

A Bibliometric Review of Stock Market Forecasting: An or Perspective

¹Dr Richa Kalpesh Saxena, ²Nivritee Singhi, ³Nrip Swar, ⁴Om Ashar, ⁵Peeyanshi Mehta, ⁶Prachet Kumar

¹Assistant Professor, Marketing & Operations, NMIMS, Anil Surendra Modi School of Commerce, Mumbai Campus, India

^{2,3,4,5,6} B.Sc. Finance, NMIMS, Anil Surendra Modi School of Commerce, Mumbai Campus, India



<https://doi.org/10.55041/ijst.v2i10.061>

Cite this Article: Singhi, N., Swar, N., Ashar, O., Mehta, P. & Kumar, P. (2026). A Bibliometric Review of Stock Market Forecasting: An or Perspective. *International Journal of Science, Strategic Management and Technology*, 02(10), 1-9. <https://doi.org/10.55041/ijst.v2i10.061>

License:  This article is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited.

Abstract

This bibliometric analysis offers a comprehensive overview of the evolving landscape of Operations Research (OR) applications in financial and stock market forecasting. This review systematically extracted and analysed 722 articles published between 2015 and 2026, following a PRISMA-guided two-stage screening of 1,495 records retrieved from Scopus. The findings reveal significant trends and insights in the OR-forecasting literature. The study finds that the field is strongly interdisciplinary, with research activity concentrated around LSTM, neural network, SVM, ARIMA, and CNN methods. The keyword and thematic analysis uncovered clusters spanning sentiment analysis, natural language processing, and technical analysis, alongside the growing use of machine learning and deep learning in forecasting applications. The findings suggest emerging gaps around the systematic comparison of forecasting techniques and the integration of structured and unstructured data. These insights offer valuable guidance for researchers and practitioners seeking to develop hybrid, standardised, and reproducible forecasting frameworks, and highlight opportunities for greater methodological integration between OR and AI-driven finance.

Keywords: Stock Market Forecasting; Operations Research, Time Series, Machine Learning, Bibliometric Analysis, Literature Review.

1. INTRODUCTION

Financial Forecasting is an important component of the global economy since it is used to predict the value of stocks, exchange rates, interest rates, corporate profit and credit risk from past and present information. The stakes of this prediction problem extend far beyond any single trading floor into pension funds, insurers, and sovereign wealth funds, which rely on such predictions to meet long-term liabilities to millions of people. Because financial data is volatile and non-linear, accuracy in this prediction carries consequences for retirement security, job stability, and public confidence in financial systems, as demonstrated by the financial crisis of 2008, a product of miscalculated and systematically mispriced risk.



This forecasting problem has a close connection with Operations Research, which provides a quantitative and decision-theoretic basis for turning raw forecasts into strategies to guide financial actions through models, optimization, and simulation under uncertainty and resource constraints. All the techniques now employed to forecast stocks are related to Operations Research, whether they are machine learning techniques such as neural networks or deep learning techniques such as LSTM networks, in the sense that optimization is, in essence, the way of training them.

In recent years this field has evolved rapidly with advances in artificial intelligence, machine learning, generative AI and large language models. Recent literature on AI and machine-learning based equity, cryptocurrency, commodity and foreign-exchange forecasting finds that methods such as neural networks consistently outperform traditional forecasting methods, even though these models are rarely connected to actual trading strategies and their interpretations do not pay enough attention to volatility (Vancsura et al., 2025), while other recent work points to uses of artificial intelligence in portfolio management, extreme-scenario analysis, policy analysis and fraud detection, alongside the observation that current techniques remain insufficient for high-confidence prediction (Zheng et al., 2023); it is this fast pace of change, with no comprehensive picture of the field keeping up with it, that motivates the present study.

