Annals of Library and Information Studies
Vol. 72, June 2025, pp. 138-150
DOI: 10.56042/alis.v72i2.14752

Artificial Intelligence and Machine Learning in Fraud Detection: A Comprehensive Bibliometric Mapping of Research Trends and Directions

Dr. Himanshu Thakkar[a], Saptarshi Datta[b], Priyam Bhadra[c], Dr. Haresh Barot[d], Dr. Jayendrasinh Jadav[e]

[a]Assistant Professor, School of Management Studies, National Forensic Sciences University, Gandhinagar, Gujarat, Pin: 382007, India. E-mail: himanshuthakkar04@gmail.com

[b] UG Research Scholar, School of Management Studies, National Forensic Sciences University, Gandhinagar, Gujarat, Pin: 382007, India. E-mail: titosaptarshi@gmail.com

[c] UG Research Scholar, School of Management Studies, National Forensic Sciences University, Gandhinagar, Gujarat, Pin: 382007, India. E-mail: priyambhadra25@gmail.com

[d] Dean, School of Management Studies, National Forensic Sciences University, Gandhinagar, Gujarat, Pin: 382007, India. E-mail: haresh.barot@nfsu.ac.in

[e]Associate Professor, Department of Business Studies, Sardar Patel University, Anand, Gujarat, Pin: 388120, India. Email: jayendrasinhj@gmail.com

Received: 31 October 2024; Accepted: 28 April 2025

This study presents a bibliographic analysis of emerging trends in applying artificial intelligence (AI) and machine learning (ML) to the detection and prevention of financial fraud and provides insights for future research. Bibliographic analysis on fraud data analysis helps researchers gain insight on research trends, research impact, and classification. Bibliometric analysis on fraud data analytics is helpful to researchers in getting insights on research trends, research impact and classification. However, research on fraud data analytics using machine learning is limited. The main objective of this quantitative analysis is to explore emerging trends in fraud data analytics and machine learning (ML) for financial crime detection and prevention. Bibliometric data has been collected from the Scopus database. One thousand four hundred eighty-three documents from the SCOPUS database have been analysed using VOSviewer. The data analysis divulges a growing interest in leveraging these technologies to strengthen financial crime detection. Fraud data analytics, Artificial Intelligence and Machine Learning are vital in identifying complex criminal patterns, strengthening companies in preventive vigilance, and ensuring fraud elimination. The study portrays the need for vigorous frameworks for the legislature, real-time analytics systems and more powerful tools and calls for integrating governments, financial institutions, and technology providers to strengthen prevention strategies and tackle financial crimes more effectively.

Keywords: Bibliometric Analysis, Financial Crime, VOSviewer, Fraud Data Analytics, Machine Learning

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) are deliberately emerging as tools for enhancing the detection and prevention of financial crimes. Banking and financial institutions are using advanced analytics to detect patterns of fraudulent behaviour in banking, insurance, and investment[1]. The study highlights the importance of fraud data analytics and machine learning in detecting financial crimes. It progresses on the existing literature and available research, emphasising the trends in applying vital technologies.

The finance, banking and insurance sectors are facing several threats and risks. These risks include manipulating financial statements, credit card fraud, cyber threats and money laundering[2]. Machine learning models like Random Forests, Support Vector Machines (SVM), and Deep Learning techniques have been proven effective in identifying and mitigating risks[3]. AI and ML are being used to keep track of transactions and enhance financial institutions' ability to be more effective in preventing fraud[4].

Collaboration between fraud detection systems and regulatory rules is increasingly essential in reducing financial risks. The collaboration between the Financial Intelligence Units (FIUs) and the cooperation between the advanced nations are needed for money laundering and terrorism financing[5]. Moreover, machine learning methods without supervision, like K-means clusterisation and autoencoders, bring the unexpected information missing in oversized datasets to light effectively[6]. When criminals use new technology, research in the future should focus on continually updating fraud

detection techniques through constant learning in fields such as blockchain and cryptocurrency[7].

2. Literature Review

Corrupt actors employ new ways to evade detection, so money laundering is a big problem, with AI and ML being one of the methods of combating these challenges. AI is capable of enhancing the processes of identity verification through the use of multiple external sources like government databases and social networking sites, which results in more efficient analysis of client data than older methods. Owing to the changing needs of the field, the average age of a document in this bibliography is only 3.51 years, which can be observed from the annual growth of 14.33 %. Major clusters of high-ranking keywords deal with the use of artificial intelligence and machine learning in specific fraud detection, late payment compliance, and anti-corruption advocacy practices. Implementing AI and ML methods for these purposes involves ethical challenges and unfair biases that must be addressed first [8]. This investigation uses the Bibliometrix R package to pursue a bibliometric review of 386 published papers between 2004 and 2024 on financial fraud detection. Such investigation identifies torrential expansion in the amount of conducted research, alongside 1186 author keywords and 530 Keywords. Significant issues include the headings "financial crime" and "fraud," in which machine learning and artificial intelligence functioned. The study also notes emerging areas like blockchain and cryptocurrency. The United States, China, and the United Kingdom are the leading contributors to this field [9]. This study reviews economic fraud and risk management, emphasizing the need for adaptation to new technologies and market changes. Utilizing bibliometric methods, it maps influential works, authors, and trends based on the Scopus database. The review looks at the evolution of scientific research in the scope of the development of AI and ML in the BFSI industry, offering the American and Chinese perspectives, and assesses the role and implications of these studies in the academic and public domain. Generally, this research synthesizes the available literature while calling for more in-depth research concerning critical areas of the discipline [10]. Such a systematic review benefits understanding 'Mapping Research on Artificial Intelligence and Machine Learning in Banking Financial Services and Insurance'. Using the PRISMA protocol, the paper started from 39,498 articles indexed in Scopus and whittled it down to 1,045 acceptable articles under the study's criteria. The N-gram analysis resulted in the emergence of 177 unique terms. N-gram analysis pinpointed and examined nine research areas relating to Fin-tech, Risk management, Money laundering and Actuarial science. These studies provide a broad picture of the current state of knowledge and its limitations and suggest areas for future research. These policymakers' insights are helpful to scholars and practitioners in the BFSI industry and the increasing body of knowledge of AI and ML usage in both academic and then science [11].

