STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **16**, August 2026, pp 1697–1715.
Published online in International Academic Press (www.IAPress.org)

# Machine Learning for Financial Market Prediction : A Systematic Literature Review

Imane Bahammou, Soumia Ziti, Karim El Bouchti*

*IPSS, Department of Computer Science, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco*

**Abstract** Artificial intelligence has recently become a powerful tool for market stock prediction. Financial markets are complex and volatile; each market behaves differently. The machine learning field hasand inputs known extensive study over the last few years. Researchers have experimented with a range of inputs for basic price prediction, employing basic market data from historical data and technical indicators to market indexes and alternative data sources such as sentiment indicators and news. However, as the goal is to generate forecasts of prices and trends, it is crucial to assess models through adequate metrics. Yet the field suffers from methodological inconsistency and a persistent disconnect between predictive accuracy and economic viability. This systematic literature review employs a structured PRISMA protocol to collect 63 studies and to examine forecasting models, evaluation metrics, alternative data integration, and real-world deployment challenges. Our analysis reveals that Long Short-Term Memory networks appear in 53% of reviewed papers and dominate the literature, and hybrid architectures account for 61% of the total. To bring conceptual clarity, we propose a four-tier taxonomy encompassing data-level, model-level, parallel-ensemble, and architectural hybrid models. The literature reveals critical gaps: 97% of papers use technical metrics while only 9% report the use of profitability, and only 4 studies apply backtesting with realistic transaction costs. Furthermore, survivorship bias is acknowledged in only three papers, while geographical coverage is heavily concentrated in the United States and Chinese markets, leaving African and Latin American markets understudied. Notably, memorizing narrow market regimes rather than generalizing exhibits overfitting risk. We conclude that the assumption of accuracy equaling profitability is unsupported by current evidence, as experiments reporting profitability and backtesting performance are largely absent, which undermines claims about the model's robustness. We advocate for standardized benchmarks, rigorous ablation studies, profitability evaluation protocols, and controlled experiments studying the effect of preprocessing and architectural synergy separately.
.

**Keywords** : Deep learning, Trading indicators, Quantitative finance, Systematic review, Machine learning, Hybrid models.

**DOI:** 10.19139/soic-2310-5070-3490

## 1. Introduction

Over the past decade, the rise of artificial intelligence has transformed the financial market. The complex nature of the market has led researchers to seek more efficient tools to predict the stock market, as technical indicators rely on fixed mathematical formulas and often struggle to capture complex non-linear relationships. To address these limitations, researchers adopted machine learning to produce models that can adapt to market dynamics. This systematic literature review captures emerging trends including transformers and attention mechanisms, and covers papers on global markets spanning diverse countries from Asia to Africa, rather than focusing solely on the US or European market. In this study, we examine commonly used machine learning models, such as neural networks, regression models, time series models, and tree-based models. as well as hybrid techniques that enhance machine learning capabilities by merging advanced algorithms, such as "Attention mechanism" or "Bee colony

*Correspondence to: Imane Bahammou (Email: imane.bahammou@um5r.ac.com). Department of Computer Science, Mohamed V University. 4, Avenue Ibn Battouta, B.P. 1014, Rabat (210093).

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2026 International Academic Press

algorithm", or connecting different types of machine learning models with each other to get the benefit of each model. We investigate five critical research questions:

- Which machine learning techniques are most commonly employed for stock market prediction models, and how do they differ from traditional technical indicators?
- How do artificial intelligence-based models incorporate alternative data?
- What evaluation frameworks are used to validate artificial intelligence-driven models, and how do they address overfitting or survivorship bias?
- What practical challenges exist in deploying artificial intelligence-driven models in live trading environments?
- How do hybrid approaches enhance the robustness of trading signals?

We examined 63 studies published between December 2018 and 2025 across different global markets, including developed markets like the United States and emerging markets such as India, South Africa, and Thailand. This review provides a comprehensive assessment of the current state of artificial intelligence-driven stock market prediction. Our findings show that recurrent neural networks, particularly Long Short-Term Memory models, dominate the literature and are dominant, being used in over 50% of the studies reviewed.