The study identified a total of 722 papers in the field of Operations Research applications in financial and stock market forecasting between 2015 and 2026 via a PRISMA-based two-phase screening process, revealing the field to be highly interdisciplinary and focused on LSTM networks, neural networks, SVM, ARIMA and CNN methods, with the primary gaps identified being the absence of bibliometric mapping of the field and the lack of comparison between methods and underutilization of combined structured and unstructured data, which motivated three research questions: (i) What are the trends in publication over time? (ii) What is the evolution of themes before and after 2020? (iii) What are the gaps in the literature? The analysis of the literature revealed that the field is highly interdisciplinary and focused on LSTM networks, neural networks, SVM, ARIMA and CNN methods as well as sentiment analysis and NLP, and that the gaps identified by this study could be addressed by current and future research in the field.

The rest of this paper is organized as follows: Section 2 outlines the research methodology, Section 3 presents the literature review that includes the previous review papers, background of the field, and the gap and objectives of the research, Section 4 reports the data analysis and results, Section 5 discusses the findings and implications of the study and Section 6 concludes the study with the limitations and further research directions.

2. RESEARCH METHODOLOGY

This section outlines the systematic procedure that was adopted to identify, screen, and select the body of literature examined in this study, and describes the bibliometric approach that was subsequently applied to the resulting dataset. Because the objective of the study is to map the structure of research on the application of Operations Research to financial forecasting rather than to test a specific hypothesis on primary data, a two-stage design was followed. The first stage



concerned literature identification and selection, guided by the PRISMA framework, adapted for bibliometric studies. The second stage concerned the bibliometric analysis of the selected corpus, conducted using the Bibliometrix R-package and its Biblioshiny web interface. Together, these two stages ensured that the dataset used for analysis was both relevant and reproducible.

2.1 Methodology Objectives

The research methodology was designed to fulfil the following objectives:

- (i) to identify a comprehensive set of publications addressing the application of Operations Research techniques to financial and stock market forecasting;
- (ii) to apply a transparent screening procedure so that the final corpus is free from duplication and topical irrelevance;
- (iii) to characterise the growth trajectory, leading sources, and most influential contributors within the field;
- (iv) to uncover the conceptual and thematic structure of the domain through science-mapping techniques, thereby identifying the major themes that define the current research agenda.

2.2 Research Design

A bibliometric research design was used to investigate the development, structure, and major research trends in the application of Operations Research to financial forecasting. Bibliometric analysis is a quantitative study of works on a specific topic based on patterns such as the quantity of publishing, authors, sources, keywords, citations, and associations between the areas.

This analysis was used to explore the intersection of Operations Research and financial forecasting, which consists of several interrelated areas, including time series modelling, stock market forecasting, and the growing application of machine learning and deep learning techniques to financial prediction problems.

The procedure followed consisted of the literature identification, screening, selection, and bibliometric analysis steps. The search period was set to 2015–2026, which allowed us to explore the research on Operations Research in financial forecasting during the selected period.

2.3 Data Collection

Data were retrieved from Scopus, which was used as the literature source for our study. Articles published in 2015–2026 were included and included five main keyword areas for our analysis: Operations Research, Financial Forecasting, Stock Market Forecasting, Time Series, and Machine Learning/Deep Learning. The keywords were combined to find the relevant documents published in the selected period: “Operations Research” AND (“Financial Forecasting” OR “Stock Market Forecasting” OR “Time Series” OR “Machine Learning” OR “Deep Learning”).

The search result of 1,495 items was retrieved for the initial bibliometric analysis of Operations Research in financial forecasting.



2.4 Trial and Error in Keyword Selection

Arriving at the final search string took several rounds of trial and error with minor modifications after each round. The first search attempt used the single term “Operations Research” on its own, which returned more than 40,000 records spanning across logistics, healthcare, manufacturing, and supply-chain optimisation which was far too broad and largely unrelated to financial applications, so a qualifying term was clearly needed.