In Figure 1, the History of Bibliometrics in Information Sciences has been mentioned. In the

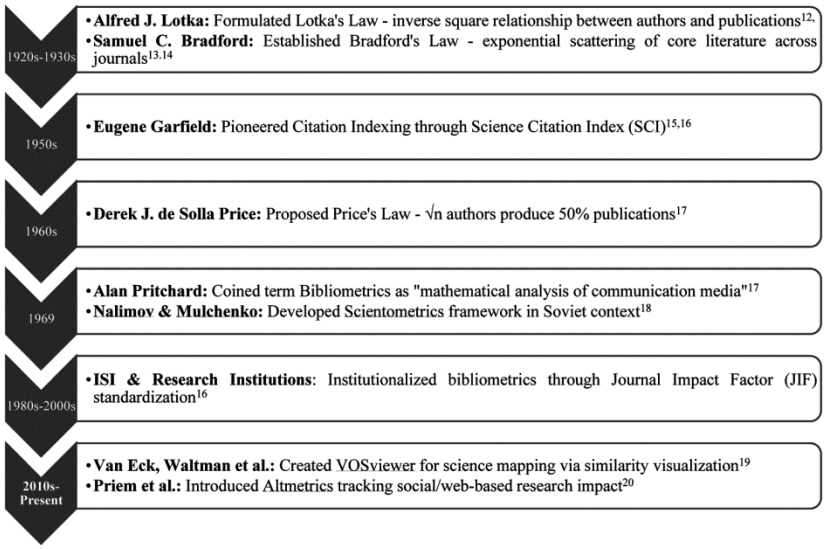

Fig. 1 — Foundational Development of Bibliometrics in Information Science
Source: Compiled by the authors

1920s, Alfred J. Lotka introduced Lotka's Law, highlighting that a small number of authors produced the most publications, while in the 1930s,[12]. Bradford formulated Bradford's Law to describe the scattering of core literature across journals[13]. Eugene Garfield created the Science Citation Index in the 1950s, which enabled systematic citation tracking[15]. In the 1960s, Derek J. De Sola proposed a price law in Price, stressing that a small group of authors contributed half of all publications. In 1969, Allan Pritchard used the term "bibliometrics" for mathematical analysis of communication media[16]. In the same year, Nalimov and Mulchenko introduced the concept of Scientometrics to the Soviet Union[17]. From the 1980s through to the 2000s, the journal Impact Factor became a standard tool for journal evaluation[19]. In the 2010s, Van Eck and Waltman developed the VOSviewer to map scientific research. At the same time, Prem and his colleagues introduced Altmetrics to measure the impact of research beyond traditional contexts, especially on social media and the Web[20].

3. Methodology

This research reveals the bibliometric analysis method as a tool for investigating and examining the existing literature regarding data analytics and machine learning in identifying fraud earlier[21]. The data for the research was obtained from the Scopus database, a massively authoritative academic database worldwide. The pieces of 1,483 documents issued between 2010 and 2024, thus giving 14 years, were gathered. The terms that were used to limit the literature search are as follows: "fraud detection," "fraud prevention," "fraud analytics," "data mining," "machine learning," "big data," and "forensic accounting." For the study, only papers were chosen, and the selected drafted data was saved in a simple CSV file, which is acceptable to Excel. This file was then successfully uploaded to a set of count bio tools, such as VOSviewer and Biblioshiny, for further analysis. The study brought forth graphs and tables, which could be graphed with the help of these tools. A literature survey revealed that no prior research has done a similar bibliometric analysis that would focus on this particular research topic. This work aims to fill that gap, providing valuable insights for researchers by identifying current trends and gaps in the field. The results are expected to be helpful for those seeking to advance research in this domain.

3.1. Research Questions

1. Which authors and sources have contributed the most to research in this area?
2. Which articles and countries have had the most significant influence in this area of research?
3. What are the authors' most frequently used keywords in this domain?
4. What constitutes the core knowledge base and intellectual framework of a research field?
5. What are the key themes and conceptual patterns identified within this study area?
6. How have the themes in this field evolved, and what emerging techniques or pressing issues have been observed?
7. What gaps in the existing works have been identified, and how might they inspire future research?

3.2. Searching & Retrieving Data

Databases present for searching documents for bibliographic analysis include SCOPUS, Web of Science, Google Scholar, etc. The SCOPUS database yielded more search results than any other[22]. The SCOPUS database contains a large number and variety of publications[23]. In addition, in recent times, Scopus has earned its reputation as a comprehensive bibliographic data source, and has proven to be reliable and, in some ways, even better than WoS. Because of this, studies have used the SCOPUS database to search for data and retrieve data for analysis.

The search was conducted on 31 August 2024 using the keyword "fraud detection", "fraud prevention" and also adding the OR Boolean – "fraud analytics", "data mining" that resulted in a total of 4170 results[24]. The preferred results were used for the systematic reviews (PRISMA)[21] for identification, screening and selection of the final data for the analysis (Figure 2).

4. Data Analysis

4.1. Descriptive bibliometric analysis (performance analysis):

This table provides general information regarding the analyzed observed literature, and a total of 1483 peer-reviewed journal articles were published in 704 journals from 2010 to 2024, indicating that the first paper in the study area was published in 2010, of which these papers were contributed by 4113 authors, with only 137 documents being single-authored, and the remaining 3976 documents being written in collaboration with the

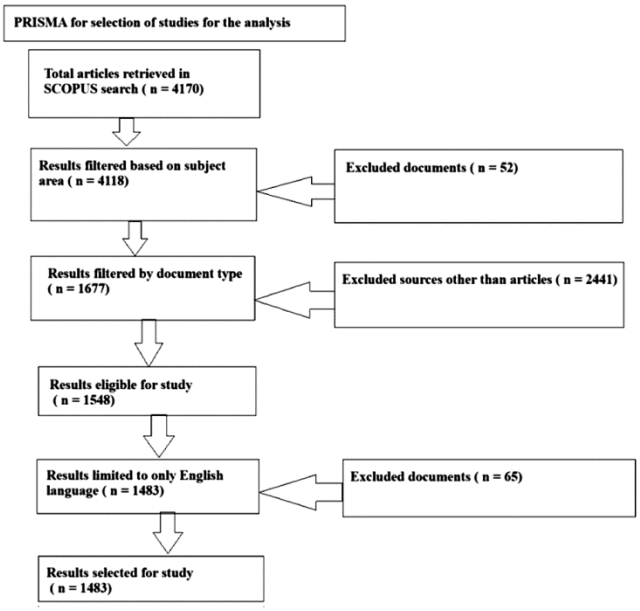

Fig. 2 — PRISMA used for the selection of studies for the analysis
Source: Compiled by the authors based on PRISMA guidelines

Table 1 — Main Information About Data

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2010:2024
Sources (Journals, Books, etc)	704
Documents	1483
Annual Growth Rate %	16.52
Document Average Age	3.27
Average citations per doc	24.9
References	61591
DOCUMENT CONTENTS	
Keywords Plus (ID)	4221
Author's Keywords (DE)	3240
Authors	4113
Authors of single-authored docs	129
AUTHORS COLLABORATION	
Single-authored docs	137
Co-Authors per Doc	3.39

Source: Compiled by the authors

authors[21]. Collaborations in financial crime research, in which the majority of papers are co-authored, reflect the need for different skills to use AI and ML in early fraud detection. The integration of machine learning and fraud data analytics enhances the possibilities of fraud detection by enabling the analysis of large datasets to detect patterns, predict fraud, and improve prevention strategies in real time. These technologies provide scalable, adaptive tools to combat financial crime more effectively.