However, stock price forecasting is a challenging task since several attributes contribute to the price fluctuation, such as political decisions, news, and company balance sheets. These factors contribute to the volatility and complexity of the market and emphasize the need to look beyond the traditional data. Among the 63 studied papers, only 9 papers integrate alternative data, such as social media data and news sentiment, with historical prices for forecasts.

In addition, through this research, we examined the evaluation framework used to validate the models. The majority of the studies rely only on technical performance metrics; most studies did not include trading strategies or check whether the models actually made a profit. The usual machine learning evaluation metrics don't match well with real financial results, showing the need for fully automated systems that can properly check financial performance.

Furthermore, the deployment of artificial intelligence-based solutions has to generate signals with minimal delay, process real-time market data, and deliver actionable insights before market opportunities disappear, underscoring the needs and the challenges faced, such as adaptability and computational efficiency. reliability, stability, and low latency.

Besides, we noticed through this review that there is a growing interest in hybrid methodologies that combine diverse techniques to leverage complementary strengths, for example, by connecting advanced algorithms such as attention mechanisms and bidirectional models, using a variation of neural networks or advanced architectural approaches, or merging statistical and machine learning models. Finally, this systematic review aims to collect and analyze existing articles in the field, focusing on machine learning techniques for forecasting the stock market prices while answering concise questions,and investigating why the assumption of accuracy equals profitability is false and the reasons behind its persistence. The 63 papers resulting from the collection and selection processes serve not only to clarify what is done but also to reveal what the field is not solving.

## 2. Research Methodology

The objective of guiding a systematic review is to screen, collect, evaluate, synthesize, and discuss relevant research articles to answer the research questions. The review should be complete and fair, or it will have no tangible value. A systematic review has certain benefits, such as producing more reliable research with less biased results by following a clear and structured methodology. Therefore, the methodology used for this systematic review will be described below.

We started with the process of selecting the main publications for this work. To achieve this task, we followed a clear research protocol, and we respected strict criteria. Three steps were proposed for the development of a systematic review:

- planning
- Conducting
- Analysis of results

*2.0.1. Planning* The first step is to define the questions to be answered by the systematic review, as well as the inclusion, exclusion, and quality criteria. The questions are listed in Table 1, and the inclusion (IC), exclusion (EC), and quality (QC) criteria are presented in 2,3,4, respectively.

Table 1. Research questions.

| ID | Research Question (RQ) |
|---|---|
| **RQ1** | What Machine learning methods are most commonly used for stock market prediction models, and how do they differ from traditional technical indicators? |
| **RQ2** | How do AI-based models incorporate alternative data? |
| **RQ3** | What evaluation frameworks are used to validate AI-driven models, and how do they address overfitting or survivorship bias? |
| **RQ4** | What are the practical challenges in deploying AI-driven models in live trading? |
| **RQ5** | How do hybrid approaches enhance the robustness of trading signals? |

Table 2. Inclusion criteria.

| ID | Inclusion criteria (IC) |
|---|---|
| IC1 | Studies using machine learning as the main technique. |
| IC2 | Studies using trading strategies. |
| IC3 | Studies conducted on multiple markets. |
| IC4 | Studies using profitability metrics. |

Table 3. Exclusion criteria.

| ID | Exclusion criteria (EC) |
|---|---|
| EC1 | Articles not published in English. |
| EC2 | Research that addresses only sentiment analysis. |
| EC3 | Research that addresses only fundamental analysis. |
| EC4 | Research that addresses only technical analysis. |

Table 4. Quality criteria.

| ID | Quality criteria (QC) |
|---|---|
| QC1 | Clear research objectives. |
| QC2 | Results clearly explained. |
| QC3 | Detailed methodology. |

