. https://doi.org/10.3837/tiis.2023.12.001

Conflict of Interest Statement: The authors declare that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

LIST OF TABLE

Table 1 */ Data Sources*

Digital Library	Number of Articles
Scopus	128
ScienceDirect	220
Emerald	219
Total	567

***Table 2** / Inclusion and Exclusion Criteria*

IC1	Year of publication between 2014-2024
IC2	For duplicate publications of the same study, only the most complete and recent version will be included
IC3	Topic relevance (title, abstract, keywords)
EC1	Papers in the form of posters, abstracts or book chapters
EC2	Review or survey type study or research
EC3	Studies not published in English

Table 3 */ Study Quality Assessment*

No	Criteria	Score	Explanation
1	Does the study mention and explain the detection technique used (AI/DM)?	0-1	1 = clearly mentioned, 0 = not mentioned
2	Does the study mention the type of dataset used (financial ratios, non-financial ratios, Dechow F-Score, Beneish M-Score, MD&A)?	0-1	1 = clearly mentioned, 0 = not mentioned
3	Did the study include evaluation metrics (accuracy, precision, recall, AUC)?	0-2	2 = complete, 1 = partial, 0 = missing
4	Do the results of the study allow the reader to evaluate the effectiveness of the technique and dataset?	0-2	2 = very informative, 1 = somewhat informative, 0 = no
5	Was the study conducted in the context of financial statements (not general transaction data)?	0-1	1 = yes, 0 = no
	Total Maximum Score	7	

***Table 4** / Data Extraction*

No	Researcher	Dataset Type	AI / Data Mining Algorithm	Accuracy
1	Kuang-Hua Hu et al. (2016)	Financial Ratios	Decision Tree (REPTree, CART, C4.5)	CART 75%, REPTreei 65%, C4.5 63%.
2	Xin-Ping Song et al. (2014)	Financial Ratio	Logistic Regression, Backpropagation Artificial Neural Network, C5.0 Decision Tree, Support Vector Machine, Ensemble of Classifiers	LR 77.9%, C5.0 DT 78.6%, BPNN and SVM 85.1% and 85.5%, ensemble of classifiers 88.9
3	Suduan Chen et al. (2014)	Financial Ratio + Non Financial Ratio	C5.0 decision tree, logistic regression, and support vector machine	C5.0 DT 93.94%, followed by LR 83.33%, and SVM 78.79%.
4	Saliha Minhas & Amir Hussain (2016)	MD&A	Stochastic Gradient Boosting), Support Vector Machine, Random Forest	SVM 88%, SGB and Boosted Logistic Regression both recorded 87%, C5 85%, Boosted Classification Trees
5	And Wang et al.	Financial Ratios	Logistic Regression, Decision Tree, Support Vector Machine, Backpropagation Neural Network, KNN + Twin Support Vector Machine Combination	BPNN 67.73%, SVM 60.52%, KNN-TSVM 60.61%, DT 59.93%, and LR 59.72%.
6	Eghbal Rahimikia et al. (2017)	Financial Ratio	Multilayer Perceptron Neural Network, Support Vector Machine, Logistic Regression	MLP in the food sector is 90.07%, textile sector 82.45%. SVM 87.47% in the food sector and 84.65% in the textile sector. LR 86.76% in the food sector and 79.13% in the textile sector.
7	Bixuan Li et al. (2024)	Financial Ratios	Decision Tree, Logistic Regression, Support Vector Machine, Random Forest, and Extreme Gradient Boosting.	XGBoost 96.94%, SVM 93.63%, RF 78.10%, LR 71.70%. DT 77.67
8	Wenjuan Li et al. (2024)	Financial Ratio	Deep Neural Network (DNN) based Autoencoder, Logistic Regression, Support Vector Machine, Decision Tree	DNN 91.7%, SVM 71.8%, LR 74.6%, DT 70.6%
9	Metawa et al.	Financial Ratio	Temporal Convolutional Network, Logistic Regression, Support Vector Machine, Random Forest, LTSM	TCN 94%, LR 82%, RF 90%, SVM 88%, and LSTM 92%.
10	Mário Papík and Lenka Papíková (2021)	Financial Ratio	Decision Tree, Random Forest	Random Forest 83.75%, Decision Tree 80.05
11	Chyan-Long Jan and David Hsiao (2018)	Financial Ratio + Non Financial Ratio	C5.0 Decision Tree, CHAID Decision Tree, ANN	CHAID+CHAID 93.47%, CHAID+ANN 81.65%, CHAID+C5.0 86.12%
12	Yeonkook J. Kim et al. (2016)	Financial Ratio + Non-Financial Ratio	Multinomial Logistic Regression, Support Vector Machine, Bayesian Network	Multinomial Logistic Regression 86.9%, SVM 85.4%, Bayesian Network 82.5%

13	Serhan Hamal and Ozlem Senvar (2021)	Financial Ratios	Random Forest, Bagging, ANN, SVM, Naive Bayes, kNN, Logistic Regression	RF 91.96%, Bagging 91.50%. LR 90.03%, KNN 89.38%, ANN 88.97%, Naive Bayes 88.56%, SVM 87.10%.
14	Kootanaee et al.	Financial Ratio	Hybrid Model (ID3 + SVM + GA), Naive Bayes, ID3 Decision Tree, Support Vector Machine	Hybrid model (ID3 + SVM + GA) 80%, Naive Bayes (78.88%), ID3 DT (75%), and SVM (73.88%).
15	Qingyang Lu et al. (2023)	Financial Ratio + Non-Financial Ratio	Logistic Regression, Support Vector Machine, Extreme Gradient Boosting, Generative Adversarial Networks+ Autoencoder, One-Class SVM	XGBoost 87.4%, SVM 87.2%, LR 86.1%, GAN + Autoencoder 67.0%, OC-SVM 69.5%
16	Chyan-Long Jan (2021)	Financial Ratio + Non-financial Ratio	Recurrent Neural Network and Long Short-Term Memory	RNN 87.18%, LSTM 94.88%
17	Jianrong Yao et al. (2019)	Financial Ratio + Non Financial Ratio	SVM, CART, BP-NN, Logistic Regression, Naive Bayes, KNN + Stepwise Regression & PCA	SVM 80.63%, CART 74.38%, LR 80%. Naïve Bayes 73.13%, BPNN 73.13%, KNN 78.13%.
No	**Researcher**	**Dataset Type**	**AI/Data Mining Algorithm**	**Accuracy**
18	Ila Dutta et al. (2017)	Financial Ratios	Decision Tree, Artificial Neural Network (ANN), Naïve Bayes, SVM, and BBN (Bayesian Belief Network)	DT 69.21%, ANN with 64.40% accuracy, Naïve Bayes 48.82%, SVM 51.42% and BBN 61.31% accuracy
19	Patricia Craja et al. (2020)	Financial Ratios	Logistic Regression, SVM, Random Forest, XGBoost, ANN, GPT-2, HAN (Hierarchical Attention Network)	XGBoost 90.83%, ANN 89.90%, RF 86.53%. HAN 84.57%, SVM 82.80%. And GPT-2 69.34%,
20	Xinyi Zheng et al (2024)	Financial Ratio	K-means clustering, Support Vector Machine, Random Forest	K-means 90.15%, SVM 86.21% and RF 80.57
21	Ali et al. (2023)	Financial Ratio	XGBoost, Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, AdaBoost	XGBoost 93.33%, SVM: 88.88% accuracy, LR: 73.88%, DT: 82%, RF: 80%, and AdaBoost: 83%
22	Wu Xiuguo et al. (2022)	Financial Ratio + MD&A	CNN, LSTM, GRU, Transformer, Random Forest, SVM, Logistic Regression, ANN, XGBoost	LTTSM 94.98%, CNN 92.53%, XGBoost 90.82%, RF and ANN 87.84% and 87.64% respectively, LR 86.26%, GRU 83.06%, Transformer 79.21%,
23	Leiruo Zhou et al. (2023)	Financial Ratio + MD&A	LightGBM + BERT, Logistic Regression, SVM, Random Forest, KNN, CNN	LightGBM + BERT 78.15%. CNN 77.92%, KNN 75.89%. LightGBM 72.71%, SVM 72.31%. RF 68.89%, LR 66.82%.
24	Byungdae An and Yongmoo Suh (2020)	Financial Ratio	Modified Random Forest (MRF), RF, Bagging, Boosting, SVM, Logistic Regression, ANN (MLP)	MRF 79.16%, RF 79.03% and ANN/MLP 78.79%
25	Wei Dong et al. (2018)	Social Media Data + Financial Ratio + MD&A	Support Vector Machine, Artificial Neural Network, Decision Tree, Logistic Regression	SVM 80.00%, Artificial Neural Network 66.17%, Decision Tree 52.38%, Logistic Regression 70.33%

26	Yi Zhan et al.	MD&A	Naive Bayes, Random Forest, Support Vector Machine (SVM)	SVM 71.39%, Random Forest 65.51% while Naive Bayes 55.72%
27	Jingyu Li et al (2024)	Financial Ratio + Non Financial Ratio + MD&A	Logistic Regression, Decision Tree, Naive Bayes, Backpropagation Neural Network. Random Forest, AdaBoost, LightGBM, XGBoost.	Logistic Regression 63.47%, Decision Tree 63.71%, Naive Bayes 62.16%, and BPNN 65.07%. RF 70.01%, AdaBoost 70.14%, LightGBM67.87%, and XGBoost 69.27%.
28	Moh. Riskiyadi (2024)	Financial Ratio	Stochastic Gradient Descent, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, Extremely Randomised Trees (ERT), AdaBoost, Gradient Tree Boosting (GTB), Neural Network	ERT 84.93%. RF model 83.93%, GTB 83.89%, AdaBoost 83.52%, Neural Network 82.63%. DT 79.42%, KNN and SGD 65.7% and 67.4% respectively. SVM 58.94
29	Suduan Chen (2016)	Financial Ratio + Non Financial Ratio	Decision Tree: CART, CHAID, Bayesian Belief Network (BBN), Support Vector Machine, Artificial Neural Networks	CHAID-CART model 87.97%, CHAID-ANN model 82.40%, followed by CHAID-BBN 81.01%, CHAID-SVM 79.05%, and CHAID-CHAID 75.28%.
30	Petr Hajek and Roberto Henriques (2017)	Financial Ratios + MD&A	Naïve Bayes (NB), Bayesian Belief Network (BBN), DTNB. CART, C4.5, JRip, LMT. SVM, Logistic Regression. Neural Networks: MLP, Voted Perceptron, Bagging, Random Forest, AdaBoostM1	BBN 90.32%, followed by DTNB 89.50%. RF and Bagging 87.50% and 87.09%. Other algorithms such as JRip (87.01%), CART (86.24%), C4.5 (86.10%), and Logistic Model Tree 85.44%, MLP 77.93%, SVM 77.95%, and AdaBoostM1 77.29%, LR 74.53%, Naïve Bayes 57.83%, and Voted Perceptron 51.16%.