A second search paired “Operations Research” with the very general term “Finance,” which brought the count down considerably but still returned a large share of papers on corporate finance, accounting, and banking regulation with no forecasting component. A third search tested “Optimization” AND “Financial Forecasting” as an alternative starting point, expecting it to be more precise, but this excluded a number of papers that used Operations Research techniques such as linear programming, simulation, and decision analysis without using the word “optimization” itself in the title, abstract, or keywords, so it was replaced with the term “Operations Research”. “Quantitative Finance” AND “Time Series” was tried next and discarded for a similar reason, since it returned mostly econometrics papers with no explicit Operations Research methodology.

Once “Operations Research” was fixed as the core term, “Financial Forecasting” and “Stock Market Forecasting” were added as parallel branches, since early results showed that different authors used one term or the other almost interchangeably; running them separately and then combined confirmed that both needed to be used, as the two sets overlapped only partially. “Time Series” was added after an initial run without it missed several highly cited Operations Research papers that focused on time-series modelling but did not use the words “forecasting” or “prediction” explicitly. “Machine Learning” and “Deep Learning” were added on, once a preliminary scan of the literature showed that a large and growing share of recent Operations Research-based forecasting work is now hybridised with these techniques, and leaving them out devalued the volume of research published from 2020 onward.

Subject-area and language limitations were also tested and language limitations directly during the search stage rather than during filtering, however this eliminated a number of relevant publications that Scopus had classed as Computer Science or Decision Sciences despite having finance-focused applications. As a result, these constraints were applied at the title-and-abstract screening step rather than during the first search, as indicated below.

2.5 Screening and Selection of Publications

After performing our initial search in Scopus, the 1,495 records were screened the 1,495 records to identify the publications relevant to our investigation. The screening process was visualised in a PRISMA flowchart adapted for bibliometric studies, which consisted of the identification, screening, and eligibility stages.

The first stage of the screening process involved the removal of duplicate records that had been retrieved more than once across our overlapping search terms. This excluded 229 records and refined the set from 1,495 to 1,266 items. The titles and abstracts were then screened of the remaining 1,266 records against our inclusion criteria, retaining only records that were finance- or

economics-relevant, written in English, and reported an application of Operations Research techniques to a forecasting problem. This step excluded a further 544 records, primarily papers that were off-topic, non-English, or published as conference abstracts, editorials, or book chapters rather than full research articles, leaving 722 records for inclusion.

722 papers were included in our bibliometric analysis of Operations Research in financial forecasting.

Figure 1. PRISMA Flow Diagram

Stage	Action	n
Identification	Records identified from Scopus	1,495
Screening	Records remaining after duplicate removal (229 duplicates excluded)	1,266
Eligibility	Records remaining after title/abstract screening (544 excluded as off-topic/ineligible)	722
Inclusion	Studies included in the bibliometric analysis	722

Table 1. Summary of the PRISMA selection process.

2.6 Data Processing

The selected bibliography was compiled the selected bibliography into the dataset and prepared it for the analysis. RStudio was used as the working environment and Bibliometrix and Biblioshiny as the tools for processing and analysing the bibliographic data. Biblioshiny is a graphical interface that allows the users to perform bibliometric analysis and visualize the research data. Publish or Perish was also used to cross-verify citation counts across sources, and Mendeley to manage and de-duplicate the exported reference lists before import. The analysis was initiated by importing the selected research papers into Biblioshiny and exploring the information before proceeding to the analysis. Data processing and verification was crucial since the quality of the dataset determined the analysis that could be performed.

2.7 Bibliometric Analysis

Using Biblioshiny in RStudio, the final dataset of 722 papers was analysed in terms of annual scientific production, thematic evolution, keyword analysis, author productivity, source and journal analysis, and discipline life-cycle analysis. The trends in the scientific production were determined, the changes in research themes and keywords were identified, and the research on Operations Research in financial forecasting life-cycle was analysed.



2.8 Limitations of the Methodology

First, the files were drawn from a single database (Scopus), and, while Scopus offers strong coverage of the finance, management science, and Operations Research literature, some relevant work indexed exclusively in other databases such as Web of Science or Science Direct may not be represented. Second, the eligibility screening, although guided by explicit inclusion criteria, involved a degree of reviewer judgement at the title/abstract stage that could not be fully automated. Third, keyword-based science mapping techniques are sensitive to inconsistencies in author keyword terminology (for example, “stock market forecasting” versus “stock price prediction”), which can fragment otherwise related concepts into separate nodes on the resulting maps.