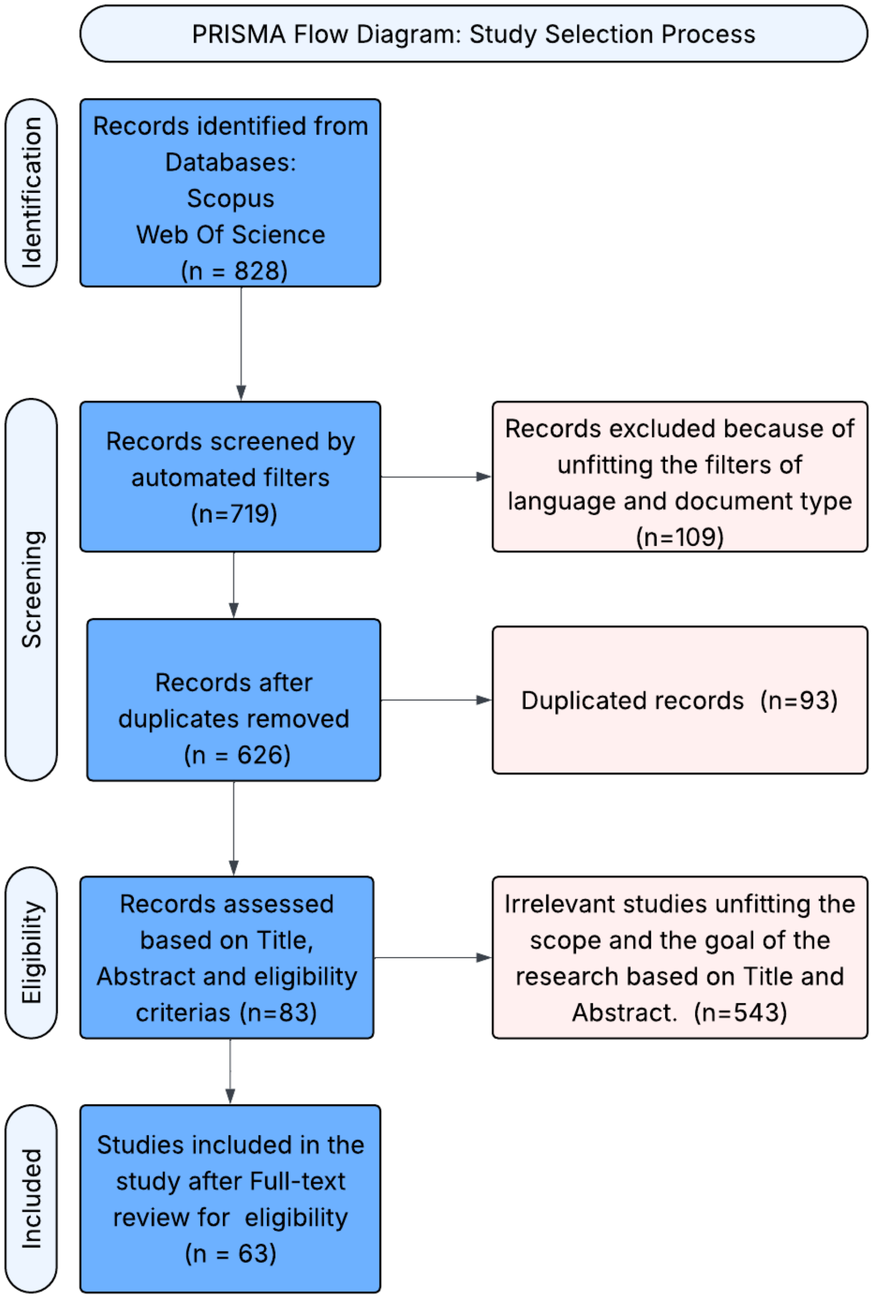


Figure 1. PRISMA Flow Diagram: Study Selection Process.