***Table 5** / Study Quality Assessment Results and Journal Quartiles*

No	Researcher	Journal	Title	Quality Score	Journal Quartile	Result
1	Kuang-Hua Hu et al. (2016)	ICIC Express Letters	Application of correlation-based feature selection and decision trees to detect the relationship between earnings management and accounting fraud	6	Q4	Accepted
2	Xin-Ping Song et al. (2014)	Journal of Forecasting	Application of Machine Learning Methods for Financial Statement Fraud Risk Assessment: Evidence from China	6	Q2	Accepted
3	SuduanChen et al. (2014)	The Scientific World Journal	A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree to forecast financial statement fraud	6	Q2	Accepted
4	Saliha Minhas & 2016	Cognitive Computing	From Spinning to Fraud Identifying Forgery in Financial Texts	7	Q1	Accepted
5	Dan Wang et al.	Journal of Network Intelligence	Financial Intelligence Forecasting Model on Regression Analysis and Support Vector Machine	7	Q3	Accepted
6	Eghbal Rahimikia et al. (2017)	International Journal of Accounting Information Systems	Detecting corporate tax evasion using hybrid intelligent systems: A case study in Iran	7	Q2	Accepted
7	BIXUAN LI et al (2024)	IEEE Access	Uncovering Financial Statement Fraud: A Machine Learning Approach With Key Financial Indicators and Real-World Applications	6	Q1	Accepted
8	Wenjuan Li et al. (2024)	Journal of Combinatorial Mathematics and Combinatorial Computing	Deep learning model-based research for anomaly detection and identification of financial fraud in corporate financial statements	7	Q4	Accepted
9	Metawa et al.	*International Journal of Energy Economics and Policy*	Fraud-Free Green Finance Using Deep Learning to Maintain Financial Statement Integrity to Enhance Capital Market Sustainability	7	Q2	Accepted
10	Mário Papík and Lenka Papíková (2021)	*Equilibrium Quarterly Journal of Economics and Economic Policy*	Application of selected data mining techniques in the detection of unintentional accounting errors	6	Q1	Accepted
11	Chyan-Long Jan and David Hsiao (2018)	ICIC EXPRESS Letter, Part B: Applications	Financial Statement Fraud Detection Using Decision Trees and Artificial Neural Networks	7	Q4	Accepted

No	Researcher	Journal	Title	Quality Score	Journal Quartile	Result
12	Yeonkook J. Kim et al. (2016)	Decision Support System	Detecting financial misstatement with fraudulent intent using multi-class cost-sensitive learning	6	Q1	Accepted
13	Serhan Hamal and Ozlem Senvar (2021)	International Journal of Computational Intelligence Systems	Comparing the performance and effectiveness of machine learning classifiers in detecting financial accounting fraud for SMEs in Turkey	6	Q2	Accepted
No	**Researcher**	**Journal**	**Title**	**Quality Score**	**Journal Quartile**	**Result**
15	Qingyang Lu et al. (2023)	Intelligent System with Applications	Assessment of corporate fraud risk in China by machine learning	6	Q1	Accepted
16	Chyan-Long Jan (2021)	Sustainability	Financial Statement Fraud Detection Using Deep Learning for Sustainable Capital Market Development under Information Asymmetry Condition	7	Q2	Accepted
17	Jianrong Yao et al. (2019)	*Sustainability*	Detecting financial statement fraud for sustainable socio-economic development in China Multi-analytical approach	7	Q2	Accepted
18	Ila Dutta et al. (2017)	Expert System with Applications	Detecting financial restatement using data mining techniques	7	Q1	Accepted
19	Patricia Craja et al. (2020)	Decision Support System	Deep Learning to Detect Financial Statement Fraud	7	Q1	Accepted
14	Kootanaee et al.	Journal of Optimisation in Industrial Engineering	A Hybrid Model Based on Machine Learning and Genetic Algorithm for Detecting Fraud in Financial Statements	6	Q3	Accepted
20	Xinyi Zheng et al. (2024)	Heliyon	Data mining algorithm in accounting fraud identification with smart city information technology	6	Q1	Accepted
21	Ali et al. (2023)	Applied Science	Robust Prediction Model for Financial Statement Fraud Based on Optimised XGBoost Ensemble Learning Technique	6	Q1	Accepted
22	Wu Xiuguo et al. (2022)	*IEEE Access*	Financial Statement Fraud Detection Analysis for Chinese Listed Companies Using Deep Learning	7	Q1	Accepted
23	Leiruo Zhou et al. (2023)	KSII Transactions on the Internet and Information Systems	Research on Fraud Detection of Financial Data of Chinese Listed Companies by Integrating Audit Opinions	7	Q3	Accepted
24	Byungdae An, Yongmoo Suh (2020)	Expert System with Applications	Identifying Financial Statement Fraud with Decision Rules Obtained from Modified Random Forest	6	Q3	Accepted
25	Wei Dong et al. (2018)	Decision Support System	Utilising Financial Social Media Data for Company Fraud Detection	7	Q1	Accepted

26	Yi Zhan et al.	Financial Research Letters	Fraudulent statement detection based on word vectors: Evidence from financial companies in China	7	Q1	Accepte d
27	Jingyu Li et al (2024)	Emerging Markets Review	Financial fraud detection for Chinese listed companies: Does abnormal manager tone matter?	7	Q1	Accepte d
28	Moh. Riskiyadi (2024)	Asian Accounting Review	Detecting future financial statement fraud using machine learning models in Indonesia: a comparative study	7	Q2	Accepte d
29	Suduan Chen (2016)	SpringerPlus	Financial statement fraud detection using hybrid data mining approach	6	Q1	Accepte d
30	Petr Hajek and Roberto Henriques (2017)	*Knowledge-Based Systems*, Elsevier	Mining corporate annual reports for intelligent financial statement fraud detection - A comparative study of machine learning methods	6	Q1	Accepte d

***Table 6** / Mapping of Research Problems and Research Objectives*

Year	Researcher	Country	Research Rationale	Research Objective
2014	Xin-Ping Song et al.	China	as fraud is a major problem in China due to its transitional economy and weak corporate governance	to investigate the risk of financial statement fraud in public companies in China
	Suduan Chen et al.	Taiwan	to investigate the risk of financial statement fraud in publicly listed companies in China	to test for financial statement fraud using a machine learning approach
	Kuang-Hua Hu et al.	Taiwan	Since earnings management is often used as a tool to commit fraud, and conventional statistical models have limitations due to the assumptions of linearity and normality. Therefore, researchers develop alternative approaches based on data mining.	To investigate the relationship between earnings management and financial statement fraud by using a combination of correlation-based feature selection (CFS) and decision tree algorithms (CART, REPTree, C4.5).
2016	Saliha Minhas & Amir Hussain	America	Because financial ratios alone are not accurate enough to detect fraud, and the increasing amount of narrative content that can be used manipulatively by management.	To detect fraud in the narrative content of financial statements (MD&A).
	Yeonkook J. Kim et al	South Korea	Prior research does not distinguish between intentional and unintentional misstatements, which risks producing misleading conclusions.	To develop a model to detect and distinguish between fraud (intentional), error (unintentional), and fair reporting using multi-class classification.
	Suduan Chen	Taiwan	Previous research is limited to 1-2 statistical methods and single-stage analysis; not accurate or comprehensive enough.	To detect financial statement fraud using a multi-stage hybrid data mining approach (CHAID, CART, BBN, SVM, ANN).
2017	Eghbal Rahimikia et al.	Iran	Due to the lack of academic research on tax evasion detection in Iran, although the Iranian National Tax Administration (INTA) faces large and complex data and the risk of non-compliance.	To detect corporate tax evasion in Iran by using a hybrid system combining MLP, SVM, and LR.
2017	Ila Dutta et al	America	Previous research focuses too much on intentional restatement (fraud), while unintentional restatement also has a large impact and is often overlooked.	To detect financial statement restatements, both intentional (fraud) and unintentional (error), using various data mining techniques.
2017	Petr Hajek and Roberto Henriques	America	Previous research did not adequately integrate financial and language data; fraud detection models were not sufficiently accurate or interpretable.	To detect financial statement fraud by combining narrative text features and financial data from annual reports using various machine learning methods.
2018	Chyan-Long Jan and David Hsiao	Taiwan	Previous research only uses 1-2 data mining techniques, which are not accurate enough to detect financial statement fraud.	To detect financial statement fraud using a two-stage model: variable selection (CART, CHAID) and classification (C5.0, ANN, CHAID).
2018	Dong et al	America	Official financial data is often delayed and prone to manipulation. There is a need for alternative data sources such as financial	To detect corporate financial fraud by analysing user opinions and interactions on financial social media

