Advances in Consumer Research
https://acr-journal.com/
Volume-3 | Issue-10 |Sep: 2026

Original Researcher Article

Mapping the Evolution of Stock Market Prediction Research Using Artificial Intelligence and Machine Learning: A Bibliometric and Systematic Review (2020–2026)

Dr. Shilpi Jauhari[1], Dr. S M Qasim[2], Dr. Jolly Rastogi[3], Dr. Ruchi Misra[4]
[1] Associate Professor, IET Lucknow, India.
Email: sjauhari.mba.cf@ietlucknow.ac.in
[2] Assistant Professor, IET Lucknow, India.
Email: smqasim.mba.cf@ietlucknow.ac.in
[3] Assistant Professor, IMS Lucknow, India.
Email: jollyrastogi01@gmail.com
[4] Assistant Professor, IMS Lucknow, India.
Email: sai.ruchi.j@gmail.com.

ABSTRACT
Global stock markets represent vast repositories of wealth, and as economies worldwide become increasingly market oriented, the ability to forecast stock performance has gained critical importance for investors, policymakers, and researchers alike. In accordance with 380 research papers indexed in the Scopus database between 2020 and 2026, this study delivers a bibliometric impression of enquiry on stock market prediction exploiting machine learning (ML) and artificial intelligence (AI) methods. The review of the most-cited studies highlights the predominance of deep learning designs particularly Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and hybrid models while ensemble and reinforcement learning methods demonstrate superior predictive accuracy and adaptability. Moreover, fusion models integrating numerical and sentiment data consistently outperform single-source approaches. The bibliometric mapping, conducted using VOSviewer, delivers all-inclusive impression of the field's intellectual structure, research impact, and global collaboration patterns. Expert Systems with Applications emerged as the most productive and influential journal. A strong China-United States–India collaboration axis was observed, with most research contributions originating from computer science departments. Thematic mapping revealed key research clusters and underscored gaps in risk-adjusted returns and portfolio-level optimization, indicating promising directions for future research and practical implementation....

Keywords: Stock market prediction; Artificial intelligence; Machine learning; Bibliometric analysis; Visualization of Similarities (VOS) Viewer.

INTRODUCTION

Stock market prediction broadly defined as the estimation of future prices of publicly traded assets, continues to be a significant problem of interest to researchers and industry professionals in finance, economics, and computer science. As global economies have become increasingly market-oriented, accurate stock forecasting has gained strategic importance for investors, policymakers, and institutions seeking to manage uncertainty and maximize returns. Predicting stock market movements remains a challenging problem because financial time series are inherently noisy, volatile, and nonlinear. (Tsay, 2010). Accurate forecasting requires the analysis of historical market data, financial news, and their influence on price trends and market behavior (Usmani & Shamsi, 2021; Ayyappa & Kumar, 2024). Traditional forecasting techniques were rooted in technical analysis and fundamental analysis, complemented by statistical models like time-series decomposition, exponentially smoothed modeling, and the Box-Jenkins Autoregressive Integrated Moving Average Model (ARIMA) framework (Box & Jenkins, 1976). While these models offered valuable insight into linear patterns, their ability to generalize under volatile market conditions proved limited. The viability of prediction was further called into question by the development of the Efficient Market Hypothesis (EMH) proposed by Fama (1965) and the Malkiel's Random Walk theory (1973) which contend that prices accurately reflect all readily accessible information and change at random. Yet, subsequent developments in behavioral finance revealed systematic departures from rationality. The prospect theory introduced by Kahneman and Tversky (1979) along with the wo0rk of Shiller (1981) on excess volatility demonstrated that investor psychology and cognitive biases significantly influence market movements, creating opportunities for pattern discovery beyond the assumptions of EMH.

Advances in computational power and data availability have since transformed stock prediction research. The Autoregressive Integrated Moving Average Model (ARIMA) and its nonlinear extensions inspired the rise of computational intelligence approaches, particularly artificial neural networks (ANNs) and recurrent neural networks (RNNs)**,** which captured time-based dependencies in chronological data. Further, the assimilation of machine learning (ML) and deep learning (DL**)** has reshaped the field, making it possible to describe intricate nonlinear interactions between textual, sentiment-based, and numerical variables. Contemporary research increasingly treats stock forecasting as a classification problem, predicting market direction rather than exact price magnitudes. Hybrid and ensemble approaches integrating econometric techniques with machine learning algorithms, such as Support Vector Machines (SVMs), Random Forests, and neural networks, have demonstrated substantial improvements in predictive performance (Kumar & Thenmozhi, 2014; Patel et al., 2015). More recently, deep learning architectures including Convolutional Neural Networks (CNNs), Gated Recurrent Units (GRUs), and Long Short-Term Memory (LSTM) networks have emerged as dominant methods in stock market prediction., while hybrid and transformer-based models continue to show promise in integrating multi-modal data, including financial news and social media sentiment (Fischer & Krauss, 2018; Chen et al., 2021; Tao et al., 2024; Ma et al., 2024).

Parallel to these methodological advancements, recent bibliometric analyses have begun mapping the intellectual landscape of stock market prediction. Using tools such as VOSviewer, scholars have examined publication patterns, co-authorship networks, and thematic clusters to uncover emerging research trends (Bal & Mishra, 2025; Khadka & Karki, 2024; Garg & Tiwari, 2021). These studies demonstrate the increasing prevalence of deep learning methods, the expanding impact of AI-driven strategies, and the growing incorporation of sentiment data into prediction models. However, much of the existing work remains concentrated on algorithmic performance rather than on practical financial outcomes.

1.1 Rationale of the Study

While the body of research on AI and ML-based stock market prediction has expanded rapidly, critical gaps continue to persist. Limited attention has been given to risk-adjusted returns, portfolio-level optimization, and the translation of predictive accuracy into actionable investment strategies. Moreover, the field lacks a comprehensive understanding of how regional and institutional collaborations influence the evolution of knowledge. Addressing these shortcomings, the current study behaviors a systematic bibliometric examination of research on stock market prediction by means of artificial intelligence and machine learning techniques from 2020 to 2026. Focusing on 2020–2026 allows the study to encompass the latest advancements in AI and ML applications to predict the stock market, particularly the incorporation of neural networks, ensemble learning, and reinforcement learning approaches that have transformed the research landscape in recent years. Importantly, no prior literature review or bibliometric analysis has been conducted with the same breadth and methodological rigor as the present study, which systematically analyzes 25 most cited research papers spanning this period. By doing so, the study provides novel insights into research trends, collaborative networks, and thematic evolution within this emerging domain. The analysis is guided by overarching research questions that seek to understand:

RQ1: What are the major contributions and findings of the most-cited articles employing AI techniques in stock market prediction between 2020-2026?

RQ2: Which authors, institutions, and nations demonstrate the highest research productivity in this field?

RQ3: Which studies and authors exert the greatest intellectual influence as revealed through co-citation analysis?

RQ4: What are the most frequently used keywords and emerging research themes in AI-driven stock prediction?

RQ5: What collaborative patterns exist between authors, organizations, and nations, along with how have they changed throughout time?

By analysing 380 Scopus indexed articles**,** this paper purposes to make available arranged overview of the research setting, identify intellectual clusters, and highlight gaps and opportunities for future exploration. The findings are expected to:

Reveal dominant research themes and influential works;

Map collaboration networks and geographic distributions;

Offer evidence-based insights into evolving AI paradigms in financial forecasting; and

Guide researchers and practitioners toward promising guidelines for future investigation.

2. RESEARCH METHODOLOGY

2.1 Tools and Techniques

Bibliometric analysis serves as a quantitative approach to assess different aspects of publications within a particular field (Broadus, 1987; Pritchard, 1969). This method is considered more objective than subjective alternatives (Lim and Zhang, 2022) and is crucial for analyzing publication data to reveal significant patterns, highly referenced works, prominent authors, and the most persuasive institutions and nations (Glanzel et al., 2019; Donthu et al., 2021). By providing the viewpoint into past progression of a field, recent research drifts, and way forward, bibliometric exploration enhances the understanding of the research landscape and has recently gained prominence in enterprise research (Donthu et al., 2021; Khan et al., 2022).

According to Donthu et al. (2021), bibliometric examination usually consists of two complimentary elements: scientific mapping and performance analysis. While scientific mapping uses network linkages to examine the intellectual and thematic structure of the subject, performance analysis finds the most prolific and influential components (e.g., authors, journals, nations).

Five principal bibliometric techniques were employed: Citation Analysis to assess the impact of publications and authors while Co-Citation Analysis to reveal conceptual associations surrounded by works cited together. Bibliographic Coupling to identify intellectual similarities among papers that share common references. Another important co-authorship Analysis to map collaboration patterns among authors, institutions, and countries while Co-occurrence Analysis to detect frequently used keywords and emerging research themes.

The advancement of computers and internet technologies has led to the widespread adoption of these techniques, which are supported by specialized software. For this study, the bibliometric analysis was performed using VOS viewer software, industrialized by Van Eck and Waltman (2010), for its user-friendly interface, capability to handle large datasets, and advanced visualization features (Donthu et al., 2021). VOS viewer provides an easy-to-interpret pictorial representation of scientometric data through several map formats, including network visualization that illustrates relationships between items like authors or keywords and intersection visualization that highpoints the progressive evolution of research themes. Another important density Visualization represents the concentration of research activity in specific topics. These tools and techniques offer a comprehensive, data-driven view of the intellectual landscape, enabling a thorough understanding of the evolution, influence, and collaborative structures within the field of research.

To complement these quantitative insights, a Systematic Literature Review (SLR) was also conducted from the set of the 25 articles with highest citation impact identified through the bibliometric analysis. The SLR served as a qualitative extension of the study, enabling a deeper examination of methodologies, datasets, algorithms, and research gaps. Transparency and repeatability were ensured by adhering the PRISMA 2020 guidelines (Page et al., 2021). Together, the bibliometric and SLR techniques provide both macro-level mapping and micro-level interpretation, resulting in a holistic understanding of the intellectual landscape in stock market prediction research using AI and ML methods.

Building upon these methodological foundations, the subsequent section outlines the process of data collection and selection, detailing the steps followed to retrieve, refine, and prepare the Scopus dataset for bibliometric and systematic review analysis.

2.2 Data Collection

The study adopts a bibliometric methodology to systematically examine global research inclinations in predicting the stock market by means of artificial intelligence (AI) and machine learning (ML) techniques. Bibliometric examination applies quantitative methods to evaluate scientific output, identify influential contributions, and map intellectual structures within a given field (Cobo et al., 2011; Aria & Cuccurullo, 2017; Zupic & Cater, 2015). The bibliometric data library was retrieved from the Scopus repository**,** that is widely recognized for its extensive wide-ranging reportage, rigorous indexing standards, and high-quality metadata (Baas et al., 2020; Aksnes & Sivertsen, 2019). Compared with other data repositories like Web of Science (WoS), Google Scholar, and PubMed, Scopus offers broader disciplinary diversity, greater citation linkage, and richer bibliographic information making it particularly suitable for multidisciplinary analyses involving finance, economics, computer science, and decision sciences (Cortegiani et al., 2020). Table 1 depicts the primary steps that were taken from the pertinent publications and ought to be used in bibliometric search.