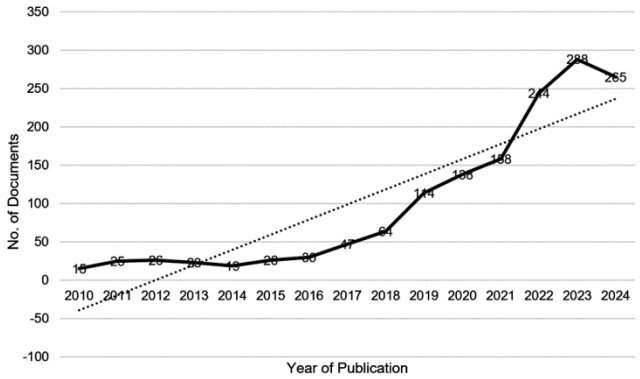

Fig. 3 — Number of published articles per year
Source: Compiled by the authors

4.2. Number of published articles per year:

Figure 3 clearly shows the yearly scientific output in the study area. This trend reflects a dynamic shift in researcher attention from 2010 to 2024. Notably, starting in 2018, there has been a significant increase in the number of publications, extending a peak in 2023. Significantly, between 2020 and October 2024, 1,093 papers were published, representing an impressive 73% of the overall output. This strong growth emphasizes the rapid advances in the study of machine learning and data analytics, which are critical in our increasingly digital world. The annual growth rate of published articles was relatively high (16.52%), as shown in the table, and the rising trendline in the figure reflects the growing interest in the field over the years, which has increased in recent years[21]. This publication surge aligns with the increasing global focus on financial crime prevention, driven by technological advancements and the need to combat more sophisticated criminal tactics. The heightened interest also suggests that researchers are responding to the growing demand for innovative solutions, mainly by integrating machine learning and data analytics, to address the evolving challenges in financial crime detection and prevention.

4.3. Most relevant sources:

The collected research articles are spread across 705 journals. Notably, 320 articles (21.57%) were concentrated in 16 journals, each contributing at least ten papers. Figure 4 highlights the most active journals regarding publication volume **(RQ1)**[21]. Among them, IEEE Access stands out as the leading journal, with 75 papers (5.05%). It is followed by Expert Systems with Applications, which accounts for 39 papers (2.62%), and Applied Sciences (Switzerland), contributing 26 papers (1.75%). The

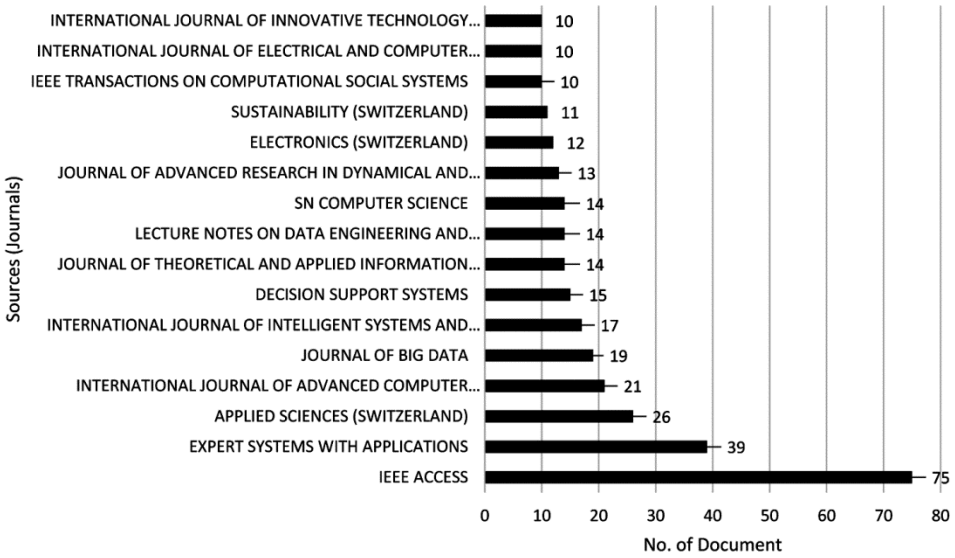

Fig. 4 — Most Relevant Sources
Source: Compiled by the authors

Table 2 — Most Cited Articles

Author	TITLE	SOURCES	TC	TCpY
(Najafabadi et al., 2015)[26]	"Deep learning applications and challenges in big data analytics"	Journal of Big Data	1736	173.6
(Johnson & Khoshgoftaar, 2019)[27]	"Survey on deep learning with class imbalance"	Journal of Big Data	1600	266.667
(Hancock & Khoshgoftaar, 2020)[28]	"Survey on categorical data for neural networks"	Journal of Big Data	310	62
(Wei et al., 2013) [29]	"Effective detection of sophisticated online banking fraud on extremely imbalanced data"	World Wide Web	233	19.417
(Shaukat et al., 2020) [30]	"Performance comparison and current challenges of using machine learning techniques in cybersecurity"	Energies	179	35.8
(Herland et al., 2018) [31]	"Big data fraud detection using multiple medicare data sources"	Journal of Big Data	121	17.286
(Johnson & Khoshgoftaar, 2019)[32]	"Medicare fraud detection using neural networks"	Journal of Big Data	68	11.333
(Zhu et al., 2021)[33]	"Intelligent financial fraud detection practices in post-pandemic era"	Innovation	67	16.75

Source: Compiled by the authors

diversity of journals reflects the field's interdisciplinary nature, showcasing the wide range of research perspectives across various disciplines.

4.4. Most cited articles:

The most cited articles were strategically determined through extensive context analysis, which demonstrates the value of quality research. Context analysis measures importance and influence and ranks publications, sources, and scholars based on the number of references [25]. The top ten most-cited documents **(RQ2)** are in Table 2. The results showed that [26] The study was the most influential and received favorable attention from scholars; it topped the most cited papers in the field, with a total of 1,736 citations (TC) and 173.6 annual TC as of October 2024, followed by [27,28,29], with 1,600, 310, and 233 citations, respectively[21]. Integrating advanced computational techniques like machine learning is pivotal in enhancing fraud detection systems. These methods enable the identification of complex patterns within large datasets, improving the detection of unusual or suspicious activities.