			social media for earlier and more responsive fraud detection.	platforms (SeekingAlpha & Yahoo Finance), using linguistic features and social networks.
2019	Jianrong Yao et al.	China	Financial statement fraud is on the rise in China, to the detriment of socio-economic development. Previous studies tend to rely solely on financial indicators.	To detect financial statement fraud in China's capital market by using six data mining techniques and two dimension reduction methods on 17 financial variables and 7 non-financial variables.
2020	Patricia Craja et al	America	Financial statement fraud has a serious global impact, but there are still few studies that utilise deep learning for feature extraction from MD&A text to detect fraud.	To detect financial statement fraud by combining financial ratios and MD&A text using the Hierarchical Attention Network (HAN) model and compare its performance with other methods.
2020	Byungdae An, Yongmoo Suh	South Korea	Financial statement fraud harms various stakeholders because companies hide adverse information. An accurate and interpretable classification model is required to detect fraud.	To detect financial statement fraud using a Modified Random Forest (MRF) model that produces explainable decision rules.
Year	**Researcher**	**Country**	**Research Rationale**	**Research Objective**
2021	Mário Papík and Lenka Papíková	Europe	While unintentional accounting errors may seem less severe, their impact on the capital markets can be comparable to fraud. It is important to evaluate whether the financial statements contain sufficient information to detect such errors.	To detect unintentional accounting errors that lead to restatement by using data mining techniques.
2021	Serhan Hamal and Ozlem Senvar	Turkey	SMEs in Turkey are highly vulnerable to fraud, and lending banks face great challenges in detecting it. There is a need to evaluate the effectiveness of various machine learning algorithms in the context of SME financial data.	To detect financial statement fraud in Turkish SMEs using seven machine learning algorithms.
2021	Kootanaee et al.	Iran	Previous fraud detection methods have low accuracy or high computational burden. A fast, accurate, and efficient hybrid model is needed to detect financial statement fraud.	To detect financial statement fraud among 151 public companies in Iran (Tehran Stock Exchange, 2014-2015) by using a combination of Improved ID3, SVM, MLPNN, and Genetic Algorithm.
2021	Chyan-Long Jan	Taiwan	Information asymmetry between management and stakeholders threatens the sustainability of the capital market. The many cases of financial statement fraud in Taiwan have damaged public trust and audit effectiveness.	To detect financial statement fraud in public companies in Taiwan using two deep learning algorithms.
2022	Wu Xiuguo and Du	China	Financial statement fraud in China is increasing along with the expansion of the capital market. Most previous studies rely	To detect financial statement fraud in 5,130 public companies in China by integrating: - Financial and non-financial ratios

	Sheng yong		solely on numerical data and ignore the potential of narrative textual information.	- Textual features of MD&A sections - Deep learning model
2022	Yi Zhan et al.	China	Public auditing in China still relies heavily on structured financial data and manual procedures. Big data approaches to text are still underutilised. Hidden information in the Chinese MD&A section may serve as a potential fraud signal.	To detect financial statement fraud in Chinese financial companies by converting MD&A text into word vectors (BoW & Word2Vec), then classifying them using machine learning algorithms.
2023	Metawa et al.	Taiwan	In the era of green finance, the integrity of financial statements is critical to ensure the sustainability of capital markets. Conventional fraud detection approaches are not effective enough in identifying manipulations in financial reporting that impact sustainable investment.	To detect financial statement fraud using a deep learning approach.
2023	Qingyang Lu et al.	China	Corporate fraud has become an important issue in China's capital market and has caused huge losses. Manual auditing is inefficient and cannot keep up with the complexity of modern financial crimes.	To develop a fraud detection system based on machine learning using data from public companies in China from 2016 to 2020.
2023	Ali et al.	MENA	Financial statement fraud significantly affects investor confidence and the sustainability of capital markets. Previous detection models have not been sufficiently accurate or efficient in handling imbalanced data.	To detect financial statement fraud (FSF) using an optimised XGBoost approach, compare it with other models (LR, DT, SVM, AdaBoost, RF), and address data imbalance using SMOTE.
2023	Leiruo Zhou, Yunlong Duan, Wei Wei	China	Financial statement fraud is a threat to the survival of China's capital market. Previous research focuses too much on numerical data and ignores audit opinion as an important source of information.	To detect financial statement fraud in 4,153 public companies in China (over the past 6 years) by combining: - Numerical indicators from financial statements - Textual features of audit opinions.
2024	Dan Wang et al.	China	Detecting financial statement fraud is a big challenge in the era of smart economy. Conventional models have low accuracy and are inefficient in handling imbalanced data.	To detect financial statement fraud in public companies in China using a hybrid model involving the selection of 60 financial ratios and a combination of K-Nearest Neighbor (KNN) and Twin Support Vector Machine (TSVM).
Year	**Researcher**	**Country**	**Research Rationale**	**Research Objective**
2024	Wenjuan Li et al.	China	Financial statement fraud in China has serious consequences for investors, creditors, and national economic stability. A deep learning-based approach is needed to detect financial anomalies with high accuracy.	To detect financial statement anomalies and fraud using a combined approach: Deep Autoencoder Neural Network and clustering model (2-step clustering + anomaly assessment).

2024	Xinyi Zheng, Mohamad Ali Abdul Hamid, Yihua Hou	China	Traditional fraud detection in China still relies on manual review, which is subjective and slow. Smart city informationisation and data mining are considered to improve the accuracy, efficiency and objectivity of fraud detection.	To detect accounting fraud in public companies in China using K-means clustering to identify abnormal clusters, and smart city information system as a data source and integrative technology.
2024	Bixuan Li et al	America	Financial statement fraud threatens market stability and investor confidence. A machine learning approach is needed that is not only accurate but also transparent and interpretable through key financial indicators.	To detect financial statement fraud using 15 financial indicators and five classification algorithms.
2024	Jingyu Li et al (2024)	China	Financial statement fraud in Chinese public companies is on the rise, as seen in cases such as Kangmei Pharmaceutical and Luckin Coffee. Previous research relies heavily on numerical data; textual indicators such as abnormal managerial tone remain extensively unexplored.	To detect financial statement fraud in Chinese companies by combining 301 indicators (financial, non-financial, and textual).
2024	Moh. Riskiyadi (2024)	Indonesia	Financial statement fraud in Indonesia has serious implications for the capital market and the economy. Manual detection methods and legacy models are no longer adequate to address the complexity of modern manipulation.	This study compares several machine learning models to detect financial statement fraud in Indonesia.

***Table** 7 / Summary of Algorithm Accuracy*

No.	Algorithm	Frequency	Minimum Accuracy	Maximum Accuracy	Average Accuracy
1	Support Vector Machine	25	51,42%	93,63%	75,62%
2	Logistic Regression	18	52,38%	90,03%	73,73%
3	Random Forest	15	65,51%	91,96%	80,30%
4	Decision Tree	14	52,38%	93,94%	74,46%
5	Naive Bayes	7	48,82%	88,56%	66,44%
6	XGBoost	6	69,27%	96,94%	88,49%
7	Artificial Neural Network	5	64,40%	89,90%	81,94%
8	K-Nearest Neighbour	5	60,61%	89,38%	73,14%
9	Artificial Neural Network Backpropagation	4	65,07%	85,10%	72,76%
10	Long Short Term Memory	3	92,00%	94,98%	93,29%
11	AdaBoost	3	70,14%	83,52%	79,00%
12	Bagging	3	77,91%	91,50%	85,50%
13	Bayesian Belief Networks	3	61,31%	90,32%	80,57%
14	TRAIN	3	74,38%	86,24%	78,54%
15	CNN	2	77,92%	92,53%	85,23%
16	LightGBM	2	67,87%	72,71%	70,29%
17	Multilayer Perceptron	2	77,93%	86,26%	82,10%
18	Neural Network	2	66,17%	82,63%	74,40%
19	CHAID + C5.0	1	86,51%	86,51%	86,51%
20	AdaBoostM1	1	77,29%	77,29%	77,29%
21	Bayesian Network	1	82,50%	82,50%	82,50%
22	Improved Classification Tree	1	82,00%	82,00%	82,00%
23	Improve	1	77,75%	77,75%	77,75%
24	C4.5 Decision Tree	1	63,00%	63,00%	63,00%
25	CHAID+ANN	1	81,65%	81,65%	81,65%
26	HELP+HELP	1	93,47%	93,47%	93,47%
27	CHAID-ANN	1	82,40%	82,40%	82,40%
28	CHAID-BBN	1	81,01%	81,01%	81,01%
29	HELP TRAIN	1	87,97%	87,97%	87,97%
30	HELPERS	1	75,28%	75,28%	75,28%
31	CHAID-SVM	1	79,05%	79,05%	79,05%
32	Decision Table / Naïve Bayes (DTNB)	1	89,50%	89,50%	89,50%
33	Deep Neural Network	1	91,70%	91,70%	91,70%
34	Highly Ruffled Tree	1	84,93%	84,93%	84,93%
35	GAN + Autoencoder	1	67,00%	67,00%	67,00%
36	GPT-2	1	69,34%	69,34%	69,34%
37	Gradient Tree Improvement	1	83,89%	83,89%	83,89%
38	GRU	1	83,06%	83,06%	83,06%
39	HAN	1	84,57%	84,57%	84,57%