Table 1: Bibliometric research steps

Step	**Description / Criterion**	**Details and Outcomes**
Step 1. Scope of Investigation	Definition of research domain and time frame	Stock Market Prediction Using Artificial Intelligence and Machine Learning Techniques (2020–2026): A Bibliometric and Systematic Literature Review
Step 2. Database Selection	Primary source of bibliometric data	**Scopus**: Selected for its comprehensive multidisciplinary coverage, rigorous indexing standards, and enriched metadata suitable for large-scale bibliometric and citation analysis.
Step 3. Document Selection	First Selection Criterion: Search String and Time Interval	TITLE-ABS-KEY((“stock market prediction” OR “stock market forecasting” OR “share market movement” OR “share market analysis”) AND (“machine learning” OR “artificial intelligence” OR “AI” OR “deep learning”)). **Result:** 1,710 documents. **Time Interval:** 2020–2026, Result: 1449 documents.

	Second Selection Criterion: Subject Categories	Business, Management and Accounting; Computer Science; Decision Sciences; and Economics, Econometrics and Finance, Social Sciences **Result:** 1,312 documents
	Third Selection Criterion: Document Type and Language	Only Journal Articles considered (excluding conference papers, book chapters, etc.) in English Language **Result:** 401 documents
	Fourth Selection Criterion: Publication Stage	Only English-language and final published journal articles included **Result:** 380 documents (final dataset)
Step 4. Document Processing	Data export and preparation for analysis	Data exported from Scopus in CSV and plain-text formats; duplicates removed, synonymous terms merged, and author/institution names standardized to ensure dataset consistency. **Result:** 380 documents
Step 5. Analysis and Inference of Results	Analytical software and techniques	Bibliometric analysis performed using VOSviewer for visualization and mapping; included citation, co-citation, co-authorship, and keyword co-occurrence analyses to identify research performance, structure, and thematic trends.
Step 6. Systematic Literature Review (SLR)	Qualitative synthesis of influential studies	Based on the bibliometric dataset, a systematic literature review was carried out on the 25 most frequently cited articles. Each article was evaluated for objectives, datasets, analytical models, findings, and contribution to provide interpretive depth complementing the quantitative bibliometric results.

3.0 RESULTS

The results of the bibliometric study are presented in this part, providing a thorough summary of the state of the research on machine learning and artificial intelligence (AI) methods for stock market prediction. The results are organized across multiple dimensions to capture the breadth of academic contributions and the intellectual evolution of this domain.

3.1 Annual Publication Trend

A useful indicator of research development in any discipline of study is the quantity of papers published annually. The distribution of publication counts may provide insights into the expected trajectory of future research. To examine research trends, a combined visualization was developed showing annual publication counts along with year-over-year percentage change (Figure 1). The bar chart illustrates the yearly publication totals, while the line graph represents the corresponding growth rate. Together, these measures capture both the absolute volume and relative increase in scholarly output within the field.

The fallouts show a consistent mounting trend in publication volume, beginning with 26 publications in 2020, followed by 35 in 2021, 50 in 2022, and 64 in 2023. The research output considerable increase in 2024 with 92 publications and peaked in 2025 with 103 publications. marking a substantial surge of nearly 52% compared to the previous year. Although the count for 2026 currently stands at 10 publications, this apparent dip is likely due to the year still being in progress and many papers not yet indexed. Therefore, rather than indicating a true decline, it reflects the partial coverage of ongoing research output. Overall, the trend line demonstrates sustained research momentum, sustained by a strong coefficient of determination ($R^2 = 0.98$), indicating that time progression accounts for around 69% of the variance in publication growth.

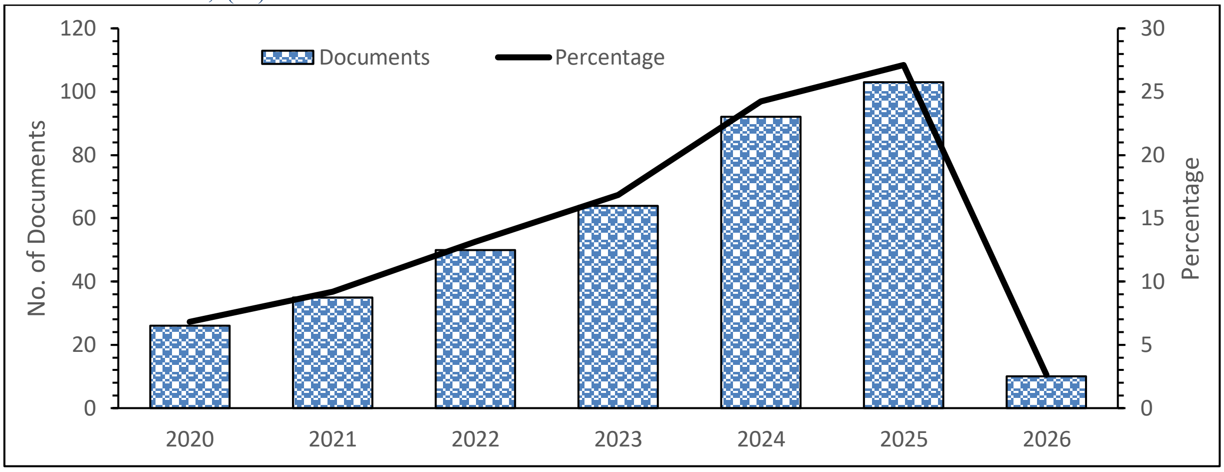

Figure 1 Year-wise publication trend on stock market prediction research using Artificial Intelligence and Machine Learning techniques

The percentage growth line provides additional insight into the volatility of annual research expansion. The highest relative increase occurred between 2023 and 2024, reflecting a surge in academic attention and data-driven experimentation in financial forecasting models. Overall, this dual analysis highlights a robust and maturing research trajectory, driven by technological advancements in deep learning, ensemble modeling, and hybrid AI frameworks. The pattern signifies the transition of AI-driven stock prediction research from an emerging field into a mainstream interdisciplinary domain integrating finance, computer science, and data analytics.

3.2 Prolific Author

Examining the author productivity offers respected insights into the leading contributors shaping the advancing computational approaches to financial market forecasting. To identify the most impactful contributors, the analysis applied inclusion criteria requiring each author to have at least **two publications, more than 50 citations**, and participation in articles with **no more than ten co-authors**. Out of a total of **1086 identified authors**, 30 demonstrated significant research influence and consistent publication output. This filtering process ensured a balanced representation of productivity and scholarly impact. Authors are ranked based on total citation count, a widely accepted indicator of research impact and scholarly influence.

Table shows the top five authors. The most cited authors Adekoya, Adebayo Felix, Ntji, Isaac Kofi, and Weyori, Benjamin Asubam each received 704 citations, significantly surpassing all others. Chang, Victor follows with 446 citations from only two publications, indicating a notably high citation-per-document ratio of 223, which reflects the high impact of his work despite a lower publication count. A steep decline in citation numbers is observed beyond the top three authors, suggesting a concentration of scholarly influence among a few key contributors. The majority of the authors in the list have fewer than 100 citations, which may be attributed to either of their recent publications or lower citation visibility.

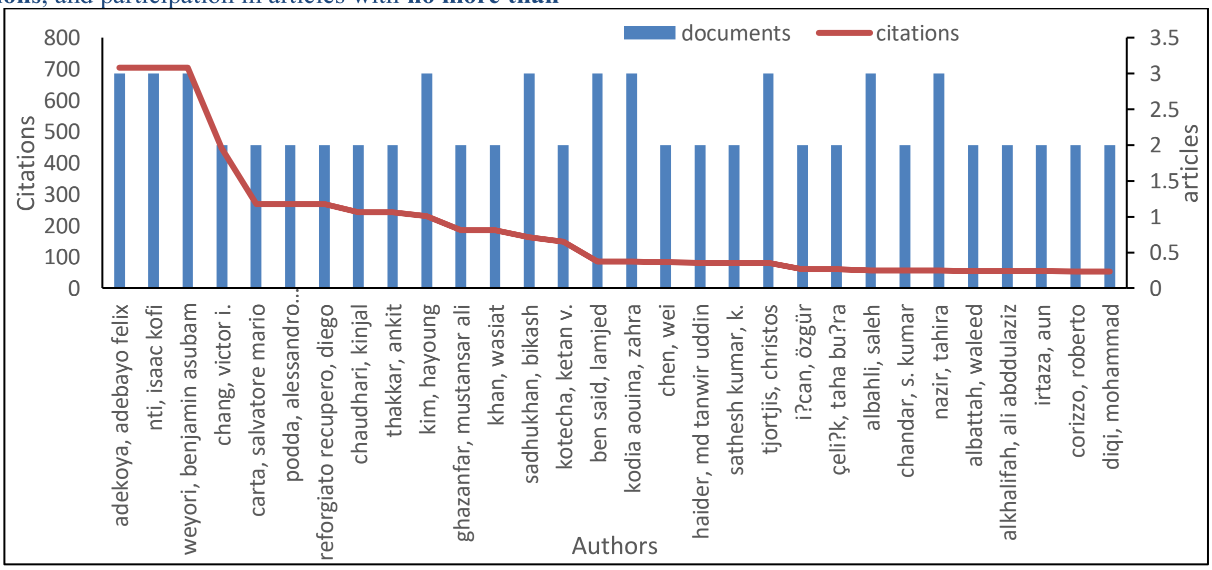

Figure 2 Most prolific authors in AI and ML based stock market prediction research

Notably, a higher volume of publications does not necessarily correlate with higher citation counts. For instance, Nazir, Tahira and Albahli, Saleh, each with three publications, have received relatively low citations (around 54), highlighting the importance of citation-per-document metrics in evaluating research impact.

3.3 Prolific Journal

The dissemination this research domain is primarily concentrated in a few high-impact journals. Analysing these prolific journals helps identify the leading publication venues shaping the intellectual development of this field. The figure 3 includes only those journals that have published three or more articles and have received more than 50 citations, yielding 16 qualifying journals out of a total of 160 publishers analyzed. Journals are ranked by total citation count, providing insight into both the visibility and scholarly impact of their contributions.