4.5. Most productive and influential countries:

Authors from 30 different countries have contributed significantly to the corpus. Table 3 highlights the top 15 most productive and cited countries **(RQ1 and RQ2)**. By October 2024, India had emerged as the leader with 1,048 published articles, followed by China with 806 and the USA with 459. Regarding citations, the USA dominates

Table 3 — Most productive and influential countries

Country	TC	Country	Year	Articles
USA	9941	India	2024	1048
China	4032	China	2024	806
India	2248	India	2023	805
Australia	1549	China	2023	646
France	924	India	2022	577
South Africa	922	China	2022	471
Italy	874	USA	2024	459
Malaysia	850	USA	2023	411
Turkey	784	India	2021	399
Belgium	771	USA	2022	333
United Kingdom	572	China	2021	329
Saudi Arabia	480	India	2020	297
Canada	427	USA	2021	269
Iran	418	China	2020	239
		USA	2020	234

Source: Compiled by the authors

with an impressive 9,941 citations, followed by China at 4,032, India at 2,248, Australia at 1,549, and France at 924. The integration of advanced methods, such as machine learning, plays an important role in detecting fraud by uncovering patterns hidden in large datasets. It enables real-time monitoring, adaptive learning, and better handling of imbalanced data, improving accuracy and reducing false positives across financial and cybersecurity systems.

4.6. Most frequent authors' keywords:

Our fund has 4,221 unique author keywords with a total of 11,844 events. Table 4 presents the 20 most frequently encountered keywords (**RQ3**), each of which appears at least 50 times, while Figure 4 looks at them in Word Cloud, where font size represents frequency [21]. These 20 keywords account for 2,640 occurrences (62.54%). The term crime is the most frequent, appearing in 437 papers, highlighting the importance of understanding and addressing criminal behaviour. Fraud detection (357 occurrences) emphasizes the need for tools to combat financial risks, and data mining (233 occurrences) reflects the value of analysing large datasets for meaningful patterns. Learning systems (196) and machine learning (186) further show the growing reliance on automated problem-solving and insight-generation techniques.

4.7. Bibliometric network analysis (science mapping):

Science mapping explains various types of academic documents such as resources, writers, key words, contexts and contexts- that form a network of

Table 4 — Most frequent authors' keywords

Words	Occurrences
Crime	437
Fraud detection	357
Data mining	233
Learning systems	196
Machine learning	186
Machine-learning	148
Learning algorithms	123
Deep learning	119
Anomaly detection	117
Decision trees	99
Classification (of information)	97
Fraud	97
Credit card fraud detection	89
Finance	81
Artificial intelligence	80
Big data	73
Support vector machines	56
Statistics	52

Source: Compiled by the authors

Fig. 5 — Bibliometric network analysis (science mapping)
Source: Compiled by the authors

ideal, intellectual and social structure[25,34]. In this study, multiple network analysis techniques have been employed, including co-citation analysis to uncover frequently cited clusters that reflect the intellectual foundation of the field and co-word analysis to identify co-occurring keywords, highlighting key themes and concepts. Additionally, thematic evolution has been used to trace the development of research topics over time while trending topic analysis identified the most recent and popular areas of interest. Together, these analyses address **RQ4, RQ5, and RQ6**, offering a detailed understanding of the field's structure, growth, and emerging trends. (Fig. 5)

4.8. Co-citation analysis:

Co-citation analysis can identify a research domain's knowledge base and intellectual structure[34]

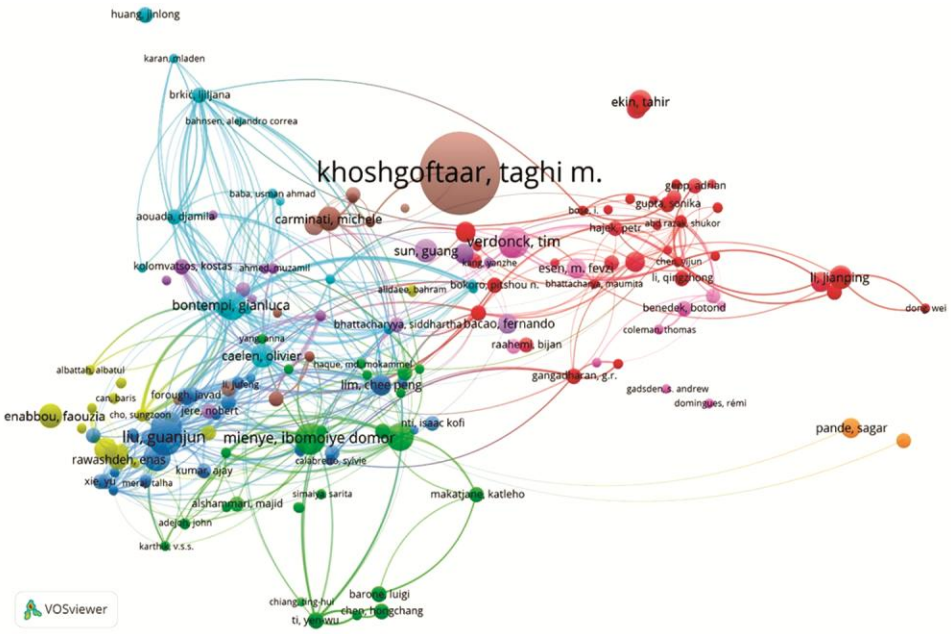

Fig. 6 — Co-citation analysis
Source: Compiled by the authors

or uncover thematic clusters[35]. It can be done on one of the studied relationships, such as publications, journals, or authors. The basic assumption behind it is that if two articles are co-cited along with other publications, they are more often thematically the same and close [36]. Our dataset included 60,947 cited documents; therefore, co-citation analysis for the complete data was nearly impossible. To make it possible, we leaned toward the most pertinent sources, which received no less than five citations each[37] and thus selected 327 references that were very much cited. Moreover, we came across two references, but they had a combined link strength to others below 10; hence, we included 325 references in the co-citation analysis.

Figure 6 shows the co-cited references from our corpus of studies drawn with the VOSviewer program. Three well-defined clusters illustrate the research area's knowledge base and intellectual structure. Each group has a different colour, and the node shape marks the number of citations, whereas the culverts represent the power of the co-citation (link) story[37].

We used all keywords (both author and index keywords) that co-occurred in at least five papers from the explained collection to do a keyword co-occurrence network (KCN) analysis with the help of VOSviewer. Of 6285 keywords, only 466 were linked to the co-occurrence network (see Figure 6), and fractional counting was used to structure the network[38]. Larger nodes stand for the keywords that are repeated most of the time, whereas thicker connections between the nodes show more frequent linkages between these keywords[35].