No.	Algorithm	Frequency	Minimum Accuracy	Maximum Accuracy	Average Accuracy
40	Hybrid (ID3 + SVM + GA)	1	80,00%	80,00%	80,00%
41	ID3 Decision Tree	1	75,00%	75,00%	75,00%
42	JRip	1	87,01%	87,01%	87,01%
43	K-means clustering	1	90,15%	90,15%	90,15%
44	LightGBM + BERT	1	78,15%	78,15%	78,15%
45	Logistic Model Tree	1	85,44%	85,44%	85,44%
46	LR + SVM + BPNN + DT	1	88,90%	88,90%	88,90%
47	Modified Random Forest	1	79,16%	79,16%	79,16%
48	Multinomial Logistic Regression	1	86,90%	86,90%	86,90%
49	One Class-SVM	1	69,50%	69,50%	69,50%
50	Recurrent Artificial Neural Network	1	87,18%	87,18%	87,18%
51	REPTree	1	65,00%	65,00%	65,00%
52	Stochastic Gradient Improvement	1	87,00%	87,00%	87,00%
53	Stochastic Gradient Descent	1	67,40%	67,40%	67,40%
54	Temporal Convolutional Network	1	94,00%	94,00%	94,00%
55	Transformers	1	79,21%	79,21%	79,21%
56	Selected perceptron	1	51,16%	51,16%	51,16%
	Total	160			

***Table 8** / Detection Model Mapping Based on Financial Ratios*

Financial Ratio / Liquidity	2014	2016	2017	2018	2021	2022	2023	2024
Current Ratio	(Song et al., 2014)	(Chen, 2016)		(Jan & Hsiao, 2018)	(Hamal et al., 2021; Kootanaee et al., 2021)	(Xiuguo et al., 2022)	(Metawa et al., 2023)	(B. Li et al., 2024)
Quick Ratio	(Song et al., 2014)	(Chen, 2016)	(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)		(Xiuguo et al., 2022)	(Metawa et al., 2023)	
Cash Ratio	(Chen et al., 2014; Song et al., 2014)				(Hamal et al., 2021)	(Xiuguo et al., 2022)	(Zhou et al., 2023)	
Operating Cash Flow Ratio	(Song et al., 2014)	(Chen, 2016)						(B. Li et al., 2024)
Current liabilities / Total assets	(Song et al., 2014)	(Chen, 2016)		(Jan & Hsiao, 2018)				
Inventory / Current assets	(Song et al., 2014)				(Hamal et al., 2021)			(Riskiyadi, 2024)
Current asset ratio		(Chen, 2016)	(Rahimikia et al., 2017)				(Ali et al., 2023)	
Cash to Total Assets					(Hamal et al., 2021; Jan, 2021)	(Xiuguo et al., 2022)	(Zhou et al., 2023; Ali et al., 2023)	
Inventory/Current Liabilities					(Kootanaee et al., 2021)			
Cash / Total Liabilities					(Kootanaee et al., 2021)		(Ali et al., 2023)	
Net Income and cash flow						(Xiuguo et al., 2022)		
Working Capital						(Xiuguo et al., 2022)	(Ali et al., 2023)	(W. Li et al., 2024)
Current assets/current liabilities						(Xiuguo et al., 2022)	(Ali et al., 2023)	(Riskiyadi, 2024)
Current assets / total assets						(Xiuguo et al., 2022)	(Ali et al., 2023)	(Riskiyadi, 2024)

Financial Ratio	2014	2016	2017	2018	2021	2022	2023	2024
Quick assets/current liabilities						(Xiuguo et al., 2022)		
Cash and deposits/current assets						(Xiuguo et al., 2022)	(Ali et al., 2023)	
Cash/net income								(Riskiyadi, 2024)
Operating cash flow / net profit								(Riskiyadi, 2024)
Working Capital / total assets						(Xiuguo et al., 2022)	(Ali et al., 2023)	(Riskiyadi, 2024)
Leverage								
Debt Ratio	(Chen et al., 2014; Song et al., 2014)	(Chen, 2016)	(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)	(Hamal et al., 2021; Jan, 2021)	(Xiuguo et al., 2022)	(Metawa et al., 2023; Zhou et al., 2023)	(Riskiyadi, 2024)
Debt to Equity Ratio	(Chen et al., 2014; Song et al., 2014)	(Chen, 2016)	(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)	(Hamal et al., 2021; Jan, 2021)	(Xiuguo et al., 2022)	(Metawa et al., 2023)	(Riskiyadi, 2024)
Interest Coverage Ratio								(B. Li et al., 2024)
Debt Service Coverage Ratio		(Chen, 2016)						
Long-term Liabilities / Total Assets	(Song et al., 2014)				(Kootanaee et al., 2021; Jan, 2021)	(Xiuguo et al., 2022)	(Metawa et al., 2023; Ali et al., 2023)	(Riskiyadi, 2024)
log(Liability) (log scale of leverage)	(Song et al., 2014)	(Chen, 2016)			(Jan, 2021)	(Xiuguo et al., 2022)	(Ali et al., 2023)	
Leverage of total assets		(Chen, 2016)				(Xiuguo et al., 2022)		
Equity Multiplier			(Rahimikia et al., 2017)			(Xiuguo et al., 2022)		
Total liabilities			(Rahimikia et al., 2017)			(Xiuguo et al., 2022)	(Metawa et al., 2023; Ali et al., 2023; Lu et al., 2023)	

Leverage						
Total liabilities / (TL + SE^{2})	(Rahimikia et al., 2017)					
Interest paid ÷ total liabilities		(Jan & Hsiao, 2018)				(Riskiyadi, 2024)
Shareholders' Debt to Total Assets			(Hamal et al., 2021)			
Debt to Shareholders to Total Liabilities			(Hamal et al., 2021)			
Debt to Shareholders to Average Inventory			(Hamal et al., 2021)			
Capital / Total Assets			(Kootanaee et al., 2021)			
Long-term capital adequacy			(Jan, 2021)		(Metawa et al., 2023)	
Short-term debt / total assets				(Xiuguo et al., 2022)		
Cash flow / total debt				(Xiuguo et al., 2022)		
Cash flow/current debt				(Xiuguo et al., 2022)		
Cash flow/equity				(Xiuguo et al., 2022)		
Equity / total assets				(Xiuguo et al., 2022)		
Current maturities of long-term debt					(Ali et al., 2023)	
IBD/TIC (Interest-bearing Debt / Total Invested Capital)						(B. Li et al., 2024)
Gear Ratio						(W. Li et al., 2024)
Efficiency						

Asset Turnover		(Chen, 2016)	(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)	(Hamal et al., 2021)		(Zhou et al., 2023)	(W. Li et al., 2024)
Inventory Turnover	(Chen et al., 2014; Song et al., 2014)	(Chen, 2016)		(Jan & Hsiao, 2018)	(Hamal et al., 2021)	(Xiuguo et al., 2022)	(Metawa et al., 2023; Ali et al., 2023)	(W. Li et al., 2024)
Accounts Receivable Growth	(Song et al., 2014)							
Inventory Growth	(Song et al., 2014)							
Inventory/Total Assets	(Chen et al., 2014; Song et al., 2014)		(Rahimikia et al., 2017)		(Hamal et al., 2021)	(Xiuguo et al., 2022)	(Ali et al., 2023)	(Riskiyadi, 2024)
Operating Capital / Total Assets	(Song et al., 2014)							
Trade receivables / Total assets	(Chen et al., 2014)	(Chen, 2016)	(Rahimikia et al., 2017)		(Hamal et al., 2021)			(Riskiyadi, 2024)
Accounts payable turnover	(Chen et al., 2014)							
Fixed Assets / Total Assets	(Song et al., 2014)	(Chen, 2016)	(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)				(Riskiyadi, 2024)
Accounts receivable			(Rahimikia et al., 2017)				(Ali et al., 2023)	
Inventory ÷ net sales				(Jan & Hsiao, 2018)				(Riskiyadi, 2024)
Average inventory ÷ total assets				(Jan & Hsiao, 2018)				
total_sales / total_assets								(Riskiyadi, 2024)
Accounts Payable to Cost of Goods Sold					(Hamal et al., 2021)			

Financial Ratio	**2014**	**2016**	**2017**	**2018**	**2021**	**2022**	**2023**	**2024**
Efficiency								
Cash Conversion Cycle Ratio					(Hamal et al., 2021)			

Accounts Receivable to Sales					(Hamal et al., 2021)			
Inventory to Sales					(Kootanaee et al., 2021)	(Xiuguo et al., 2022)	(Ali et al., 2023)	(Riskiyadi, 2024; B. Li et al., 2024)
Accounts Receivable/Sales					(Kootanaee et al., 2021)	(Xiuguo et al., 2022)		
Sales/Fixed Assets					(Kootanaee et al., 2021)		(Ali et al., 2023)	
Cost of Goods Sold/Sales					(Kootanaee et al., 2021)			(Riskiyadi, 2024)
Cost of Goods Sold / Total Assets					(Kootanaee et al., 2021)			
Working capital turnover ratio						(Xiuguo et al., 2022)		
Tangible asset ratio						(Xiuguo et al., 2022)	(Ali et al., 2023)	
Net Inventory							(Lu et al., 2023)	
Net Fixed Assets							(Lu et al., 2023; Ali et al., 2023)	
Net Intangible Assets							(Lu et al., 2023)	
Operating costs confirmed by the company							(Lu et al., 2023)	
Accounts receivable							(Ali et al., 2023)	
Total assets							(Ali et al., 2023)	
Shrinkage							(Ali et al., 2023)	
Accounts receivable turnover days							(Zhou et al., 2023)	
Accounts Payable Turnover Ratio								(W. Li et al., 2024)