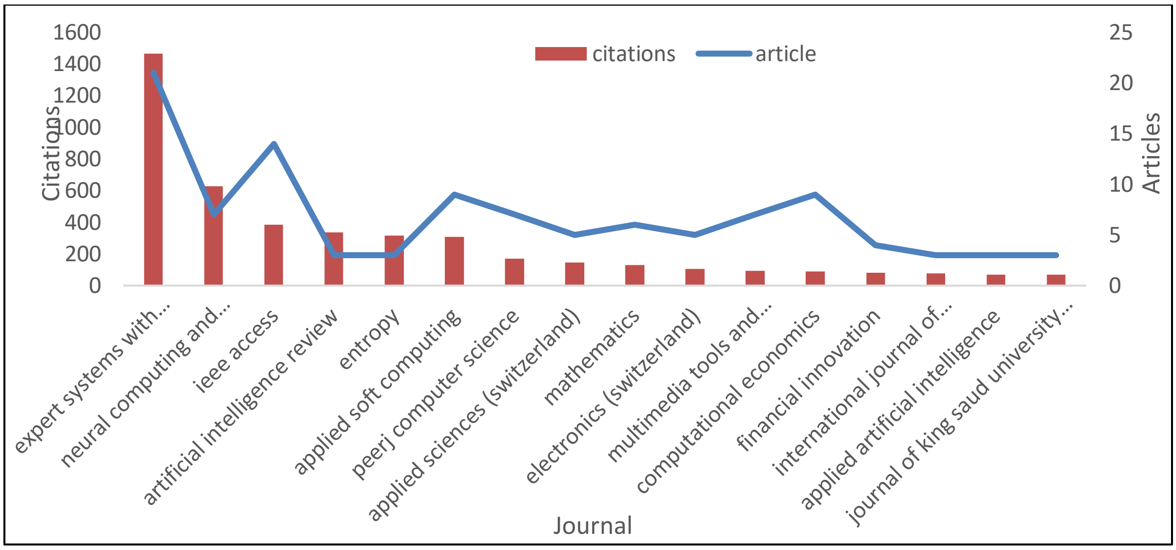

Figure 3 Most prolific journals publishing research on AI- and ML-based stock market prediction

Among the journals, Expert Systems with Applications emerges as the most influential outlet, with 21 articles and a remarkable 1,463 citations, reflecting both substantial research output and high impact. This is followed by Neural Computing and Applications, which, despite publishing only 7 articles, has garnered 629 citations, suggesting a high citation-to-article ratio and indicating that its published work is particularly influential. IEEE Access also demonstrates a strong presence with 14 publications and 389 citations, contributing both quantity and visibility to the field. Notably, journals such as Artificial Intelligence Review and Entropy have only 3 publications each, yet have achieved 337 and 315 citations respectively. This further underscores the point that publication volume does not necessarily correlate with impact, and that journals with fewer, high-quality contributions may exert significant scholarly influence. Conversely, several journals including Applied Sciences, Mathematics, and Electronics (Switzerland) exhibit moderate to high article counts but relatively lower citation totals, which may suggest either emerging research areas, narrower scopes, or more recent publications still gaining traction. Overall, the data highlight the importance of citation-to-article ratio as a complementary metric to publication count when assessing journal impact. While a few titles dominate in both output and citations, many others demonstrate influence through select high-impact publications. This decoration reproduces the specialized nature of the field, where impactful work is often focused in a restricted number of highly relevant journals.

3.4 Prolific Contributor Country

The global distribution of publications reveals that research on AI and ML based stock market prediction is geographically diverse, with notable contributions from both developed and emerging economies. A list of nations with ten or more publications and more than fifty citations in this field is shown in Figure 4. Out of a total of 68 countries, thirteen meet these thresholds. Each country's entirety of citations is used to rank the ranking, which reflects the impact and relevance of research in this domain. Among the listed countries, India leads in terms of the number of publications (119), followed by China (57). China, however, has the most citations (2,599), with India coming in second (1,904). The United States, with only 22 publications, has garnered 1015 citations. Notably, China and the United States exhibit the highest citations per article, with 45.5 and 46.13, respectively, indicating a strong research impact. In contrast, India shows a lower average impact per paper, with 16 citations per article. These citation-to-publication ratios highlight the significant influence and relevance of work originating from China and the United States in this research area.

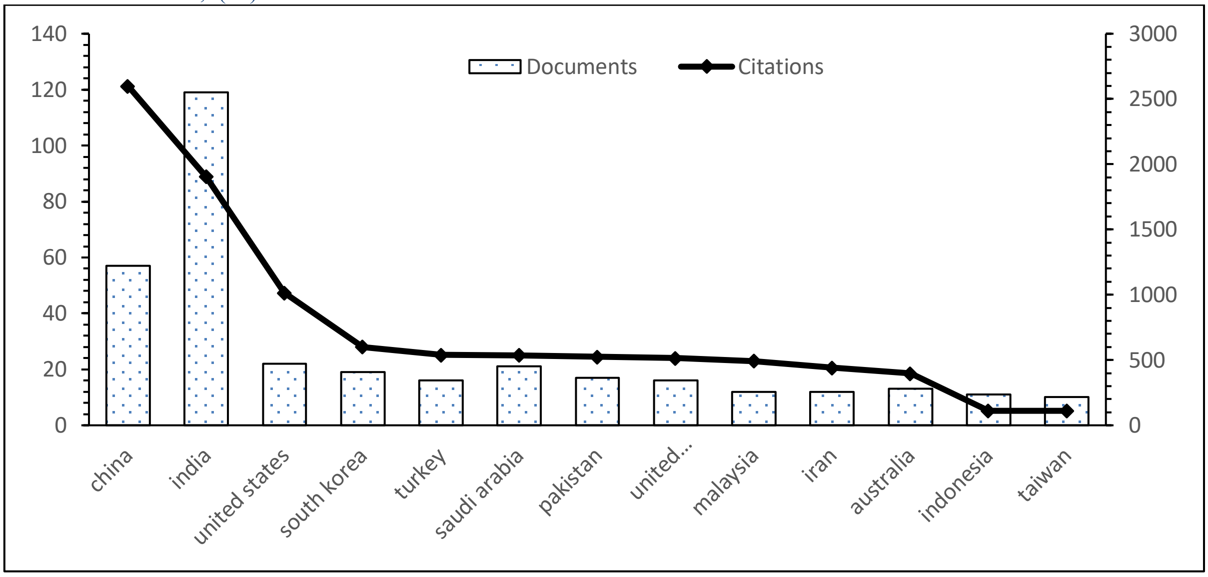

Figure 4 Top contributing countries by total citations and number of publications

4.0 Bibliographic Coupling

Bibliographic coupling refers to measuring how closely two authors are related based on the references they cite in their publications. This approach assumes that references cited together share thematic connections. It reflects similarity in research interests, background literature, and field**.** According to Caputo et al. (2021), strong bibliographic coupling relationships suggest that authors occupy more central positions in the citation network and are more actively engaged in academic discussions. Degree centrality refers to the extent of an author's direct relational ties within a research network (Donthu et al., 2021).

4.1 Bibliographic Coupling Authors

The concentration of bibliographic coupling among authors is presented in Figure 5. Out of 1086 authors, 71 met the criteria of publishing at least two publications and at least five citations.

The table 2 shows the 10 most cited authors. The most cited authors are Adekoya, Adebayo Felix; Nti, Isaac Kofi; and Weyori, Benjamin Asubam, each with 704 citations, three documents and a total link strength of 75.

Table 2: Top bibliographic coupling authors in AI and ML based stock market prediction research

S.No	Author	Documents	Citations	Total link strength
1	adekoya, adebayo felix	3	704	75
2	nti, isaac kofi	3	704	75
3	weyori, benjamin asubam	3	704	75
4	chang, victor i.	2	446	8
5	carta, salvatore mario	2	268	51
6	podda, alessandro sebastian	2	268	51
7	reforgiato recupero, diego	2	268	51
8	chaudhari, kinjal	2	242	34
9	thakkar, ankit	2	242	34
10	kim, hayoung	3	229	46

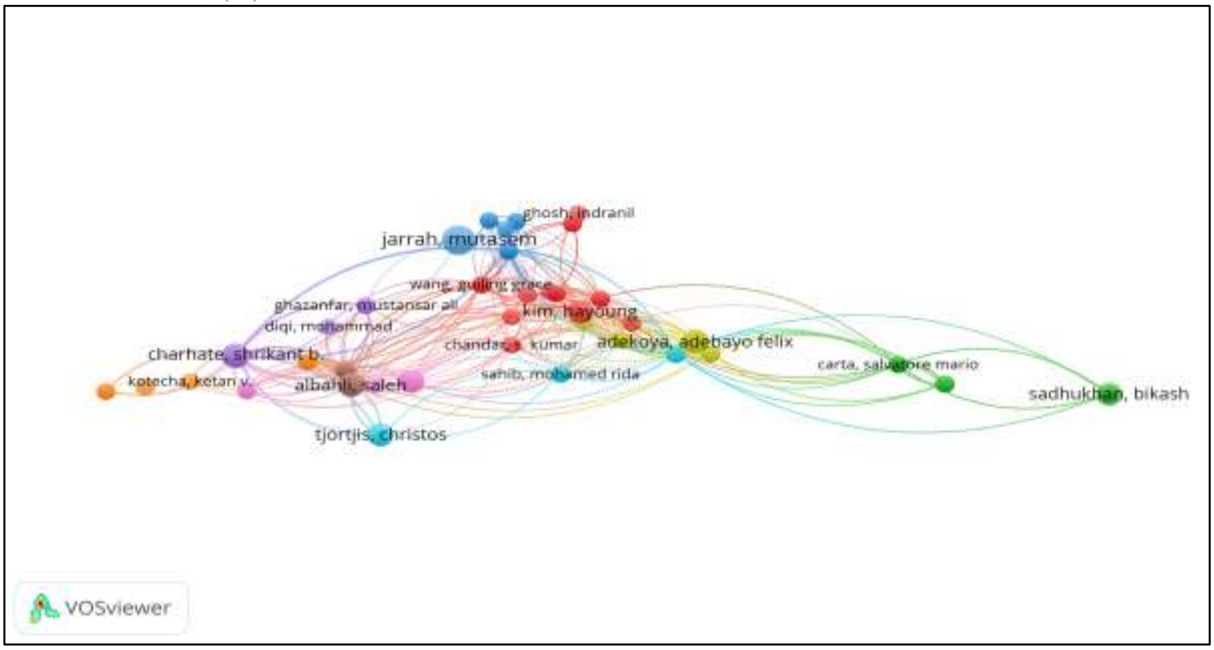

Figure 5 Bibliographic coupling author network in AI and ML based stock market prediction research

The visualization reveals multiple interconnected clusters, indicating the presence of collaborative research groups rather than isolated scholarly efforts. Node size reflects publication influence, while link strength represents co-authorship intensity.

The network demonstrates moderate centralization, with a small set of authors occupying bridging positions between clusters. These scholars function as intellectual connectors, facilitating cross-group knowledge exchange and enhancing interdisciplinary integration. The main research hub represented with, Adekoya, Adebayo Felix and Nti, Isaac Kofi and weyori, benjamin asubam as prominent contributors. The presence of distinct color-coded clusters suggests geographically or institutionally aligned research teams, likely centered around specialized methodological expertise such as deep learning, hybrid forecasting models, or sentiment analysis..