3. LITERATURE REVIEW

3.1 Background

Over the last ten years, AI and deep learning have played a big part in research on financial markets. This shows up in stock price prediction, currency forecasting, and lately in studies on cryptocurrency trends too. Part of the reason is simply that more data and more computing power became available, and researchers started drifting away from purely statistical forecasting toward learning-based systems that could pick up non-linear, high-dimensional patterns in market behaviour that older models were never built to catch. The result is a literature that has grown quickly and spread across computer science, finance, and data science venues at once but nobody has really stepped back and mapped it as a single field.

A bibliometric approach fits this kind of situation reasonably well. Rather than trying to read and synthesise a handful of representative papers, it works from publication patterns, keyword structures, and thematic clustering across hundreds of records at once, which is really the only practical way to get a structured picture when the literature is this large and this fast-moving. That is the approach taken here, applied to a dataset of 722 bibliographic records, with the aim of building a clearer account of where the field currently stands.

3.2 Review of Existing Review Articles

Since this study is itself a review of a review-type literature, it was necessary to first collect and examine existing review articles written on AI and deep learning in financial forecasting. Six such reviews were identified in the dataset. Sezer, Gudelek and Ozbayoglu (2020) reviewed how deep learning models, such as the Long Short-Term Memory (LSTM) network, a type of model designed to learn patterns in sequential or time-based data, have been applied to financial forecasting between 2005 and 2019. Kumbure, Lohrmann, Luukka and Porras (2022) reviewed machine learning techniques and the kinds of data used as inputs for stock market prediction. Xing, Cambria and Welsch (2018) focused on forecasting models that use natural language, such as news articles or social media posts, as a source of information.

Hu, Zhao and Khushi (2021) reviewed deep learning methods specifically for forex and stock price prediction. Htun, Biehl and Petkov (2023) reviewed the methods used to select or build useful input features for prediction models. Thakkar and Chaudhari (2021) reviewed a decade of fusion



approaches, meaning models that combine more than one method or data type to improve prediction.

Taken together, these six reviews show that most existing work has been done using narrative or systematic methods, and each review looks at only one part of the field, such as architecture, input data, or language-based methods. None of them uses a bibliometric method to map the field as a whole, and none traces how these different parts of the field, such as model architecture, input data, and language-based methods, relate to or build on one another over time. This gap in the literature is the basis for the present study.

		AuthorScope of Review Type of Study
Sezer, Gudelek & Ozbayoglu (2020)	Deep learning methods (LSTM, CNN, hybrid models) used for financial time-series prediction, 2005–2019	Systematic review
Kumbure, Lohrmann, Luukka & Porras (2022)	Machine learning techniques and the types of input data used in stock market forecasting	Systematic review
Xing, Cambria & Welsch (2018)	Use of natural language (news, social media text) for financial forecasting	Narrative review
Hu, Zhao & Khushi (2021)	Deep learning approaches applied to forex and stock price prediction	Systematic review Narrative review
Htun, Biehl & Petkov (2023)	Feature selection and extraction techniques used for stock prediction models	Systematic review
Thakkar & Chaudhari (2021)	A decade of fusion approaches (combining multiple models/data types) for stock prediction	

Table 2. Summary of type of study.

3.3 Objectives of the Study

Based on the gap identified above, this study aims to conduct a bibliometric review of AI and deep learning research in financial market forecasting, using publication and keyword data to trace how the field has grown and changed in focus since it began. Newer techniques covered in the more recent literature include the Graph Neural Network (GNN), a model that learns from data connected in a network, such as how one company's stock price relates to another's (Cheng, Yang, Xiang & Liu, 2022), and Explainable AI (XAI), a set of methods that make a model's predictions understandable to humans rather than treating it as a "black box" (Arsenault, Wang & Patenaude,

2025). Both of these are fairly new to the field, and none of the six reviews discussed above cover them. This is one more reason an updated, structural look at the literature is needed.

