

ROLE OF BIG DATA ANALYTICS IN FINANCIAL FRAUD DETECTION- A BIBLIOMETRIC ANALYSIS

Shivani Abrol¹ and Monika Gupta²

DOI:<https://doi.org/10.58426/cgi.v5.i1.2023.82-111>

Abstract

Using data analytics and machine learning to combat fraud is a strategy that many businesses have already considered. Fraud may be detected, investigated, and prevented with the aid of big data analytics and machine learning. The purpose of this research is to systematically review the 219 Scopus-indexed publications in context of data analytics in detecting financial crime during 1999 to 2022. The findings indicate that a significant portion of the literature focuses on the utilization of big data analytics, specifically machine learning and deep learning techniques, for the purpose of detecting credit fraud or financial statement fraud. Previous studies have primarily concentrated on the utilization of hybrid technology in the realm of financial fraud detection, thereby indicating its potential as a promising avenue for future research. This study highlights the prominent research gap existing for a predictive model that can issue a warning as soon as a vulnerability for fraudulent behavior is noted. Moreover, findings highlight the accentuated need for data-driven financial investment model and stock market anomalies in context of data analytics and text mining, along with key future research agenda.

Keywords: Financial Fraud Detection, Big Data, Big Data Analytics, Machine Learning, Deep Learning, Bibliometric Analysis

JEL Classification: G3, G320, C8, C55, D83

¹ Associate Professor, Department of Commerce, Zakir Husain Delhi College, University of Delhi
Email Id: Shivani.abrol@zh.du.ac.in

² Associate Professor, Department of Commerce, Motilal Nehru College, University of Delhi
Email Id: monikagupta@mln.du.ac.in

1. Introduction

Unstructured data is the norm in the modern day, and it typically features high levels of volume, velocity, variety, and variability. The rapid pace at which new data is being generated makes it challenging to analyze using conventional methods. Now that so much data is readily available, businesses in virtually every sector are concentrating on finding ways to leverage it to their advantage. Therefore, it is vital that we analyze this pile of data in order to find answers to several business issues. “Big data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes” (IBM). Pattern recognition, machine learning, predictive analytics, text analytics, deep learning, data modelling, data mining, statistics, and computational linguistics are some of the advanced analytics techniques that companies can use on their own or in conjunction with their existing enterprise data to uncover previously undiscovered insights. The Association of Certified Fraud Examiners (ACFE) report says that “the average company loses more than \$1.5 million to fraud” (Association of Certified Fraud Examiners, 2022). Beyond the obvious monetary losses, fraud can also negatively impact customer satisfaction, company image, operational problems, and more. We can no longer rely on tried-and-true techniques for detecting fraud. The methods used to manage fraud risk need to develop in parallel with the increasing sophistication of criminal attacks. Data analytics and machine learning aid in the identification and detection of fraud by reducing costs and increasing income, saving time and effort by automating fraud analysis, stopping fraud in its tracks, increasing the percentage of successful fraud detections by minimizing false alarms, and also helping to locate and correct flaws in systems or business processes. Data mining is used to find previously unknown connections and patterns in large datasets. The methods of data analysis focus mostly on the extraction of numerical and statistical features of the data. Using these methods, you can gain a deeper understanding of the processes underlying your data and make more informed decisions as a result. Capturing, storing, analyzing, and correctly visualizing massive amounts of diverse data is made possible by big data analytics. It is helpful to construct a predictive model that can issue a warning as soon as a vulnerability for fraudulent behaviour is identified.

2. Literature Review

Many academic studies have concentrated on the use of big data analytics to uncover fraudulent activity and create safeguards against it. The first study “Using data mining techniques in fiscal fraud detection” by (Bonchi et al., 1999) is the conference paper from Lecture Notes in Computer Science. This paper presents a case study that exemplifies how classification-based techniques can be utilized to aid in the process of developing audit procedures. Ngai et al., (2011) provides a comprehensive analysis of data mining methodologies used for detecting financial fraud, with a focus on their systematic and discernible application. This paper provides a taxonomy of financial fraud, encompassing four distinct categories: bank fraud, insurance fraud, securities and commodities fraud, and other forms of financial fraud that are closely associated. Furthermore, the document delineates six distinct categories of data mining methodologies, namely classification, regression, clustering, prediction, outlier detection, and visualization. The findings suggest that data mining techniques have been extensively employed in the detection of insurance fraud, with a growing emphasis on corporate fraud and credit card fraud in recent years. Additionally, various review papers have been published that study the literature available on the role of data analytics in decision-making or fraud detection. Specifically, Cockcroft & Russell, (2018) analysed 47 journals in the fields of information systems, accounting, and finance that were published between 2007 and 2016, providing a glimpse of academic research into big data in these areas and pointing to potential areas for future research in accounting and finance. Risk and security, data visualization, predictive analytics, data management, and data quality were recognized as having a dearth of research in the field of accounting and finance related to big data. The study also explored the use of big data for fraud protection. West & Bhattacharya, (2016) provides a thorough examination of the existing body of research on the detection of financial fraud, specifically focusing on data mining methods and computational intelligence (CI)-based techniques. This study analysed more than fifty scientific literature sources, primarily published between 2004 and 2014. The focus of the analysis was on empirical studies that specifically examined financial fraud detection based on computational intelligence (CI) methods. A research gap was identified in the existing literature, as none of the review articles examined the relationship between fraud types, CI-based detection algorithms, and their performance. Putra et al., (2022) did a comprehensive literature review on the topic of fraud prevention as a mediator between internal audit, risk

management, whistleblowing systems, and big data analytics to prevent financial crime. Pejić Bach et al., (2019) conducted a systematic literature review on text mining for big data analysis in the financial sector. The study concluded that information from a wide variety of semi-structured and unstructured data sources is now within reach, with the advancements in big data technologies. Recent studies and practical applications have shown that such data might be useful in decision-making. Since the focus of these publications is limited to a small subset of possible scenarios, research in the field of the role of big data analytics in fraud detection is fragmented and inconsistent. Thus, it is important to perform a comprehensive and all-encompassing assessment of the topic of the role of big data analytics in financial fraud detection in a global setting. Moreover, there is a need to perform a more objective and systematic assessment of the body of literature on this topic, as previous reviews have used a more subjective and traditional literature review approach.

3. Objective of the Study

This study presents a bibliometric literature review of the existing literature on the topic of the role of data analytics in financial fraud detection. A comprehensive and objective evaluation of the literature on the role of data analytics in financial fraud detection is required to contribute to the current field of research by providing a synthesis of existing literature, highlighting trends, and themes, identifying knowledge structure, and suggesting areas for future research. 219 papers related to the topic role of big data analytics in detecting financial crime were analyzed and evaluated through bibliometric analysis keeping in mind the following objectives:

1. To determine the most productive countries and scholars publishing on the topic of the role of big data analytics in financial fraud detection, and their collaboration.
2. To identify the most popular source of publication related to the topic of research.
3. To find the most influential publications on the role of big data analytics in financial fraud detection.
4. To identify the potential thrust area in the field of the role of big data analytics in financial fraud detection.
5. To comprehend the knowledge base or intellectual framework of the topic of the role of big data analytics in financial fraud detection.

The objective of the paper is to thoroughly examine the topic of the role of big data analytics in financial fraud detection to comprehend its historical development, current research trends, and prospective research directions. When a subject of study is still relatively new and developing, it is crucial to understand its historical development and its potential future directions. The central topics of different eras and how they changed through time, as well as how the complete body of study is classified at regular intervals can be easily understood with the help of bibliometric analysis. The study would also help in identifying the topics that have been examined extensively, those that have been studied less, and those that are just beginning to be investigated. This will aid in establishing a path for future study in the field of the role of data analytics in financial fraud detection. Furthermore, this study provides insights to corporates regarding the various big data techniques that can be utilized for the purpose of mitigating and identifying instances of financial fraud. This study provides a comprehensive review of potential future prospects of the data and statistical techniques.