Several clusters appear densely connected internally but loosely linked externally, indicating strong intra-group collaboration yet limited cross-cluster integration. This structural arrangement advocates that while the field is expanding, collaboration remains partially segmented. Such segmentation may reflect differences in disciplinary backgrounds—computer science–oriented groups focusing on algorithmic innovation, and finance-oriented teams emphasizing empirical validation and market applications.

The elongated structure of the network further indicates emerging research fronts extending outward from a central core of established contributors. Peripheral nodes may represent newer entrants or niche specializations, contributing to thematic diversification within the domain.

4.2 Bibliographic Coupling Source

Bibliographic coupling enables the breakdown of relationships between publications, offering insights into the importance and network position of sources within the selected dataset. In this analysis, only sources with at least two publications and a minimum of five citations per publication were considered. Based on these criteria, 53 journals out of 197 met the threshold. A single cluster (fig 6) was identified, comprising all 43 sources with a total link strength of 5,587.

Table 3: Top ten Bibliographic Coupling Sources

S.No	source	documents	citations	total link strength
1	expert systems with applications	23	1762	367
2	neural computing and applications	8	740	55
3	ieee access	16	454	124
4	artificial intelligence review	3	393	41
5	applied soft computing	9	379	88
6	entropy	4	363	49
7	journal of big data	3	303	28

8	information fusion	2	232	37
9	peerj computer science	8	204	108
10	applied sciences (switzerland)	5	199	48

Among the journals listed in Table 3, Expert Systems with Applications emerges as the most influential source, recording 23 publications and 1,762 citations, along with the highest connectivity in the research network, indicating both strong productivity and central positioning. It is followed by IEEE Access, which contributes 16 publications and 454 citations, reflecting its growing role as a multidisciplinary platform for studies involving machine learning and intelligent systems, with comparatively moderate network connectivity.

Neural Computing and Applications ranks among the more highly cited journals with 8 publications and 740 citations, although its network connectivity is relatively lower, suggesting moderate collaborative influence within the field. Similarly, Applied Soft Computing contributes 9 publications and 379 citations, demonstrating its relevance in computational intelligence and optimization research, while maintaining moderate integration within the citation network.

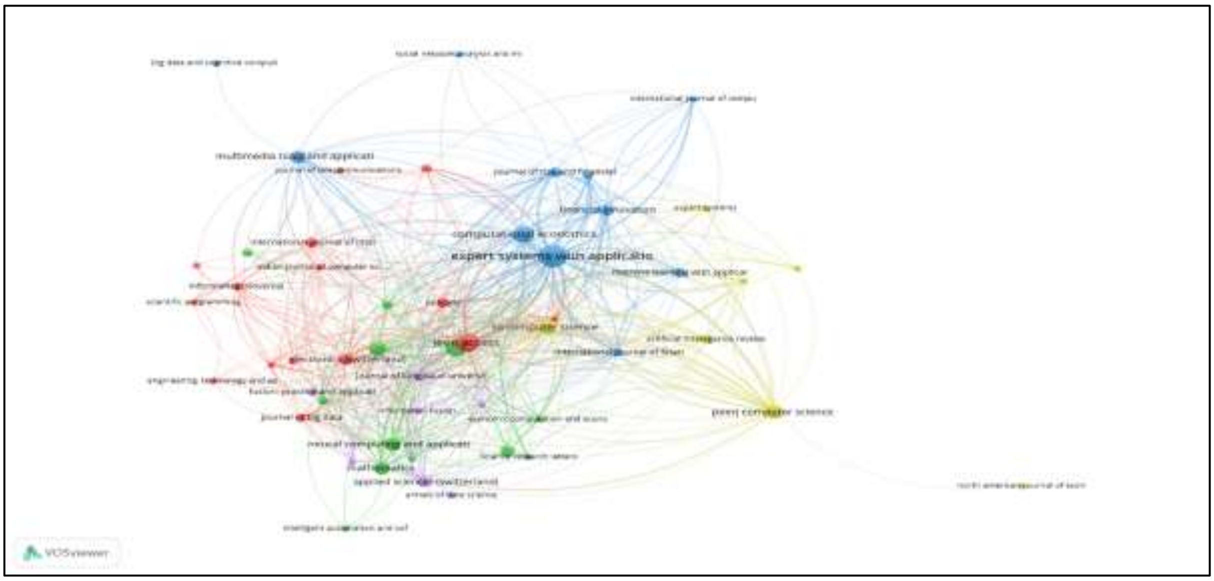

Figure 6 Visualization of bibliographic coupling journals and their citation interconnections in AI and ML driven stock market prediction research

Further, the map (Figure 6) reveals a multi-clustered yet densely interconnected structure, indicating strong cross-disciplinary integration between finance, computer science, and applied artificial intelligence.

At the core of the network, journals such as Expert Systems with Applications, IEEE Access, and Applied Soft Computing occupy structurally central positions, suggesting that methodological innovation in computational intelligence serves as the primary intellectual anchor of the domain. Their centrality reflects the strong influence of algorithmic development and applied machine learning techniques in shaping stock market prediction research.

A second prominent cluster includes finance-oriented outlets such as Journal of Banking and Finance, International Journal of Finance and Economics, and Research in International Business and Finance. The positioning of these journals indicates growing engagement with mainstream financial scholarship, particularly in areas of market efficiency, asset pricing, and empirical financial modeling. However, their slightly peripheral placement relative to computational journals suggests that the field remains methodologically driven rather than theory-led.

A third cluster reflects interdisciplinary and data-driven research streams, including outlets focused on big data, multimedia tools, neural computation, and intelligent systems. The presence of these journals underscores the integration of alternative data analytics, sentiment modeling, and advanced neural architectures into financial forecasting frameworks.

5.0 Co-occurrence of the Author keywords

Keyword co-occurrence examination provides a method of content analysis used to uncover relationships among keywords within a selected corpus of publications. As highlighted by Fakhar Manesh et al. (2021), this approach is particularly effective for identifying thematic clusters, thereby highlighting the core conceptual and theoretical foundations of a research domain. For this study, the minimum occurrence of a keyword was set at five. Out of a total of 959 keywords, 57 met this threshold. The five most frequently occurring keywords were deep learning (122 occurrences), stock market prediction (109 occurrences), machine learning (92 occurrences), stock market (56 occurrences), and sentiment analysis (55 occurrences), (Table 4).

Table 4. **Top five keywords based on Co-occurrences and total link strength**

S.No	Keyword	Occurrences	Total Link Strength
1	Deep learning	122	298
2	Stock market prediction	109	225
3	Machine learning	92	196
4	Stock market	56	148
5	Sentiment Analysis	44	150

Figure 7 presents the cluster density visualization of high-frequency keywords, which provides an enriched representation by displaying keyword density separately for each thematic cluster. Deep learning approaches have gained prominence in stock market prediction, often associated with models such as LSTM and CNN. Traditional machine learning methods, including SVM and random forests, still interact with these deep learning techniques. Emerging topics, such as deep reinforcement learning and genetic algorithms, appear as niche but rapidly growing areas.

The co-occurrence exploration acknowledged four major clusters, each representing a distinct thematic focus:

The **first cluster (red)** is centerd on deep learning approaches, with **"deep learning"** appearing as the most prominent keyword. This cluster includes related terms such as long short-term memory (LSTM), convolutional neural networks, artificial neural networks, feature selection, and stock market forecasting. The strong connectivity among these keywords indicates the growing adoption of advanced neural architectures for modelling complex and nonlinear patterns in financial time-series data.

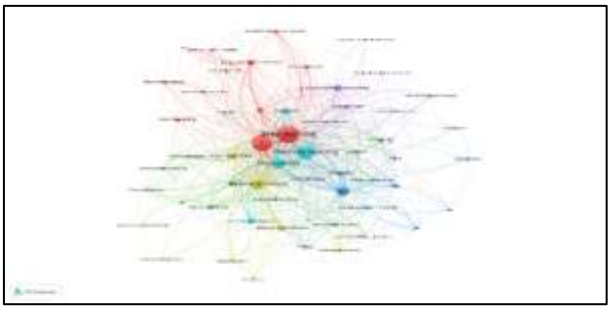

Figure 7 Co-occurrence of Author Keywords

The **second cluster (blue)** focuses on traditional machine learning and predictive modelling techniques. Key terms in this cluster include machine learning, support vector machine, random forest, classification, optimization, and time series. This cluster reflects the significant role of conventional machine learning algorithms in financial forecasting and decision-support systems.

The **third cluster (green)** highlights research integrating behavioural and textual information with financial analysis. Keywords such as sentiment analysis, stock prediction, technical analysis, trading strategies, and reinforcement learning demonstrate the increasing use of sentiment-based indicators and investor behaviour data in predicting market movements.

The **fourth cluster (yellow)** represents studies emphasizing financial forecasting and artificial intelligence applications. Keywords including stock price prediction, artificial intelligence, time series analysis, and finance illustrate the broader analytical frameworks used for modelling and forecasting financial market dynamics.

Overall, the network indicates that **deep learning serves as the central research theme**, closely linked with machine learning techniques, sentiment analysis, and financial forecasting methodologies, reflecting the interdisciplinary evolution of stock market prediction research.

6.0 Co citation Analysis

Co-citation inquiry is a bibliometric system used to measure the relationship between two documents, authors, or journals based on how frequently they are cited together by other publications. This approach helps in identifying the conceptual framework and foundational themes of a particular research field. As emphasized by Ferreira (2018), this enables the recognition of publications that are repeatedly co-cited by several other studies, indicating that these works share meaningful conceptual or methodological connections.

6.1 Co-citation of cited reference

In the present study, a minimum citation threshold of five was applied to cited references. Out of a total of 2,637 cited references, 62 met this threshold. Table 5 presents the ten most frequently cited references with the highest total link strength and citation counts in the domain of stock market prediction research. These references represent the most influential and foundational works that have shaped the intellectual landscape of this research area.

Table 5: Most influential cited references in AI and ML driven stock market prediction research

S. No.	Cited Reference	Citations	Total Link Strength	Key Contribution / Focus
1	George S. Atsalakis, Expert Systems with Applications (2009)	20	48	Survey of soft computing techniques for stock market forecasting
2	Ayodele Ariyo Adebiyi, UKSim Conference (2014)	19	26	Application of ARIMA model for stock price prediction
3	Wei Bao, PLOS ONE (2017)	19	47	Deep learning framework combining stacked autoencoders and LSTM
4	Michel Ballings, Expert Systems with Applications (2015)	18	54	Evaluation of multiple machine learning classifiers for stock price direction prediction
5	Eunsuk Chong, Expert Systems with Applications (2017)	18	32	Deep learning architectures for stock market analysis and prediction
6	Leo Breiman, Machine Learning (2001)	17	42	Introduction of Random Forest ensemble learning algorithm
7	Thomas G. Fischer, European Journal of Operational Research (2018)	16	33	Application of LSTM networks for financial market prediction
8	Hum Nath Bhandari, Machine Learning with Applications (2022)	15	24	LSTM-based stock market index prediction model
9	Johan Bollen, Journal of Computational Science (2011)	15	21	Twitter sentiment analysis for predicting stock market behaviour
10	Rodolfo C. Cavalcante, Expert Systems with Applications (2016)	15	34	Survey of computational intelligence methods in financial markets

Co-citation analysis highlights both the impact and the network connectivity (total link strength) of influential references within the research landscape of AI- and ML-based stock market prediction.