Based on this gap, the review is built around three research questions:

- i: What do publication trends look like, in terms of yearly growth and the mix of topics, in AI and deep learning research on financial market forecasting?
- ii: What are the main themes in this literature, and how has the focus shifted between the years before 2019 and from 2020 onward?
- iii: What gaps still remain; in how different research strands connect, in interpretability work, and in real-world testing, and what do these suggest for future research?

4. DATA ANALYSIS AND RESULTS

4.1 Annual Scientific Production: Growth Trend

The publication output was seventy eight articles in 2015. It has fluctuated at a substantially higher plateau since peaking at ninety articles in 2024. The 2026 figure shows a partial year of indexing and was excluded from trend fitting.

Linear and exponential models were both fitted to the 2015–2025 series to identify which growth form best describes the data. The linear model produced an R-squared value of 0.35 while the exponential model produced an R-squared value of 0.32. The linear model was fitted in units whereas the exponential model was fitted on the log-linearised series which is the standard approach for comparing exponential fits. The linear model therefore offers a better statistical fit than the exponential model though neither exceeds a moderate fit. Both R-square values are well below 0.5.

This weak fit for both models is itself a finding. Rather than smooth compounding growth typical of a rapidly emerging niche the financial forecasting field shows a step change followed by an irregular saw tooth pattern of expansion and contraction. This pattern is consistent with a maturing but still-volatile research area, where output responds to bursts of interest (for example new neural architectures, crypto-asset booms) rather than a stable underlying growth rate. For the report it is more defensible to describe the trend as “non-monotonic/irregular growth” than to claim a strong linear or exponential law.

Figure 2. Annual scientific production on financial forecasting (2011–2026) with linear and exponential trendlines.

4.2 Thematic Map (Strategic Diagram)

The thematic strategic diagram shows each keyword cluster by centrality against density. It shows how different themes interact with each other and how tightly knit they are. Eight clusters

appeared. ‘Machine Learning’ sits in the ‘motor theme’ quadrant with centrality and high density showing that it is a well-developed theme that is crucial to the fields’ current state. ‘Financial Forecasting’ sits close to the origin on the central axis and high on density acting as the point of connection of the domain. ‘Deep Learning’ and ‘Financial Time Series Forecasting’ fall into the basic/transversal quadrant, with centrality but lower density meaning they are widely referenced cross-cutting techniques rather than tightly bounded sub-fields. ‘Stock Market’ and ‘Financial Time-Series Forecasting’ appear as niche themes with low centrality and high density and are well-developed but comparatively self-contained strands. ‘Forecasting’ and ‘Stock Market Forecasting’ fall in the emerging/declining quadrant suggesting these are either fading generic terms being absorbed into more specific vocabulary such as LSTM, GRU and transformer-based forecasting.

Figure 3. Thematic strategic diagram: centrality vs. density of research themes.

4.3 Thematic Evolution (2015–2019 vs. 2020–2026)

Comparing the two subperiods, a clear conceptual shift is observed. In the period from 2015 to 2019 keyword clusters revolve around statistical methods, such as ARIMA, neural networks and financial time-series forecasting. By the period from 2020 to 2026 the main destination clusters have become learning and financial forecasting themselves absorbing terms like LSTM, CNN, ensemble learning and market-efficiency concepts. The strongest continuity flows from forecasting in 2015–2019 to financial forecasting in 2020–2026 carrying keywords such as genetic algorithms and algorithmic trading. This shows that financial forecasting persists, with methodological refinement. A second strong flow moves from machine learning in 2015–2019 into learning in 2020–2026 confirming a shift from traditional machine learning toward deep neural architectures. The stock market forecasting theme, from the period splits into forecasting, deep learning and machine learning in the later period indicating that a previously unified applied theme has diversified into method-specific streams.