4. Methodology

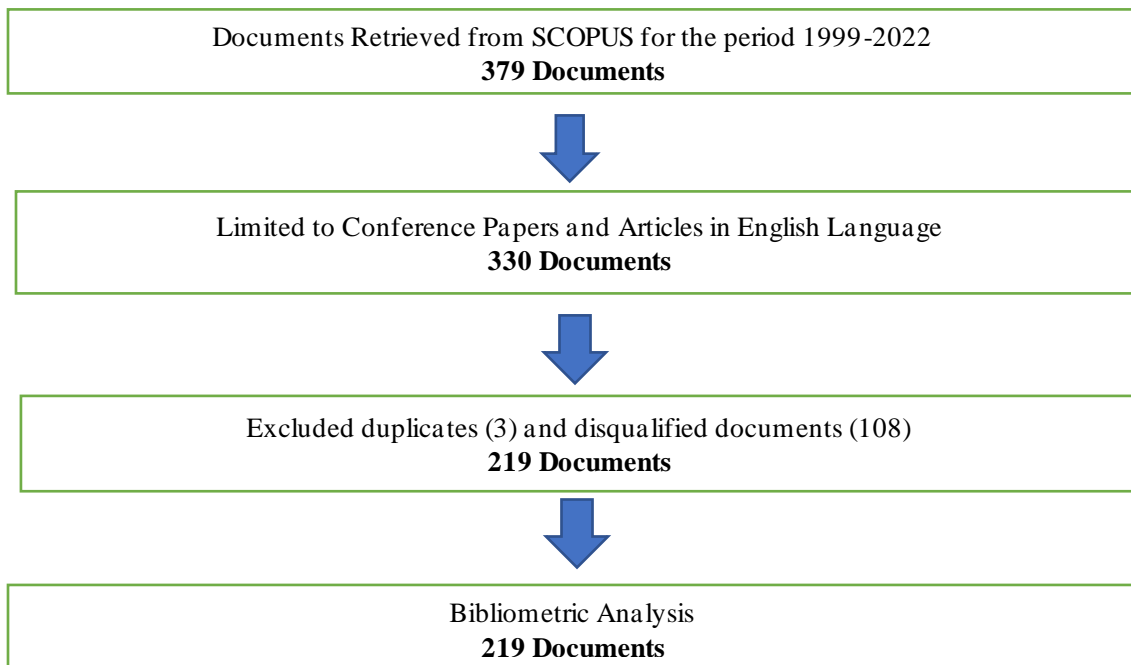
Different scholars have used many different approaches to the literature review including content analysis, meta-analysis, and bibliometric analysis. The goals of a review serve as the primary criteria for choosing an appropriate review approach. The paper takes a more comprehensive view of how current research (the role of data analytics in detecting financial crime) has developed through time and how different fields have contributed to that progress. For this reason, a bibliographic literature review is conducted. Bibliometrics is a quantitative method for reviewing literature that utilizes a variety of categorization and representation strategies to analyze the bibliographic data submitted by various researchers in a field in order to precisely represent and summaries all relevant writing (Pawar & Paluri, 2022).

Data Search and retrieval

The meta-data on the topic of the role of data analytics in financial fraud detection has been retrieved from SCOPUS on 21st November 2022. The SCOPUS is the largest database offering metadata statistics on peer-reviewed literature (Emich et al., 2020) and also Web of Science, contained largely duplicate entries. The search string (“Big Data” OR "Big Data Analytics" OR "Data Mining" OR "Data Processing" OR " Data Analytics" OR “Machine Learning” OR “Artificial Intelligence” OR “AI” OR “Deep Learning”) AND ("Financial Fraud" OR "Fiscal

Fraud" OR "Financial Corruption" OR " Financial Embezzlement" OR " Financial Crime")) in the title, abstract, keyword sections yielded 379 documents for the period 1999 to 2022. The conference papers and articles written in English were filtered and as result, only 330 documents were left. At the later stage, after screening the title, abstract, and document if needed, 108 documents were disqualified to be included in the list as they either not directly related to the topic under study or are review papers. We considered conference papers as they were discussing the new paradigm in this domain and excluding them would negate the possibility of doing an indepth study. As shown in Figure 1, after removing 3 duplicates, a list of 219 documents was finalized for conducting bibliometric analysis.

Figure 1: Publication Selection Process



Source: Author's Computation

5. Data Analysis

The retrieved documents were analyzed through the open software Biblioshiny (Bibliometrix R package) and VOSviewer. Both the techniques of bibliometric analysis namely performance analysis and science mapping were used for the analysis, adding keyword analysis and co citation analysis will reveal evolution of this theme and future scope of research. The performance analysis examines the contribution of research constituents e.g. authors, institutions, countries, and publication sources, on the other hand, science mapping explains relationships between different

research constituents. Publications and citations are two of the most often used metrics, with the former serving as a measure of productivity and the latter for relevance and influence. The performance of research components can also be measured by using alternative metrics, such as the citation per publication or the h-index, which takes into account both citations and publications. Using science mapping techniques, researchers investigate the conceptual and organizational relationships between research constituents. Science mapping uses methods such as citation analysis, co-citation analysis, bibliographic coupling, co-word analysis, and co-authorship analysis. When paired with network analysis, such methods effectively display the bibliometric and intellectual structure of the study area in question (Donthu et al., 2021).

6. Findings and Discussions

This section explains the results of a bibliometric analysis conducted on 219 publications retrieved for the study. Table 1 provides the main information on the meta-data collected.

The publication span of the meta-data ranges from 1999 to 2022. The meta-data consists of 93 articles and 126 conference papers collected from 151 sources. It is evident from Table 1 that the first publication was a conference paper in the year 1999 with the title “Using data mining techniques in fiscal fraud detection”. The annual growth rate of publication is 18 per cent. The average citation score per document is 12.41. As far as the authors are concerned, there are 616 authors, and 20 authors publish 20 single-author documents. The average co-author per document is 3 and international co-authorship is approximately 16 per cent.

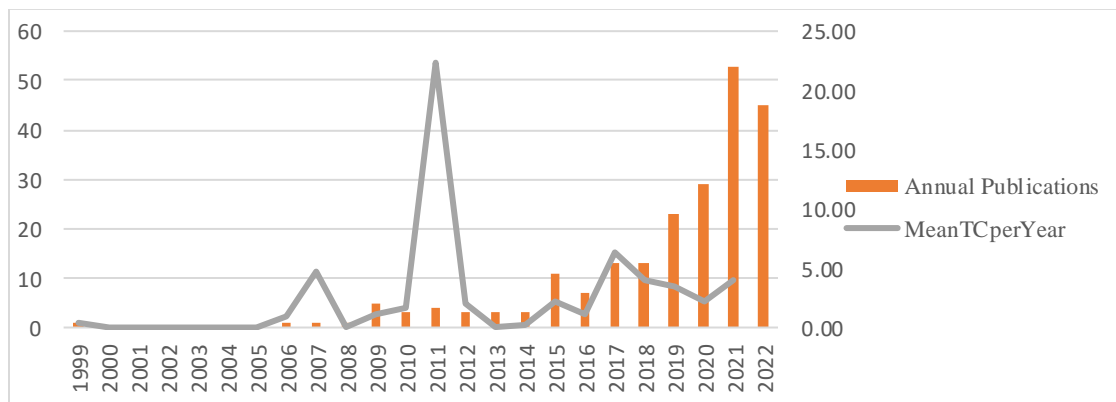
Table 1: Main Information of Meta-data

Description	Results
Timespan	1999:2022
Sources (Journals, Books, etc)	151
Documents	219
DOCUMENT TYPES	
Article	93
Conference Paper	126
Annual Growth Rate per cent	18
Document Average Age	3.24
Average citations per doc	12.41
References	6297

DOCUMENT CONTENTS	
Keywords Plus	1178
Author's Keywords	562
AUTHORS	
Authors	616
Authors of single-authored docs	20
AUTHORS COLLABORATION	
Single-authored docs	20
Co-Authors per Doc	3.23
International co-authorships per cent	15.98

Source: Retrieved from Biblioshiny R package

Figure 2: Pareto Chart for Publications and Average total citations per year



Source: Self-compiled from Biblioshiny R package

It is evident from Figure 2 that after a gap of six years following the first publication in the year 1999, there was only a single publication in the years 2006, 2007, and 2008. The number of publications stayed below ten until 2014, and thereafter under twenty until 2018. It is only after 2019, the rate of publication accelerated. The reason may be the growing popularity and application of big data analytics in accounting and auditing. The maximum number of 53 publications occurred in 2021. This analysis is combined with keyword analysis or evolution map, to trace the development of research in this domain, highlighting which constructs were important in early stages and which are relevant at present.

Mean Total Citations per year (MeanTCPerYear) is the highest for the year 2011 as shown in Figure 2. This indicates that documents published in 2011 have earned a greater number of citations. Three of the four papers published in the year 2011 contributed more to the increase in

the number of citations per year. The authors of these documents are Ngai et al., 2011, Ravisankar et al., 2011, and Zhou & Kapoor, 2011 with total citations of 604, 268, and 100 respectively. This signifies the research conducted by these three authors establishes a foundation for future research in the field being investigated.

Countries and their collaborations

Table 2 displays the leading nations together with their overall number of publications and worldwide citations. In terms of the number of publications, China ranked top (n=174), followed by India (n=116) and the United States of America (n=58). Bangladesh, France, and Indonesia each have twelve publications to their credit. Countries like Poland and Saudi Arabia have published less than 10 as shown in Table 2.

Table 2: Top Countries

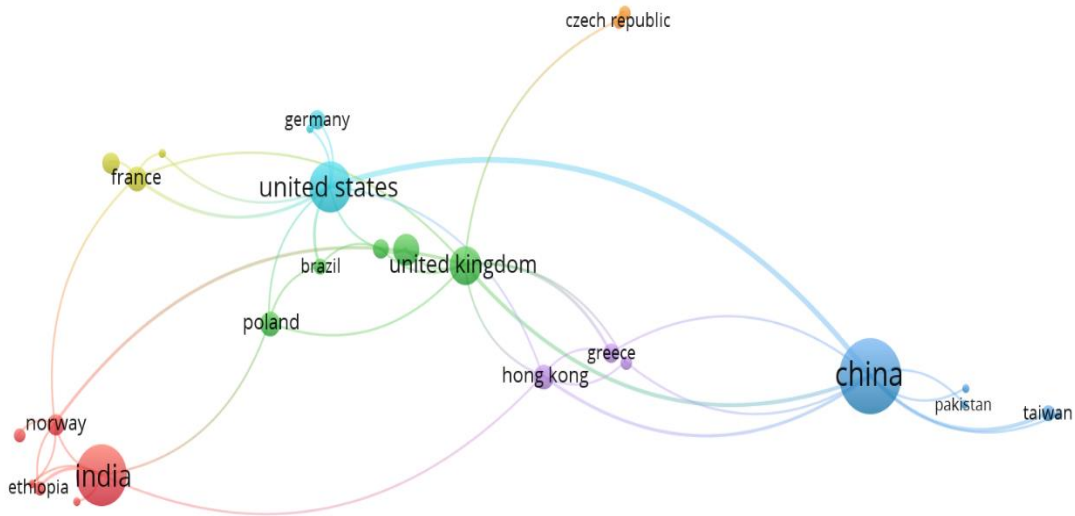
Order	Country	Total Publication	Per cent	Total Citation	Per cent
1	China	174	26.32	758 (#1)	39.81
2	India	116	17.55	442 (#2)	23.21
3	US	58	8.77	137 (#3)	7.2
4	UK	29	4.39	64 (#7)	3.36
5	Australia	22	3.33	135 (#4)	7.1
6	Italy	22	3.33	11	0.5
7	South Africa	21	3.18	3	0.1
8	Germany	17	2.57	1	0
9	Morocco	15	2.27	43 (#9)	2.25
10	Bangladesh	12	1.82	0	0
11	France	12	1.82	100 (#6)	5.25
12	Indonesia	12	1.82	0	0
13	Spain	10	1.51	1	0
14	Poland	9	1.36	46 (#8)	2.41
15	Saudi Arabia	7	1.06	14 (#10)	0.7
16	Czech Republic	2	0.3	114 (#5)	5.98

Source: Self-compiled from Biblioshiny R package

Whereas in terms of global citations, China topped the list (n=758), followed by India (n=442) and the United States (n=137). In comparison to the United States, Australia has about the same

number of citations for around half as many publications. Bangladesh and Indonesia were not cited at all.

Figure 3: Countries' Collaborations Network Diagram



Source: Research Output

The important point of discussion is that the Czech Republic has only two publications in credit, but bang 5th position in terms of global citations with n=114.

Total of 67 countries published on the topic of the role of big data analytics in financial fraud detection. 27 countries collaborated on the aforesaid topic as shown in Figure 3. The countries are divided into 7 clusters. A country is represented by a node, and the size of the node corresponds to the total number of publications from that country. The collaboration between countries is represented by the lines connecting the nodes, with the thickness of the lines indicating the depth of the collaboration. The biggest cluster consists of 6 countries-Ethiopia, Ghana, India, Kuwait, Nigeria, and Norway. With 38 publications, India also has collaborative relationships with six countries and a total link strength of seven. The second cluster includes Australia, Brazil, Canada Poland, and United Kingdom at the centre. The number of collaborations with the United Kingdom is ten for 15 publications with an overall link strength is fourteen. In cluster three, there are 5 countries, namely, China, Japan, Pakistan, Philippines, and Taiwan. China collaborated for 58 documents with nine nations, including the United States, the United Kingdom, Hong Kong, Pakistan, Greece, Taiwan, Serbia, and the Philippines. Cluster 5 (Greece, Hongkong, Serbia) and

Cluster 6 (Germany, Thailand, and the US) include three nations each. The United States worked on 26 documents with nine other nations. Cluster 7 has Portugal and the Czech Republic. As depicted in Figure 3, the Czech Republic worked alongside the United Kingdom and Portugal only. A limited number of nations participated in the collaborative efforts pertaining to the subject of investigation. With the exception of China, India, the United Kingdom, Australia, and the United States, the prevalence of cross-border collaborations is notably limited. India has engaged in collaborative efforts for a total of 38 documents; however, it is noteworthy that none of these collaborations involve the countries of the United States, the United Kingdom, or Australia.

Publication Resources

Table 3 displays the top-producing publication sources out of a total of 151 sources of publications.

Table 3: Most Productive Sources

S No.	Sources	Publishing House/Country	No. of Publications	Total Citations	Year of First Publication	Journal/Conference Proceedings	Current Cite Score
1	ACM International Conference Proceeding Series	Association for Computing Machinery/US	14	17	2015	Conference Proceedings	1
2	Advances in Intelligent Systems and Computing	Springer/US	9	35	2017	Conference Proceedings	Discontinued from Scopus
3	Procedia Computer Science	Elsevier/UK	9	174	2015	Journal	3.6
4	Lecture Notes in Computer Science	Springer/US	7	33	1999	Conference Proceedings	2.1
5	IEEE Access	Institute of Electrical and Electronics	6	19	2019	Journal	6.7

		Engineers Inc./US					
6	Communications in Computer and Information Science	Springer/US	4	4	2018	Conference Proceedings	0.9
7	Decision Support Systems	Elsevier/UK	4	982	2011	Journal	11.3
8	CEUR Workshop Proceedings	RWTH Aachen University/ Germany	3	2	2014	Conference Proceedings	1.1
9	Expert Systems with Applications	Elsevier/UK	3	37	2021	Journal	12.2
10	Journal of Financial Crime	Emerald/UK	3	22	2019	Journal	1.5
11	Lecture Notes in Networks and Systems	Springer/US	3	11	2021	Conference Proceedings	0.7

Source: Self-compiled from Biblioshiny R package

The top publication sources as shown in Table 3 published 65 documents in total which constitutes approximately 30 per cent of total publications. The most valuable source is the ACM International Conference Proceeding Series, which contains 14 documents on the issue of big data analytics in detecting financial fraud. Having published 9 documents each, Advances in Intelligent Systems and Computing, and Procedia Computer Science Journal tie for second place. Additionally, 8 other sources publish more than three articles each. All the publication houses belong to either US or UK. The oldest source of publication is Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 1999. However, in terms of citations Decision Support Journal received the highest total citation (n=982). Advances in Intelligent Systems and Computing (Conference Proceedings), one of eleven sources, was removed from Scopus in 2022. Expert Systems with Applications and Decision Support Systems are the two Journals with the highest current citation score of 12.2 and 11.3. The data indicates that conferences play a significant role in this study field because conference proceedings make

up half of the top publication resources. Increasing the number of conferences in this field of study is a wise step because they enable constructive interactions between academics and the development of new collaborative partnerships.

Authors Productivity

The list of researchers who have produced three or more papers is shown in Table 4.

Table 4: Most Productive Authors

S No.	Authors	Institution/Country	No. of Publications	h_index	g_index	m_index	Total Citations	Year of First Publication
1	Maumita Bhattacharya	Charles Sturt University/ Australia	5	4	5	0.5	51	2015
2	Jarrod West	Charles Sturt University/ Australia	5	4	5	0.5	51	2015
3	Yijun Chen	Guangdong University of Foreign Studies/China	4	2	4	0.167	607	2011
4	Ashish Jain	Indian Institute of Management Lucknow/India	4	2	4	0.5	26	2019
5	Linli Wang	The Hunan University of Finance and Economics/China	4	2	4	0.4	38	2018
6	Zhen Wang	Ant Financial Services Group/ China	4	2	4	0.5	85	2019
7	Sharat Sourabh Akhoury	Council for Scientific and Industrial Research/ South Africa	3	2	3	0.25	35	2015
8	Abhishek Gupta	Bharathidasan Institute of Management/India	3	2	2	1	5	2021
9	Shouxi Li	Shanghai University/China	3	1	2	0.5	8	2021
10	Mingxi Liu	Chinese Academy of Sciences/China	3	1	3	0.333	12	2020
11	Stephen Obakeng Moepya	University of Johannesburg/ South Africa	3	2	3	0.25	35	2015

12	Fulufhelo V Nelwamondo	University of Johannesburg/ South Africa	3	2	3	0.25	35	2015
13	Jiahao Zhang	Shanghai University/ China	3	1	3	0.2	30	2018
14	Zhiping Zhang	Chongqing College of Finance and Economics/China	3	1	1	0.5	2	2021

Source: Self-compiled from Biblioshiny R package

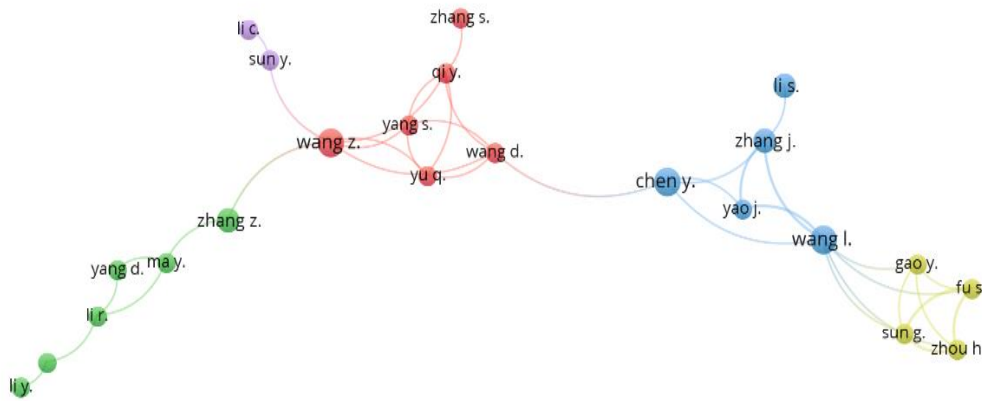
The top productive scholars include Maumita Bhattacharya and Jarrod West, both from Charles Sturt University/ Australia with 5 publications each. Seven scholars are from China. Yijun Chen, Linli Wang, and Zhen Wang with 4 papers each; and Shouxi Li, Mingxi Liu, Jiahao Zhang, and Zhiping Zhang with 3 papers each in their credit. From India, there are two authors namely Ashish Jain (n=4) and Abhishek Gupta (n=3). Rest two, Sharat Sourabh Akhoury (n=3) and Fulufhelo V Nelwamondo (n=3) are from South Africa. Yijun Chen, Guangdong University of Foreign Studies, China received the highest number of citations (n=607) while Maumita Bhattacharya and Jarrod West have the highest h-index. Maumita Bhattacharya and Jarrod West both collaborated on the same studies and published 3 papers in 2015 and 2 papers in 2016. However, Yijun Chen published his first paper in 2011 and then recently in 2020 (n=2) and 2021 (n=3). Linli Wang started in the year 2018 and since then he is persistently publishing with one article each in 2020, 2021, and 2022. Mingxi Liu also published consistently for 3 years (2020, 2021 and 2022). Abhishek Gupta, Shouxi Li, and Zhiping Zhang are the recent contributors.

The authors' collaboration network with at least two publications has shown in Figure 4. Out of 616 authors, only 66 authors meet the criterion. However, out of these 66 authors, only 23 authors are largely connected as shown in Figure 4.

Scholars are represented by individual nodes, with node size according to the scholar's publication output; the thickness of connections between scholars' nodes symbolizes the quality of their working relationship. Different hues represent distinct clusters of scholars who work closely together. Five clusters can be identified by their distinct colour. There are two big clusters with 6 items each. Four documents were co-authored by Linli Wang and seven others, with a total of nine links. Zhen Wang has also been a co-author on 4 documents with a total of 6 other people. The

potential use of big data analytics to combat financial fraud has gained more attention in recent years only. As a result, there is less teamwork among authors.

Figure 4: Authors Collaboration Network



Source: Research Output

Most Cited Document

The top ten cited documents along with other relevant information are depicted in Table 5. Six articles and 4 conference papers with a total of 1656 citations are on the list as shown in Table 5. These documents have been published from 2007-2017. Six documents receive more than 100 citations. The article, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of the literature” (Ngai et al., 2011) in Decision Support Systems Journal (2011) received 604 citations, the highest number. “This paper presents the inaugural, methodical, discernible, and all-encompassing examination of data mining methodologies employed in the realm of financial fraud detection. The paper presents a classification of four distinct categories of financial fraud, namely bank fraud, insurance fraud, securities and commodities fraud, and other related financial fraud. Additionally, it outlines six classes of data mining techniques, including classification, regression, clustering, prediction, outlier detection, and visualization. The results indicate that data mining techniques have been widely utilized in the identification of insurance fraud, with additional focus on corporate fraud and credit card fraud in recent times. The study's findings suggest that future investigations into the utilization of data mining methods for addressing financial fraud detection issues should

incorporate considerations of cost sensitivity” (Ngai et al., 2011). The affiliation of Ngai E.W.T. is Hong Kong Polytechnic University, Hongkong. Among the conference papers, Credit card fraud detection using machine learning techniques: A comparative analysis (Awoyemi et al., 2017) from Proceedings of the IEEE International Conference on Computing, Networking, and Informatics, ICCNI 2017, receives a total of 191 citations. The oldest paper in the list is a conference paper on the topic “A Review of Data Mining-Based Financial Fraud Detection Research” (Yue et al., 2007) in International Conference on Wireless Communications, Networking, and Mobile Computing, WiCOM 2007 got 71 citations. However, three articles in the list belong to the same year of publication, 2011 with a total of 972 citations, and surprisingly, these three documents are published by the same journal, Decision Support Systems. This backs up the findings of Figure 2, which show that 2011-published publications have received more citations, and the findings of Table 3, which rank Decision Support Systems Journal as the most influential source in terms of highest citation. From the top 10 list, two documents belong to India and China each, the rest from other countries including the US, Australia, Czech Republic, France, Hongkong, and Nigeria. The identification and prevention of credit card fraud in online transactions have been significantly aided by the utilization of data mining, as evidenced by the high citation rates of relevant studies. The effectiveness of fraud detection in credit card transactions is heavily impacted by the sampling methodology employed on the dataset, the selection of variables, and the detection technique(s) utilized. Additionally, the research places significant emphasis on the utilization of hybrid methodologies in the detection and identification of financial fraud. “Hybrid data mining techniques that incorporate multiple classifiers can be utilized for the purpose of detecting financial fraud. Furthermore, the integration of text mining algorithms for sentiment analysis of textual descriptions in financial statements can be merged with data mining algorithms to assess the financial components within said statements” (Ravisankar et al., 2011). Hajek & Henriques, 2017 also supports that “both financial statements and textual information in annual reports can be effectively employed for the purpose of identifying non-fraudulent firms. However, it is essential to consider non-annual report data in order to effectively detect and recognize fraudulent companies. The discovery holds significant implications for the process of variable selection in the development of intelligent systems designed to detect financial statement fraud”.

Table 5: Most Cited Documents

Title	Authors	Source	Document Type	First Author Institution/Country	Year Of Publication	Total Citations	Yearly average citations
The application of data mining techniques in financial fraud detection: A classification framework and an academic review of the literature (Ngai et al., 2011)	E.W.T. Ngai; Yong Hu ; Y.H. Wong; Yijun Chen ; Xin Sun	Decision Support Systems	Article	Hong Kong Polytechnic University/ Hongkong	2011	604	50.33 (#1)
Detection of financial statement fraud and feature selection using data mining techniques (Ravisankar et al., 2011)	P. Ravisankar; V.Ravi, G. Raghava Rao; I.Bose	Decision Support Systems	Article	Institute for Development, Research in Banking/India	2011	268	22.33 (#3)
Credit card fraud detection using machine learning techniques: A comparative analysis (Awoyemi et al., 2017a)	John O. Awoyemi ; Adebayo O. Adetunmbi ; Samuel A. Oluwadare	Proceedings of the IEEE International Conference on Computing, Networking, and Informatics, ICCNI 2017	Conference Paper	The Federal University of Technology Akure/ Nigeria	2017	191	31.83 (#2)
Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods (Hajek & Henriques, 2017)	Hajek P., Henriques R.	Knowledge-Based Systems	Article	University of Pardubice/Czech Republic	2017	114	19 (#6)
Big data techniques in auditing research and practice: Current trends and future opportunities (Gepp et al., 2018)	Adrian Gepp; Martina K.Linnenluecke; Terrence J. O'Neill; Tom Smith	Journal of Accounting Literature	Article	Bond University/ Australia	2018	102	20.4 (#4)

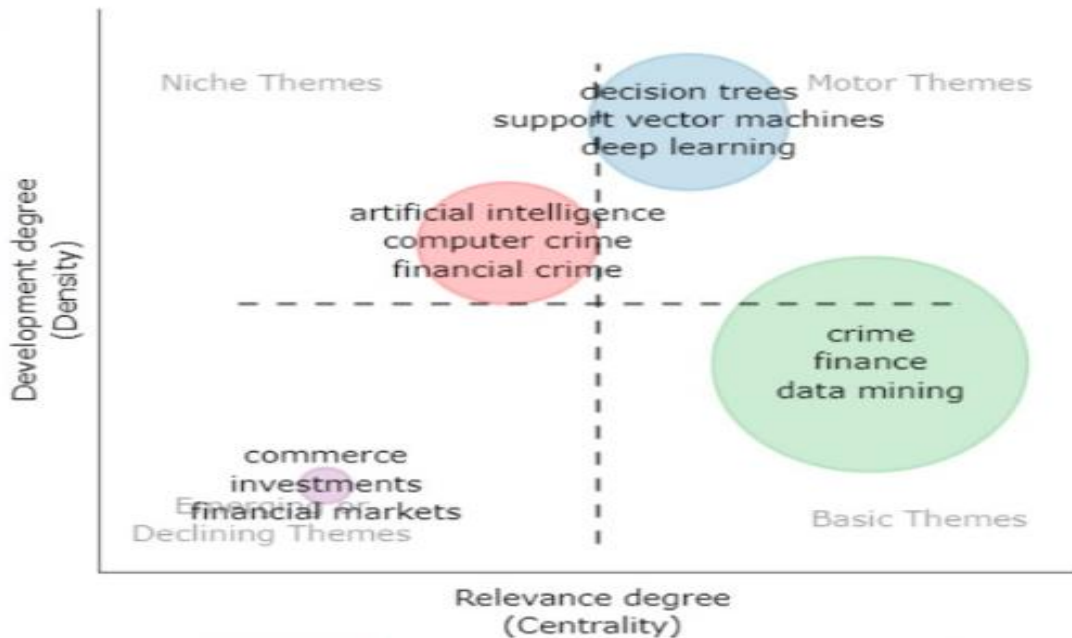
Detecting evolutionary financial statement fraud (Zhou & Kapoor, 2011)	Wei Zhou; Gaurav Kapoor	Decision Support Systems	Article	Information Systems and Technologies/ France	2011	100	8.33 (#8)
A semi-supervised graph attentive network for financial fraud detection (Wang et al., 2019)	Daixin Wang; Yuan Qi; Jianbin Lin; Peng Cui; Quanhui Jia; Zhen Wang; Yanming Fang; Quan Yu; Jun Zhou; Shuang Yang	Proceedings - IEEE International Conference on Data Mining, ICDM	Conference Paper	Tsinghua University/China	2019	80	20 (#5)
Impact of big data analytics on the banking sector: Learning for Indian Banks (Srivastava & Gopalkrishnan, 2015)	Utkarsh Srivastava; Santosh Gopalkrishnan	Procedia Computer Science	Conference Paper	Symbiosis International University/India	2015	75	9.38 (#7)
A review of data mining-based financial fraud detection research (Yue et al., 2007)	Dianmin Yue; Xiaodon Wu; Yunfan Wang; Yue Li ; Chao-Hsien Chu	2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007	Conference Paper	Hebei University of Technology/ China	2007	71	4.44 (#9)
Statistical methods for fighting financial crimes (Sudjianto et al., 2010)	Agus Sudjianto; Sheela Nair; Ming Yuan; Aijun Zhang ; Daniel Kern; Fernando Cela-Díaz	Technometrics	Article	Bank of America/ United States	2010	51	3.92 (#10)

Source: Self-compiled from Biblioshiny R package

Thematic Analysis

The keywords plus were used to create a strategic thematic map in Biblioshiny R, which was then used to analyze the field's theme development as shown in Figure 5. To better visualize the major themes and issues in a field of study, researchers can develop a strategic map by conducting a co-word analysis on the most frequently used keywords in that study's field (Cobo et al., 2011). Strategic diagrams of research fields, as described by (Cobo et al., 2011) are two-dimensional graphs of density and centrality rank values. By mining the keywords, we may determine two measures of importance: (a) density, or the number of connections inside the research theme, and (b) centrality, or the number of connections between the research theme and other themes. The strategic map further displays four primary quadrants based on the intersection of high and low values along these two dimensions (density and centrality) named as themes. The motor themes are located in the upper right quadrant, the transversal or basic themes are located in the lower right quadrant, the peripheral or Niche themes are located in the upper left quadrant, and the emerging or disappearing themes are located in the lower left quadrant (Aparicio et al., 2019a; Cobo et al., 2011).

Figure 5: Strategic Thematic Map



Source: Research Output

Motor Theme

As shown in Figure 5, the upper right quadrant has one cluster of keywords which includes decision tree, machine learning, deep learning, support vector machines, learning algorithm, learning system, credit card frauds, big data, and others. According to previous studies (Aparicio et al., 2019; Murgado-Armenteros et al., 2015), topics that emerge in this quadrant are highly central to the subject and densely represented. These well-developed concepts serve as the basis for the structuring of the research field. The publication time of this theme spread from 2015-2022. The publications on this theme mainly emphasize the use of big data analytics, data mining, deep learning data analytics, or machine learning in detecting financial fraud or credit card fraud. Research on this theme mainly includes credit card fraud detection using machine learning (Awoyemi et al., 2017b), credit card fraud detection using pipeling and ensemble learning (Bagga et al., 2020) detection of fraudulent financial reports with machine learning techniques (Seemakurthi et al., 2015), and fraud data analytics tools and techniques (Makki et al., 2017), and application of data mining technology in financial fraud identification (S. Yao, 2021) to name few. The most popular study under this theme is credit card fraud detection using machine learning techniques: a comparative analysis. “This study examines the efficacy of naïve Bayes, k-nearest neighbour, and logistic regression algorithms when applied to credit card fraud data that exhibits a high degree of skewness. The skewed data is subjected to a hybrid technique involving both under-sampling and oversampling. The three techniques are implemented on both the raw and pre-processed data. The comparative analysis reveals that the k-nearest neighbour algorithm outperforms both the naïve Bayes and logistic regression methods”(Awoyemi et al., 2017a). In conclusion, these concepts have been clearly defined and extensively developed, serving as the foundation of the research field.

Basic Theme

The right lowered quadrant includes base themes. The themes located in the bottom right corner have a high centrality but low density This is because most studies have been conducted on these topics. There is a need for additional study because these themes have strong external connections but weak internal ties. The cluster in this quadrant includes crime, finance, data mining, big data,

financial fraud, financial fraud detection, learning system, machine learning, big data, and learning algorithms. The research in this quadrant concentrates on the topic of the role of machine learning, data mining, hybrid data mining methods, imputation, and data analytics in detecting financial statement fraud. The studies on this theme address the issue of the detection of financial statement fraud using data mining techniques (Hajek & Henriques, 2017; Ravisankar et al., 2011; West & Bhattacharya, 2015), detecting evolutionary financial statement fraud (Zhou & Kapoor, 2011), financial statement fraud detection model based on hybrid data mining methods (J. Yao et al., 2018), and the role of imputation in detecting fraudulent financial reporting (Moepya et al., 2016). The most cited paper is Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques (Ravisankar et al., 2011). “This paper outlines the utilization of intelligent techniques for the purpose of forecasting instances of financial statement fraud within corporate entities. In this study, an analysis is conducted on a dataset consisting of 202 Chinese companies. Various stand-alone techniques, such as Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), and Logistic Regression, are applied to examine the dataset. The findings, as determined by the area under the curve (AUC), demonstrate that the Probabilistic Neural Network (PNN) exhibited superior performance. The significance of anticipating financial fraud lies in its potential to mitigate substantial monetary losses arising from embezzlement. The study exhibits a significant advancement in a particular domain, indicating that alongside the data mining methods utilized in this study, a combination of hybrid data mining techniques and text mining algorithms for sentiment analysis of textual descriptions within financial statements can be employed with data mining algorithms to assess the financial components present in these statements”(Ravisankar et al., 2011). In summary, extensive research has been conducted on various aspects related to this theme, yet further investigation is warranted.

Niche Theme

The themes in this quadrant have a large number of internal ties but a low number of exterior ties, placing them in the third quadrant. Topics that fall into the category of ‘peripheral’ are those that are both narrow in scope and of low significance to the overall body of knowledge (Aparicio et

al., 2019b). The cluster as shown in Figure 5 in the niche quadrant includes artificial intelligence, computer crime, financial crime, money laundering, and data handling. The studies for this theme focus on the application of a semi-supervised graph attentive network for financial fraud detection (Wang et al., 2019), statistical methods for fighting financial crimes (Sudjianto et al., 2010), signature-based methods for financial fraud detection (Edge & Falcone Sampaio, 2009), Eva: visual analytics to identify fraudulent events (Leite et al., 2018) among others. The research conducted within this thematic area focuses on the utilization of various data analysis techniques for the identification of financial crimes across diverse domains. These include the application of semi-supervised graph algorithms in online transactions, the utilization of big data analytics within the banking sector, the implementation of signature-based methods for detecting financial fraud, and the utilization of visual analytics, among others.

Emerging or Disappearing Theme

The keywords in the lowest left theme include topics on commerce, investments, and financial markets, text mining, and knowledge-based system. The research for this theme aims to concentrate on the application of the computational intelligent hybrid model for detecting disruptive trading activity (Zhai et al., 2017), modelling and simulation of financial investment model based on data mining (Pan, 2021), analyzing stock market fraud cases using a linguistics-based text mining approach (Zaki & Theodoulidis, 2013), financial fraud identification of listed companies based on text data mining (Liu et al., 2020), financial discussion boards irregularities detection system using information extraction (Owda et al., 2017) The most cited study by (Panigrahi, 2011)“proposes a framework known as Knowledge-driven Internal Fraud Detection (KDIFD) for the purpose of detecting instances of internal financial fraud. The proposed framework advocates for an approach centred on processes, taking into account the tacit knowledge base of forensic auditors as well as computer-based data analysis and mining techniques”. The proposed framework has the potential to enhance the efficiency of auditors in detecting instances of internal financial fraud. Both the internal and external ties are weak for themes in this section. For this reason, topics in this area of the publishing landscape tend to receive limited coverage. To conclude, the aforementioned research endeavours are currently developing,

with a growing emphasis on the utilization of data analytics and text mining techniques in the stock market irregularities and the financial investment model that is informed by data analysis.

Intellectual Framework

To better grasp, the intellectual framework of the subject of the role of big data analytics in financial fraud detection, a co-citation network of cited references and referenced sources has been developed using VOSviewer. The intellectual framework is defined as an “organized map of salient features of a knowledge base and it reveals the disciplinary composition and tradition of research in a knowledge domain” (Shafique, 2013). The intellectual framework of a topic of study is often a useful tool for figuring out how different academic specializations contribute to that study.

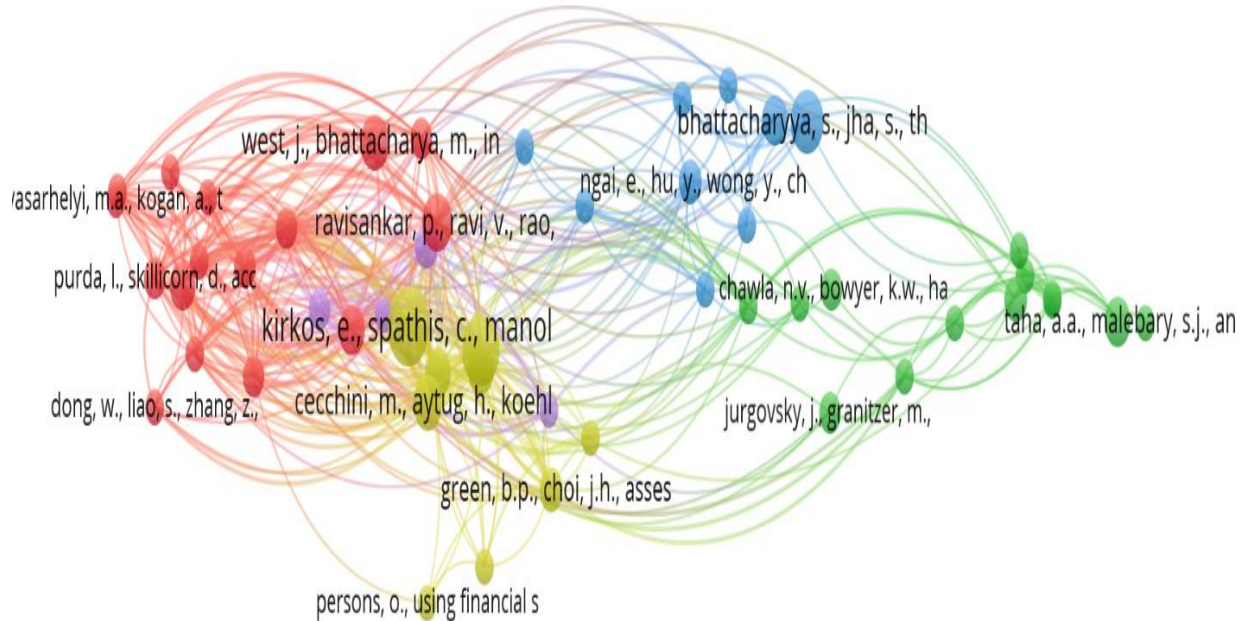
Co-citation of Cited References

The number of times that two secondary sources are mentioned together in a third reference is one of the factors considered in a citation analysis (Small, 1973). The disciplinary makeup of a research field can be ascertained by a study of co-citations of cited references (Liu et al., 2015). The origins of a research field can be traced through the connections made between cited works (Vogel, 2012). Out of 6276 cited references, 56 meet the criterion of a minimum of 4 citations each. Out of 56, 51 cited references are connected. The co-citation network of 51 cited references is shown in Figure 6.

There are 5 clusters as shown in Figure 7 (cluster 1 includes 16 references, cluster 2 with 13 references, Clusters 3 and 4 with 9 references each, and cluster 5 with 4 references) with 414 links and a total of 637 link strength. Every citation is represented by a node, and the size of the node indicates how often that citation appears. The link represents the co-citation of the connected nodes. The largest cluster represented by red colour consists of 16 references; Ravisankar et al., Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques (2011), West & Bhattacharya, Intelligent Financial Fraud Detection: A Comprehensive Review (2016), and Dechow et al., Predicting Material Accounting Misstatements (2011) as the most cited references. This group of papers examines the practice of using data mining and analytics to detect

instances of financial statement fraud or other forms of deceptive financial reporting. The period of this cluster varies from 1997-2018.

Figure 6: Co-Citation of Cited references



Source: Research Output

The second-largest cluster represented by the green colour in the network as shown in Figure 7 consists of 13 cited references. This cluster is based on Kirkos et al., Data Mining Techniques for The Detection of Fraudulent Financial Statements (2007), Taha & Malebary, An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine (2020), and Jurgovsky et al., Sequence Classification for Credit-Card Fraud Detection (2018). These citations are for studies that focus on preventing or detecting financial or credit card fraud by utilizing data analytics.