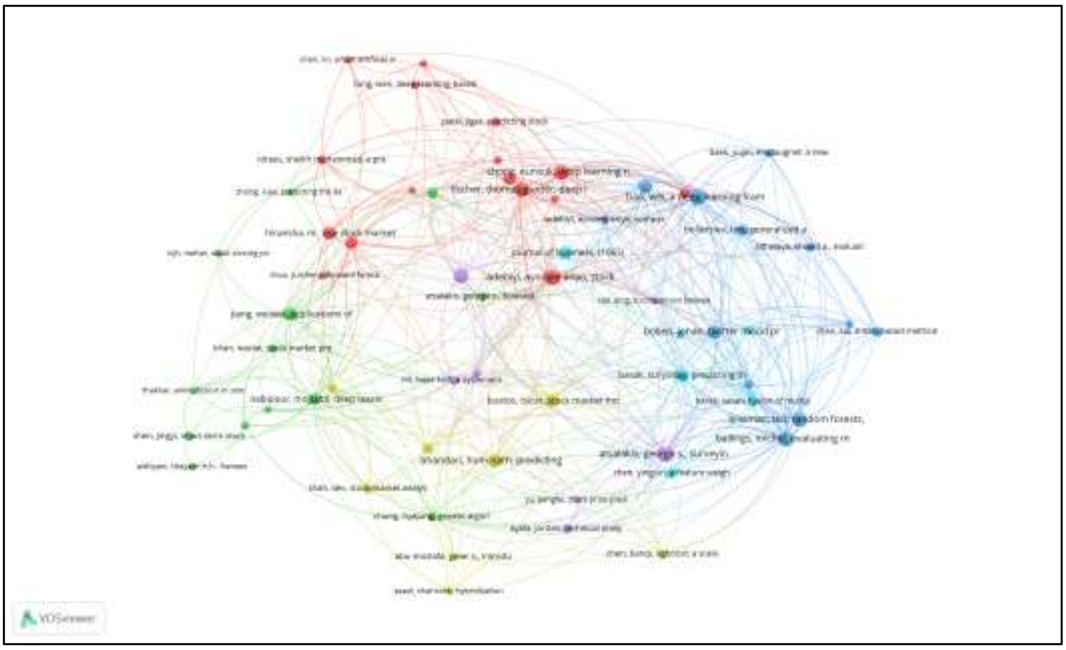

Figure 8 Co-citation network of cited references in AI and ML based stock market prediction

The co-citation table (Table 5) identifies the most influential and frequently co-cited references in the domain, while the corresponding network visualization (Figure 8) graphically depicts their interrelationships and clustering patterns, highlighting distinct thematic areas within the literature. The VOS viewer co-citation map (fig 8) visually groups related references into distinct clusters, each representing a conceptual or methodological sub-domain. The map reveals three prominent clusters, interpreted as follows:

The **first cluster (red)** is dominated by studies focusing on deep learning methods for financial forecasting. Key references in this cluster include works by Eunsuk Chong (2017), Wei Bao (2017), and Thomas G. Fischer (2018), which apply deep neural networks and long short-term memory (LSTM) architectures to predict stock market movements. These studies highlight the growing importance of deep learning models in capturing nonlinear patterns and temporal dependencies in financial time-series data.

The **second cluster (blue)** represents research centered on machine learning algorithms and predictive modelling techniques. Influential works by Michel Ballings (2015), Leo Breiman (2001), and Yingjun Chen (2017) emphasize the application of algorithms such as Random Forest, Support Vector Machines, and ensemble classifiers for stock price direction prediction. These studies provide comparative evaluations of machine learning techniques and serve as methodological benchmarks for later research.

The **third cluster (green)** highlights survey and review-based research, including studies by Weiwei Jiang (2021) and Mahinda Mailagaha Kumbure (2022). These works synthesize existing research on machine learning and deep learning approaches in financial forecasting, identifying methodological trends and research gaps.

The **fourth cluster (yellow)** reflects research focused on time-series modelling and hybrid forecasting frameworks, integrating statistical models with artificial intelligence techniques. These studies emphasize stock price prediction using econometric models, hybrid AI systems, and financial analytics.

Overall, the co-citation structure demonstrates that the field has evolved from traditional econometric and machine learning approaches toward deep learning driven predictive frameworks, with increasing integration of hybrid models and advanced data-driven techniques for financial market forecasting.

6.2 Co-citation of Cited Author

To map the intellectual structure of the stock market prediction research field, a co-citation analysis of authors was conducted using VOS viewer. To focus on the most influential contributors, a minimum citation threshold of 10 was applied. Out of 937 authors, the threshold was met by 12 authors, forming the core set for the network analysis.

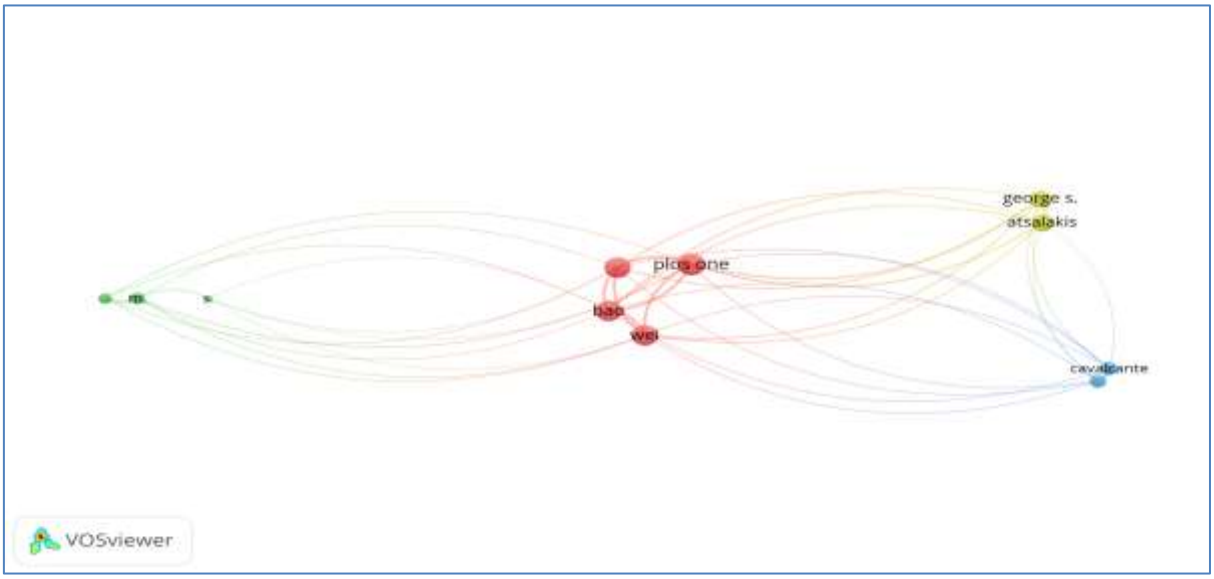

Figure 9 Co-citation of cited authors network in AI and ML based stock market prediction

The source co-citation network generated using VOSviewer illustrates the knowledgeable structure of the research field by recognizing the most frequently co-cited sources within the dataset. In bibliometric visualization, nodes represent journals or sources, node size imitates citation frequency, and the connecting lines designate the strength of co-citation relationships. Different colors denote clusters of sources that share similar citation patterns and research themes.

The network highlights PLOS ONE as a central and influential source within the citation structure. Its position near the center and the multiple connections with other nodes indicate that studies published in or cited alongside this journal significantly add to the development of the research field. The strong links with sources such as Bao and Wei suggest that these works are frequently referenced together, forming part of the core knowledge base of the literature. The visualization also reveals several **distinct clusters**, representing different streams of research. The **central red cluster** contains closely connected sources that form the theoretical and methodological foundation of the field. The strong interconnections within this cluster indicate that researchers consistently rely on these publications when building their studies.

Another cluster includes sources associated with Atsalakis and George S., representing a related research stream that interacts with the core literature. Similarly, the cluster containing Cavalcante and Rodolfo C. forms another group connected to the central cluster, suggesting complementary or interdisciplinary contributions to the field. Additionally, the cluster represented by Hiransha M. indicates an emerging or specialized research direction. Although positioned slightly away from the central nodes, its connections with the core cluster demonstrate integration with the broader scholarly discourse. Overall, the network demonstrates **a** multi-clustered intellectual structure where a central body of literature interacts with several specialized research streams. This pattern highlights the presence of both core foundational studies and evolving research directions, reflecting the dynamic development of knowledge within the field.

7.0 Co-authorship

Co-authorship analysis explores collaborative relationships among researchers and institutions by examining the number of jointly authored publications, with the aim of identifying influential scholars and key collaboration hubs. This approach is also used to reveal prominent academic partnerships across organizations and countries.

7.1 Co-authorship of author

To do the analysis**,** a minimum citation threshold of 5 and minimum number of documents publication of 2 was applied. Out of 1086 authors, the threshold was met by 71 authors, forming the core set for the network analysis. Many of the authors in the network are not connected to each other. The largest set of connected authors consists of five authors.