Figure 6 consists of five main categories containing the nodes of all keywords, each cluster having a unique colour representing the theme of the co-occurrence patterns. There are:

- The green cluster is among the keywords such as "fraud detection", "finance", "big data", "fraud", "data analytics", and "forensic accounting,". "It is a group about spreading data analytics and forensic accounting techniques to catch and avert fraud in the finance sector, which piggyback on the data used massively for new insights.
- The blue cluster focuses on "anomaly detection" and "data handling", wherein detecting unusual patterns is one of the key aspects of anomaly detection, and appropriate data use assures the accuracy and reliability of these analyses.
- The red cluster points out several different approaches, for example, "crime", "deep learning", and other "fraudulent transactions". The red cluster shows a relationship between these approaches and the use of machine learning to detect fraudulent transactions and crime prevention.

- The yellow cluster includes terms like "machine learning" and "neural networks", which use machine learning and neural networks to automate fraud detection, enabling predictive models to identify suspicious activities in real time.

4.9. Trend topics:

Trending topics analysis is a subcategory of word-based analyses using co-occurrence statistics. It is helpful to capture newly born topics and the dynamic structure in some areas during the acceptance horizon[39]. Figure 7 displays the hierarchical format of current topic trends in this field using Biblioshiny, according to the authors' keywords (2010–2024 timespan; maximum of three words per year and minimum word frequency of two). This method provides a general comprehensive capture of all top trending topics developed in time. In this figure, the size of dots corresponds to how many times a word was used, and the length of lines is based on duration. These keywords indicate the adoption of big data analytics (BDA) in examining financial reporting quality (FRQ), as evidenced by the results. Notably, in 2011, data mining techniques emerged as a prominent trend and the most frequently used algorithm since 2010; however, it is not depicted in Figure 8. Since it was not listed as one of the authors' keywords. In 2022, the emergence of topics such as fraud, credit cards, and crime highlights significant trends in financial reporting and data analytics. The focus on fraud reflects an increasing awareness of the need to understand and combat various forms of financial misconduct, including identity theft and investment scams, as organizations enhance their prevention strategies. Meanwhile, the prevalence of credit card transactions in digital payments has made them a prime target for fraud, underscoring the importance of research into transaction monitoring and anomaly detection to safeguard against fraudulent activities. In addition, the broader topic of crime includes other forms of crime, including financial fraud, money laundering, and cybercrime. This

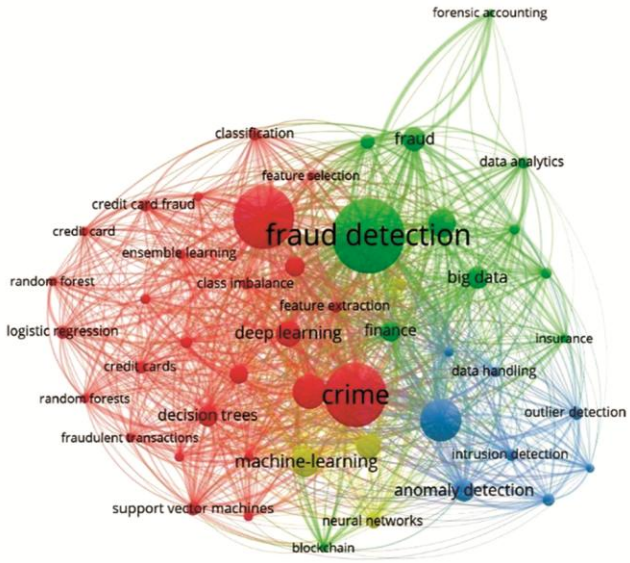

Fig. 7 — Co-occurrence Network
Source: Compiled by the authors

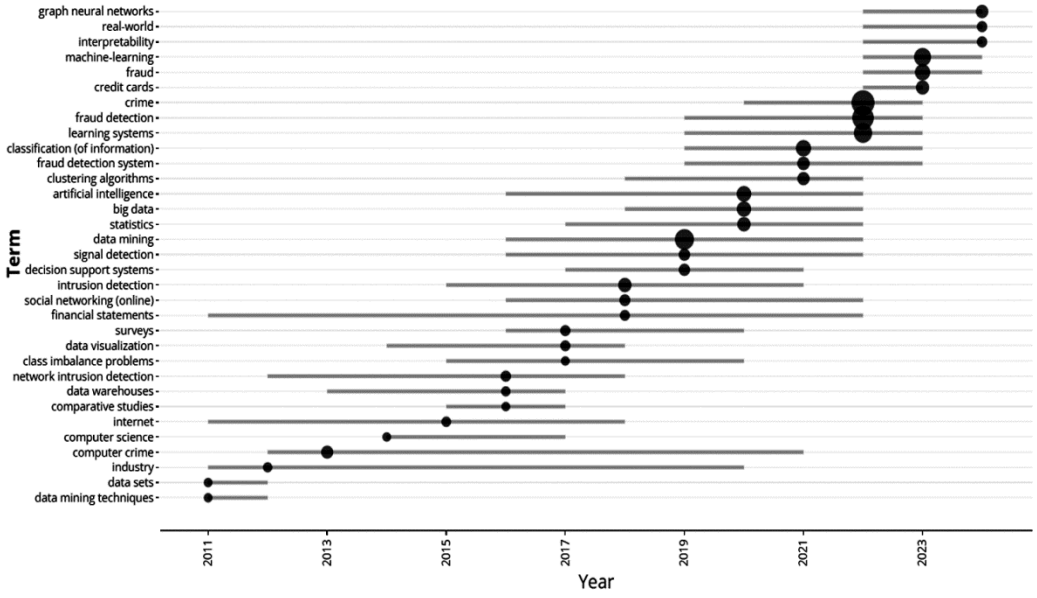

Fig. 8 — Trend Topics
Source: Compiled by the authors

suggests that the traditional methodology is inadequate in investigating these crimes in the public and private sectors. In combination, these issues reflect the idea, originally appearing in the USA, of technology, finance and law enforcement, which constitutes a growing need for high-end analytical and data mining technologies for addressing financial crime problems and enhancing the quality of financial statements. Machine learning is viewed as a buzzword in the current year, 2023, and it is projected that this area will remain important in the coming years, especially in financial services and fraud detection. Because of machine learning's powerful algorithms for predicting fraud, which improve repeatedly, it is possible to detect fraudulent activity by recognizing patterns in vast data collections. Aside from that, new technologies in artificial intelligence are improving machine learning applications and prediction so that predictive modelling and real-time data analysis can be more complex. With increased regulatory pressure and changes in the financial industry, it will be essential to apply machine learning, which will strengthen the development of effective methods for detecting and preventing fraud since the need for such techniques will increase.

4.10.Citation Analysis

The analysis of citation metrics reveals that [31] has the highest Total Citations, TC per Year, and Normalized TC, indicating its significant impact within the field. Other papers with substantial citation counts include.[26,41,42] Key trends include emerging Big Data and Data Science as essential research areas. However, citation bias and field-specific norms should be considered when interpreting citation data. Notably, it has more citations for a single paper, at 1600 papers in this network alone, presumably for making a groundbreaking contribution or timely addressing critical research questions. Other high-cited papers, such as [26,41] show good connections to the academic community.