4.4 Conceptual Structure

The keyword co-occurrence matrix was analysed with a Multiple Correspondence Analysis. The Multiple Correspondence Analysis showed the space of the field in two dimensions. A single-cluster solution was obtained, showing that all 54 retained keywords merge into one cluster instead of several disconnected ones. This indicates that financial forecasting research is strongly cohesive. Financial forecasting research does not split into different streams of thought, instead it forms one broadly interconnected knowledge domain.

Dimension 1 separates statistical terms and optimisation-heavy terms on one side from modern deep-architecture terms on the other. On the pole of Dimension 1 terms such as multilayer perceptron, artificial neural network, genetic algorithm and particle swarm optimisation appear. On the pole of Dimension 1 terms such as recurrent neural network, CNN, GRU, transformer and LSTM appear. Dimension 2 separates text and sentiment driven approaches on one side from

signal decomposition and hybrid time series techniques on the other. On the pole of Dimension 2 terms such as text mining, natural language processing, big data and predictive analytics appear. On the pole of Dimension 2 terms such as empirical mode decomposition, ARIMA and support vector regression appear. This structure suggests that the methodological identity of financial forecasting research is defined along two axes: model architecture versus deep and data type textual or sentiment, versus numerical or time-series.

Figure 4. Conceptual structure map (MCA) of financial forecasting keywords.

4.5 Most Productive Authors

Author productivity is very focused. Five author names have more than five articles in the dataset. The two entries for “Nayak S C” and “Nayak Sarat Chandra” probably belong to the same researcher. If we combine them this author would have about 17 articles, the most in the field. After that Cambria Erik (8 articles) and Behera H.S. (7 articles) are the most prolific followed by Misra B.B. (6). This focus with the small number of articles even for the top authors shows that the field has not yet formed a few dominating research groups. Instead, productivity is spread over authors, which is typical for an area that is still growing and mixing different disciplines.

4.6 Keyword Co-occurrence Network

The keyword co-occurrence network after filtering to 50 nodes is divided into four clusters that match the themes. A “Financial Forecasting Core” cluster (financial forecasting, neural networks, recurrent/convolutional networks), a “Deep Learning & ML” cluster (deep learning, machine learning, LSTM, sentiment analysis), a “Stock Market / Finance” cluster (stock market, forecasting, artificial intelligence, finance) and a “Time Series & Hybrid Models” cluster (financial time series forecasting, ARIMA, genetic algorithm, particle swarm optimisation). “Deep learning” and “machine learning” are the two highest-PageRank nodes in the network showing that they are the backbone that links more specific forecasting methods to finance topics.

Figure 5. Keyword co-occurrence network, coloured by cluster.

4.7 Summary of Results

The financial forecasting literature shows a field that has expanded irregularly — a pattern better captured by a linear than an exponential trend. Thematically, the domain is grounded by machine learning and financial forecasting as core themes, with deep learning architectures acting as a widely spread transversal technique. Thematic evolution confirms a methodological migration from classical statistical and shallow ML techniques toward deep-learning-based forecasting, while the MCA conceptual map shows the field remains conceptually unified rather than fragmented, organised along a classical-to-deep-architecture axis and a numerical-to-textual data axis. Author productivity is dispersed rather than concentrated, and the co-occurrence network



confirms deep learning and machine learning as the connective hubs of the knowledge structure. Collectively, these results position financial forecasting as a maturing, technique-driven interdisciplinary field still undergoing active methodological transition.

5. DISCUSSION AND IMPLICATIONS

The analysis of the bibliography suggests that the field of financial forecasting has become a very interdisciplinary one, and significant effort has been given in the use of computational and artificial intelligence techniques. The data represented in the literature is heavily concentrated on financial forecasting, machine learning, deep learning, and forecasting the stock market and financial time-series, which shows how researchers are using data-informed approaches to solve the challenges and uncertainties of the financial markets. LSTM, neural networks, support vector machines, ARIMA, CNN and other machine-learning and deep-learning methods are all used in one way or another, reflecting the diversity of methods used in the field. Meanwhile, the prevalence of sentiment analysis, NLP-based analysis, technical analysis and feature selection and optimisation methods suggest a move beyond traditional price-based or statistical approaches.