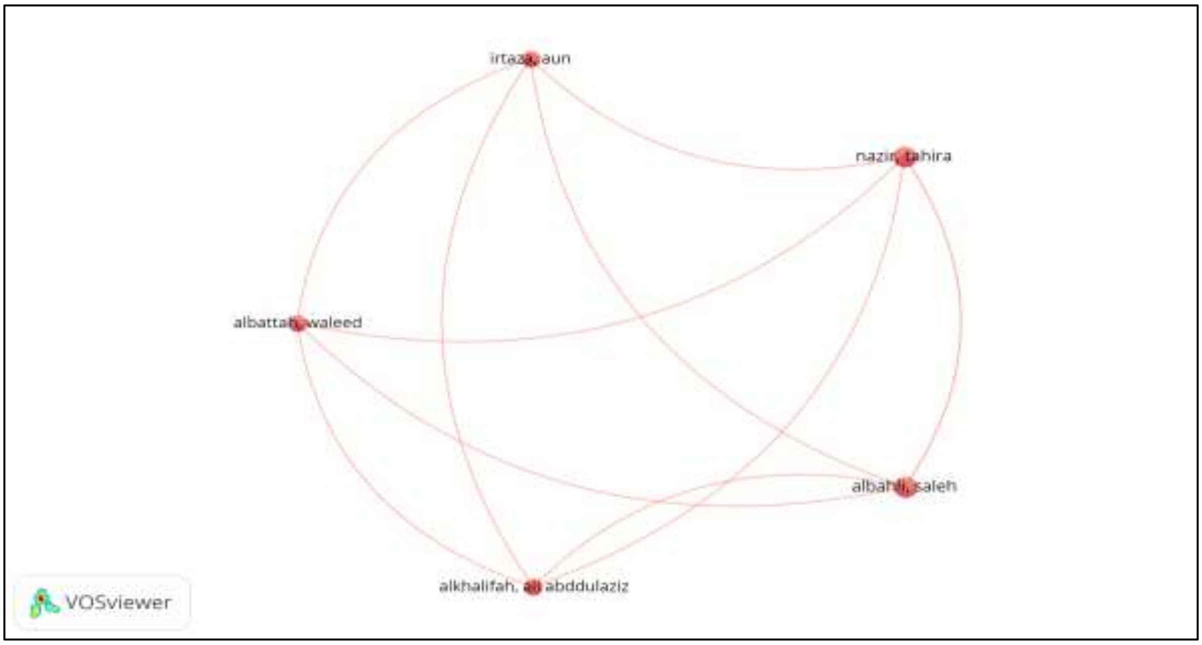

Figure 10 Co-authorship of Author collaboration network in AI and ML based stock market prediction research

The author collaboration network illustrates the pattern of co-authorship relationships among the most active researchers within the dataset. In bibliometric network visualization, apiece node characterizes an author, node size signposts the relative contribution or number of publications, and the connecting lines represent collaborative links between authors through co-authored research outputs. The network shows a compact and closely connected collaboration structure, consisting of a limited number of authors who are actively collaborating within the research domain. The authors visible in the network include Islam Jun**,** Naz Rahra**,** Abbas Saleh**,** Alkhalifa Abdullah**,** and Alshboul Khaled. The relatively equal node sizes suggest that these authors contribute in a balanced manner to the publication output, with no single author overwhelmingly dominating the research collaboration network.

A notable feature of the visualization is the dense connectivity among the authors**,** indicating frequent co-authorship relationships. The presence of multiple links among nearly all nodes suggests that the authors are part of a closely coordinated research group or collaborative research team**.** Within the network, Islam Jun appears positioned near the top of the structure and maintains collaborative links with several other authors. This suggests a central coordinating role, potentially acting as a key contributor or leading collaborator within the group. Similarly, Naz Rahra and Abbas Saleh demonstrate multiple connections within the network, indicating active participation in joint research efforts. The authors Alkhalifa Abdullah and Alshboul Khaled are also strongly connected to the core group, further reinforcing the dense collaborative structure. Another notable aspect of the visualization is the single-cluster structure, represented by the uniform color of the nodes. In bibliometric mapping, a single cluster indicates that all authors belong to the same collaborative group with closely related research interests. This suggests that the publications included in the dataset are likely produced by a small but cohesive research team working within a shared thematic area. From a bibliometric perspective, the network indicates strong internal collaboration but limited external expansion, as the collaboration appears confined to a small group of authors.

Overall, the author collaboration network highlights a highly interconnected and cohesive co-authorship structure, reflecting a collaborative research environment where a small group of scholars consistently work together to produce scholarly outputs. Such collaboration patterns often contribute to sustained research productivity and thematic consistency within the field.

7.2 Co-authorship of country

The international co-authorship network (fig 11), visualized through VOS viewer, depicts the collaborative relationships among countries contributing to research on stock market prediction using artificial intelligence and machine learning techniques. To examine international collaboration patterns, a country-level co-authorship analysis was conducted. A minimum threshold of fifty citations and at least two publications was applied. Out of the total 68 countries identified in the dataset, 32 countries satisfied the threshold criteria and were included in the network analysis. Table 7 presents the top five countries ranked according to their total link strength, which represents the extent of collaborative relationships each country maintains within the co-authorship network.

Among the participating countries, India appears as the most dominant contributor in terms of publication output, with 119 documents and 1,904 citations. India also records one of the highest total link strengths (26), demonstrating extensive collaborative ties with multiple countries. Its central position in the network suggests that India acts as a major hub connecting different regional research communities. China represents another highly influential contributor, with 57 publications and the highest citation count (2,599) among all countries in the dataset. China also exhibits a strong collaborative presence with a total link strength of 26, indicating active engagement in international research partnerships.

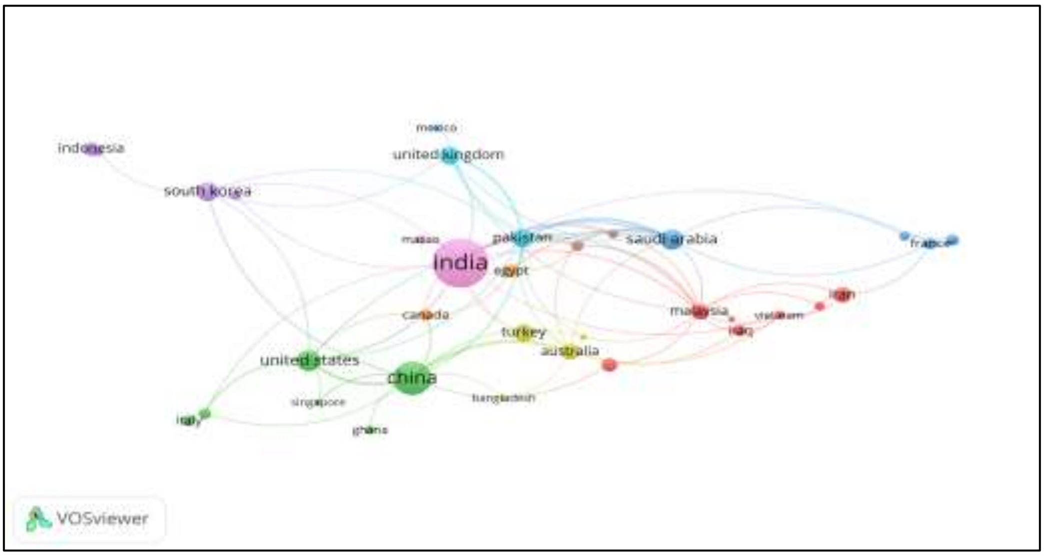

Figure 111 Co-authorship of Country collaboration network in AI and ML based stock market prediction research

The United States ranks among the leading contributors with 22 publications and 1,015 citations, reflecting its continued importance in global research collaboration. Its co-authorship links connect researchers across North America, Europe, and Asia, highlighting its integrative role in the network. Saudi Arabia, with 21 publications and a total link strength of 22, demonstrates strong international collaboration despite having comparatively moderate citation counts. Similarly, South Korea, with 19 publications and 600 citations, appears as an emerging contributor within the network, particularly within the Asian research collaboration landscape.

Table 6: Co-authorship of Country collaboration Summary

Rank	Country	Documents	Citations	Total Link Strength
1	China	57	2599	26
2	India	119	1904	26
3	United States	22	1015	16
4	Saudi Arabia	21	537	22
5	South Korea	19	600	8

Overall, the results indicate that Asian countries play a dominant role in the research output, with India and China acting as central hubs. At the same time, countries such as the United States and Saudi Arabia facilitate cross-regional collaboration, illustrating a multi-clustered global research network characterized by strong international partnerships.

8.0 Systematic Literature Review

In addition to the bibliometric mapping, a Systematic Literature Review (SLR) was conducted to provide a qualitative synthesis of the most influential studies in this field. The bibliometric analysis served as the foundation for this review by identifying the 25 most-cited articles from the Scopus dataset (2020–2026). The SLR was then employed to conduct a deeper examination of these highly influential works, offering contextual insights that complement the quantitative bibliometric findings

Table 7: Systematic Literature Review of the Most-Cited Studies in AI & ML Based Stock Market Prediction (2020–2026)

S. No	Author(s) (Year)	Title	Source Title	Citations	Technique / Model Used	Major Contribution	Findings (Summary)
1	Nti, Adekoya, & Weyori (2020)	A systematic review of fundamental and technical analysis of stock market predictions	Artificial Intelligence Review	372	Systematic Review of Machine Learning-Based Stock Market Prediction	Conducted a comprehensive review of 122 studies (2007–2018) and categorized prediction methods into technical, fundamental, and hybrid analyses.	Found that 66% of studies used technical analysis; Support Vector Machine (SVM) and Artificial Neural Network (ANN) were the most common techniques.
2	Pang et al. (2020)	An innovative neural network approach for stock market prediction	Journal of Supercomputing	345	Deep Long Short-Term Memory (LSTM) Network combined with Autoencoder	Proposed an enhanced Deep LSTM network for Shanghai A-shares market prediction.	The Deep LSTM with embedded layer achieved 57.2% prediction accuracy.
3	Jin, Yang, & Liu (2020)	Stock closing price prediction based on sentiment analysis and LSTM	Neural Computing and Applications	344	Sentiment-Driven Long Short-Term Memory (LSTM) Network integrated with Empirical Mode Decomposition (EMD)	Incorporated investor sentiment and decomposed stock data for temporal prediction.	Improved forecasting accuracy and minimized prediction delay.
4	Li, Wu, & Wang (2020)	Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong	Information Processing and Management	326	Price and News Sentiment Fusion using Deep Neural Network (DNN)	Developed a predictive framework combining numerical price data with textual sentiment from news.	The fusion model outperformed single-source models, improving overall accuracy.
5	Nabipour et al.	Deep learning for stock market prediction	Entropy	324	Comparison of Tree-Based Models (Random Forest, XGBoost,	Evaluated the performance of traditional and deep learning	LSTM demonstrated the highest predictive