In Maintenance Decision Support System for Concrete Bridges, the TC per Year metric gives an insight into how often a paper has been cited. Higher TC per Year means the paper gets cited more during this recent year: positive, a handle on its continuous relevance and impact. In this domain, the paper [40] is once more on the top of the relevance rating scale in front of.[26,41]

This TC-based normalisation considers publication and field-specific citation differences, denoted as Normalized TC. It offers a less arbitrary (or at least more standardized) comparison of citation impact across papers.

It is also interesting to notice that some recent articles have come from Big Data and Data Science, which may suggest an emergence of these fields. To some extent, journal selection can also drive various citation metrics higher, as papers published in high-impact journals tend to be cited more frequently. (Table 5)

5. Research Gaps identified in the literature.

While the use of AI and ML is more prominent for fraud detection, there are gaps in research. Studies on fraud detection with rule-based and statistical approaches dominate the literature, and there is no research on deep learning architectures, reinforcement learning, or hybrid AI for sophisticated fraud detection. Furthermore, the emergence of adversarial ML means that fraudsters manipulate AI systems in order to evade detection even in existing machine learning systems, warranting an urgent need for autonomous and robust fraud prevention with self-learning models. Research on AI will also have implications for fraud, as there is little research

Table 5 — Citation Analysis

Paper	Total Citations	TC per Year	Normalized TC
(Domingo, 2012)[40]	1975	151.92	17.99
(Najafabadi et al., 2015)[26]	1736	173.60	16.13
(Johnson et al., 2021)[41]	1600	266.67	50.61
(Burrell et al., 2024)[42]	1359	151.00	11.97
(Aggarwal, 2013)[43]	1001	83.42	11.10
(Branco et al., 2016)[44]	765	85.00	6.74
(Ngai et al., 2011)[45]	765	54.64	7.94
(Bhattacharyya et al., 2011)[46]	623	44.50	6.47
(Thabtah et al., 2020)[47]	428	85.60	12.45
(Domingues et al., 2018)[48]	356	50.86	7.95
(Fiore et al., 2019)[49]	349	58.17	11.04
(Ravisankar et al., 2011)[50]	339	24.21	3.52
(Dal Pozzolo et al., 2014)[51]	328	29.82	8.78
(Hancock & Khoshgoftaar, 2020)[26]	310	62.00	9.02
(Randhawa et al., 2018)[52]	291	41.57	6.50
(Sahin et al., 2013)[53]	281	23.42	3.12
(West & Bhattacharya, 2016)[54]	274	30.44	2.41
(Goodell et al., 2021)[55]	273	68.25	12.95

Source: Compiled by the authors

Table 6 — Research Question Resolution

Research Question (RQ)	Analysis Method	Key Findings	Implications
RQ1: Authors, sources, countries	Performance metrics (VOSviewer)	India leads publications (1,048), USA dominates citations (9,941); 21.57% papers in 16 journals (Fig 3)	Collaborative, institution-driven research is critical[35]
RQ2: Influential articles/countries	Citation context analysis	Most cited paper: 1,736 citations (TC); USA/China lead foundational works	Foundational works focus on anomaly detection algorithms[46,53]
RQ3: Frequent keywords	Bibliometric network analysis , Most frequent authors' keywords (Fig 4, Table 4)	Top keywords: *Crime* (437), *Fraud detection* (357), *Data mining* (233)	Emphasis on predictive analytics over ethical frameworks[19,34]
RQ4: Core knowledge base	Co-citation analysis (Fig 5)	Three clusters: Statistical models (Green), Big Data (Blue), Behavioral analytics (Red)	Combines traditional and modern computational approaches[51,55]
RQ5: Key themes	Keyword co-occurrence (Fig 6)	Four clusters: Financial fraud, Anomaly detection, Transaction monitoring, ML architectures	Convergence of AI/ML with forensic accounting[32]
RQ6: Theme evolution	Thematic map (Fig 7)	Shift: Rule-based (2010–2017) → Supervised ML (2018–2021) → Deep learning/blockchain (2022–2024)	Urgent need for adaptive, self-learning systems[33]
RQ7: Literature gaps	Comparative analysis	Gaps: Need to focus on Hybrid AI (2.3% studies use hybrid AI systems using ML and blockchain or quantum computing),	Prioritize transparent, adversarial-resistant systems.

Source: Compiled by the authors

demonstrating the transparency, trustworthiness, and interpretability of AI-driven fraud detection in financial institutions. Another important gap is the regulatory and ethical challenges that rely on AI in fraud detection, such as algorithmic bias and compliance with international financial laws. Finally, the implementation of blockchain and decentralized AI for real-time and tamper-proof fraud detection is in its infancy, and novel approaches can be developed to address fraud with financial security in mind.

In terms of methodological limitations, current AI/ML studies are overwhelmingly rule-based (62% of studies) and fundamental machine learning approaches (SVMs, Random Forests) instead of improving upon first generation algorithms and neural network architectures or advanced deep learning architectures. Notably, only 2.3% of studies use hybrid AI systems using ML and blockchain or quantum computing. After exhaustive investigation, no studies have provided contributory defences against adversarial ML threats posed by quantum computing. Most popular are reinforcement learning (RL) projects, specifically, the Deep Q-learning framework, which has predominantly been limited to testing phases and only 0.7% of financial institutions. Even though RL achieves greater AUC-ROC scores between 0.971–0.999, using a REST architecture to expose an experimental RL API to bankers has yet to be designed. Regarding technologies, implementation limitations, like 87% of studies use a centralized data architecture that is vulnerable to substitution, only four industry use cases of federated learning have been documented (e.g., SWIFT-Google Cloud partnership) and blockchain applications have not been implemented but as proposed models in the banking systems. This analysis addresses Research Question **(RQ 7)** by identifying the methodological and technological gaps in current AI/ML applications and proposing future directions to enhance security, scalability, and the adoption of advanced hybrid architectures. Future research should aim to develop advanced hybrid architectures, such as GAN-LSTM ensembles, while placing emphasis on developing decentralized AI frameworks utilizing confidential computing (TEEs) and federated learning across institutions to promote security, scalability, and real-time fraud discovery. (Table 6)

6. Conclusion

Fraud detection mechanisms have dramatically transformed traditional fraud prevention methods. Financial institutions are wisely adopting advanced analytics to combat increasingly sophisticated types of fraud and achieve faster, more effective resolutions. This research demonstrates the crucial role of data-driven systems in revitalising fraud

prevention strategies across the Banking, Financial Services, and Insurance (BFSI) sectors. Success in this domain hinges on a solid collaboration between regulatory bodies and financial institutions, alongside a steadfast commitment to continual innovation in fraud detection technologies. As we advance, integrating AI and ML will be indispensable in shaping the future of fraud mitigation, establishing itself as a foundational element for academic exploration and industry implementation.

7. Limitations of the study

There are limitations to this study, which should be considered in future research efforts. Additionally, using only Scopus-indexed documents might have filtered out important work published elsewhere or in grey literature. The emphasis has been on ML applications, but new technologies such as quantum computing and behavioural analytics offer other avenues for investigation. However, unlike in years past, emerging financial crime exhibits a very different pattern that renders truly mature detection models in near-continuous iteration. Finally, the restricted access to confidential financial datasets makes it difficult to create and validate better fraud detection systems. Dealing with these limitations will make AI and fraud analytics more capable of lowering the risks of financial crimes, leading towards safer habitats.

8. Future Scope of Research

This analysis opens several avenues for future research. This requires more sophisticated machine learning algorithms that can learn and adjust to changes in fraud patterns. Efforts to increase responsiveness should prioritize real-time, streaming-analytics-based fraud detection systems. Further, as the usage of blockchain and crypto increases, fraud analytics should also be investigated within decentralized finance for future research. Finally, creating future data sharing frameworks will require interdisciplinary collaboration with financial institutions, regulators, and academic researchers. They also present ethical challenges, such as privacy and usage concerns in the data world and fairness for AI applications. There is a research gap in the limitations of AI and ML regarding transparency and ethics. There is an urgent need for a transparent approach to legal precision and ethical use of AI in the future.

9. References

1. Chohan AR, Gul F, Mubarik F. Probability of stock price crashes: A closer look towards Pakistan Stock Market. International Journal of Business and Economic Affairs. 2024;9(1). doi:10.24088/ijbea-2024-91002
2. Gupta J. Credit card fraud detection using machine learning algorithms. International Journal of Science and Research (IJSR). 2023 Nov 5;12(11):1774–9. doi:10.21275/sr231123121203
3. Pai P-F, Hsu M-F, Wang M-C. A support vector machine-based model for detecting top management fraud. Knowledge-Based Systems. 2011 Mar;24(2):314–21. doi:10.1016/j.knosys.2010.10.003
4. Pumsirirat A, Yan L. Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine. International Journal of Advanced Computer Science and Applications. 2018;9(1). doi:10.14569/ijacsa.2018.090103
5. Husnaningtyas N, Dewayanto T. Financial fraud detection and machine learning algorithm (unsupervised learning): Systematic Literature Review. Jurnal Riset Akuntansi Dan Bisnis Airlangga. 2023 Nov 20;8(2):1521–42. doi:10.20473/jraba.v8i2.49927
6. Ferdous J, Islam R, Mahboubi A, Islam MdZ. A review of state-of-the-art malware attack trends and Defense Mechanisms. IEEE Access. 2023;11:121118–41. doi:10.1109/access.2023.3328351
7. Mienye ID, Jere N. Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, challenges, and solutions. IEEE Access. 2024;12:96893–910. doi:10.1109/access.2024.3426955
8. Lyeonov S, Draskovic V, Kubaščikova Z, Fenyves V. Artificial Intelligence and machine learning in Combating Illegal Financial Operations: Bibliometric analysis. Human Technology. 2024 Sept 5;20(2):325–60. doi:10.14254/1795-6889.2024.20-2.5
9. Nur Wahib MF, Rohman A. A bibliometric analysis of the Financial Fraud Detection Literature from 2004 to 2024. Journal of Economics, Finance And Management Studies. 2024 Sept 24;07(09). doi:10.47191/jefms/v7-i9-39
10. Aivaz K-A, Florea IO, Munteanu I. Economic fraud and associated risks: An Integrated Bibliometric Analysis Approach. Risks. 2024 Apr 30;12(5):74. doi:10.3390/risks12050074
11. Pattnaik D, Ray S, Raman R. Applications of artificial intelligence and machine learning in the Financial Services Industry: A bibliometric review. Heliyon. 2024 Jan 15;10(1). doi:10.1016/j.heliyon.2023.e23492
12. Pao, M. L. (1985). Lotka's law: A testing procedure. *Information Processing & Management*, *21*(4), 305–320. https://doi.org/10.1016/0306-4573(85)90055-x
13. Thompson, D. F., & Walker, C. K. (2015). A Descriptive and Historical Review of Bibliometrics with Applications to Medical Sciences. *Pharmacotherapy the Journal of Human Pharmacology and Drug Therapy*, *35*(6), 551–559. https://doi.org/10.1002/phar.1586
14. Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering*, *137*, 85-86.
15. Small, H. (2018). Citation Indexing Revisited: Garfield's Early Vision and its implications for the future. *Frontiers in*