*2.0.2. Conducting the review.* This step consists of extracting relevant publications and selecting the articles based on the criteria defined previously. Therefore, we used the Scopus and Web of Science databases to extract articles, as they are leading academic references. We employed the following keywords as search descriptors: "Stock Market," "Deep Learning," "Forecasting," "Prediction," "Machine Learning," and "Technical Analysis." To maximize the number of relevant publications addressing these topics, we also incorporate plausible variants of these terms in the query. Accordingly, the final search string adopted for Scopus and Web of Science databases is: **((”Stock Market”) OR (”Equity Market”) OR (”Share Market”) OR (”Stock Exchange”) OR (”Finance”)) AND (”Deep Learning”) AND (”Machine Learning”) AND ((”Technical Analysis”) OR (”Technical Indicators”) OR (”Candlestick Analysis”) OR (”Candlestick Technique”) OR (”Quantitative Analysis”)) AND ((”Forecasting”) OR (”Predicting”))** While searching both databases, we use the previously defined string and keywords for filtering the search. Additionally, to set a limitation for data collection, we used ”articles” and ”proceedings”. The total number of publications on each platform using the search string is 828 articles. This process resulted in 735 articles.

After reading the abstracts and conclusions, we used platforms such as SSRN, IEEE Xplore, and ACM Library are used for downloading the full-text papers, examine them, and finish the analysis process. We applied the inclusion and exclusion criteria, resulting in 83 articles. Among these, 63 met the quality criteria.

Finally, this study is based on the remaining 63 articles after excluding those that did not meet the exclusion, inclusion, and quality criteria established in the previous tables. After reading the selected articles, we filled a table with all relevant information for each work. Such as the market, timeframe, data sources, dataset, attributes, the model used for prediction (predictor), the models used to compare with the predictor (comparisons), whether they used hybrid models or not, profitability metrics, performance metrics, whether they use alternative data, and what kind of models they used, neural networks or statistical models, and whether they use data integration techniques and what the addressed problematics are in the paper. In addition, we inspected the key findings and limitations of each paper. We also analyzed whether the authors used trading strategies and profitability metrics.

## 3. Results

The third step consists of analyzing the selected articles. Through the lens of technical approaches: Technical approaches, model assessment, validation, and robustness, real-world challenges, and hybrid models.

### *3.1. Analysis based on technical approaches*

We aim to analyze prediction techniques of the stock market prices and trends. This analysis examines machine learning techniques used for this task.

From Tables 5, 6, and 7, LSTM networks dominate the model landscape, accounting for 34 studies using LSTM as a single model or as a part of a hybrid model. This prevalence suggests that researchers find value in LSTM’s ability to capture temporal dependencies in financial time series data. This paper [40] explicitly acknowledges that LSTM struggles with volatility; the authors soften this by preprocessing using Empirical Mode Decomposition. The model is tested on BTC, which has the highest volatility in the dataset, yet the EMD-TI-LSTM has showed MAPE improvement of 78% on NASDAQ and 29.88% on BTC. The inherent ”random walk” nature of the financial market is acknowledged as a limit to predictability where near-perfect accuracy is neither realistic nor sustainable in the long-term. Overfitting is a recurring limitation that is identified across several articles. A primary driver of overfitting is the use of a relatively narrow interval of time or a small number of observations [30] reports LSTM achieving 92.45% accuracy across Indian NSE stocks from 2014-2018, studies focusing only on a few stocks. models trained on specific short periods risk of memorizing specific market regimes rather than learning generalizable patterns. For instance, [36, 30, 6] each focused on specific markets, such as the Iranian, Indian, and Indonesian markets, but without testing the robustness of these models on different market regimes and geographies, which shows that the high-value metrics are not indicative of true predictive power because models learn microstructure rather than market dynamics.

LSTM models show moderate to high overfitting risk across the literature. For instance, [2] demonstrates a notable train-test performance gap, producing 98.1% training accuracy versus 91.97% testing accuracy. [37] identifies that prediction models are limited to following the patterns in the data set without being able to adapt during anomalous periods, which is a sign of overfitting to historical noise. Additionally, [8] noted that no model works consistently across different stocks and market conditions, suggesting that models overfit to specific market regimes. However, the literature suggests strategies to address this issue, for example, ensemble approaches [50], can significantly reduce overfitting. Another suggested strategy is hyperparameter optimization via a metaheuristic algorithm [34] that uses the Sparrow Search Algorithm, making a 10.72% $R^2$ improvement. [15] used teaching and learning-based optimization, achieving an $R^2$ of 0.915 to prevent architecture-level overfitting. The consensus is that LSTM produces tangible results when working with time series data. However, pure LSTM is highly susceptible to overfitting, but hybrid architectures with regularization and optimization can mitigate this issue and bring the error to an acceptable level.