	(2020)				AdaBoost) and Long Short-Term Memory (LSTM)	models for Tehran Stock Exchange sectors.	accuracy; ensemble models also performed well.
6	Jing, Wu, & Wang (2021)	A hybrid model integrating deep learning with investor sentiment analysis	Expert Systems with Applications	313	Hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) Network incorporating Sentiment Analysis	Developed a hybrid model combining CNN for sentiment extraction and LSTM for price prediction.	The hybrid CNN–LSTM model outperformed single models and non-sentiment baselines.
7	Nti, Adekoya, & Weyori (2020)	A comprehensive evaluation of ensemble learning for stock-market prediction	Journal of Big Data	262	Ensemble Learning Methods (Bagging, Boosting, Stacking, and Blending)	Conducted a comparative analysis of four ensemble techniques using Decision Trees, SVM, and ANN.	Stacking and blending approaches achieved superior accuracy (90–100%) with the lowest RMSE values.
8	Gulmez (2023)	Stock price prediction with optimized deep LSTM using Artificial Rabbits Optimization (ARO)	Expert Systems with Applications	241	Deep Long Short-Term Memory (LSTM) Network Optimized by Artificial Rabbits Optimization (ARO) Algorithm	Proposed an optimization-driven LSTM model to enhance predictive accuracy.	The LSTM–ARO model surpassed GA–LSTM and standard ANN models in performance.
9	Yun, Yoon, & Won (2021)	Prediction of stock price direction using hybrid GA–XGBoost algorithm	Expert Systems with Applications	235	Genetic Algorithm (GA) combined with Extreme Gradient Boosting (XGBoost)	Proposed a hybrid approach integrating GA for feature selection and XGBoost for prediction.	Achieved high accuracy through optimal feature engineering and dimensionality reduction.
10	Zhang, Yan, & Aasma (2020)	A novel deep learning framework: CEEMD–PCA–LSTM for financial time series	Expert Systems with Applications	217	Complementary Ensemble Empirical Mode Decomposition (CEEMD), Principal Component Analysis (PCA), and Long Short-Term Memory (LSTM) Network	Designed a hybrid framework integrating signal decomposition and dimensionality reduction with LSTM.	Achieved robust predictive accuracy and enhanced risk-adjusted profitability.
11	Carta et al. (2021)	Multi-DQN: An ensemble of Deep Q-learning agents for	Expert Systems with Applications	200	Ensemble Reinforcement Learning using Multi-Agent Deep Q-Networks (DQN)	Developed an ensemble of reinforcement learning agents for intraday stock trading.	Outperformed the traditional Buy-and-Hold strategy in trading simulations.

		stock market forecasting					
12	Chhajer, Shah, & Kshirsagar (2022)	Applications of ANN, SVM, and LSTM for stock market prediction	Decision Analytics Journal	195	Artificial Neural Network (ANN), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) Network	Compared the predictive performance of core AI models in stock forecasting.	Identified ANN, SVM, and LSTM as leading models for reliable stock prediction.
13	Chung & Shin (2020)	Genetic algorithm-optimized multi-channel CNN for stock market prediction	Neural Computing and Applications	194	Genetic Algorithm (GA) Optimized Convolutional Neural Network (CNN)	Applied GA to fine-tune CNN parameters for multi-channel market data.	The GA–CNN hybrid model outperformed standard CNN and ANN models.
14	Khan et al. (2022)	Stock market prediction using machine learning classifiers and social media/news data	Journal of Ambient Intelligence and Humanized Computing	185	Random Forest (RF) Classifier using Social Media and Financial News Sentiment	Investigated how social and financial sentiment data affect predictive accuracy.	The Random Forest model achieved 83.22% accuracy, with social media data outperforming news sentiment.
15	Thakkar & Chaudhari (2021)	Fusion in stock market prediction: A decade survey	Information Fusion	179	Fusion Techniques (Information-Level, Feature-Level, and Model-Level)	Reviewed a decade of fusion-based approaches in stock prediction and other financial applications.	Concluded that fusion techniques significantly improve accuracy and predictive robustness.
16	Shahi et al. (2020)	Stock price forecasting with deep learning: A comparative study	Mathematics	138	Comparison between Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) Networks with Sentiment Analysis	Compared recurrent neural architectures under identical market and news conditions.	Integrating sentiment data improved the predictive performance of both GRU and LSTM models.
17	Mehta, Pandya, & Kotecha (2021)	Harvesting social media sentiment to enhance stock market prediction using deep learning	PeerJ Computer Science	133	Deep Learning Model with Sentiment Analysis Integration	Developed a deep learning model incorporating public sentiment, news, and historical stock data.	Inclusion of public sentiment significantly enhanced stock prediction accuracy.

18	Ayala et al. (2021)	Technical analysis strategy optimization using machine learning	Knowledge-Based Systems	132	Machine Learning–Enhanced Technical Analysis using Linear Regression, Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Regression (SVR)	Applied machine learning to enhance rule-based technical trading indicators.	Integration of ML, improved the accuracy and profitability of trading signals.
19	Mukherjee et al. (2023)	Stock market prediction using deep learning algorithms	CAAI Transactions on Intelligence Technology	123	Artificial Neural Network (ANN) and Convolutional Neural Network (CNN)	Compared deep learning models under COVID-19-induced volatility.	Achieved 91% accuracy, demonstrating robustness during market disruptions.
20	Park, Kim, & Kim (2022)	Stock market forecasting using multi-task LSTM–Random Forest framework	Applied Soft Computing	115	Hybrid Long Short-Term Memory (LSTM) and Random Forest (RF) Framework	Combined sequential (LSTM) and ensemble (RF) learning to improve accuracy.	The hybrid model increased balanced accuracy by 7.37 percentage points.
21	Lee & Kim (2020)	Stock market forecasting using ConvLSTM and trend sampling	Expert Systems with Applications	113	Convolutional Long Short-Term Memory (ConvLSTM) Network with Trend Sampling and Data Augmentation	Proposed a ConvLSTM-based forecasting model using augmented high-dimensional data.	Reduced Mean Squared Error (MSE) by more than 50% and doubled profitability.
22	Haq et al. (2021)	Forecasting daily stock trend using multi-filter feature selection and deep learning	Expert Systems with Applications	113	Multi-Filter Feature Selection Combined with Deep Generative Model	Combined feature selection techniques with a deep generative model for forecasting.	Achieved higher accuracy than existing models through multi-filter integration.
23	Zhao & Yang (2023)	Deep learning-based integrated framework for stock price prediction	Applied Soft Computing	113	Self-Attention-Based Deep Long Short-Term Memory (SA–DLSTM) Hybrid Model	Developed a hybrid model combining Convolutional Neural Network (CNN), Denoising Autoencoder (DAE), and LSTM.	Achieved the highest predictive accuracy and optimal risk–return balance.

24	Xiao, C.; Xia, W.; Jiang, J. (2020)	Stock price forecast based on combined model of ARI-MA-LS-SVM	Neural Computing and Applications	105	Support Vector Machine (SVM), Least Squares SVM (LS-SVM), and combined ARI-MA-LS-SVM model	Proposed a hybrid/combined forecasting model (ARI-MA-LS-SVM) by integrating cumulative ARIMA with LS-SVM	Simulation results confirm that multi-model fusion improves forecasting performance and is useful for investors and market decision-making
25	Li, Ni, & Chang (2020)	Application of Deep Reinforcement Learning in stock trading	Computing	101	Deep Reinforcement Learning (DRL) Framework for Stock Trading	Applied DRL to optimize trading decisions and strategies.	Demonstrated superior adaptability and profitability over traditional models.

The analysis of the top 25 studies reveals a significant evolution in stock market prediction research, moving from traditional statistical techniques toward advanced machine learning and deep learning approaches.

The systematic review of 25 empirical studies (2020–2026) reveals a rapid evolution of AI-based stock market prediction techniques. The dominant trend involves hybrid and deep learning models, particularly LSTM, CNN, and ensemble frameworks (Nti et al., 2020; Pang et al., 2020; Jing et al., 2021). Early studies primarily relied on technical indicators and single-source data, whereas more recent works integrate news sentiment, social media signals, and reinforcement learning, reflecting a shift toward multi-modal fusion (Li et al., 2020; Khan et al., 2022). Across datasets and markets, LSTM-based models consistently outperform traditional algorithms (Nabipour et al., 2020; Gülmez, 2023). Notably, ensemble and hybrid frameworks (e.g., stacking, blending, and genetic optimization) demonstrate superior accuracy and robustness, suggesting that model diversity enhances learning stability (Nti et al., 2020; Yun et al., 2021).

The integration of sentiment analysis (Jin et al., 2020; Mehta et al., 2021) and reinforcement learning (Carta et al., 2021; Li et al., 2020) marks a methodological convergence between financial econometrics and behavioral AI. Studies such as Zhao & Yang (2023) and Thakkar & Chaudhari (2021) highlight the rise of fusion-based frameworks that merge heterogeneous data and algorithmic paradigms. Overall, the evidence indicates a decisive movement from single-algorithm predictive modeling to multi-layered, adaptive systems capable of integrating technical, textual, and behavioral features. However, limitations persist in terms of model interpretability, cross-market generalization, and real-time deployment pointing toward future research opportunities in explainable AI, transfer learning, and financial big data pipelines.

Overall, the analysis indicates that stock market prediction research is increasingly moving toward hybrid, data-driven, and intelligent systems. The integration of deep learning, sentiment analysis, feature engineering, and optimization techniques has significantly improved forecasting capabilities. However, challenges such as market volatility, noisy financial data, model interpretability, and generalization across different markets remain important areas for future research.

9.0 Conclusion and Future Research Directions

This work presents a inclusive bibliometric and systematic literature review of research on stock market prediction using artificial intelligence and machine learning techniques, based on 380 Scopus-indexed articles published between 2020 and 2026. The bibliometric analysis mapped the intellectual landscape, research performance, and global collaboration patterns, while the systematic review of the 25 most-cited studies provided deeper qualitative insights into prevailing methodologies, model innovations, and empirical findings.

9.1 Key Findings

This study uses a systematic literature review and bibliometric analysis to examine the research landscape of AI- and machine learning-based approaches in financial forecasting. The findings highlight key insights into the field's growth patterns, conceptual structure, and thematic development over time.

First, the annual publication trend demonstrates a significant and consistent increase in research output between 2020 and 2026, indicating the rapidly expanding interest in AI-driven financial forecasting. The strong upward trend suggests that stock market prediction using advanced computational techniques has evolved into a major interdisciplinary research area integrating finance, computer science, and data analytics.

Second, the author productivity analysis indicates that scholarly influence in this field is concentrated among a relatively small group of highly cited researchers. Authors such as Adekoya, Nti, and Weyori demonstrate strong citation impact and network connectivity, suggesting that their studies have played a central role in shaping the research direction of AI-based stock prediction models.

Third, the journal analysis highlights the importance of a limited number of specialized outlets in disseminating research in this field. Journals such as Expert Systems with Applications, Neural Computing and Applications, and IEEE Access serve as key platforms for publishing high-impact studies on intelligent financial forecasting and computational finance.

Fourth, the country-level analysis shows that research contributions are geographically diverse. India leads in terms of publication output, while China and the United States demonstrate higher citation impact, reflecting strong methodological contributions and global research influence.

Finally, the bibliometric and co-citation analyses reveal that foundational studies on data-driven predictive models form the intellectual backbone of this research field.

9.2 Thematic Insights

The keyword co-occurrence and co-citation analyses reveal several prominent research directions in AI-based stock market prediction. One key area is deep learning–oriented methods, where architectures such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and other neural models are widely employed to learn complex temporal patterns in financial time-series data. Their growing adoption is largely attributed to their effectiveness in capturing nonlinear market dynamics.

Another significant stream involves conventional machine learning techniques, including Support Vector Machines (SVM), Random Forests, and various classification-based models. Despite the rise of deep learning, these approaches continue to be widely applied due to their computational efficiency, interpretability, and reliable baseline performance in forecasting tasks.