AR Proportion (Account Receivables ÷ Total Assets)								(W. Li et al., 2024)
PY_COGS (liability/HPP)								(Riskiyadi, 2024)
SA_TE (sales / total_equity)								(Riskiyadi, 2024)
Profitability								
Gross Margin				(Jan & Hsiao, 2018)		(Xiuguo et al., 2022)	(Metawa et al., 2023)	
ROA	(Chen et al., 2014; Song et al., 2014)	(Chen, 2016)	(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)	(Hamal et al., 2021; Jan, 2021)	(Xiuguo et al., 2022)	(Metawa et al., 2023; Ali, 2023)	(W. Li et al., 2024)
Return on Equity (ROE)			(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)	(Hamal et al., 2021; Jan, 2021)	(Metawa et al., 2023; Ali, 2023; Zhou, 2023)		
Net profit before tax ratio	(Chen et al., 2014)	(Chen, 2016)						
EBITDA	(Chen et al., 2014)							(W. Li et al., 2024)
Net Profit Margin	(Song et al., 2014)							
EBIT to Total Assets	(Song et al., 2014)					(Xiuguo et al., 2022)		
Financial Ratio	**2014**	**2016**	**2017**	**2018**	**2021**	**2022**	**2023**	**2024**
Profitability								
Net Income / Income from Operations	(Song et al., 2014)	(Chen, 2016)						
Gross profit margin	(Song et al., 2014; Chen, 2014)	(Chen, 2016)	(Rahimikia et al., 2017)	(Jan & Hsiao, 2018)		(Xiuguo et al., 2022)		(W. Li et al., 2024)
Return on assets before tax, interest, dep.		(Chen, 2016)						
Return on Assets (EBIT version)			(Rahimikia et al., 2017)		(Hamal et al., 2021)			

ROA/Return on Operations of Operations					(Kootanaee et al., 2021)			
Gross profit			(Rahimikia et al., 2017)					
Net profit			(Rahimikia et al., 2017)			(Xiuguo et al., 2022)	(Metawa et al., 2023; Lu et al., 2023; Ali, 2023)	(Riskiyadi, 2024)
Retained Earnings to Assets			(Rahimikia et al., 2017)					(Riskiyadi, 2024)
Net profit rate				(Jan & Hsiao, 2018)	(Hamal et al., 2021)			(Riskiyadi, 2024)
Gross profit rate (*Gross Profit ÷ Revenue)*				(Jan & Hsiao, 2018)			(Ali et al., 2023)	(Riskiyadi, 2024)
Net Profit / Cost of Goods Sold					(Kootanaee et al., 2021)			
Operating Profit/Sales					(Kootanaee et al., 2021)	(Xiuguo et al., 2022)		(Riskiyadi, 2024)
Earnings Before Interest and Taxes/Sales (EBIT Margin)					(Kootanaee et al., 2021)			
EBIT / Total Assets					(Kootanaee et al., 2021)	(Xiuguo et al., 2022)		(Riskiyadi, 2024)
EBIT/Current Liabilities					(Kootanaee et al., 2021)			
Operating Cost/Sales					(Kootanaee et al., 2021)			(Riskiyadi, 2024)
Earnings before interest and tax						(Xiuguo et al., 2022)		
Return on invested capital						(Xiuguo et al., 2022)		

Net profit realised by the company							(Lu et al., 2023)	
Total profit realised by the company							(Lu et al., 2023)	
Amortisation							(Ali et al., 2023)	
Operating cost ratio							(Zhou et al., 2023)	
Debt to Asset Ratio Growth								(B. Li et al., 2024)
AR Change Ratio								(B. Li et al., 2024)
Return on Net Assets								(W. Li et al., 2024)
Net Interest Rate								(W. Li et al.),
Financial Ratios	**2014**	**2016**	**2017**	**2018**	**2021**	**2022**	**2023**	**2024**
Market Value								
Dividend Yield							(Metawa et al., 2023; Ali, 2023)	
Earnings per Share (EPS)							(Metawa et al., 2023; Ali, 2023)	
Ratio (P/E)						(Xiuguo et al., 2022)	(Metawa et al., 2023; Ali, 2023)	
Cash flow/cash dividend						(Xiuguo et al., 2022)		
Total earnings per share						(Xiuguo et al., 2022)		
Income per share						(Xiuguo et al., 2022)		
Cash dividend per share						(Xiuguo et al., 2022)		
Price-book ratio						(Xiuguo et al., 2022)		
Others (Size/Growth								
Operating expense ratio	(Chen et al., 2014)			(Jan & Hsiao, 2018)				
Income from Operations / Total Assets	(Song et al., 2014)							

Efficiency of Revenue to Assets	(Song et al., 2014)						
log(Total Assets) (Company Size)	(Song et al., 2014)	(Chen, 2016)	(Rahmikia et al., 2017)				
Revenue Growth from Business	(Song et al., 2014)				(Hamal et al., 2021)		
log(Main Operation Cost)	(Song et al., 2014)				(Hamal et al., 2021; Kootanaee et al., 2021)		
Operating cash flow ÷ net sales				(Jan & Hsiao, 2018)			
Sales growth rate		(Chen, 2016)		(Jan & Hsiao, 2018)	(Hamal et al., 2021)	(Xiuguo et al., 2022)	(Zhou et al., 2023)
Operating cash flow ÷ current liabilities				(Jan & Hsiao, 2018)			
Proportion of cash to total assets		(Chen, 2016)					
Total asset growth rate				(Jan & Hsiao, 2018)		(Xiuguo et al., 2022)	
Accounts receivable growth					(Hamal et al., 2021)		
Other Fixed Assets to Total Assets					(Hamal et al., 2021)		
Other extraordinary expenses and losses to net sales					(Hamal et al., 2021)		
Doubtful Accounts Receivable to Total Assets					(Hamal et al., 2021)		
Prepaid Expenses for Future Months					(Hamal et al., 2021)		

Financial Ratio	2014	2016	2017	2018	2021	2022	2023	2024
to Total Assets								
Prepaid Expenses for Future Years to Total Assets					(Hamal et al., 2021)			
Other (Size/Growth								
Net cash flow from operations					(Jan, 2021)		(Metawa et al., 2023)	
Capital accumulation ratio						(Xiuguo et al., 2022)		
Asset inflation and incremental ratio						(Xiuguo et al., 2022)		
Profit growth rate						(Xiuguo et al., 2022)	(Zhou et al., 2023)	
Operation lever						(Xiuguo et al., 2022)		
Sales						(Xiuguo et al., 2022)	(Metawa et al., 2023; Ali, 2023)	
Inventory						(Xiuguo et al., 2022)	(Ali et al., 2023)	
R&D expenditure as a percentage of revenue							(Metawa et al., 2023)	
Net Cash Flow from Operating Activities							(Lu et al., 2023)	
Net Increase in Cash and Cash Equivalents							(Lu et al., 2023)	
Amortisation of Intangible Assets							(Lu et al., 2023)	
Amortisation of Long-term Deferred Charges							(Lu et al., 2023)	

Change in current assets	(Ali et al., 2023)
Cash return	(Ali et al., 2023)
Change in income tax payable	(Ali et al., 2023)

***Table 9** / Detection Model Mapping Based on Financial Ratios*

Non-Financial Ratios	2014	2016	2018	2021	2022	2023	2024
Shareholding ratio of major shareholders	(Chen et al., 2014)	(Chen, 2016)		(Jan, 2021)		(Metawa et al., 2023)	
Shareholding ratio of directors and supervisors	(Chen et al., 2014)	(Chen, 2016)	(Jan & Hsiao, 2018)	(Jan, 2021)		(Metawa et al., 2023)	
Whether the chairman concurrently serves as CEO	(Chen et al., 2014)	(Chen, 2016)			(Xiuguo & Shengyong, 2022)		
Board size	(Chen et al., 2014)	(Chen, 2016)			(Xiuguo & Shengyong, 2022)		
Share guarantee for directors and supervisors	(Chen et al., 2014)	(Chen, 2016)		(Jan, 2021)		(Metawa et al., 2023)	
Number of external supervisors		(Chen, 2016)					
Audited by BIG 4		(Chen, 2016)	(Jan & Hsiao, 2018)			(Lu et al., 2023)	
Number of directors and supervisors			(Jan & Hsiao, 2018)				
Audit committee			(Jan & Hsiao, 2018)				
Number of audit committee members			(Jan & Hsiao, 2018)				
Restatement of financial statements.			(Jan & Hsiao, 2018)				
Total Accruals		(Hu et al., 2016)					
President director percentage				(Jan, 2021)		(Metawa et al., 2023)	
Ownership concentration index CR1					(Xiuguo & Shengyong, 2022)		
Ownership concentration index CR5					(Xiuguo & Shengyong, 2022)		