One important limitation is the lack of empirical comparisons of forecasting techniques. The fact that there are many individual approaches does not provide any indication of whether a specific methodology is more likely to deliver superior performance in any particular dataset, asset, market environment or time horizon. Systematic comparisons with consistent evaluation metrics and forecasting periods can thus be made in the future using the same datasets. The other gap is related to the integration process of various methodological approaches. Further research can explore hybrid forecasting methods that integrate the strengths of different approaches, such as machine learning, deep learning, and statistical models and optimisation techniques.

The bibliometric findings also point to opportunities for more integration of structured financial information and unstructured textual information. The role of sentiment analysis and natural language processing in the research landscape indicates that financial forecasting will increasingly be able to utilise information from news and other textual sources, in addition to market variables. It is desirable to explore further the application of sentiment, technical indicators, trading data and past price action into a single forecasting model. Moreover, the recent developments of new architectures like transformer-based models offer opportunities for comparative studies with traditional models like LSTM, CNN, ARIMA and support vector machines.

5.1 Implications

The results are of interest to researchers and practitioners. The thematic concentration identified can be used to further research into deep learning, machine learning, time-series forecasting, sentiment analysis, optimisation and technical analysis. The bibliometric patterns can assist researchers to locate existing methodological fields and opportunities for interdisciplinary research. Practitioners may benefit from the developments of financial forecasting models that offer analytical tools to interpret market information and aid in forecasting-oriented applications. In assessing model predictions, however, it is necessary to take into account not only the accuracy



of predictions but also their robustness and the data needed for prediction, and the forecasting context. The results also indicate that methodological approaches should be integrated.

Financial markets produce many different types of information, which means that future forecasting frameworks can explore the possibility of using a mix of statistical methods, machine-learning algorithms, deep-learning architectures, optimisation techniques, and textual analysis. This integration can give a more holistic view of financial forecasting than modelling on any one piece of information.

5.2 Limitations

The results of the bibliometric analysis have several limitations in their interpretation. First, the conclusions are based on the bibliographic dataset, and on the fact that it is not exhaustive, so that publications not included in the dataset may not be represented in the identified research patterns. Second, bibliometric analysis can reveal links between publications, keywords, and research themes but cannot prove the predictive power of each forecasting model. Third, the analytic approach of keyword-based thematic analysis may be affected by different terminologies that are used for similar concepts among writers. As a result, some research may be listed in different thematic categories. Lastly, the observed research trends are not necessarily viewed as a complete picture of the financial-forecasting research field; they are merely the literature that can be found within the selected dataset.

5.3 Future Research Agenda

Based on the identified findings, research gaps, implications and limitations, future research can be directed towards the development of more integrated and robust financial forecasting frameworks. Focus can be shifted towards hybrid approaches that integrate statistical forecasting, machine learning, deep learning and optimisation methods.

Specifically, researchers can explore the comparative and combined use of LSTM, CNN, neural networks, support vector methods, ARIMA and optimisation algorithms within specific financial markets and time scales. The potential of combining sentiment-based and financial/technical metrics is also significant and warrants further exploration. Future research could explore the impact of the addition of textual information on forecasting accuracy as well as its robustness under various market conditions. The study can be expanded to investigate various feature-selection and optimisation methods that can be used to select the most significant features and minimise feature redundancy in financial datasets.

Also, the rising adoption of newer deep-learning architectures opens up the possibility of comparing new approaches with existing forecasting models. Future research can go beyond comparing the accuracy of individual models to establish standardised, reproducible and cross-market assessments, where researchers can examine the robustness of models in different datasets, on different financial instruments and for various time periods. In sum, the research programme that emerges through the bibliometric analysis suggests a more hybrid, data-integrated,



comparative, and robust future research agenda on financial forecasting, acknowledging the methodological and data-related shortcomings of current research.