The 9 references-based blue-coloured cluster has Bhattacharyya et al., Data Mining for Credit Card Fraud: A Comparative Study (2011) and Olszewski, Fraud Detection Using Self-Organizing Map Visualizing the User Profiles (2014) in its centre. The other cluster with 9 items represented with

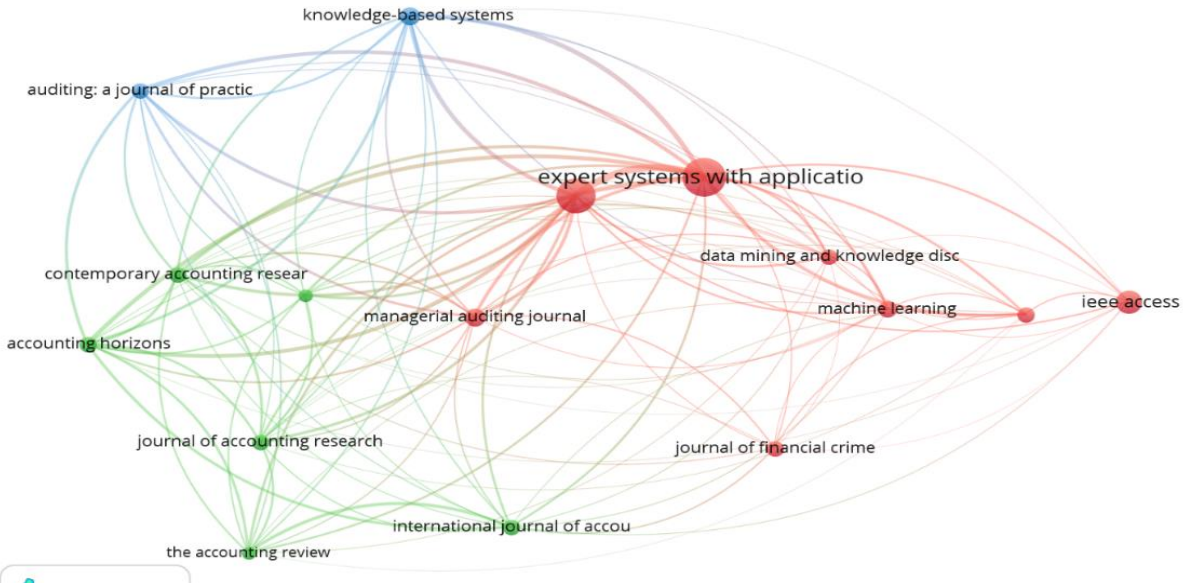
yellow colour in the network is represented by Kirkos et al., *Data Mining Techniques for the Detection of Fraudulent Financial Statements* (2007), Kotsiantis, *Forecasting Fraudulent Financial Statements Using Data Mining* (2006), Cecchini et al., *Detecting Management Fraud in Public Companies* (2010) and Spathis, *Detecting False Financial Statements Using Published Data: Some Evidence From Greece* (2002) as its central point of attraction. The smallest cluster of 4 items includes Ravisankar et al., *Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques* (2011) with 5 citations, and Beneish, *The Detection of Earnings Manipulation* (1999) with 4 citations. The foregoing discussion makes clear that the most common contributing themes of the issues in the study is the identification of credit card fraud or financial statement fraud using various statistical methods, data analysis tools, and data mining.

Co-citation of Sources

VOSviewer is used to construct a co-citation network of sources to examine the subfields that have contributed to the development of the study area as a whole. Out of the total of 3807 cited sources, 18 met the criterion of a minimum of 20 citations. The co-citation network of 18 cited sources with 109 links and 3686 total link strength is shown in Figure 7. The largest cluster constitutes 8 cited sources with Expert System with Application (176 citations, 15 links, and 1433 total link strength) and Decision Support System (144 citations, 15 links, and 1317 total link strength) as nodal points. Other sources include IEEE Access, Data Mining and Knowledge Discovery, Information Sciences, Machine Learning, Journal of Financial Crime, and Managerial Auditing Journal. Cluster 2 consists of 6 sources Contemporary Accounting Research, Accounting Horizons, Journal of Accounting Research, The Accounting Review, International Journal of Accounting, and Journal of Finance.

This cluster can be named the Accounting discipline cluster as all sources are related to accounting. In the smallest cluster, there are only two sources – Auditing: A Journal of Practice & Theory and Knowledge Base System. The above analysis depicts that Expert System with Application is the most productive and impactful source with maximum citations.

Figure 7: Co-Citation of Cited Sources



Source: Research Output

It is evident from Figure 7 that three domains or disciplines mainly contribute to the research topic under study – Accounting, Auditing, and Technology. Technology has laid down the foundation of the topic in research and later the technology is blended with the area of accounting.

7. Conclusion

To the best of our knowledge, this is the first comprehensive bibliometric study in the field of the role of big data analytics in financial fraud detection based on a tenure of twenty-three years to understand the length and breadth of global research. The study investigates the role of big data analytics in financial fraud detection to understand its evolution, present state of research, and prospects. Meta-data of 219 publications were analyzed through VOSviewer and Biblioshiny R. Meta-data consists of 93 articles and 126 conference papers from 151 sources. Although the topic of the role of big data analytics in financial fraud detection is an emerging topic of research, the analysis shows that it has attracted academic interest since 2015 and in the year 2019-2022, the publications have increased manifold. The average publication growth rate is 18 per cent as shown in Table 1. However, the spark on this topic started in the year 1999 when the first document

“Using data mining techniques in fiscal fraud detection” (conference paper) by Bonchi et al. in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) was published. China and India, both are at the top in terms of the number of publications (China =174, India=116) and worldwide citations (China =758, India=442). Even while the United States, United Kingdom, Australia, Canada, and Japan often make substantial contributions to international literature, their impact in this area is surprisingly noticeable. Additionally, very few countries collaborated on the topic under research. Except for China, India, the UK, Australia, and the US, there are very few cross-border collaborations in terms of the number of documents and links. Table 4 depicts that the most productive and influential scholars are from Australia, China, India, and South Africa. Additionally, 616 authors and only 23 authors are largely connected as shown in Figure 4. The developing field of "Role of big data analytics in Financial Fraud Detection" needs substantial contributions from academics in developed nations working together with other academics.

As depicted in Table 3, the most productive sources are ACM International Conference Proceeding Series and Advances in Intelligent Systems and Computing. In addition, half of the top ten most productive sources (Table 3) and four of the top 10 influential documents (Table 5) in this discipline are conference papers, illustrating the great influence of conference presentations in this domain. Considerably, increasing the number of conferences in this field of study is a good strategy because such gatherings provide excellent opportunities for academics to exchange ideas and establish new networks of cooperation. Among the most prestigious and influential academic journals, those from Springer, Elsevier, and Emerald are the most common. These publishing houses are either American or British. There is a need for other publishers to step forward and publish in this area of study.

Thematic analysis as shown in Figure 5, reveals deep learning, machine learning, decision tree, and big data analytics in credit card fraud and fiscal fraud detection as motor themes. However, there is still a need for research on the topic of big data analytics and financial reporting or financial statement fraud detection. There is a notable gap in the current literature regarding the application of hybrid technology in the context of detecting financial fraud. This suggests that further

investigation into this area holds promise for future research endeavours. Moreover, data analytics and text mining are becoming increasingly relevant in discussing stock market anomalies and the data-driven financial investment model. Hence, it is recommended that future studies employ diverse research methodologies in order to gain a deeper understanding of the evolution of the research topic and to identify additional gaps and potential avenues for future research. In addition, the findings of this research shed light on the numerous big data strategies that corporations might employ to prevent and detect financial fraud. The study looks ahead at where the data and statistical methods could go from here. The examination of cited references reveals that the main concerns raised by the research are related to the detection of credit card and financial statement fraud using a variety of statistical techniques, data analysis tools, and data mining. The topic of research is interdisciplinary in nature and gaining relevance for accounting, auditing, and knowledge management as key areas of expertise that contribute to the body of literature in this domain. Initially, technological advances paved the way for academic inquiry into this subject, and later, technological advancements were integrated into the discipline of accounting and finance with a greater applicability.