Relationships among the top 10 shareholders					(Xiuguo & Shengyong, 2022)		
Proportion of independent board members			(Jan & Hsiao, 2018)		(Xiuguo & Shengyong, 2022)		
Number of employees					(Xiuguo & Shengyong, 2022)		
Supervisor size					(Xiuguo & Shengyong, 2022)		
Number of senior supervisors					(Xiuguo & Shengyong, 2022)		
Total annual salary of directors, - supervisors and senior supervisors					(Xiuguo & Shengyong, 2022)		
Total annual salary of the top 3 directors, supervisors, and senior supervisors					(Xiuguo & Shengyong, 2022)		
Non-Financial Ratio	**2014**	**2016**	**2018**	**2021**	**2022**	**2023**	**2024**
Total annual salary of the top 3 directors					(Xiuguo & Shengyong, 2022)		
Total annual salary of the top 3 senior supervisors					(Xiuguo & Shengyong, 2022)		
Standard and unqualified auditor's report					(Xiuguo & Shengyong, 2022)		
Auditors from overseas accounting firms or not						(Lu et al., 2023)	
Total stock transfer turnover obtained by weighted average						(Lu et al., 2023)	
Total stock transfer volume obtained by weighted average						(Lu et al., 2023)	
Average turnover of stock transfers						(Lu et al., 2023)	

Reason for resignation	(Lu et al., 2023)
Audit Opinion on Financial Statements	(Zhou et al., 2023)
Management capability	(J. Li et al., 2024)
Internal control	(J. Li et al., 2024)
political connections	(Riskiyadi, 2024)
audit quality	(Riskiyadi, 2024)

Table 10 */ Predictive Model Mapping: Dechow F-Score, Beneish M-Score, Altman Z-Score, and MD&A*

Dechow F-Score	
Year	**Frequency**
2016	(Kim et al., 2016)
2017	(Dong et al., 2018; Dutta et al., 2017; Hajek, 2017)
2020	(Craja et al., 2020)

Beneish M-Score	
Year	**Frequency**
2016	(Kim et al., 2016)
2017	(Dong et al., 2018)
2021	(Papík & Papíková, 2024)
2023	(Ali et al., 2023)

Altman Z-Score	
Year	**Frequency**
2019	(Yao et al., 2019)
2022	(Xiuguo & Shengyong, 2022)
2024	(W. Li et al., 2024)

Narrative text (MD&A)	
Year	**Frequency**
2016	(Minhas & Hussain, 2016)
2017	(Hajek & Henriques, 2017)
2018	(Dong et al., 2018)
2020	(Craja et al., 2020)
2022	(Zhang et al., 2022)
2024	(J. Li et al., 2024)