**Evaluating Market Volatility Using LSTM:** Regime changes cause the market's volatility. Literature shows many strategies used to address market volatility, first through multisource data integration and specialized hybrid architectures. [34] Mitigates volatility through sentiment indexing from forum posts combined with SSA-optimized LSTM. Additionally, papers [32, 17, 38] show that sentiment from news headlines, tweets, and psychological sentiment classifications improves accuracy under volatile conditions by capturing exogenous shocks.

Table 5. Model-level hybrid models

| Authors | Year | Predictor |
|---|---|---|
| [37] | 2022 | LSTM-AE |
| [58] | 2023 | TCN + Transformer Encoder |
| [13] | 2022 | ABC-LSTM |
| [1] | 2022 | CNN-LSTM |
| [16] | 2022 | PSO / AutoML-DNN |
| [8] | 2023 | LSTM-CNN and CNN-LSTM + Attention |
| [24] | 2023 | ResNLS |
| [33] | 2023 | Transformer + Time2Vec + multiresolution analysis |
| [53] | 2023 | Attention CNN-BiLSTM → LGBM (CNN-BiLSTM) |
| [51] | 2023 | ConvLSTM + Knowledge Graph + GCN |
| [39] | 2024 | Meta-learning + XGBoost |
| [15] | 2023 | HPT-HCLSTM |
| [44] | 2024 | ARIMA-LSTM / ARIMA-GRU |
| [9] | 2022 | Rolling-window RNN |
| [57] | 2022 | ARIMA-LSTM |
| [20] | 2023 | PCA-BP neural network |
| [54] | 2023 | Adversarial game LSTM + Attention |
| [23] | 2023 | HFS-GRU |
| [10] | 2023 | Multi-layer composite model |
| [61] | 2024 | LSTM + sliding window + convolutional filters |

### *3.2. Analysis based on Alternative data sources*

As this study focuses on AI models for stock market forecasting, it is common to go beyond classical market data, such as OHLCV and technical indicators, as mentioned in Table 9. Around 14% of studies combine historical market data with alternative data sources. Sentiment analysis is the most common alternative data type integrated, as seen in Table 9 Technical indicators were frequently retained alongside alternative data as mentioned in [18, 58, 13, 49, 21].

Our analysis revealed several distinct categories of alternative data being leveraged, such as News-based data [18], incorporating 25 headlines of the day with classical market data, while [58] and [27] developed a specialized model for integrating news and textual information. The following papers extract sentiment and other attributes from social media, most of the time from Twitter, such as [7], which analyzes sentiment polarity, sentiment subjectivity, hashtags, and other specific attributes. Researchers in [13] collected tweets about each company alongside classical market data and technical analysis. On the other hand, these studies [34, 49] leverage sentiment information from different sources; the first integrated sentiment indicators with classical market data, the second incorporated research report ratings and market sentiment indicators. Some studies utilize multiple alternative

Table 6. Parallel-ensemble-method models

| Authors | Year | Predictor |
|---|---|---|
| [45] | 2023 | Sig-DNN / Sig-RF |
| [46] | 2022 | Incremental learning vs Offline-Online learning |
| [59] | 2021 | two parallel RNNs combined via a fully connected layer |
| [6] | 2020 | LSTM vs XGBoost |
| [36] | 2020 | Multi-algorithm parallel ensemble |
| [50] | 2023 | RF + XGBoost + LSTM |
| [26] | 2020 | MDL |
| [18] | 2020 | ML vs DL parallel comparison |
| [60] | 2022 | Multi-objective tree ensemble |
| [63] | 2021 | RF vs DNN / XGBoost / SVM / LSTM |
| [55] | 2019 | Motif-CNN |
| [28] | 2023 | Stacking ensemble of neural networks |

Table 7. Architectural hybrid models

| Authors | