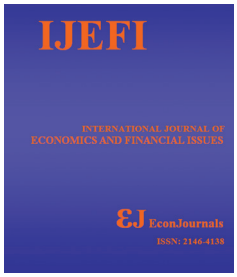


**International Journal of Economics and Financial Issues**

ISSN: 2146-4138

available at http: www.econjournals.com

**International Journal of Economics and Financial Issues, 2024, 14(4), 299-315.**

# Uncovering the Dynamics in the Application of Machine learning in Computational Finance: A Bibliometric and Social Network Analysis

**Samuel-Soma M. Ajibade[1,2]*, Muhammed Basheer Jasser[2], David Olayemi Alebiosu[2], Ismail Ahmed Al-Qasem Al-Hadi[3], Ghassan Saleh Al-Dharhani[3,4], Farrukh Hassan[2], Bright Akwasi Gyamfi[5]**

[1]Department of Computer Engineering, Istanbul Ticaret University, Istanbul, Turkey, [2]Department of Computing and Information Systems, School of Engineering and Technology, Sunway University, Bandar Sunway, 47500 Selangor Darul Ehsan, Malaysia, [3]Institute of Computer Science and Digital Innovation, UCSI University, Kuala Lumpur Malaysia, [4]Faculty of Information Science, and Technology, UKM University, Bangi, Selangor, Malaysia, [5]School of Management, Sir Padampat Singhania University, Bhatewar, Udaipur 313601, India. *Email: asamuel@ticaret.edu.tr

**Received:** 13 March 2024 **Accepted:** 16 June 2024 **DOI:** https://doi.org/10.32479/ijefi.16399

**ABSTRACT**

This paper examined the research landscape on the applications of machine learning in finance (MLF) research based on the published documents on the topic indexed in the Scopus database from 2007 to 2021. Consequently, the publication trends on the published documents data were examined to determine the most prolific authors, institutions, countries, and funding bodies on the topic. Next, bibliometric analysis (BA) was employed to analyse and map co-authorship networks, keywords occurrences, and citations. Lastly, a systematic literature review was carried out to examine the scientific and technological developments in the field. The results showed that the number of published documents on MLF research has soared tremendously from 5 to 398 between 2007 and 2021, which signifies an enormous increase (~7,900%) in the subject area. The high productivity is partly ascribed to the research activities of the most research-active academic stakeholders namely Chihfong Tsai (National Central University in Taiwan) and Stanford University (United States). However, the National Natural Science Foundation of China (NSFC) is the most active funder in the United States and has the largest number of published documents. BA analysis revealed high collaboration rates, published documents, and citations among the stakeholders. Keywords occurrence analysis revealed that MLF research is a highly inter- and multidisciplinary area with numerous hotspots and themes ranging from systems, algorithms and techniques to the security and crime prevention in Finance using ML. Citation analysis, the most prominent (and by extension the most prestigious) source titles on MLF are IEEE Access, *Expert Systems with Applications* and *ACM International Conference Proceedings Series* (ACM-ICPS). The systematic literature review revealed the various areas and applications of MLF research, particularly in the areas of predictive/forecasting analytics, credit assessment and management, as well as supply chain, carbon trading, neural networks, and artificial intelligence, among others. It is expected that MLF research activities and their impact on the wider global society will continue to increase in the coming years.

**Keywords:** Machine Learning, Financial Industry, Publication Trends, Financial Access, Industrial Growth, Bibliometric Analysis
**JEL Classifications:** F3, G0, G1, G4, O16

## 1. INTRODUCTION

The historical origins of machine learning (ML) date back to the early 1950s and 1960s when researchers sought to reproduce human learning using computer programs (Ghoddusi et al., 2019, Ajibade et al., 2023a). ML is reported to have evolved from the twin fields of conventional statistics and artificial intelligence (AI) to create effective data-based models (Edgar and Manz, 2017,

Ajibade et al., 2023c). Over the years, the quest has grown into one of the hottest computational science areas. As a result, ML has resulted in the growth and development of various techniques that collect, process, and utilise data for various applications. Given this, numerous researchers have proposed numerous definitions of the concept based on various theoretical and empirical perspectives in the scientific literature. According to Mitchell and Mitchell (1997), ML broadly speaking is any field of research dedicated to the full comprehension of knowledge based on developed techniques that learn or leverage data to enhance the performance of some designated tasks.

In general, ML has been described as a field of study that employs computational algorithms to transform experimental data into usable models (Edgar and Manz, 2017, Rabbi et al., 2022). Likewise, ML has been defined as a multi- and interdisciplinary field adapted from AI principles with the objective of "teaching" computers how to seek relationships in data (Subasi, 2020, Belyadi and Haghighat, 2021). In principle, ML can be defined as the capacity of any computer-based system to obtain and incorporate knowledge through extensive observations (Woolf, 2009, Ajibade et al., 2020). According to Woolf (2009), the definition extends to the system's independent capability to enhance and expand the frontiers of the acquired knowledge through the process of learning new knowledge as opposed to being programmed. Furthermore, the definition highlights ML as an integrated process or procedure that collects, and processes large volumes of data typically utilized to generate models.

Given the broad definitions, ML has been developed and implemented for various applications, particularly in the current global era of big data and analytics. For example, ML is used for detecting, understanding, and predicting irregular behaviour or patterns in various cyber-based phenomena, processes, and systems around the globe (Edgar and Manz, 2017). It is also used to leverage modern platforms for processing parallel data (Madiajagan and Raj, 2019). In addition, ML provides an integrated structure for establishing intelligent domains for making informed decisions (Bonetto and Latzko, 2020). The concept is also widely applied in forecasting, quality assessments, or performance optimization employing different algorithms (Subasi, 2020, Belyadi and Haghighat, 2021, Ajibade et al., 2020). With the use of ML, computers can also be trained to perform varied tasks such as calculations, clustering, and identifying trends in data. In addition, ML can be used for problem-solving tasks and processes such as collecting, classification, regression, and determination of association rules (Chanal et al., 2021).

Given its versatility, ML has created novel opportunities for various innovations in academia, industry, as well as policy and governance. One such area in that ML has shown great promise in recent times is the areas of economics, energy, and finance. The application of ML in finance has been widely reported in the literature. For example, ML has been widely applied in the finance related topics and areas such as signal processing and portfolio selection (Kim and Boyd, 2008, Jayeola et al., 2022), criteria identification and classification (Xu et al., 2008), and scenario (e.g., supply chain finance, credit risk assessments) prediction and forecasting (Zhu et al., 2016, 2017, 2019; Ma and Lv, 2019; Vaidya et al., 2023). Other researchers have employed ML in quantitative finance topics and areas such as fast pricing, hedging, and fitting of derivatives (De Spiegeleer et al., 2018, Zaidi et al., 2023), as well as asset management (Rundo et al., 2019), and developing various financial products (Gan et al., 2020). ML algorithms deployed with quantum computers have also been applied to portfolio optimization (Alcazar et al., 2020), whereas Boughaci and Alkhawaldeh (2020) demonstrated the applicability of ML for credit scoring and predicting bankruptcy in finance. With the growing calls for the integration of sustainability strategies in all spheres of human endeavour, ML has also been applied to the area of carbon finance. The study by Nguyen et al. (2021), sought to examine the corporate carbon footprints for risk analysis in climate finance using ML. The study demonstrated the applicability of ML in predicting the carbon emissions of corporations to enhance risk analyses by potential stakeholders.

Further review of the literature on machine learning in finance (or MLF) research showed that the research landscape on the subject area is broad with numerous publications on the theories/foundations, themes, and applications existing in various databases. For example, the search on the Elsevier database revealed 2,418 documents based on the "TITLE-ABS-KEY" search criteria. Despite a large number of published documents on the subject area, there have been just two thematic reviews on the subject by Kumar et al. (2021) and Goodell et al. (2021). In the study by Kumar et al. (2021), the application of ML in digital credit scoring in rural finance was examined using a systematic literature review. The findings also showed that the study is focused on the application of ML in a branch of finance and its application in agriculture. In contrast, Goodell et al. (2021) presented a bibliometric analysis (BA) of the application of ML (as a branch of AI) in finance with a focus on highlighting the fundamental themes and research clusters on the subject area. In the study, the authors adopted co-citation and coupling analyses to examine the MLF research but excluded vital BA analysis techniques such as co-author (CA), keyword occurrence (KO), and citation (CT) analyses. The absence of CA, KO, and CT, creates a research gap which limits comprehensive analysis of the research landscape on MLF research.

Therefore, the main objective of this paper is to critically examine the research landscape on MLF research based on published documents on the subject area indexed in the Elsevier Scopus database from 2007 to 2021. In addition, the publication trends, major stakeholders, and funding organisations on the topic will be comprehensively examined. Lastly, the systematic literature review will be adopted to examine the growth, development, and applications of MLF research. It is also envisaged that the study will provide answers to the following pertinent research questions:

1. What is the current status of publication trends and major stakeholders related to MLF research?
2. How has research productivity, collaborations, and citations impacted the area of MLF research?
3. Where are the current hotspots, future directions, and motivating factors for MLF research?

The responses to these relevant issues will offer crucial understandings of the research landscape, scientific growth, and

technological developments in MLF research. The findings will also significantly benefit current and future researchers by offering an overview of the current status and future developments in the subject area of MLF research.

## 2. METHODOLOGY

The objective of this paper is to examine the current and future trends in the application of Machine Learning in Finance (MLF) through bibliometric analysis. Consequently, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) approach (Zhang et al., 2017, Al-Omari et al., 2022, Ajibade et al., 2023e) was adopted to detect, select, and assess the data on the published documents on the MLF topic from the Scopus website. Figure 1 presents the schematic diagram for the PRISMA scheme used for the identification, screening, and analysis of MLF and related documents on the topic in the scientific literature. First, the search query "Machine Learning" AND "Finance" was developed based on the title keywords before executing the search in the Scopus database to recover related publications from 2007 to 2021. A period of 15 years was selected to obtain sufficient related documents, typically >5 published documents per year and >200 publications in total, which is required for conducting a complete bibliometric analysis (Rogers et al., 2020).

The executed search results recovered a total of 1,925 document results based on the TITLE-ABS-KEY search criterion, which was subjected to screening to remove irrelevant documents. The selected screening procedure was executed using the "LIMIT-TO" and "EXCLUDE" search query functions and Boolean operators "AND" and "OR" as shown in the Search query code below:

```
TITLE-ABS-KEY ("machine learning" AND "Finance")
AND PUBYEAR > 2006
AND PUBYEAR < 2022
AND (LIMIT-TO (PUBSTAGE,"final"))
AND (LIMIT-TO (DOCTYPE,"cp")
OR LIMIT-TO (DOCTYPE,"ar")
OR LIMIT-TO (DOCTYPE,"re"))
AND (LIMIT-TO (LANGUAGE, "English"))
AND (EXCLUDE (LANGUAGE,"Chinese"))
AND (LIMIT-TO (SRCTYPE,"p")
OR LIMIT-TO (SRCTYPE,"j"))
```

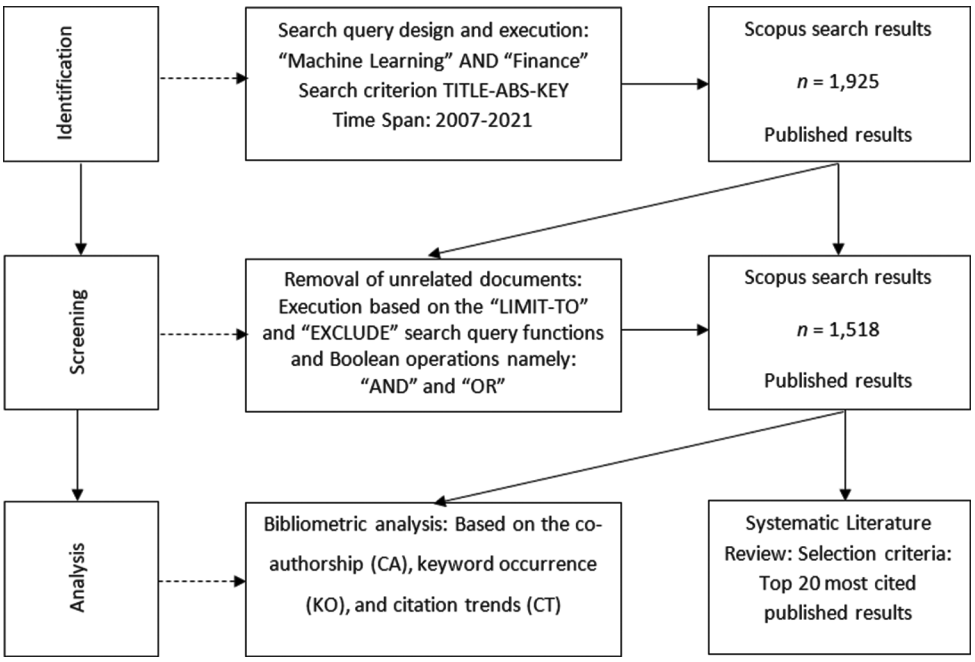
As can be seen, the screening process limited the final search results to the Document types (Conference Proceedings, Articles, and Reviews), Source type (Conference Proceedings, and Journal), and Language (English). The final number of published documents after screening become 1,518 documents, which were consequently analysed to evaluate the publication trends on the MLF. Next up, the bibliometric analysis was conducted to examine co-authorship (CA), keyword occurrence (KO), and citation trends (CT) networks on the topic. The selected conditions for CA, KO, and CT analyses are defined in section V of the paper.

## 3. RESULTS AND DISCUSSION

### 3.1. Publication Trends

The critical analysis of the publication trends on MLF research was carried out from 2007 to 2021 to examine the level of output on the subject over time. Figure 2 shows the plot of the total number of published documents published annually over the 15 years under examination. As can be observed, the number of published documents increased progressively (as evident in the gradually rising curve) from 5 to 398 between 2007 and 2021. The data reveals that there was a 7,860% increase in the number of publications on the subject, which signal an enormous increase in scientific interest in the topic. However, a more critical analysis of the publication trends data based on 5-year intervals was carried out. The findings show that interest in the topic was rather low in Phase I (2007-2011) based on the low total number of publications

**Figure 1:** Schematic for extraction, screening, and analysis of publications on MLF Research

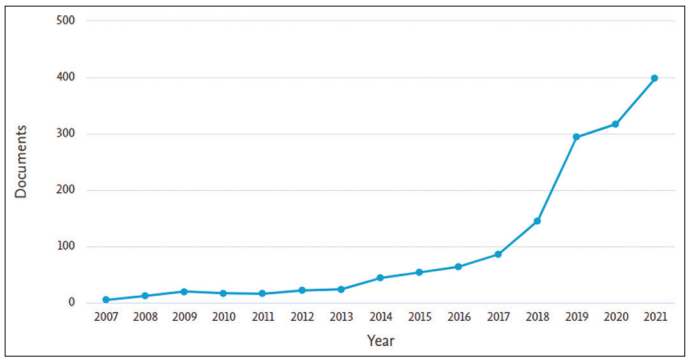


($n$ = 70). In contrast, the total number of publications is 208 and 1,240 for Phase II (2012-2016) and Phase III (2017-2021), respectively. The observation shows that there was a 197.14% and 1,671.43% increase from PI to PII and PI to PIII, respectively, show interest in the topic increased considerably from 2007 to 2012 and 2017.

The steep rise in the number of published documents on MLF after 2017 could be accredited to several factors notably the growing global importance of big data (BD). In principle, BD is defined as large sets or volumes of complex data that have been generated at various speeds and ambiguity and as such cannot be processed through conventional algorithms or programmes for data processing (Krishnan, 2013, Ajibade et al., 2022). BD can be derived from numerous sources ranging from traffic cameras, weather satellites, and Internet of Things (IoT) devices to social media (SAP, 2022). It is characteristically comprised of three important factors or 3 V's namely, variety, volume, and velocity – a definition that is largely based on the early definition proposed by Doug Laney in the early 2000s (SAP, 2022). In current times, BD has become an important tool for utilising large volumes of data to detect patterns and analyse problems (Zheng et al., 2013). According to SAP, big data is used by many companies to improve processes, products, and policy decisions worldwide (SAP, 2022). Given its importance, the world of finance has also embraced BD to drive the growth and development of businesses around the world. Likewise, the rising prominence of BD has prompted increased research interest in the application of ML in finance, which is evident in a large number of publications on the subject as shown in Figure 2. Further analysis also shows that the increase in published documents is also evident in the variety of document types on the subject. Figure 3 shows the distribution of document types on MLF in the Scopus database over the period examined in the study.

As can be seen, "Conference proceedings" (CP) which represent 793 published documents account for the largest share of the total publications. The lead of CP is closely followed by "Articles" (AR), which account for 666 published documents compared to 59 for "Reviews". Based on the results, it can be perceptively deduced that CP is the most preferred document type for researchers working on MLF research topics worldwide. The reason for this observation may be due to the relative ease and speed of disseminating scholarly findings at conference meetings

**Figure 2:** Published documents per year on MLF Research (2007-2021)




or workshops organised by peers in any field when compared to the publishing of articles. Typically, the process of publishing scientific findings in articles requires peer review, which although a crucial step in the scholarly process, could be rather lengthy and arduous for the researchers/scientists working in the industry (Manchikanti et al., 2015, Ajibade et al., 2023d).

Nonetheless, the publication of scholarly findings as articles in journals is still considered *de rigueur* for researchers and scientists in academia. Hence, the objective is to publish in peer-reviewed, prestigious, and high-impact journals, which are typically associated with academic excellence and scholarly prestige in the selected subject area. Likewise, researchers in the field of MLF typically publish their findings in selected journals or source titles. Table 1 shows the top 10 journals preferred by scholars in the field of MLF based on data from the Scopus database.

As seen, the top 10 journal sources that publish works on MLF largely comprise Conference Proceedings and Journals each accounting for 50% of the total. Further analysis shows that the top 10 journal sources account for 237 published documents or approximately 15.61% of the total publications on the subject area, which is significant considering that 115 sources have published at least 2 published documents. Hence, the sources in Table 1 can be considered the most prolific and by extension the

**Figure 3:** Distribution of document types on MLF research (2007-2021)

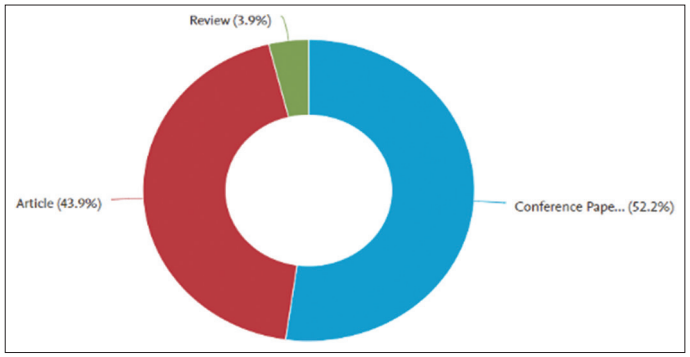



**Table 1: Top 10 journal sources for MLF research**

| Source Title | TP | %TP | Source Type |
|---|---|---|---|
| ACM International Conference Proceeding Series | 55 | 3.62 | Conference Proceedings |
| IEEE Access | 43 | 2.83 | Journal |
| Expert Systems with Applications | 32 | 2.11 | Journal |
| Procedia Computer Science | 26 | 1.71 | Conference Proceedings |
| *Ceur* Workshop Proceedings | 18 | 1.19 | Conference Proceedings |
| Journal of Physics Conference Series | 17 | 1.12 | Conference Proceedings |
| Applied Soft Computing Journal | 13 | 0.86 | Journal |
| Neurocomputing | 13 | 0.86 | Journal |
| Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining | 11 | 0.72 | Conference Proceedings |
| European Journal of Operational Research | 9 | 0.59 | Journal |

TP: Total publications; %TP: Percentage total publications

most prestigious on the subject considering the high preference to publish in them. In particular, the *ACM International Conference Proceeding Series* (ACM-ICPS) and the *IEEE Access* are the most sought-after conference proceedings and journals, on the subject area respectively. The ACM-ICPS is published by the Association for Computing Machinery headquartered in New York, USA. According to its 2021 Scopus metrics, the ACM-ICPS has a CiteScore 1.0, SJR 0.232 and SNIP 0.310, while it is grouped into various Computer Science categories (e.g., software, Computer Vision and Pattern Recognition, and Human-Computer Interaction). However, the *IEEE Access* journal is published by the Institute of Electrical and Electronics Engineers headquartered in New York, USA. Based on its 2021 Scopus metrics, the journal has Citescore 6.7, SJR 0.927, and SNIP 1.326 while it is grouped into Engineering, Computer Science, and Material Science categories in the Scopus database.

The analysis shows that the prestige and reputation of the source type are critical factors for researchers seeking to publish their works. Hence, the more prestigious a source title, the higher the likelihood of the researchers particularly the top players in the field to publish in them. Likewise, the reputation of the source types also plays a crucial role in the future citation rates as well as the research impact of the published documents. To further examine this submission, the top 50 most cited published documents on the MLF research were examined as shown in Table 2. The selection criterion was the retrieval of published documents with over 100 citations to date.

The findings show that the top 50 most cited published documents on the MLF have a combined 12,111 citations (or 242.22 on average). The most cited is "Discovering governing equations from data by sparse identification of nonlinear dynamical systems" by (Brunton et al., 2016), whereas the most cited is "Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods" by (Hajek and Henriques, 2017). Further analysis showed that the most common document type is articles with 32 published documents (or 64% of the top 50 most cited), whereas conference proceedings and reviews each account for 9 each (or 18% of the top 50 most cited). The analysis of the source type revealed that the top sources are *Expert Systems with Applications* (6 published documents), *European Journal of Operational Research* (4 published documents), *Applied Soft Computing Journal* (3 published documents), and *Neurocomputing* (2 published documents). The findings confirm the earlier submission (s*ee* Table 1) that Expert Systems with Applications is a top source for MLF research studies, which may also explain its reputation, citation profile, and attraction by the top researchers in the subject area. Section II will identify and highlight the top researchers or scientists working in the subject area of MLF Research based on data retrieved from the Scopus database from 2007 to 2021.

### 3.2. Top Authors and Institutions

The analysis of the top researchers and institutions on MLF research was examined based on information retrieved from the Scopus database between 2007 and 2021. The objective is to analyse

**Table 2: Top 50 most cited published documents on MLF Research (2007-2021)**

| References | Title | Source Title | Cited by | Document Type |
|---|---|---|---|---|
| Brunton et al. (2016) | Discovering governing equations from data by sparse identification of nonlinear dynamical systems | Proceedings of the National Academy of Sciences of the United States of America | 1220 | Article |
| Fischer and Krauss (2018) | Deep learning with long short-term memory networks for financial market predictions | European Journal of Operational Research | 719 | Article |
| Hirschberg and Manning (2015) | Advances in natural language processing | Science | 525 | Review |
| Gabrel et al. (2014) | Recent advances in robust optimization: An overview | European Journal of Operational Research | 511 | Review |
| Patel et al. (2015) | Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques | Expert Systems with Applications | 478 | Article |
| Dwivedi et al. (2021) | Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy | International Journal of Information Management | 436 | Article |
| Nelson et al. (2017) | Stock market's price movement prediction with LSTM neural networks | Proceedings of the International Joint Conference on Neural Networks | 352 | Conference Paper |
| Tsai and Wu (2008) | Using neural network ensembles for bankruptcy prediction and credit scoring | Expert Systems with Applications | 342 | Article |
| Dosilovic et al. (2018) | Explainable artificial intelligence: A survey | 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings | 336 | Conference Paper |
| Gunning and Aha (2019) | DARPA's explainable artificial intelligence program | AI Magazine | 330 | Article |

*(Contd...)*

**Table 2:** ***(Continued)***

| References | Title | Source Title | Cited by | Document Type |
|---|---|---|---|---|
| François-Lavet et al. (2018) | An introduction to deep reinforcement learning | Foundations and Trends in Machine Learning | 315 | Article |
| Cavalcante et al. (2016) | Computational Intelligence and Financial Markets: A Survey and Future Directions | Expert Systems with Applications | 303 | Article |
| Siami-Namini et al. (2019) | A Comparison of ARIMA and LSTM in Forecasting Time Series | Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018 | 283 | Conference Paper |
| Chen et al. (2015) | A LSTM-based method for stock returns prediction: A case study of China stock market | Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015 | 282 | Conference Paper |
| Patel et al. (2015) | Predicting stock market index using fusion of machine learning techniques | Expert Systems with Applications | 273 | Article |
| Gunning et al. (2019) | XAI-Explainable artificial intelligence | Science Robotics | 262 | Article |
| Hernández-Lobato et al. (2014) | Predictive entropy search for efficient global optimization of black-box functions | Advances in Neural Information Processing Systems | 258 | Conference Paper |
| Krauss et al. (2017) | Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500 | European Journal of Operational Research | 241 | Article |
| Sezer et al. (2020) | Financial time series forecasting with deep learning: A systematic literature review: 2005–2019 | Applied Soft Computing Journal | 229 | Article |
| Malekipirbazari and Aksakalli (2015) | Risk assessment in social lending via random forests | Expert Systems with Applications | 228 | Article |
| Baek and Chilimbi (2010a) | Green: A framework for supporting energy-conscious programming using controlled approximation | Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI) | 210 | Conference Paper |
| Lin et al. (2012) | Machine learning in financial crisis prediction: A survey | IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews | 193 | Review |
| Hajian et al. (2016) | Algorithmic bias: From discrimination discovery to fairness-aware data mining | Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining | 183 | Conference Paper |
| Shen et al. (2015) | Forecasting exchange rate using deep belief networks and conjugate gradient method | Neurocomputing | 179 | Article |
| Stellato et al. (2020) | OSQP: an operator splitting solver for quadratic programs | Mathematical Programming Computation | 174 | Article |
| Pagolu et al. (2017) | Sentiment analysis of Twitter data for predicting stock market movements | International Conference on Signal Processing, Communication, Power and Embedded Systems, SCOPES 2016 - Proceedings | 169 | Conference Paper |
| Grace et al. (2018) | Viewpoint: When will ai exceed human performance? Evidence from ai experts | Journal of Artificial Intelligence Research | 164 | Review |
| Huang (2012) | A hybrid stock selection model using genetic algorithms and support vector regression | Applied Soft Computing Journal | 161 | Article |
| Ma et al. (2018) | Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning | Electronic Commerce Research and Applications | 158 | Article |
| Androutsopoulou et al. (2019) | Transforming the communication between citizens and government through AI-guided chatbots | Government Information Quarterly | 149 | Article |
| Chen and Hao (2017) | A feature-weighted support vector machine and K-nearest neighbour algorithm for stock market indices prediction | Expert Systems with Applications | 148 | Article |

*(Contd...)*

**Table 2:** ***(Continued)***

| References | Title | Source Title | Cited by | Document Type |
|---|---|---|---|---|
| Jain et al. (2016) | New opportunities for materials informatics: Resources and data mining techniques for uncovering hidden relationships | Journal of Materials Research | 142 | Review |
| Li and Hoi (2014) | Online portfolio selection: A survey | ACM Computing Surveys | 142 | Review |
| Liang et al. (2016) | Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study | European Journal of Operational Research | 136 | Article |
| Jadhav et al. (2018) | Information gain directed genetic algorithm wrapper feature selection for credit rating | Applied Soft Computing Journal | 133 | Article |
| Zhang et al. (2021) | Towards augmented kernel extreme learning models for bankruptcy prediction: Algorithmic behaviour and comprehensive analysis | Neurocomputing | 131 | Article |
| Singh and Srivastava (2017) | Stock prediction using deep learning | Multimedia Tools and Applications | 131 | Article |
| Hariri et al. (2019) | Uncertainty in big data analytics: survey, opportunities, and challenges | Journal of Big Data | 126 | Article |
| Kim and Kim (2014) | Investor sentiment from internet message postings and the predictability of stock returns | Journal of Economic Behavior and Organization | 126 | Article |
| Krittanawong et al. (2019) | Deep learning for cardiovascular medicine: A practical primer | European Heart Journal | 117 | Review |
| Baek and Chilimbi (2010b) | Green: A framework for supporting energy-conscious programming using controlled approximation | ACM SIGPLAN Notices | 117 | Conference Paper |
| Liang et al. (2015) | The effect of feature selection on financial distress prediction | Knowledge-Based Systems | 116 | Article |
| Orús et al. (2019) | Quantum computing for finance: Overview and prospects | Reviews in Physics | 115 | Review |
| Ghoddusi et al. (2019) | Machine learning in energy economics and finance: A review | Energy Economics | 114 | Article |
| Rhif et al. (2019) | Wavelet transform application for/in non-stationary time-series analysis: A review | Applied Sciences (Switzerland) | 114 | Review |
| Zhu et al. (2019) | Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach | International Journal of Production Economics | 113 | Article |
| Gupta and Raskar (2018) | Distributed learning of deep neural network over multiple agents | Journal of Network and Computer Applications | 112 | Article |
| Tsai (2014) | Combining cluster analysis with classifier ensembles to predict financial distress | Information Fusion | 107 | Article |
| Harrigan et al. (2021) | Quantum approximate optimization of non-planar graph problems on a planar superconducting processor | Nature Physics | 104 | Article |
| Hajek and Henriques (2017) | Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods | Knowledge-Based Systems | 104 | Article |

the research landscape on the subject area based on the major stakeholders. As can be seen in Figure 4, the top 10 authors on MLF research have published ≥ 4 or 49 (or 4.9 on average or 3.23% of TP) documents over time. The top or most prolific author (defined as the researcher with the highest number of publications) on MLF research is *Chihfong Tsai* (CT) with 7 published documents. CT is based at the National Central University (Taiwan) and has published notable works which have been cited a combined 894 times over the years. The most notable published work "Using neural network ensembles for bankruptcy prediction and credit scoring" published in *Expert Systems with Applications* has gained 342 citations to date. The 2nd most prolific author on MLF is *Chienfeng Huang* (CH) who is also based in Taiwan. The National University of Kaohsiung-based author has 6 published documents on MLF research which have a combined 226 citations. The most cited published document by CH with 161 citations is "A hybrid stock selection model using genetic algorithms and support vector regression". In contrast, Stephen P. Boyd who is the 3rd most prolific researcher on MLF has 5 published documents on MLF research. The works of the Stanford University (Palo Alto, United States) based researcher have gained a total of 282 citations, of which the most notable is "OSQP: an operator splitting solver for quadratic programs" published in *Mathematical Programming Computation* with 174 citations to date.

In contrast, the top or most prolific institution (defined as the organisation with the highest number of publications) on MLF research is Stanford University (United States) with 17 published documents which have gained 1298 citations today, as shown in Figure 5. The leadership of Stanford University ((SU), (USA)) on MLF research can be largely attributed to the productivity of researchers such as Julia Hirschberg, Christopher D. Manning, Woongki Baek, Trishul M. Chilimbi, Stephen S. Boyd and their co-workers. The leadership of SU is closely followed by the Chinese Academy of Sciences (China) and Tsinghua University (China) with 16 and 15 published documents, respectively. Overall, the analysis showed that MLF research is generally dominated by researchers and institutions based on three continents North America, Europe, and Asia. This observation may indicate that the productivity of these researchers and institutions could be ascribed to factors ranging from access to research funding, and other forms of academic resources to institutional or national research policies. This submission can be validated by examining the top funding organisations and countries actively involved in MLF research as will be examined in Section III.