LIST OF FIGURE

Figure 1 */ SLR Process Stages*

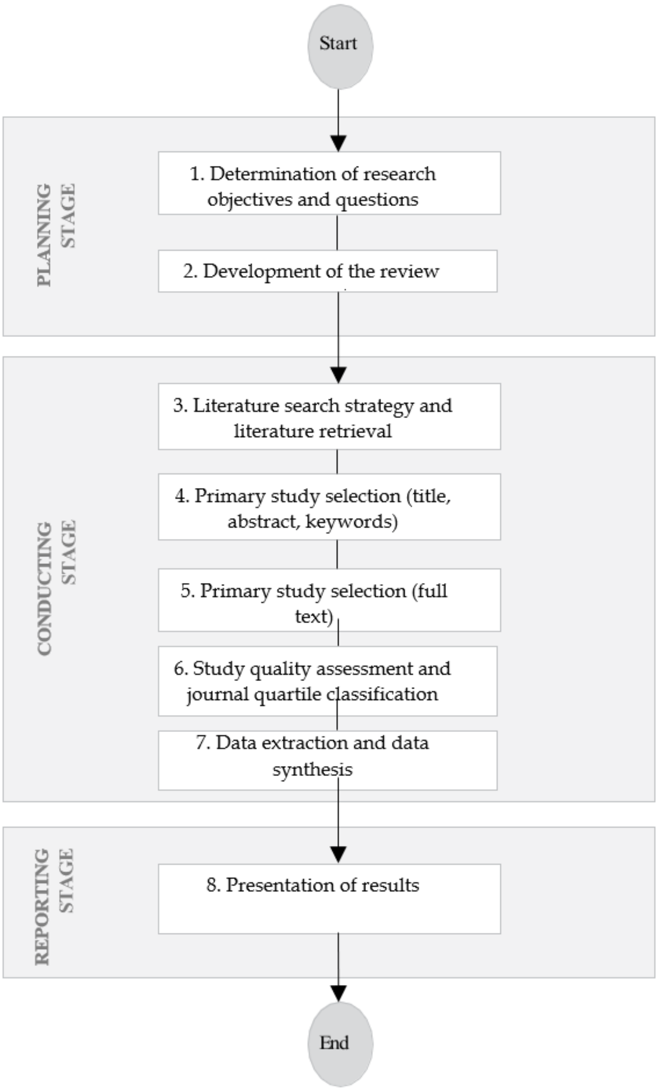

***Figure 2** / PRISMA Diagram*

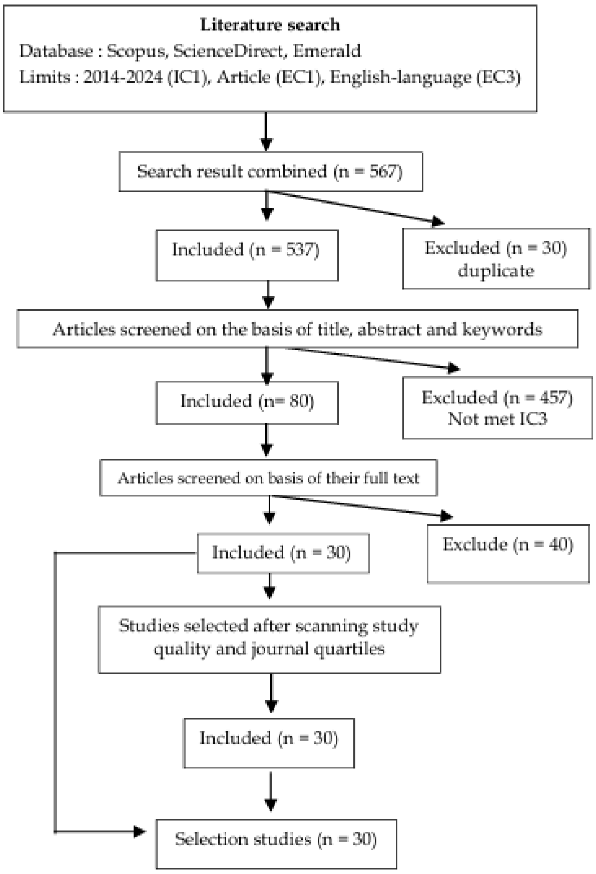

***Figure 3** / Frequency of Articles by Year of Publication*

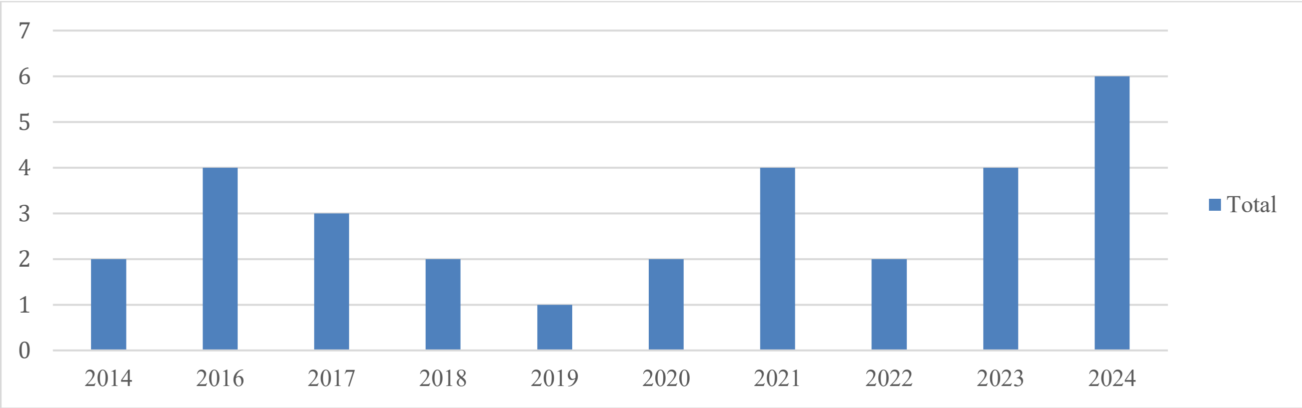

Figure 4 */ Frequency of Articles by Country*

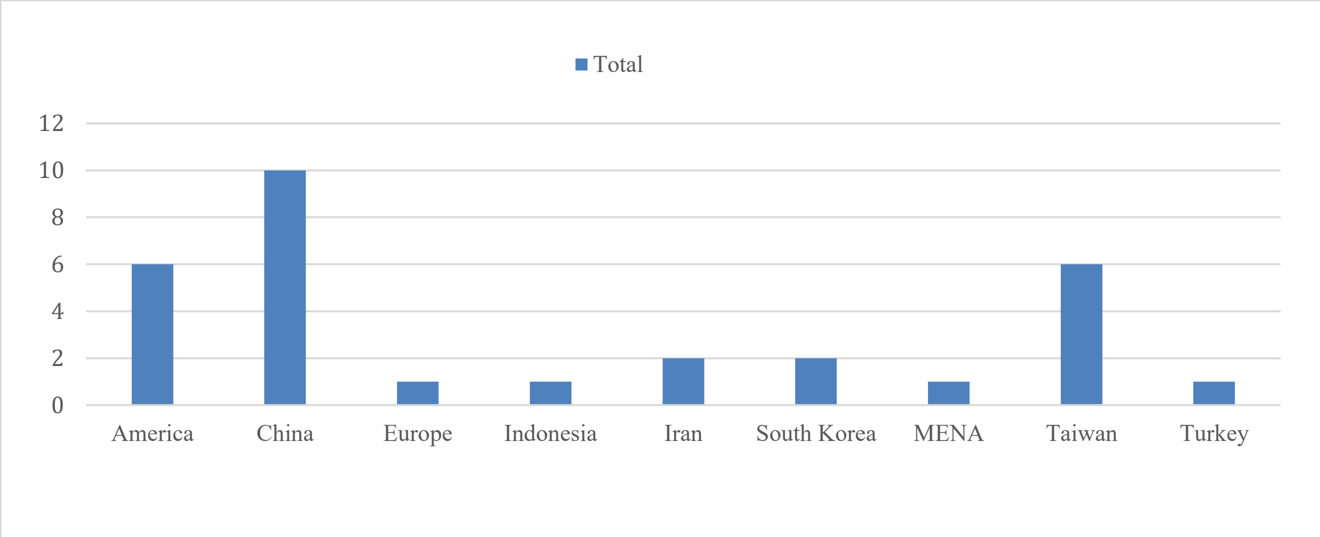

Figure 5 */ Types of Learning Algorithms*

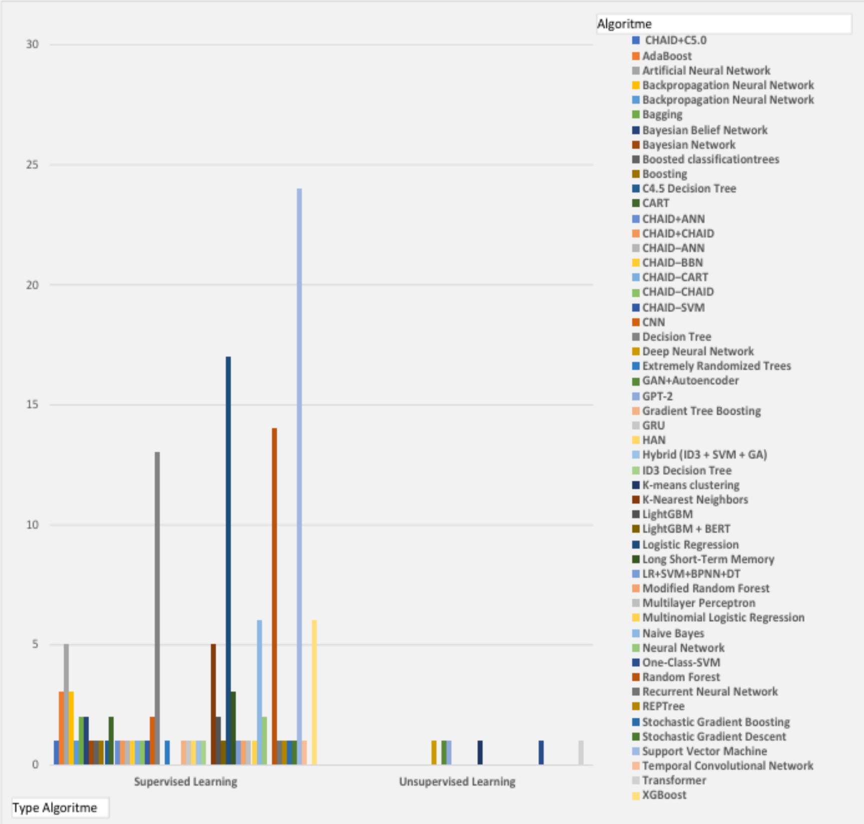

***Figure 6** / Types of Algorithms*

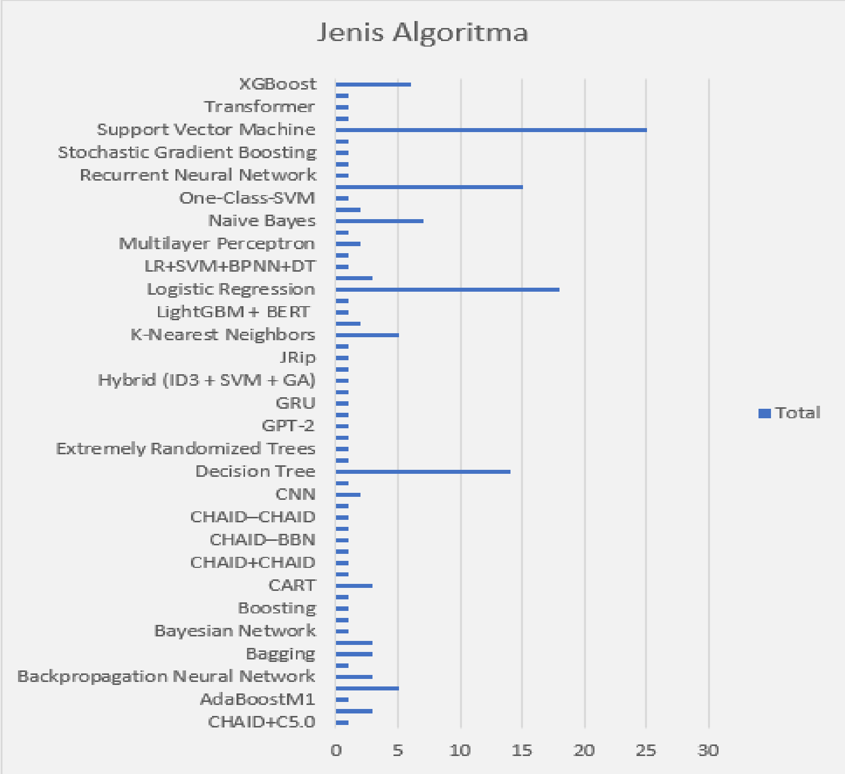