8. Limitation and Future Scope of the Study

This paper concludes with a comprehensive analysis of the topic of the role of big data analytics in financial fraud detection, but it is not exempt from a few of the limitations of traditional bibliometric evaluations. Initially, it must be reiterated that bibliometric analyzes are useful for measuring and mapping a study field, but cannot replace literature reviews. Secondly, the metadata from SCOPUS is considered for the study, the other researchers can use another database such as Web of Science or Google Scholar, PubMed, and their combined data. Thirdly, the metadata is analyzed using two open software namely VOSviewer and Biblioshiny R. The researchers can use other software available (BibExcel, Pajek, Gephi, HistCite). Finally, research may incorporate citation, co-citation or bibliographic coupling and additional networks, such as decision trees and bibliographic networks in their analysis to identify key clusters.

References

1. Aparicio, G., Iturralde, T., & Maseda, A. (2019a). Conceptual structure and perspectives on entrepreneurship education research: A bibliometric review. *European Research on Management and Business Economics*, 25(3), 105–113. <https://doi.org/10.1016/j.iedeen.2019.04.003>
2. Association of Certified Fraud Examiners. (2022). *Occupational Fraud 2022: A Report to the Nations*. <https://legacy.acfe.com/report-to-the-nations/2022>
3. Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017a). Credit card fraud detection using machine learning techniques: A comparative analysis. *2017 International Conference on Computing Networking and Informatics (ICCNI)*, 1–9. <https://doi.org/10.1109/ICCNI.2017.8123782>
4. Bagga, S., Goyal, A., Gupta, N., & Goyal, A. (2020). Credit card fraud detection using pipeling and ensemble learning. *Procedia Computer Science*, 173, 104–112. <https://doi.org/10.1016/j.procs.2020.06.014>
5. Bonchi, F., Giannotti, F., Mainetto, G., & Pedreschi, D. (1999). Using Data Mining Techniques in Fiscal Fraud Detection (pp. 369–376). https://doi.org/10.1007/3-540-48298-9_39
6. Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
7. Cockcroft, S., & Russell, M. (2018). Big Data Opportunities for accounting and finance practice and research. *Australian Accounting Review*, 28(3), 323–333. <https://doi.org/10.1111/auar.12218>
8. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
9. Edge, M. E., & Falcone Sampaio, P. R. (2009). A survey of signature based methods for financial fraud detection. *Computers & Security*, 28(6), 381–394. <https://doi.org/10.1016/j.cose.2009.02.001>
10. Emich, K. J., Kumar, S., Lu, L., Norder, K., & Pandey, N. (2020). Mapping 50 years of small group research through small group research. *Small Group Research*, 51(6), 659–699. <https://doi.org/10.1177/1046496420934541>
11. Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature*, 40(1), 102–115. <https://doi.org/10.1016/j.acclit.2017.05.003>
12. 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>
13. Leite, R. A., Gschwandtner, T., Miksch, S., Kriglstein, S., Pohl, M., Gstrein, E., & Kuntner, J. (2018). EVA: Visual Analytics to Identify Fraudulent Events. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 330–339. <https://doi.org/10.1109/TVCG.2017.2744758>
14. Liu, Hualing, Tao, Long, & Liu, Mengyao. (2020). Research on financial fraud identification of listed companies based on text data mining. In R. Su (Ed.), *2020 International Conference on Image, Video Processing and Artificial Intelligence* (p. 92). SPIE. <https://doi.org/10.1117/12.2581398>
15. Makki, S., Haque, R., Taher, Y., Assaghir, Z., Ditzler, G., Hacid, M.-S., & Zeineddine, H. (2017). Fraud data analytics tools and techniques in big data era. *2017 International Conference on Cloud and Autonomic Computing (ICCAC)*, 186–187. <https://doi.org/10.1109/ICCAC.2017.26>
16. Moepya, S. O., Akhoury, S. S., Nelwamondo, F. V., & Twala, B. (2016). The role of imputation in detecting fraudulent financial reporting. *International Journal of Innovative Computing, Information and Control*, 12(1), 333–356
17. Murgado-Armenteros, E. M., Gutiérrez-Salcedo, M., Torres-Ruiz, F. J., & Cobo, M. J. (2015). Analysing the conceptual evolution of qualitative marketing research through science mapping analysis. *Scientometrics*, 102(1), 519–557. <https://doi.org/10.1007/s11192-014-1443-z>
18. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559–569. <https://doi.org/10.1016/j.dss.2010.08.006>

19. Owda, M., Lee, P. S., & Crockett, K. (2017). Financial discussion boards irregularities detection system (FDBs-IDS) using information extraction. 2017 Intelligent Systems Conference (IntelliSys), 1078–1082. <https://doi.org/10.1109/IntelliSys.2017.8324262>
20. Pan, H. (2021). Modeling and simulation of financial investment model based on data mining. 2021 International Conference on Aviation Safety and Information Technology, 840–844. <https://doi.org/10.1145/3510858.3511404>
21. Pawar, P. V., & Paluri, R. A. (2022). Big data analytics in logistics and supply chain management: a review of literature. *Vision: The Journal of Business Perspective*, 097226292210916. <https://doi.org/10.1177/09722629221091655>
22. Pejić Bach, M., Krstić, Ž., Seljan, S., & Turulja, L. (2019). Text mining for big data analysis in financial sector: a literature review. *Sustainability*, 11(5), 1277. <https://doi.org/10.3390/su11051277>
23. Putra, I., Sulistiyono, U., Diah, E., Rahayu, S., & Hidayat, S. (2022). The influence of internal audit, risk management, whistleblowing system and big data analytics on the financial crime behavior prevention. *Cogent Economics & Finance*, 10(1). <https://doi.org/10.1080/23322039.2022.2148363>
24. Ravisankar, P., Ravi, V., Raghava Rao, G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50(2), 491–500. <https://doi.org/10.1016/j.dss.2010.11.006>
25. Seemakurthi, P., Zhang, S., & Qi, Y. (2015). Detection of fraudulent financial reports with machine learning techniques. 2015 Systems and Information Engineering Design Symposium, 358–361. <https://doi.org/10.1109/SIEDS.2015.7117005>
26. Srivastava, U., & Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. *Procedia Computer Science*, 50, 643–652. <https://doi.org/10.1016/j.procs.2015.04.098>
27. Sudjianto, A., Nair, S., Yuan, M., Zhang, A., Kern, D., & Cela-Díaz, F. (2010). Statistical methods for fighting financial crimes. *Technometrics*, 52(1), 5–19. <https://doi.org/10.1198/TECH.2010.07032>
28. Wang, D., Qi, Y., Lin, J., Cui, P., Jia, Q., Wang, Z., Fang, Y., Yu, Q., Zhou, J., & Yang, S. (2019). A semi-supervised graph attentive network for financial fraud detection. 2019 IEEE International Conference on Data Mining (ICDM), 598–607. <https://doi.org/10.1109/ICDM.2019.00070>
29. West, J., & Bhattacharya, M. (2015). Mining financial statement fraud: An analysis of some experimental issues. 2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA), 461–466. <https://doi.org/10.1109/ICIEA.2015.7334157>
30. Yao, J., Zhang, J., & Wang, L. (2018). A financial statement fraud detection model based on hybrid data mining methods. 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), 57–61. <https://doi.org/10.1109/ICAIBD.2018.8396167>
31. Yao, S. (2021). Application of data mining technology in financial fraud identification. 2021 4th International Conference on Information Systems and Computer Aided Education, 2919–2922. <https://doi.org/10.1145/3482632.3487540>
32. Yue, D., Wu, X., Wang, Y., Li, Y., & Chu, C.-H. (2007). A review of data mining-based financial fraud detection research. 2007 International Conference on Wireless Communications, Networking and Mobile Computing, 5514–5517. <https://doi.org/10.1109/WICOM.2007.1352>
33. Zaki, M., & Theodoulidis, B. (2013). Analyzing financial fraud cases using a linguistics-based text mining approach. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2353834>
34. Zhai, J., Cao, Y., Yao, Y., Ding, X., & Li, Y. (2017). Computational intelligent hybrid model for detecting disruptive trading activity. *Decision Support Systems*, 93, 26–41. <https://doi.org/10.1016/j.dss.2016.09.003>
35. Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision Support Systems*, 50(3), 570–575. <https://doi.org/10.1016/j.dss.2010.08.007>