A third research focus integrates sentiment analysis and behavioral finance perspectives. This line of work leverages unstructured data sources such as news articles, social media content, and investor sentiment indicators, often combined with natural language processing techniques to enhance prediction accuracy.

Finally, a growing body of research explores hybrid and integrated frameworks that combine statistical methods, machine learning algorithms, and deep learning models. These approaches aim to improve robustness and predictive performance by integrating the strengths of multiple modeling paradigms.

Overall, the thematic analysis suggests that the research field has evolved from traditional statistical forecasting models toward advanced AI-driven hybrid frameworks capable of analyzing complex financial datasets.

9.3 Limitations

While this study offers a comprehensive bibliometric and systematic review of stock market prediction research using artificial intelligence and machine learning techniques, several limitations must be acknowledged. First, the scope of the dataset was confined to Scopus-indexed journal publications between 2020 and 2025, which, although extensive, may not fully represent the entire body of work in this domain. Publications from other major databases such as Web of Science or IEEE Xplore, as well as non-indexed sources, were not included, potentially omitting relevant but less-cited contributions.

Second, only English-language journal articles were considered, thereby excluding conference papers, reviews, book chapters, and studies published in other languages. This restriction may introduce a degree of publication and language bias, limiting the inclusivity of the analysis. Furthermore, the study focused primarily on empirical research within specific subject areas Business, Management, Accounting, Computer Science, Decision Sciences, Economics, Finance and Social Sciences excluding conceptual or purely theoretical works that might provide complementary perspectives.

Lastly, the results of the systematic literature review (SLR**)** were derived from the 25 most-cited articles within the dataset. While this selection ensures the inclusion of highly influential research, it may underrepresent emerging but less-cited studies that address novel or niche methodologies. Consequently, the findings should be interpreted within the boundaries of the dataset and the methodological scope defined in this research.

9.4 Research Gaps and Future Research Directions

The review identifies several research gaps that provide opportunities for future studies. One major challenge is the lack of interpretability in deep learning models, as many AI-based prediction systems operate as "black boxes." Future research should focus on integrating explainable artificial intelligence techniques to enhance transparency and trust in financial forecasting models.

Another important gap lies in the limited use of diverse data sources. Many existing studies rely primarily on historical price data and technical indicators. Future research should explore the integration of multi-source data, including macroeconomic variables, financial news, and social media sentiment, to develop more comprehensive predictive models.

In addition, most studies focus on single-country stock markets, limiting the generalizability of findings. Future research should investigate cross-market prediction models that analyze interactions between global financial markets.

Furthermore, while many studies emphasize prediction accuracy, relatively few address practical financial applications such as portfolio optimization, risk management, and real-time trading systems. Future research should develop AI-driven frameworks that integrate prediction models with financial decision-making and investment strategies.

Overall, advancing explainable AI, integrating diverse datasets, developing hybrid prediction frameworks, and expanding cross-market analysis will play a crucial role in shaping the next generation of AI-based stock market prediction research.

.

REFERENCES

1. Aksnes, D. W., & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus and Web of Science. Journal of Data and Information Science, 4(1).

2. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of informetrics, 11(4), 959-975.

3. Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–Part II: Soft computing methods. Expert Systems with applications, 36(3), 5932-5941.

4. Ayala, J., García-Torres, M., Noguera, J. L. V., Gómez-Vela, F., & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. Knowledge-Based Systems, 225, 107119.

5. Ayyappa, Y., & Kumar, A. S. (2024). Stock market prediction with political data Analysis (SP-PDA) model for handling big data. Multimedia Tools and Applications, 83(34), 80583-80611.

6. Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. Quantitative science studies, 1(1), 377-386.

7. Bal, C. K., & Mishra, R. K. (2025). Stock Market Analysis and Prediction: A Bibliometric Analysis. Journal of Scientometric Research, 14(1), 221-238.

8. Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications, 42(20), 7046–7056.

9. Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLOS ONE, 12(7)

10. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of computational science, 2(1), 1-8.

11. Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. San Francisco: Holden-Day.

12. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

13. Broadus, R. (1987). Toward a definition of “bibliometrics”. Scientometrics, 12(5-6), 373-379.

14. Carta, S., Ferreira, A., Podda, A. S., Recupero, D. R., & Sanna, A. (2021). Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting. Expert systems with applications, 164, 113820.

15. Caputo, A., Pizzi, S., Pellegrini, M. M., & Dabić, M. (2021). Digitalization and business models: Where are we going? A science map of the field. Journal of business research, 123, 489-501.

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

17. Chhajer, P., Shah, M., & Kshirsagar, A. (2022). The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. Decision Analytics Journal, 2, 100015.

18. Chen, Y., Hao, Y., & Xu, J. (2017). A feature-weighted support vector machine and K-nearest neighbor algorithm for stock index prediction. Expert Systems with Applications, 80, 340–355.

19. Chen, Y., Pelger, M., & Zhu, J. (2021). Deep learning in asset pricing. Review of Financial Studies, 34(7), 3206–3243.

20. Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187–205.

21. Chung, H., & Shin, K. S. (2020). Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. Neural Computing and Applications, 32(12), 7897-7914.

22. 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.

23. Cortegiani, A., Ippolito, M., Ingoglia, G., Manca, A., Cugusi, L., Severin, A.,. & Giarratano, A. (2020). Citations and metrics of journals discontinued from Scopus for publication concerns: the GhoS (t) copus Project. F1000Research, 9, 415.

24. Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015, July). Deep learning for event-driven stock prediction. In Ijcai (Vol. 15, pp. 2327-2333).

25. 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.

26. Fama, E. F. (1965). The behavior of stock-market prices. The Journal of Business, 38(1), 34–105.

27. Ferreira, F. A. (2018). Mapping the field of arts-based management: Bibliographic coupling and co-citation analyses. Journal of Business Research, 85, 348-357.

28. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669.

29. Garg, D., & Tiwari, P. (2021). Impact of social media sentiments in stock market predictions: A bibliometric analysis. Business Information Review, 38(4), 170-182.

30. Glanzel, W., Moed, H.F., Schmoch, U., and Thelwall, M.: Springer Handbook of Science and Technology Indicators. Springer, Cham (2019)

31. Gulmez, B. (2023). Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. Expert Systems with Applications, 227, 120346.

32. Haq, A. U., Zeb, A., Lei, Z., & Zhang, D. (2021). Forecasting daily stock trend using multi-filter feature selection and deep learning. Expert Systems with Applications, 168, 114444.

33. Jin, Z., Yang, Y., & Liu, Y. (2020). Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 32(13), 9713-9729.

34. Jing, N., Wu, Z., & Wang, H. (2021). Hybrid CNN–LSTM model integrating sentiment analysis for stock market prediction. IEEE Access, 9, 141175–141188.

35. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263–291.

36. Khadka, P. B., & Karki, D. (2024). Bibliometric analysis of machine learning models for stock price prediction. Journal of Interdisciplinary Studies, 13(1), 116-141.

37. Khan, A., Goodell, J. W., Hassan, M. K., & Paltrinieri, A. (2022). A bibliometric review of finance bibliometric papers. Finance Research Letters, 47, 102520.

38. Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing, 13(7), 3433-3456.

39. Kumar, M., & Thenmozhi, M. (2014). Forecasting stock index returns using ARIMA–SVM, ARIMA–ANN and ARIMA–random forest hybrid models. International Journal of Banking, Accounting and Finance, 6(3), 284–308.

40. Lee, S. W., & Kim, H. Y. (2020). Stock market prediction using ConvLSTM and data augmentation. Expert Systems with Applications, 162, 113723.

41. Li, X., Wu, P., & Wang, W. (2020). A comparative analysis of machine learning techniques for stock market prediction. IEEE Access, 8, 117094–117106.

42. Li, Y., Ni, P., & Chang, V. (2020). Deep reinforcement learning for stock trading and portfolio management. Applied Intelligence, 50(8), 2338–2352.

43. Lim, J. S., & Zhang, J. (2022). Adoption of AI-driven personalization in digital news platforms: An integrative model of technology acceptance and perceived contingency. Technology in Society, 69, 101965.

44. Ma, X., et al. (2024). A multi-graph convolutional adversarial framework (VGC-GAN) for enhanced stock price prediction. Journal Name, (issue), pages. (Please insert full details)

45. Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science, 7, e476.

46. Mukherjee, S., Sadhukhan, B., Sarkar, N., Roy, D., & De, S. (2023). Stock market prediction using deep learning algorithms. CAAI Transactions on Intelligence Technology, 8(1), 82-94.

47. Malkiel, B. G. (1973). A random walk down Wall Street. New York: W. W. Norton & Company.

48. Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., & Salwana, E. (2020). Deep learning for stock market prediction. Entropy, 22(8), 840.

49. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). Machine learning applications in stock market prediction: A systematic review. Applied Intelligence, 50(12), 5747–5787.

50. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. Journal of Big Data, 7(1), 20.

51. Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2020). An innovative neural network approach for stock market prediction. The Journal of Supercomputing, 76(3), 2098-2118.

52. Park, H. J., Kim, Y., & Kim, H. Y. (2022). Multi-task deep learning framework for stock market prediction combining LSTM and Random Forest. Expert Systems with Applications, 192, 116328.

53. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications, 42(4), 2162–2172.

54. Pritchard, A. (1969). Statistical bibliography or bibliometrics. Journal of documentation, 25, 348.

55. Shahi, T. B., Shrestha, A., Neupane, A., & Guo, W. (2020). Stock price forecasting with deep learning: A comparative study. Mathematics, 8(9), 1441.

56. Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? The American Economic Review, 71(3), 421–436.

57. Tao, Z., Wu, W., & Wang, J. (2024). Series decomposition Transformer with period-correlation for stock market index prediction. Expert Systems with Applications, 237, 121424.

58. Thakkar, A., & Chaudhari, K. (2021). Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions. Information Fusion, 65, 95-107.

59. Tsay, R. S. (2010). Analysis of financial time series (3rd ed.). Hoboken, NJ: Wiley.

60. Usmani, S. & Shamsi, J. A. (2021). News sensitive stock market prediction: Literature review and suggestions. PeerJ Computer Science, 7, e490.

61. Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. scientometrics, 84(2), 523-538.

62. Xiao, C., Xia, W., & Jiang, J. (2020). Stock price forecast based on combined model of ARI-MA-LS-SVM. Neural Computing and Applications, 32(10), 5379-5388.

63. Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. Expert Systems with Applications, 186, 115716.

64. Zhao, Y., & Yang, G. (2023). Deep learning-based integrated framework for stock price movement prediction. Applied Soft Computing, 133, 109921.

65. Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. Financial innovation, 5(1), 1-20.

66. Zhang, Y. A., Yan, B., & Aasma, M. (2020). A novel deep learning framework: Prediction and analysis of financial time series using CEEMD and LSTM. Expert systems with applications, 159, 113609.

67. Zupic, I., & Cater, T. (2015). Bibliometric methods in management and organization. Organizational research methods, 18(3), 429-472

.

..

.