6. CONCLUSION

The bibliometric analysis provided a snapshot of the current state of financial forecasting research, highlighting the significant growth in the field and the strong interdisciplinary nature of the literature, particularly in the use of algorithms such as LSTM, CNN, SVM and ARIMA to model financial and stock market data, moving beyond contemporary and traditional forecasting methods, and identifying the shift toward sentiment analysis and NLP approaches to forecasting. The analysis also outlined the gaps in current research, including questions about the effectiveness of forecasting compared to real-world trading decisions, forecasting methods in abnormal market conditions, and the role of generative AI in forecasting in the future (Vancsura, Tatay & Bareith, 2025), providing a better understanding of financial forecasting literature.

Although the paper does not address every domain equally, it does have real-world value in more than one way: academically it contributes to the field of Operations Research and AI-driven finance, an area of great interest due to the direct economic applications of improved financial forecasting (Nair & Mohandas, 2015); and from a managerial perspective, the paper's discussion of accuracy and resilience under volatile conditions resonates with the use of data-driven forecasting models as components of corporate decision support systems (Yang, 2025) and the increasing trend of asset-management companies treating forecasting and portfolio decisions as a unified process (Sosabowski, 2018).

Certain practical limitations arose during the writing of the research paper: the selection of the keywords—financial forecasting is described in many different ways in the literature and it took repeated searching, reviewing and adjusting of the keywords to collect relevant papers for the bibliometric review—which would take time away from the practical interpretation of the results; secondly, there were problems with the coverage of the databases, as some research papers were indexed under the wrong terms, and some were not fully recorded, lacking author information or keywords at all; and thirdly, the large number of research papers made handling the citations a very tedious task: Mendeley helped keep the library under control, but it took a lot of time to remove the duplicate entries.

REFERENCES

- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
<https://doi.org/10.1016/j.joi.2017.08.007>
- Arsenault, P.-D., Wang, S., & Patenaude, J.-M. (2025). A survey of explainable artificial intelligence (XAI) in financial time series forecasting. *ACM Computing Surveys*, 57(10), 1–37.
<https://doi.org/10.1145/3729531>



- Cheng, D., Yang, F., Xiang, S., & Liu, J. (2022). Financial time series forecasting with multi-modality graph neural network. *Pattern Recognition*, 121, Article 108218. <https://doi.org/10.1016/j.patcog.2021.108218>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation*, 9, Article 26. <https://doi.org/10.1186/s40854-022-00441-7>
- Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. *Applied System Innovation*, 4(1), Article 9. <https://doi.org/10.3390/asi4010009>
- Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197, Article 116659. <https://doi.org/10.1016/j.eswa.2022.116659>
- Nair, B. B., & Mohandas, V. P. (2015). Artificial intelligence applications in financial forecasting: A survey and some empirical results. *Intelligent Decision Technologies*, 9(2), 99–140. <https://doi.org/10.3233/IDT-140211>
- 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, Article 106181. <https://doi.org/10.1016/j.asoc.2020.106181>
- Sosabowski, J. (2018). Financial forecasting and portfolio optimization in the 21st century. In *Coresource 4* (pp. 135–137). <https://doi.org/10.1002/9781119444510.ch33>
- 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. <https://doi.org/10.1016/j.inffus.2020.08.019>
- Vancsura, L., Tatay, T., & Bareith, T. (2025). Navigating AI-driven financial forecasting: A systematic review of current status and critical research gaps. *Forecasting*, 7(3), Article 36. <https://doi.org/10.3390/forecast7030036>
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: A survey. *Artificial Intelligence Review*, 50(1), 49–73. <https://doi.org/10.1007/s10462-017-9588-9>
- Yang, A. (2025). Big data-driven corporate financial forecasting and decision support: A study of CNN-LSTM machine learning models. *Frontiers in Applied Mathematics and Statistics*, 11, Article 1566078. <https://doi.org/10.3389/fams.2025.1566078>