***Figure** 7 / Types of Datasets*

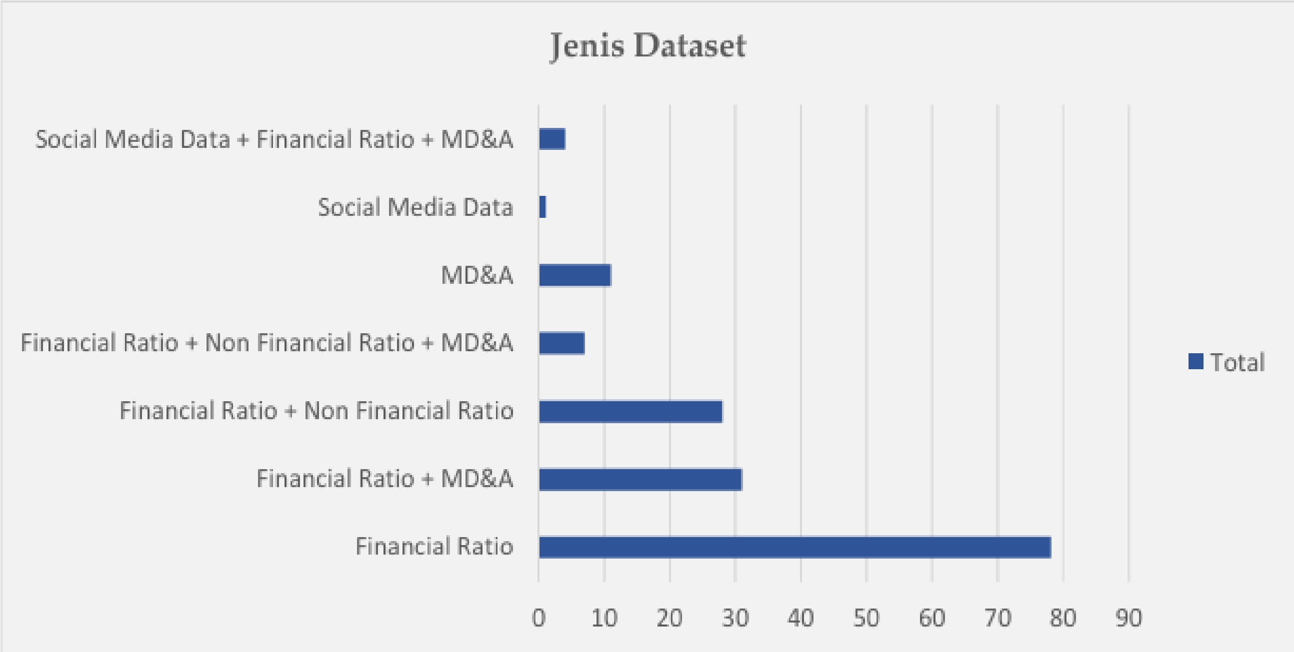

***Figure 8** / Dataset Structure*

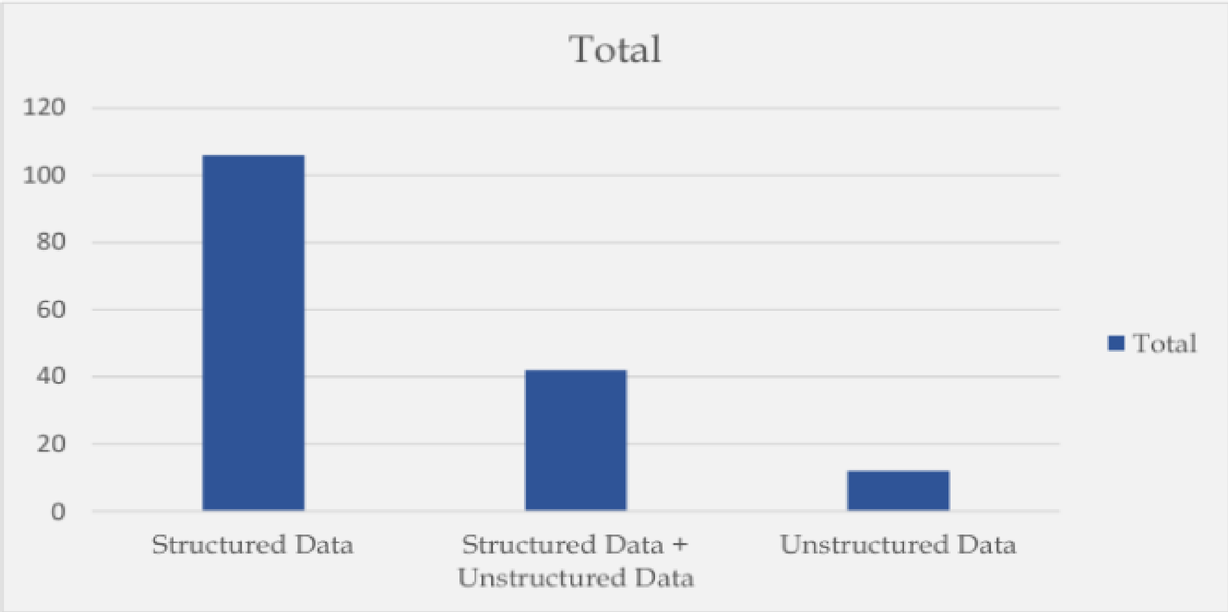

***Figure 9** / Algorithm Accuracy*

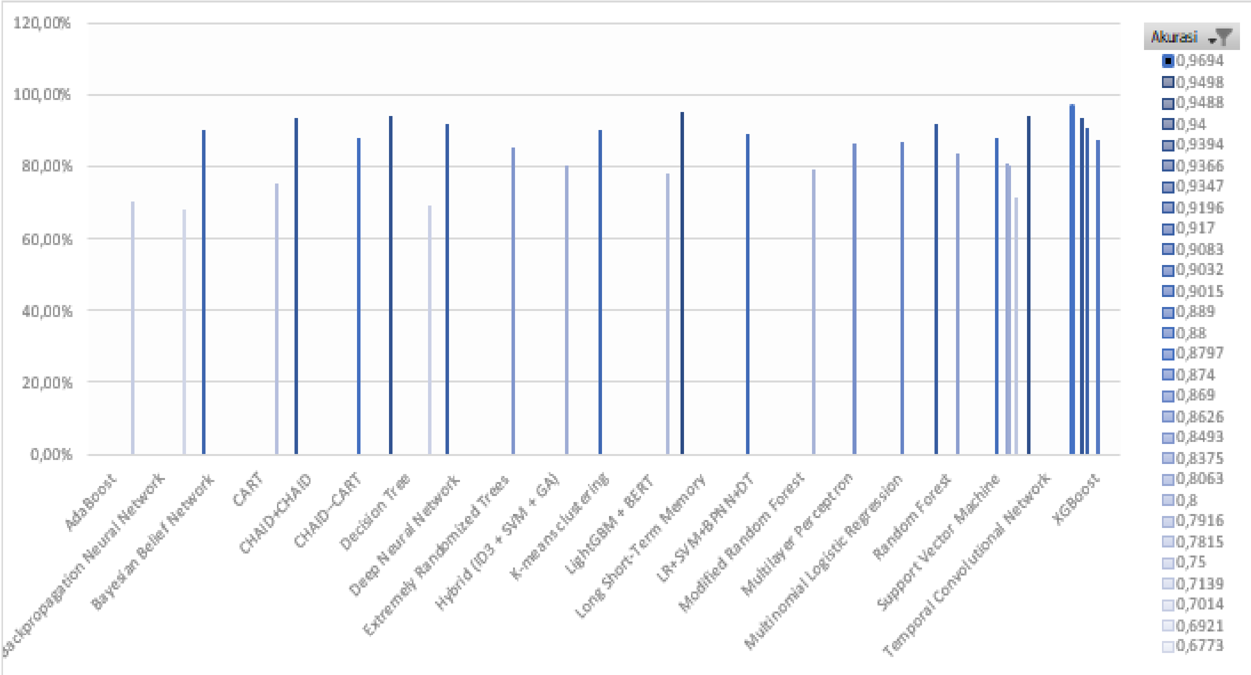

Figure 10 */ Dataset Accuracy*

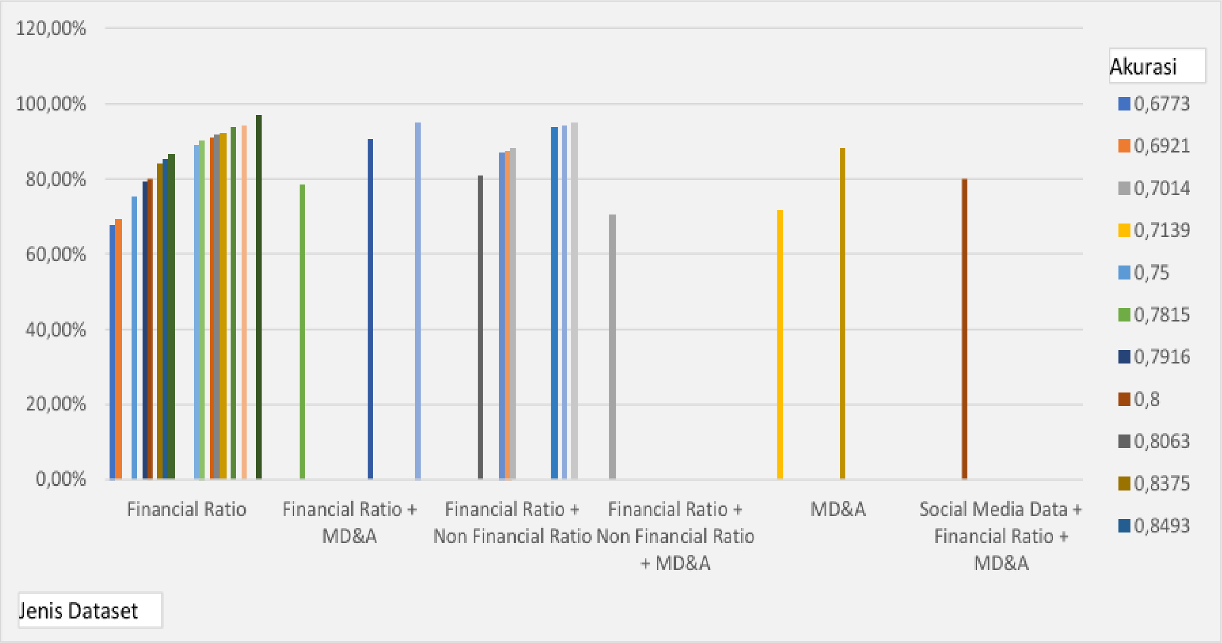

Figure 11 */ Average Accuracy of the Top 10 Most Frequently Used Algorithms*

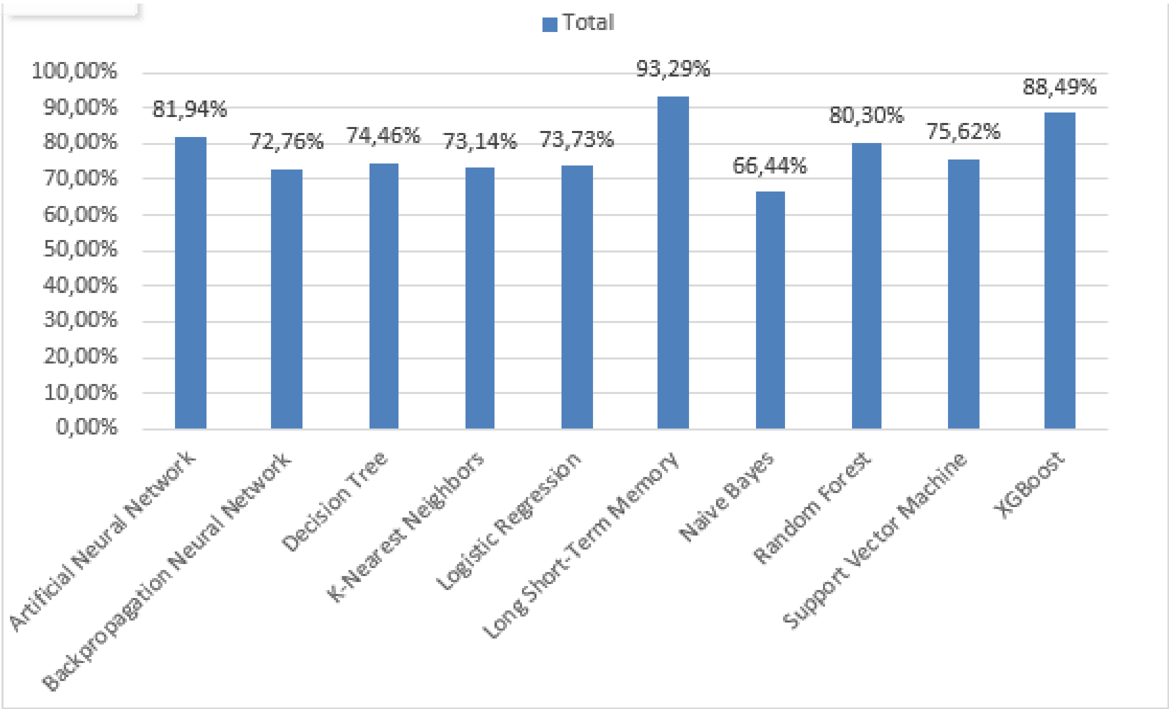

***Figure 12** / Trends in the Use of Accounting-Based Detection Models*