## 4. TOP FUNDING ORGANISATIONS AND COUNTRIES

The analysis of the top funding organisations and countries was carried out to examine the impact of financial support and locations on MLF research over the time frame examined in this study. Figures 6 and 7 present the plots of the top 10 funding organisations and countries actively involved in MLF research between 2007 and 2021.

As can be seen in Figure 6, the top 10 organisations that active fund MLF research have each funded 10 or more published documents over the years. However, the most active funder is the National Natural Science Foundation of China (NSFC) with 99 published documents on MLF. The National Science Foundation (USA) and the Horizon 2020 Framework Programme are the 2nd and 3rd most active funders with 46 and 20 published documents each.

As shown in Figure 7, the funding activities of these organisations have resulted in numerous publications and other forms of research outputs in various countries such as the United States, China, India, the United Kingdom, and Taiwan among others. Various studies have shown that access to funding plays a significant role in the growth and development of scientific and technological research (Scholten et al., 2021, Laudel, 2006). In addition, funding reportedly enhances collaboration between research groups and teams across national and regional boundaries, which results in research networks that can significantly promote research growth and development (Ubfal and Maffioli, 2011, Ebadi and Schiffauerova, 2013). To examine the level of collaborative

**Figure 4:** Top 10 most active researchers on MLF research (2007-2021)

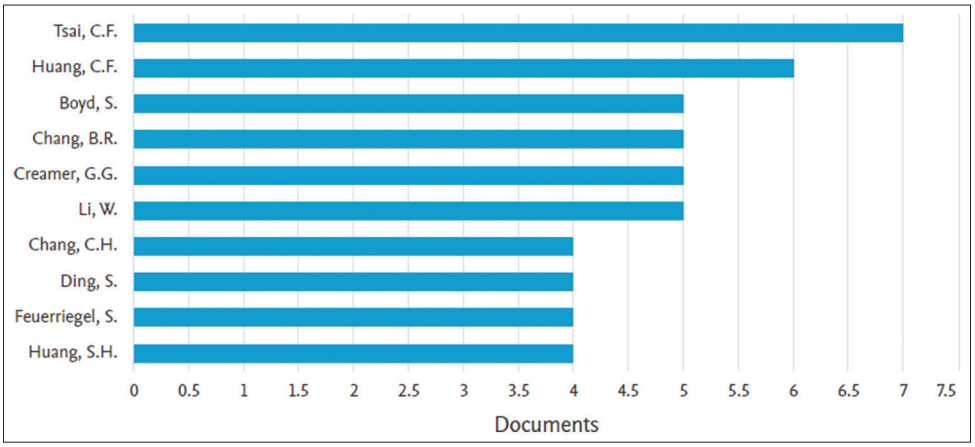



**Figure 5:** Top 10 most active research institutions on MLF research (2007-2021)

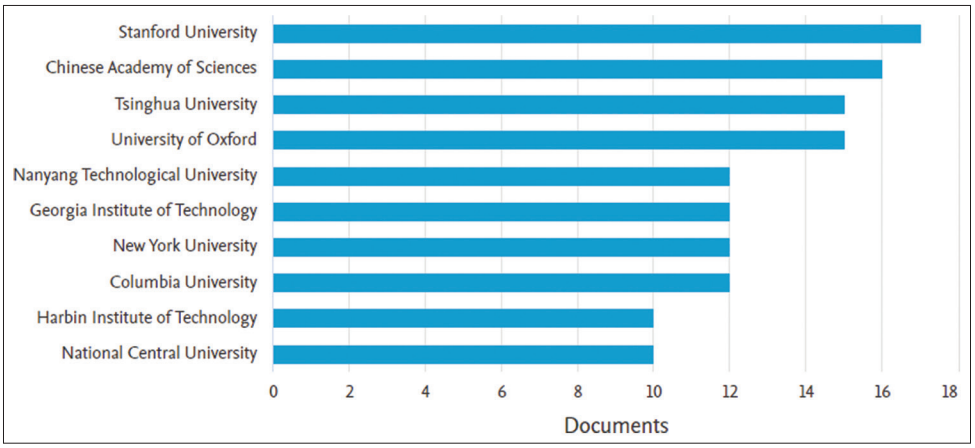


**Figure 6:** Top 10 largest funding organisations for MLF research (2007-2021)

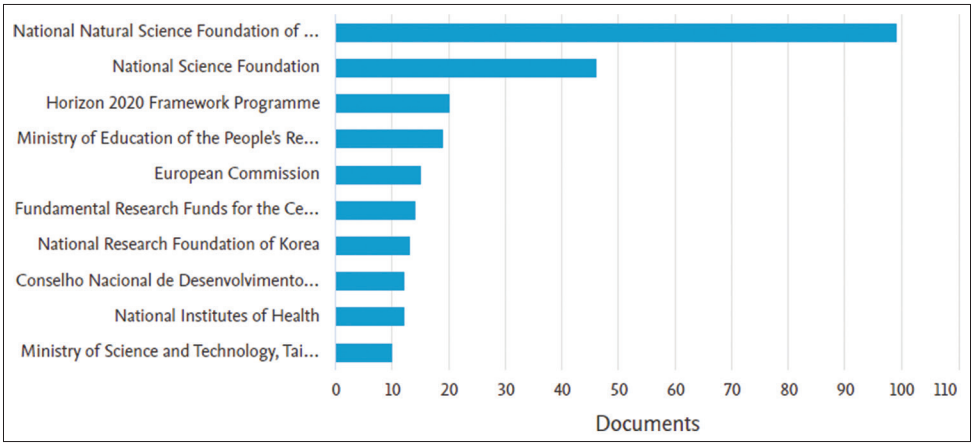
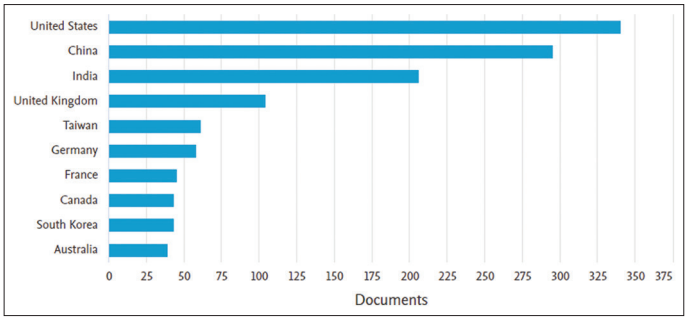



**Figure 7:** Top 10 most active research nations on MLF research (2007-2021)




research and technological cooperation between the top nations actively researching various topics and themes on ML in finance, the bibliometric analysis will be adopted. Section IV presents the results of the bibliometric analysis of MLF research based on the co-authorship, keywords, and citations on the published documents in the subject area.

## 5. BIBLIOMETRIC ANALYSIS (BA)

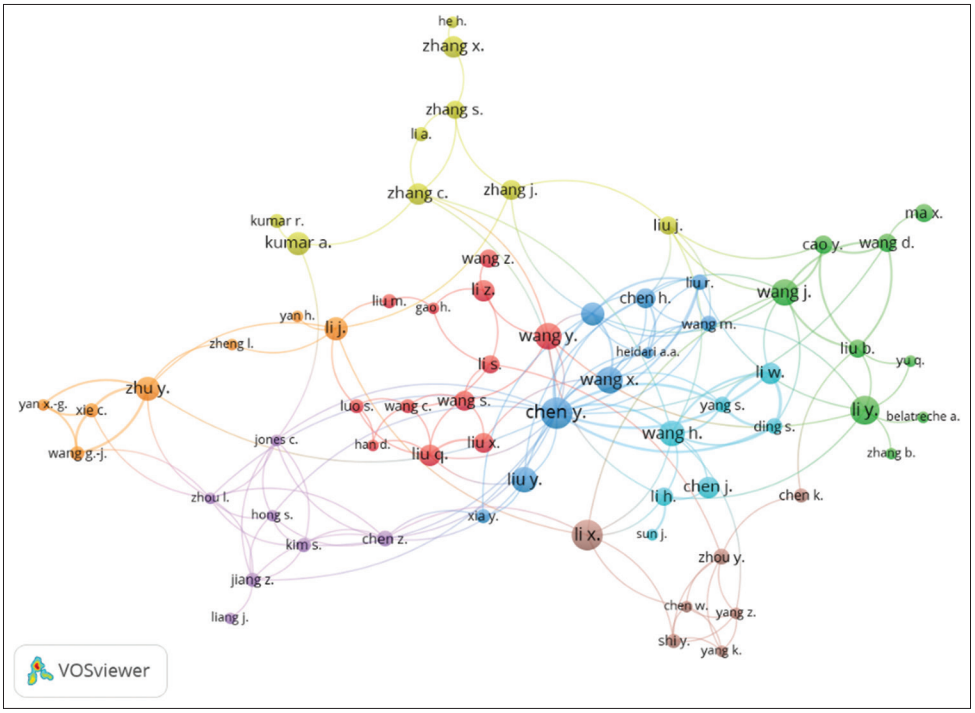
BA is a widespread numerical method used to analyse the research landscape, scientific developments, and technological growth of numerous subject areas (Donthu et al., 2021). Likewise, BA is depicted as an information-based methodology for the mapping and interpretation of published materials such as conference proceedings, articles, reviews, and books among others related to any field of study (Nyakuma et al., 2021, Donthu et al., 2021). It has also been used to quantitatively identify, filter, and assess information on published research materials on various topics or subject areas (Halepoto et al., 2022, Rogers et al., 2020). In principle, the process of BA involves gathering, sorting, and examining the bibliographic data on published documents recovered from scientific databases such as the Elsevier Scopus, Web of Science, CrossRef, PubMed, and PubMed to mention but a few (Wong et al., 2022). The recovered data is subsequently analysed using bibliographic software such as VOSviewer, Bibliometrix, Bibexcel, Pajek, Gephi, SciMat, Sci2, and UCINET among others (Van Eck and Waltman, 2010, Aria and Cuccurullo, 2017, Ajibade et al., 2023b). The BA technique has been successfully adopted to examine the research landscape on numerous subject areas ranging from energy, climate change, business, medicine, and management to mention but a few. In this paper, BA is used to examine the co-authorship, keywords, and citation networks on MLF research to analyse the research landscape on the subject area.