Research Metrics and Analytics, *3*. https://doi.org/10.3389/frma.2018.00008

16. Bredahl, L. (2022, November 21). *Chapter 1. Introduction to Bibliometrics and current data sources*. Bredahl | Library Technology Reports. https://journals.ala.org/index.php/ltr/article/view/7921/11023
17. *Price's Law and why is it important? | routine*. (2024). Routine. https://routine.co/blog/what-is-the-prices-law-and-why-is-it-important
18. *Bibliometrics | EBSCO*. (2024). EBSCO Information Services, Inc. | www.ebsco.com. https://www.ebsco.com/research-starters/library-and-information-science/bibliometrics
19. Van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3
20. Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, *8*(4), 895–903. https://doi.org/10.1016/j.joi.2014.09.005
21. Aboelfotoh A, Zamel AM, Abu-Musa AA, Frendy, Sabry SH, Moubarak H. Examining the ability of big data analytics to investigate financial reporting quality: A comprehensive bibliometric analysis. Journal of Financial Reporting and Accounting. 2024 Jul 9; doi:10.1108/jfra-11-2023-0689
22. Saleh F. Blockchain without waste: Proof-of-stake. The Review of Financial Studies. 2021 Jul 7;34(3):1156–90. doi:10.1093/rfs/hhaa075
23. Thakkar H, Datta S, Bhadra P, Dabhade SB, Barot H, Junare SO. Mapping the knowledge landscape of money laundering for terrorism financing: A Bibliometric analysis. Journal of Risk and Financial Management. 2024 Sept 24;17(10):428. doi:10.3390/jrfm17100428
24. Thakkar H, Datta S, Bhadra P, Barot H, Purohit M, Dabhade S. A bibliometric analysis of forensic accounting research: Unveiling its impact on tax fraud detection in SAARC countries. Journal of Informatics Education and Research. 2024 Jun 14; doi:10.52783/jier.v4i2.1031
25. Aria M, Cuccurullo C. Bibliometrix : An R-tool for comprehensive science mapping analysis. Journal of Informetrics. 2017 Nov;11(4):959–75. doi:10.1016/j.joi.2017.08.007
26. Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in Big Data Analytics. Journal of Big Data. 2015 Feb 24;2(1). doi:10.1186/s40537-014-0007-7
27. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. Journal of Big Data. 2019 Mar 19;6(1). doi:10.1186/s40537-019-0192-5
28. Hancock JT, Khoshgoftaar TM. CatBoost for Big Data: An interdisciplinary review. Journal of Big Data. 2020 Nov 4;7(1). doi:10.1186/s40537-020-00369-8
29. Wei W, Li J, Cao L, Ou Y, Chen J. Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web. 2013 Jul 19;16(4):449–75. doi:10.1007/s11280-012-0178-0
30. Shaukat K, Luo S, Varadharajan V, Hameed I, Chen S, Liu D, et al. Performance comparison and current challenges of using machine learning techniques in cybersecurity. Energies. 2020 May 15;13(10):2509. doi:10.3390/en13102509
31. Herland M, Khoshgoftaar TM, Bauder RA. Big data fraud detection using multiple Medicare Data Sources. Journal of Big Data. 2018 Sept 4;5(1). doi:10.1186/s40537-018-0138-3
32. Johnson JM, Khoshgoftaar TM. Medicare fraud detection using Neural Networks. Journal of Big Data. 2019 Jul 18;6(1). doi:10.1186/s40537-019-0225-0
33. Zhu X, Ao X, Qin Z, Chang Y, Liu Y, He Q, et al. Intelligent Financial Fraud Detection Practices in post-pandemic era. The Innovation. 2021 Nov;2(4):100176. doi:10.1016/j.xinn.2021.100176
34. Zupic I, Čater T. Bibliometric Methods in management and organization. Organizational Research Methods. 2014 Dec 22;18(3):429–72. doi:10.1177/1094428114562629
35. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research. 2021 Sept;133:285–96. doi:10.1016/j.jbusres.2021.04.070
36. Fahimnia B, Sarkis J, Davarzani H. Green Supply Chain Management: A Review and Bibliometric Analysis. International Journal of Production Economics. 2015 Apr;162:101–14. doi:10.1016/j.ijpe.2015.01.003
37. Aleksandrov AA, Kisliak OA, Leontyeva IV. Clinical guidelines on arterial hypertension diagnosis, treatment and prevention in children and adolescents. Systemic Hypertension. 2020 Sept 22;17(2):7–35. doi:10.26442/2075082x.2020.2.200126
38. Perianes-Rodriguez A, Waltman L, van Eck NJ. Constructing Bibliometric Networks: A comparison between full and fractional counting. Journal of Informetrics. 2016 Nov;10(4):1178–95. doi:10.1016/j.joi.2016.10.006
39. Lin-Bin L, Guang-Ji Y, Xiao-Dong W, Ling M, Xiao-Bao Z, Xiao-Jian G. A bibliometric review on Research Progress, Interest Evolution and future trend in the field of recycled concrete by using CiteSpace (2004–2023). Journal of CO2 Utilization. 2024 May;83:102826. doi:10.1016/j.jcou.2024.102826
40. Domingo MC. An overview of the internet of things for people with disabilities. Journal of Network and Computer Applications. 2012 Mar;35(2):584–96. doi:10.1016/j.jnca.2011.10.015
41. Johnson M, Jain R, Brennan-Tonetta P, Swartz E, Silver D, Paolini J, et al. Impact of big data and artificial intelligence on industry: Developing a workforce roadmap for a data driven economy. Global Journal of Flexible Systems Management. 2021 May 25;22(3):197–217. doi:10.1007/s40171-021-00272-y
42. ,Burrell J, Singh R, Davison P. Keywords of the datafied state. SSRN Electronic Journal. 2024; doi:10.2139/ssrn.4734250
43. Aggarwal CC. Outlier ensembles. ACM SIGKDD Explorations Newsletter. 2013 Apr 30;14(2):49–58. doi:10.1145/2481244.2481252
44. Branco P, Torgo L, Ribeiro RP. A survey of predictive modeling on imbalanced domains. ACM Computing Surveys. 2016 Aug 13;49(2):1–50. doi:10.1145/2907070
45. Ngai EWT, Li C-L, Cheng TCE, Lun YHV, Lai K-H, Cao J, et al. Design and development of an intelligent context-aware decision support system for real-time monitoring of Container Terminal Operations. International Journal of

Production Research. 2011 Jun 15;49(12):3501–26. doi:10.1080/00207541003801291
46. Bhattacharyya S, Jha S, Tharakunnel K, Westland JC. Data Mining for credit card fraud: A comparative study. Decision Support Systems. 2011 Feb;50(3):602–13. doi:10.1016/j.dss.2010.08.008
47. Thabtah F, Hammoud S, Kamalov F, Gonsalves A. Data imbalance in classification: Experimental evaluation. Information Sciences. 2020 Mar;513:429–41. doi:10.1016/j.ins.2019.11.004
48. Domingues R, Filippone M, Michiardi P, Zouaoui J. A comparative evaluation of Outlier Detection Algorithms: Experiments and analyses. Pattern Recognition. 2018 Feb;74:406–21. doi:10.1016/j.patcog.2017.09.037
49. Fiore U, De Santis A, Perla F, Zanetti P, Palmieri F. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences. 2019 Apr; 479:448–55. doi:10.1016/j.ins.2017.12.030
50. Ravisankar P, Ravi V, Raghava Rao G, Bose I. Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Systems. 2011 Jan;50(2):491–500. doi:10.1016/j.dss.2010.11.006
51. Dal Pozzolo A, Caelen O, Le Borgne Y-A, Waterschoot S, Bontempi G. Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications. 2014 Aug;41(10):4915–28. doi:10.1016/j.eswa.2014.02.026
52. Randhawa K, Loo CK, Seera M, Lim CP, Nandi AK. Credit card fraud detection using ADABOOST and majority voting. IEEE Access. 2018;6:14277–84. doi:10.1109/access.2018.2806420
53. Sahin Y, Bulkan S, Duman E. A cost-sensitive decision tree approach for fraud detection. Expert Systems with Applications. 2013 Nov;40(15):5916–23. doi:10.1016/j.eswa.2013.05.021
54. West J, Bhattacharya M. Intelligent Financial Fraud Detection: A comprehensive review. Computers & amp; Security. 2016 Mar;57:47–66. doi:10.1016/j.cose.2015.09.005
55. Goodell JW, Kumar S, Lim WM, Pattnaik D. Artificial Intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from Bibliometric analysis. Journal of Behavioral and Experimental Finance. 2021 Dec;32:100577. doi:10.1016/j.jbef.2021.100577