## 6. CO-AUTHORSHIP ANALYSIS

The CA analysis is considered a practical method for analysing the level of collaboration between authors, institutions and countries actively involved in any area of research (Sampaio et al., 2016). In addition, the growth and development of any area of research or scientific endeavour is also an integral function of the level of collaboration between various actors and stakeholders (Kumar, 2015, Lundberg et al., 2006, Dadhich et al., 2023). Therefore, it is crucial to examine the levels of co-authorships, or collaboration by extension, on any given area of research to ascertain the level of research growth and scientific development. In this paper, the author- and country-based co-authorship networks on MLF were developed using VOSviewer based on published documents data from Scopus. Figures 8 and 9 show the network visualisation maps of the author- and country-based co-authorships on MLF research from 2007 to 2021.

The author-based co-author analysis was performed based on the minimum number of 2 documents per author each with 50 or more citations. The results showed that a total of 4,152 authors have published one or more documents on the MLF, although only 150 fulfilled the set search criteria for further analysis. The final results of the analysis presented in Figure 8 show that a total of 67 authors have co-authored publications on the subject area during the time frame examined in the study. Furthermore, the highest number of co-authored published documents (15) and total link strength (TLS) (33) is by Chen (blue cluster), whereas the highest number of citations is by Brunton and Kutz both with 1207 citations. The results also showed that there are 8 clusters, 67 connected researchers, 168 links and a TLS of 209. The largest cluster of co-authors consists of researchers such as Gao, Han, and Wand among others, while the smallest cluster is made up of Chen et al. Based on the co-authored outputs, links, and TLS recorded, Chen (blue cluster) who is based at Anhui University, Hefei (China) is the most influential collaborator on MLF research

**Figure 8:** Visualisation map of the author-based co-authorship network on MLF Research (2007-2021)

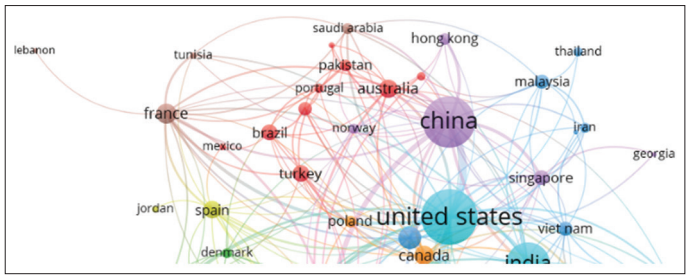



worldwide. Overall, the co-authorship analysis showed that there is a high level of collaboration between the various authors involved in MLF research worldwide. This observation may be due to the growing relevance of ML as an important tool for the comprehensive analysis of big data, which is required for various uses such as enhanced decision-making, process efficiency, and performance in many organisations and countries worldwide.

Given its growing importance, many nations around the world have set up programmes, schemes, and strategies to explore and exploit the use of ML and related tools such as IoT for the analysis of complex (big) data for applications in many aspects of policy and governance. Hence, the mandate has been placed on the academia and industry in these countries, which has resulted in significant research over the years. To this end, numerous researchers, scientists, and academics within and outside the borders of many nations have sought to collaborate and explore solutions to numerous problems related to MLF. The resulting product has been numerous co-authored published documents, which not only serve as a measure of collaboration but also the level of growth and development of the field. To further examine this, the country-based co-authorship networks on MLF were examined, as shown in Figure 9.

Likewise, the country-based co-author analysis was performed based on the minimum number of 2 documents per country each with 50 or more citations. The results showed that a total of 130 countries have published 2 or more papers although only 50 fulfilled the set criteria for co-authorship analysis. In addition, the largest set of connected items, which indicates the highest number of co-authorship links between countries is 48, which generated 9 clusters, 226 links and a TLS of 494. Based on the analysis, the country with the highest number of co-authored published documents, highest citations, and TLS is the United

**Figure 9:** Visualisation map of the country-based co-authorship network on MLF Research (2007-2021)




States. The findings reveal that the US is the most prominent research nation in terms of output, citations, and collaborations as also reported earlier (*see* Figure 7). The dominance of the US is to be expected as it is widely recognised as the global leader in information technologies such as ML, and IoT as well as a financial powerhouse. Based on data from the International Monetary Fund, the United States is the world's largest economy with a minimum gross domestic product of US$20.49 trillion, which represents approximately 25% of the global economy (WorldAtlas, 2022). Hence, the high productivity of the United States on MLF research could be reasonably surmised as the product of national interests aimed at developing and implementing technologies that not only has research but as well a societal impact. The research impact of any field can also be examined through keyword occurrence, as examined in the next section of the paper.

## 7. KEYWORDS ANALYSIS

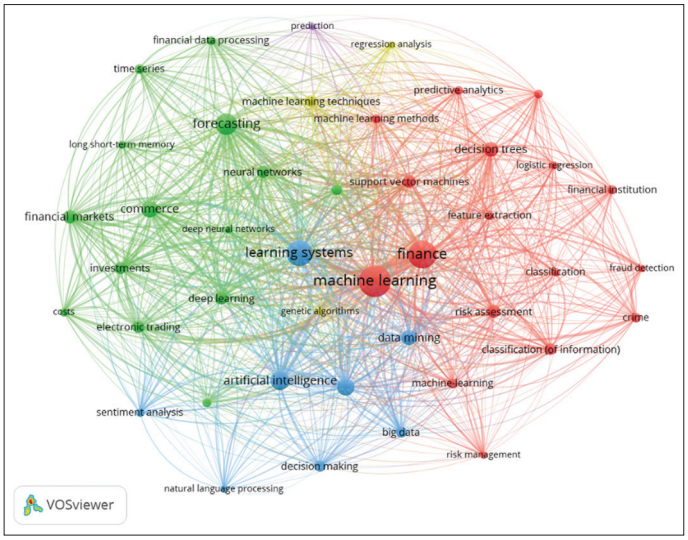
The keyword occurrence is another important aspect of the BA process that examines the research impact of the level of relevance of any given subject area. In addition, the process helps researchers

to map the hotspots, current and potential areas of interest for future research on any given subject area (Han et al., Qian et al., 2019). In this study, the keyword occurrence analysis of MLF was carried out using the designated feature in VOSviewer using the title keywords. The analysis was based on the minimum 50 occurrences of a keyword, which resulted in 9,023 results of which 43 fulfilled the set criteria for further analysis. Subsequently, the visualisation map of the keyword occurrences on MLF Research was developed as shown in Figure 10.

As expected, the most occurrent keywords are "machine learning" and "finance" which occurred 855 and 694 times respectively. However, the TLS of "finance" (3,431) was found to be higher when compared to ML (3,242). The results also show a total of 5 clusters, the largest (red) of which consists of keywords such as Classification, Crime, Decision Trees, Finance, and Risk assessment to mention but few. The categorisation of the keywords gives rise to research hotspots or themes, which for this study indicates 5 major themes namely;

1. Applications of ML for Security, crime prevention and risk analysis in Finance (red cluster),
2. Various branches of ML in Finance (green cluster),
3. Systems and technologies of ML in Finance (blue cluster),
4. Prevention and forecasting using ML in Finance (purple cluster),
5. Regression analysis in MLF (yellow cluster).

**Figure 10:** Visualisation map of the keyword occurrences on MLF Research (2007-2021)




Furthermore, the results showed the keyword occurrence map has 865 links and a TLS of 17,819. Overall, the findings indicate that MLF has an extensive network of core keywords which are linked to several secondary keywords, which invariably indicates the field has a high research impact. In addition, the diversity of the keywords deduced during the analysis indicates MLF is highly multidisciplinary, which could provide an important foundation for the growth and development of the subject area.

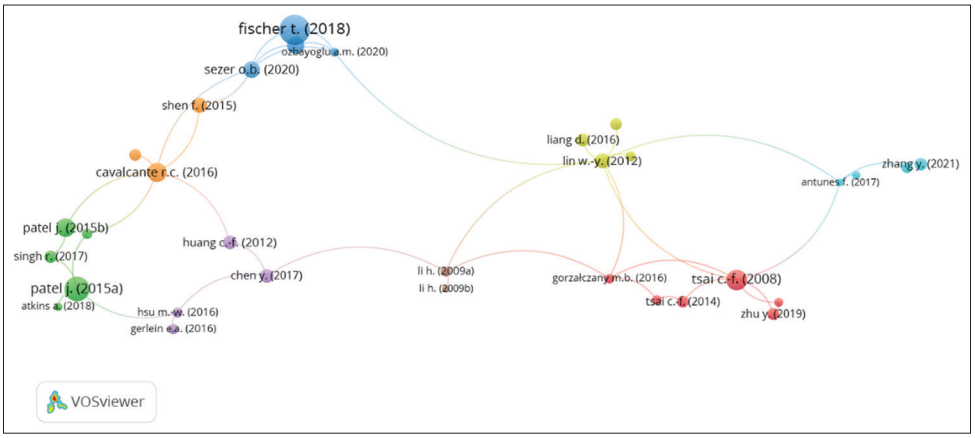
## 8. CITATIONS ANALYSIS

The analysis of the document and source-based citations on MLF research was carried out based on the data on the published documents from 2007 to 2021 using VOSviewer. The software was used to develop the network visualisation maps for citations gained by published documents and journal sources related to MLF research. Figure 11 shows the visualisation map for the network of document-based citations on MLF Research.

The analysis was based on the set criteria of a minimum of 50 citations per document, of which 109 documents out of 1518 fulfilled the set criteria for further analysis. As shown in Figure 11, the largest set of connected documents is 32, which generated 8 clusters along with 44 links. The largest (red) cluster consists of studies such as Gorzalczany et al., (2016), Plawaiak et al., (2020), and Tsai et al. (2008), whereas the smallest cluster consists of the works of Li et al. (2009a) and Li et al., (2009b). In contrast, the citation analysis of the journal sources was carried out based on a minimum of 2 published documents each with 50 or more citations, of which 56 out of the possible 929 sources fulfilled the set criteria for further analysis. Figure 12 shows the visualisation map for the source-based citations network on MLF Research.

The results indicate there are 38 connection papers, which generated 10 clusters, 88 links and a TLS of 162. The largest (red) cluster is comprised of the journals Soft Computing, Energy Economics, and Journal of Banking & Finance, whereas

**Figure 11:** Visualisation map for the document-based citations network on MLF Research (2007-2021)



the smallest cluster consists of Electronic Commerce Research & Applications, and *Physica* A: Statistical Mechanics and its Applications. However, the source with the highest number of cited documents (55) is *ACM International Conference Proceedings Series*, whereas *Expert Systems with Applications* has the highest number of citations (2760) as well as TLS (67). Hence, it can be logically deduced that the sources *ACM International Conference Proceedings Series* (ACM-ICPS) and *Expert Systems with Applications* (ESA) are the most prominent sources of publications for MLF research. According to the details on the journal's website, the ESA journal aims to publish world-class research findings on expert and intelligent systems that can be applied in academia, industry, or government. The Elsevier-published journal has a high impact factor (IF = 8.665), Citescore (CS = 12.2), SCImago Journal Rank (SJR = 2.070), and Source Normalized Impact per Paper (SNIP = 2.985). It is indexed in various categories such as General Engineering, as well as Computer Science sub-themes such as Computer Science Applications and Artificial Intelligence. Based on the journal metrics ESA is considered an important and high-impact publication that could greatly advance the researcher's prestige and career when studies have been published in the journal.

## 9. SYSTEMATIC LITERATURE REVIEW

The research landscape on the current developments in MLF research was also examined through a systematic literature review of the benchmark publications on the subject area. The document selection for SLR was based on the criteria; (i) Published documents with 5 or more citations; (ii) published documents with the keywords "machine learning and "finance" in their TITLE only; (iii) indexed in Scopus between 2007 and 2021; (iv) document type (articles and conference proceedings). Table 3

**Figure 12:** Visualisation map for the source-based citations network on MLF Research (2007-2021)

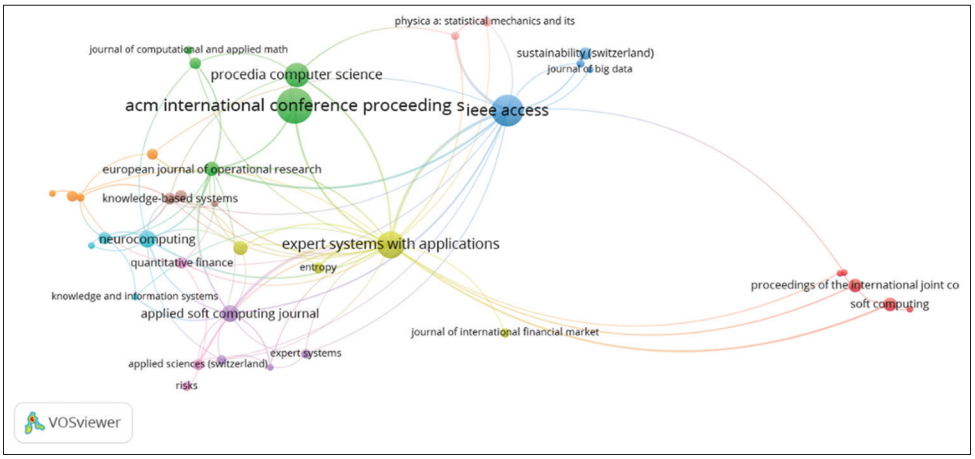



**Table 3: Summary of findings of benchmark publications on MLF Research**

| References | Study Objectives | Summary of Study findings |
|---|---|---|
| Kim and Boyd (2008) | Propose a minimax-based formula for applying ML in finance. | The study demonstrated the potential use and applications of ML in finance using hybrid concepts of robust Fisher linear discriminant and robust portfolio selection. The authors demonstrated that saddle points exist that can be efficiently used for convex computing and optimization. |
| Xu et al. (2008) | Proposes a novel approach for identifying Chinese language text in finance through ML. | The study proposed an innovative approach for the selection of feature items that could be applied to enhance the identification of financial texts using ML. The results showed that the use of the SVMs classifier on the real-world corpora improves selection, which validates the efficacy of LR (likelihood ratio) as a dependable metric for selecting informative features. The findings also showed higher performance that ensured 80% removal of unrelated terms. |
| Zhu et al. (2016) | Demonstrate the use of ML in forecasting SME credit risk in supply chain finance using ML. | The study demonstrated that the proposed combined or collective ML approach comprising Random Subspace-Real AdaBoost (RS-RAB) for forecasting credit risk displayed excellent performance. Hence the approach was deemed appropriate for predicting credit risk in supply chain finance particularly small and medium-sized enterprises (SMEs) in China. |
| Zhu et al. (2017) | Presents a comparative study of ML techniques applied in forecasting credit risk in supply chain finance for SMEs in China. | The study demonstrated the effectiveness of utilising 6 ML techniques for predicting credit risk in SMEs in China. The findings showed that IEML (integrated ensemble ML) methods acquire better performance than IML (individual ML) and EML (ensemble ML) methods. |
| De Spiegeleer et al. (2018) | Proposes ML as a tool for rapid derivative pricing, fitting, and hedging in quantitative finance. | The study showed a practical approach for deploying ML approaches for problem-solving in quantitative finance. Hence, Gaussian regression-based ML was deployed for rapid analysis and resolution of classical finance problems although the approach resulted in a loss of accuracy. |

*(Contd...)*

**Table 3:** ***(Continued)***

| References | Study Objectives | Summary of Study findings |
|---|---|---|
| Zhu et al. (2019) | Proposes an enhanced hybrid ensemble ML method for predicting credit risk in supply chain finance of SMEs. | The results showed that the enhanced hybrid ensemble ML method (termed RS-MultiBoosting) is effective for predicting SMEs credit risks for small sample sizes. Furthermore, the findings showed that an even better performance for assessing the financing ability can be obtained by using "traditional factors" like the current or quick ratio of SMEs. Likewise, the use of specific factors like trade goods features could greatly improve SCF. |
| Ma and Lv (2019) | Proposes the use of ML to predict credit risk in internet finance | The study adopted three archetypal test functions to simultaneously compare the performance of ML IA and logic-based prediction algorithms based for financial credit risk prediction and assessment. The results showed that ML algorithms can be utilised and improved for application in predicting credit risk in the financial sector. |
| Cai et al. (2019) | Examined the potential for using ML and expert judgement for the analysis of emergent themes in accounting and finance research. | The study demonstrated the comparative use of two ML algorithms for keyword detection as an effective approach for qualitative data analysis and systematic literature reviews. The results showed that automated analysis is effective for large quantities of text and provides a standardized and non-biased method way of examining the literature. Yet, the human researcher or manual approach showed better results for analysing current issues and future trends in the literature. |
| Gan et al. (2020) | Examined the application of ML solutions to problems such as financial product pricing in finance. | The paper proposed an innovative and model-free ML approach for quickly and accurately determining the price arithmetic and geometric average options in finance. |
| Alcazar et al. (2020) | Investigated the application of classical and quantum models of ML in finance. | The study constructed scenarios from the probabilistic version of the popular problem of portfolio optimization in finance using time-series data for pricing from asset subsets of the S&P 500 index of the stock market. The objective was also to solve questions related to real-world classical data sets using quantum performance model approaches. The results showed that the quantum models showed superior performance on typical instances (when compared to classical approaches) even with recognised RBMs (restricted Boltzmann machines) training and under similar resources and parametric terms. |
| Boughaci and Alkhawaldeh (2020) | Examined the application of suitable ML techniques predicting bankruptcy and for credit scoring in banking and finance. | In the study, ML methods for predicting bankruptcy and credit scoring were examined for applications in banking and finance using various real-life datasets. The results showed that the ML techniques were able to successfully generate credit scores for applicants, which could help banks and other financial institutions make informed decisions. |
| Nguyen et al. (2021) | Investigated an ML technique for forecasting the carbon footprint for the analysis of the climate finance risk of corporate entities. | This paper adopted ML to enhance the forecasting of corporate carbon emissions for enhanced risk assessments by investors. The two-step framework proposed by the authors was aimed at combining projections from multiple base learners as the safest approach for emission prediction. The findings showed an accuracy gain of about 30% based on mean absolute error when associated with present models. However, the accuracy of the prediction could be further enhanced by integrating extra predictors and firm discoveries in specific areas. |
| Hansen and Borch (2021) | Examined the uncertainty of absorption and multiplication using ML-based finance. | The study examined the potential of using ML for analysing and translating uncertainty into controllable risk. The findings showed that ML could also be used to establish a novel and powerful kind of uncertainty (termed critical model uncertainty, CMU) in addition to absorbing uncertainty in finance. The CMU is defined as the incapacity to describe ML models (e.g., neural networks) or how and why they reach their projections and conclusions. Lastly, the authors recommended that the discussion about the ambiguity of absorption and multiplication associated with ML models required further investigation in the subject area of finance. |

presents a summary of the major findings of the benchmark papers ($n$ = 13), on MLF research from 2007 to 2021.

## 10. CONCLUSIONS

The research landscape on the applications of machine learning in finance (MLF) research was critically examined through publication trends, bibliometric analysis, and systematic literature review of the published documents on the topic indexed in the Scopus database from 2007 to 2021. The publication trends analysis revealed an immense increase (7,860%) in the number of published documents on the subject area from 2007 to 2021. The findings indicate significant interest in MLF research, which may be due to the growing need for robust algorithms to process and analyse big or complex data rapidly and accurately. Big data is becoming an increasingly important aspect of doing business, which is implemented by companies to not only improve processes, products, and services but also to make informed policy decisions or legislations in governments worldwide. Further analysis showed the subject area consists of numerous authors, institutions, and countries who are enthusiastically researching various aspects and applications of ML in Finance. The findings also indicate that financial support may be widely available and accessible to researchers in the field, which has motivated the high rate of research activities and productivity by MLF researchers.

Bibliometric analysis showed revealed high rates of collaboration among authors, institutions, and countries as well as high rates of published documents and citations among these major stakeholders. The keywords occurrence analysis revealed a large diversity of keywords and research hotspots/themes that indicate MLF is a highly multidisciplinary subject area. Citation analysis, on the other hand, showed that the publications metrics play an important role in the choice of source title (i.e., journal or proceedings) selected by researchers for publishing their research works. For MLF research, the most prominent (and by extension the most prestigious) source titles are the *ACM International Conference Proceedings Series* (ACM-ICPS), IEEE Access, and *Expert Systems with Applications*. Lastly, the systematic literature review revealed that MLF research is expected to continue in its growth and development trajectory with its research themes expected to branch out and or leverage various areas and applications of MLF research such as financial data analytics, the credit assessment, risk management, supply chain, carbon trading, as well as neural networks, predictive analytics, artificial intelligence, and data mining among others. Overall, the study showed that the application of ML in finance is a broad, multidisciplinary, and impactful area of research that will remain relevant in the coming years, as shown by the findings of this paper.

## 11. ACKNOWLEDGMENTS

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number C2023-54-01/HĐ-KHCN.

## REFERENCES

Ajibade, S.S., Zaidi, A., Al Luhayb, A.S.M., Adediran, A.O., Voumik, L.C., Rabbi, F. (2023a), New insights into the emerging trends research of machine and deep learning applications in energy storage: A bibliometric analysis and publication trends. International Journal of Energy Economics and Policy, 13, 303-314.

Ajibade, S.S.M., Ahmad, N.B., Shamsuddin, S.M. (2020), A Data Mining Approach to Predict Academic Performance of Students Using Ensemble Techniques. In: Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) Held in Vellore, India. Vol. 1. Springer. p749-760.

Ajibade, S.S.M., Ayaz, M., Ngo-Hoang, D.L., Tabuena, A.C., Rabbi, F., Tilaye, G., Bassey, M.A. (2022), Analysis of Improved Evolutionary Algorithms Using Students' Datasets. In: 2022 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), IEEE. p180-185.

Ajibade, S.S.M., Bashir, F.M., Dodo, Y.A., Dayupay, J.P., De La Calzada, L.M., Adediran, A.O. (2023b), Application of Machine Learning in Energy Storage: A Scientometric Research of a Decade. In: International Conference on Information and Software Technologies, Springer. p124-135.

Ajibade, S.S.M., Bekun, F.V., Adedoyin, F.F., Gyamfi, B.A., Adediran, A.O. (2023c), Machine learning applications in renewable energy (MLARE) research: A publication trend and bibliometric analysis study (2012-2021). Clean Technologies, 5, 497-517.

Ajibade, S.S.M., Flores, D.D.C., Ayaz, M., Dodo, Y.A., Areche, F.O., Adediran, A.O., Oyebode, O.J., Dayupay, J.P. (2023d), Application of Machine Learning In Renewable Energy: A Bibliometric Analysis of a Decade. In: 2023 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), IEEE. p173-179.

Ajibade, S.S.M., Zaidi, A., Bekun, F.V., Adediran, A.O., Bassey, M.A. (2023e), A research landscape bibliometric analysis on climate change for last decades: Evidence from applications of machine learning. Heliyon, 9, e20297.

Alcazar, J., Leyton-Ortega, V., Perdomo-Ortiz, A. (2020), Classical versus quantum models in machine learning: Insights from a finance application. Machine Learning: Science and Technology, 1, 035003.

Al-Omari, B., Ahmad, T., Al-Rifai, R.H. (2022), SARS-CoV-2 and COVID-19 research trend during the first two years of the pandemic in the united arab emirates: A PRISMA-compliant bibliometric analysis. International Journal of Environmental Research and Public Health, 19, 7753.

Androutsopoulou, A., Karacapilidis, N., Loukis, E., Charalabidis, Y. (2019), Transforming the communication between citizens and government through AI-guided chatbots. Government Information Quarterly, 36, 358-367.

Aria, M., Cuccurullo, C. (2017), Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11, 959-975.

Baek, W., Chilimbi, T.M. (2010a), Green: A Framework for Supporting Energy-conscious Programming Using Controlled Approximation. In: ACM SIGPLAN 2010 Conference on Programming Language Design and Implementation (PLDI), Toronto. p198-209.

Baek, W., Chilimbi, T.M. (2010b), Green: A framework for supporting energy-conscious programming using controlled approximation. ACM SIGPLAN Notices, 45, 198-209.

Belyadi, H., Haghighat, A. (2021), Machine Learning Guide for Oil and Gas Using Python. Amsterdam: Elsevier.

Bonetto, R., Latzko, V. (2020), Chapter 8: Machine learning. In: Fitzek, F.H.P., Granelli, F., Seeling, P., editors. Computing in Communication Networks. Cambridge: Academic Press.

Boughaci, D., Alkhawaldeh, A.A.K. (2020), Appropriate machine learning techniques for credit scoring and bankruptcy prediction in banking and finance: A comparative study. Risk and Decision Analysis, 8, 15-24.

Brunton, S.L., Proctor, J.L., Kutz, J.N., Bialek, W. (2016), Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences of the United States of America, 113, 3932-3937.

Cai, C.W., Linnenluecke, M.K., Marrone, M., Singh, A.K. (2019), Machine learning and expert judgement: Analyzing emerging topics in accounting and finance research in the Asia-Pacific. Abacus, 55, 709-733.

Cavalcante, R.C., Brasileiro, R.C., Souza, V.L.F., Nobrega, J.P., Oliveira, A.L.I. (2016), Computational intelligence and financial markets: A survey and future directions. Expert Systems with Applications, 55, 194-211.

Chanal, P.M., Kakkasageri, M.S., Manvi, S.K.S. (2021), Security and privacy in the internet of things: Computational intelligent techniques-based approaches. In: Recent Trends in Computational Intelligence Enabled Research. Amsterdam: Elsevier.

Chen, K., Zhou, Y., Dai, F. (2015), A LSTM-based method for stock returns prediction: A case study of China stock market. In: Luo, F., Ogan, K., Zaki, M.J., Haas, L., Ooi, B.C., Kumar, V., Rachuri, S., Pyne, S., Ho, H., Hu, X., Yu, S., Hsiao, M.H.I., Li, J., editors. 3rd IEEE International Conference on Big Data, IEEE Big Data 2015. Piscataway: Institute of Electrical and Electronics Engineers Inc. p2823-2824.

Chen, Y., Hao, Y. (2017), A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. Expert Systems with Applications, 80, 340-355.

Dadhich, M., Rathore, S., Gyamfi, B.A., Ajibade, S.S.M., Agozie, D.Q. (2023), Quantifying the dynamic factors influencing new-age users' adoption of 5G Using TAM and UTAUT models in emerging country: A multistage PLS-SEM Approach. Education Research International, 1, 5452563.

De Spiegeleer, J., Madan, D.B., Reyners, S., Schoutens, W. (2018), Machine learning for quantitative finance: Fast derivative pricing, hedging and fitting. Quantitative Finance, 18, 1635-1643.

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.#

Dosilovic, F.K., Brcic, M., Hlupic, N. (2018), Explainable artificial intelligence: A survey. In: Vrdoljak, B., Tijan, E., Grbac, T.G., Sruk, V., Cicin-Sain, M., Ribaric, S., Skala, K., Koricic, M., Mauher, M., Gros, S., Pale, P., Janjic, M., editors. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics. Piscataway: MIPRO, Institute of Electrical and Electronics Engineers Inc. p210-215.

Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P.V., Janssen, M., Jones, P., Kar, A.K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., & Williams, M.D. (2021), Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994.

Ebadi, A., Schiffauerova, A. (2013), Impact of funding on scientific output and collaboration: A survey of literature. Journal of Information and Knowledge Management, 12, 1350037.

Edgar, T., Manz, D. (2017), Research Methods for Cyber Security. Rockland: Syngress.

Fischer, T., Krauss, C. (2018), Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270, 654-669.

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M.G., Pineau, J. (2018), An introduction to deep reinforcement learning. Foundations and Trends in Machine Learning, 11, 219-354.

Gabrel, V., Murat, C., Thiele, A. (2014), Recent advances in robust optimization: An overview. European Journal of Operational Research, 235, 471-483.

Gan, L., Wang, H., Yang, Z. (2020), Machine learning solutions to challenges in finance: An application to the pricing of financial products. Technological Forecasting and Social Change, 153, 119928.

Ghoddusi, H., Creamer, G.G., Rafizadeh, N. (2019), Machine learning in energy economics and finance: A review. Energy Economics, 81, 709-727.

Goodell, J.W., Kumar, S., Lim, W.M., Pattnaik, D. (2021), Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance, 32, 100577.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., Evans, O. (2018), Viewpoint: When will ai exceed human performance? Evidence from ai experts. Journal of Artificial Intelligence Research, 62, 729-754.

Gunning, D., Aha, D.W. (2019), DARPA's explainable artificial intelligence program. AI Magazine, 40, 44-58.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.Z. (2019), XAI-Explainable artificial intelligence. Science Robotics, 4, eaay7120.

Gupta, O., Raskar, R. (2018), Distributed learning of deep neural network over multiple agents. Journal of Network and Computer Applications, 116, 1-8.

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.

Hajian, S., Bonchi, F., Castillo, C. (2016), Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDD), Association for Computing Machinery. p2125-2126.

Halepoto, H., Gong, T., Memon, H. (2022), A Bibliometric analysis of antibacterial textiles. Sustainability, 14, 11424.

Han, Z., Jiang, C., Hu, F. (2019), The current status, hotspots and prospects of cross-cultural human resource management-visual analysis based on citespace. Journal of Global Economy, Business and Finance, 4(7), 1-7.

Hansen, K.B., Borch, C. (2021), The absorption and multiplication of uncertainty in machine-learning-driven finance. British Journal of Sociology, 72, 1015-1029.

Hariri, R.H., Fredericks, E.M., Bowers, K.M. (2019), Uncertainty in big data analytics: Survey, opportunities, and challenges. Journal of Big Data, 6, 44.

Harrigan, M.P., Sung, K.J., Neeley, M., Satzinger, K.J., Arute, F., Arya, K., Atalaya, J., Bardin, J.C., Barends, R., Boixo, S., Broughton, M., Buckley, B.B., Buell, D.A., Burkett, B., Bushnell, N., Chen, Y., Chen, Z., Ben Chiaro, C., Collins, R., & Babbush, R. (2021), Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. Nature Physics, 17, 332-336.

Hernández-Lobato, J.M., Hoffman, M.W., Ghahramani, Z. (2014), Predictive entropy search for efficient global optimization of black-box functions. In: Ghahramani, Z., Welling, M., Lawrence, N.D., Cortes, C., Weinberger, K.Q., editors. 28th Annual Conference on Neural Information Processing Systems 2014, Neural Information Processing Systems Foundation. p918-926.

Hirschberg, J., Manning, C.D. (2015), Advances in natural language processing. Science, 349, 261-266.

Huang, C.F. (2012), A hybrid stock selection model using genetic algorithms and support vector regression. Applied Soft Computing Journal, 12, 807-818.

Jadhav, S., He, H., Jenkins, K. (2018), Information gain directed genetic algorithm wrapper feature selection for credit rating. Applied Soft Computing Journal, 69, 541-553.

Jain, A., Hautier, G., Ong, S.P., Persson, K. (2016), New opportunities for materials informatics: Resources and data mining techniques for uncovering hidden relationships. Journal of Materials Research, 31, 977-994.

Jayeola, O., Sidek, S., Sanyal, S., Hasan, S.I., An, N.B., Ajibade, S.S.M., Phan, T.T.H. (2022), Government financial support and financial performance of SMEs: A dual sequential mediator approach. Heliyon, 8, e11351.

Kim, S.H., Kim, D. (2014), Investor sentiment from internet message postings and the predictability of stock returns. Journal of Economic Behavior and Organization, 107, 708-729.

Kim, S.J., Boyd, S. (2008), A minimax theorem with applications to machine learning, signal processing, and finance. SIAM Journal on Optimization, 19, 1344-1367.

Krauss, C., Do, X.A., Huck, N. (2017), Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259, 689-702.

Krishnan, K. (2013), Data Warehousing in the Age of Big Dat. Lithgow: Newnes.

Krittanawong, C., Johnson, K.W., Rosenson, R.S., Wang, Z., Aydar, M., Baber, U., Min, J.K., Wilson Tang, W.H., Halperin, J.L., Narayan, S.M. (2019), Deep learning for cardiovascularmedicine: A practical primer. European Heart Journal, 40, 2058-2069.

Kumar, A., Sharma, S., Mahdavi, M. (2021), Machine learning (Ml)

technologies for digital credit scoring in rural finance: A literature review. Risks, 9, 192.
Kumar, S. (2015), Co-authorship networks: A review of the literature. Aslib Journal of Information Management, 67, 55-73.
Laudel, G. (2006), The art of getting funded: How scientists adapt to their funding conditions. Science and Public Policy, 33, 489-504.
LI, B., Hoi, S.C.H. (2014), Online portfolio selection: A survey. ACM Computing Surveys, 46, 1-36.
Liang, D., Lu, C.C., Tsai, C.F., Shih, G.A. (2016), Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. European Journal of Operational Research, 252, 561-572.
Liang, D., Tsai, C.F., Wu, H.T. (2015), The effect of feature selection on financial distress prediction. Knowledge-Based Systems, 73, 289-297.
Lin, W.Y., Hu, Y.H., Tsai, C.F. (2012), Machine learning in financial crisis prediction: A survey. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 42, 421-436.
Lundberg, J., Tomson, G., Lundkvist, I., SK?r, J., Brommels, M. (2006), Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. Scientometrics, 69, 575-589.
Ma, X., Lv, S. (2019), Financial credit risk prediction in internet finance driven by machine learning. Neural Computing and Applications, 31, 8359-8367.
Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., Niu, X. (2018), Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electronic Commerce Research and Applications, 31, 24-39.
Madiajagan, M., Raj, S.S. (2019), Parallel computing, graphics processing unit (GPU) and new hardware for deep learning in computational intelligence research. In: Deep Learning and Parallel Computing Environment for Bioengineering Systems. Amsterdam: Elsevier.
Malekipirbazari, M., Aksakalli, V. (2015), Risk assessment in social lending via random forests. Expert Systems with Applications, 42, 4621-4631.
Manchikanti, L., Kaye, A.M., Boswell, M.V., Hirsch, J.A. (2015), Medical journal peer review: Process and bias. Pain Physician, 18, E1-E14.
Mitchell, T.M., Mitchell, T.M. (1997), Machine Learning. New York: McGraw-Hill.
Nelson, D.M.Q., Pereira, A.C.M., De Oliveira, R.A. (2017), Stock market's price movement prediction with LSTM neural networks. In: 2017 International Joint Conference on Neural Networks, (IJCNN), Institute of Electrical and Electronics Engineers Inc. p1419-1426.
Nguyen, Q., Diaz-Rainey, I., Kuruppuarachchi, D. (2021), Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. Energy Economics, 95, 105129.
Nyakuma, B.B., Wong, S., Mong, G.R., Utume, L.N., Oladokun, O., Wong, K.Y., Ivase, T.J.P., Abdullah, T.A.T. (2021), Bibliometric analysis of the research landscape on rice husks gasification (1995-2019). Environmental Science and Pollution Research, 28, 49467-49490.
Orús, R., Mugel, S., Lizaso, E. (2019), Quantum computing for finance: Overview and prospects. Reviews in Physics, 4, 100028.
Pagolu, V.S., Reddy, K.N., Panda, G., Majhi, B. (2017), Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. In: 2016 IEEE International Conference on Signal Processing, Communication, Power and Embedded System, (SCOPES), Institute of Electrical and Electronics Engineers Inc. p1345-1350.
Patel, J., Shah, S., Thakkar, P., Kotecha, K. (2015), Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications, 42, 259-268.
Qian, J., Law, R., Wei, J. (2019), Knowledge mapping in travel website studies: A scientometric review. Scandinavian Journal of Hospitality and Tourism, 19, 192-209.
Rabbi, F., Ayaz, M., Dayupay, J.P., Oyebode, O.J., Gido, N.G., Adhikari, N., Tabuena, A.C., Ajibade, S.S.M., Bassey, M.A. (2022), Gaussian Map to Improve Firefly Algorithm Performance. In: 2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC), IEEE. p88-92.
Rhif, M., Abbes, A.B., Farah, I.R., Martínez, B., Sang, Y. (2019), Wavelet transform application for/in non-stationary time-series analysis: A review. Applied Sciences, 9, 1345.
Rogers, G., Szomszor, M., Adams, J. (2020), Sample size in bibliometric analysis. Scientometrics, 125, 777-794.
Rundo, F., Trenta, F., Di Stallo, A.L., Battiato, S. (2019), Machine learning for quantitative finance applications: A survey. Applied Sciences, 9, 5574.
Sampaio, R.B., Fonseca, M.V.D.A., Zicker, F. (2016), Co-authorship network analysis in health research: Method and potential use. Health Research Policy and Systems, 14, 34.
SAP. (2022), What is Big Data? USA: SAP Software Inc. Available from: https://bit.ly/3vtj0tq [Last accessed on 2022 Oct 14].
Scholten, W., Franssen, T.P., Van Drooge, L., De Rijcke, S., Hessels, L.K. (2021), Funding for few, anticipation among all: Effects of excellence funding on academic research groups. Science and Public Policy, 48, 265-275.
Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M. (2020), Financial time series forecasting with deep learning: A systematic literature review: 2005-2019. Applied Soft Computing, 90, 106181.
Shen, F., Chao, J., Zhao, J. (2015), Forecasting exchange rate using deep belief networks and conjugate gradient method. Neurocomputing, 167, 243-253.
Siami-Namini, S., Tavakoli, N., Siami Namin, A. (2019), A Comparison of ARIMA and LSTM in Forecasting Time Series. In: Wani, M.A., Sayed-Mouchaweh, M., Lughofer, E., Gama, J., Kantardzic, M., editors. 17th IEEE International Conference on Machine Learning and Applications, (ICMLA), Institute of Electrical and Electronics Engineers Inc. p1394-1401.
Singh, R., Srivastava, S. (2017), Stock prediction using deep learning. Multimedia Tools and Applications, 76, 18569-18584.
Stellato, B., Banjac, G., Goulart, P., Bemporad, A., Boyd, S. (2020), OSQP: An operator splitting solver for quadratic programs. Mathematical Programming Computation, 12, 637-672.
Subasi, A. (2020), Practical Machine Learning for Data Analysis Using Python. Cambridge: Academic Press.
Tsai, C.F. (2014), Combining cluster analysis with classifier ensembles to predict financial distress. Information Fusion, 16, 46-58.
Tsai, C.F., Wu, J.W. (2008), Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Systems with Applications, 34, 2639-2649.
Ubfal, D., Maffioli, A. (2011), The impact of funding on research collaboration: Evidence from a developing country. Research Policy, 40, 1269-1279.
Vaidya, S., Suri, A., Batla, V., Keshta, I., Ajibade, S.S.M., Safarov, G. (2023), A computer-aided feature-based encryption model with concealed access structure for medical Internet of Things. Decision Analytics Journal, 7, 100257.
Van Eck, N., Waltman, L. (2010), Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84, 523-538.
Wong, S.L., Mong, G.R., Nyakuma, B.B., Ngadi, N., Wong, K.Y., Hernández, M.M., Armenise, S., Chong, C.T. (2022), Upcycling of plastic waste to carbon nanomaterials: A bibliometric analysis (2000-

2019). Clean Technologies and Environmental Policy, 24, 739-759.
Woolf, B.P. (2009), Chapter 7: Machine learning. In: Building Intelligent Interactive Tutors. San Francisco, CA: Morgan Kaufmann. p221-297.
Worldatlas. (2022), The World's Largest Economies. United States of America: WA Publishers Inc, USA. Available from: https://bit.ly/3THuTZx [Last accessed on 2022 Oct 14].
Xu, J., Ding, Y., Wang, X., Wu, Y. (2008), Genre Identification of Chinese Finance Text Using Machine Learning Method. 2008 IEEE International Conference on Systems, Man and Cybernetics, (SMC), Singapore. p455-459.
Zaidi, A., Ajibade, S.S.M., Musa, M., Bekun, F.V. (2023), New insights into the research landscape on the application of artificial intelligence in sustainable smart cities: A bibliometric mapping and network analysis approach. International Journal of Energy Economics and Policy, 13, 287-299.
Zhang, Y., Huang, J., Du, L. (2017), The top-cited systematic reviews/meta-analyses in tuberculosis research: A PRISMA-compliant systematic literature review and bibliometric analysis. Medicine (Baltimore), 96, e4822.
Zhang, Y., Liu, R., Heidari, A.A., Wang, X., Chen, Y., Wang, M., Chen, H. (2021), Towards augmented kernel extreme learning models for bankruptcy prediction: Algorithmic behavior and comprehensive analysis. Neurocomputing, 430, 185-212.
Zheng, Z., Zhu, J., Lyu, M.R. (2013), Service-generated big data and big data-as-a-service: An overview. In: 2013 IEEE International Congress on Big Data, IEEE, p403-410.
Zhu, Y., Xie, C., Wang, G.J., Yan, X.G. (2016), Predicting China's SME credit risk in supply chain finance based on machine learning methods. Entropy, 18, 195.
Zhu, Y., Xie, C., Wang, G.J., Yan, X.G. (2017), Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. Neural Computing and Applications, 28, 41-50.
Zhu, Y., Zhou, L., Xie, C., Wang, G.J., Nguyen, T.V. (2019), Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach. International Journal of Production Economics, 211, 22-33.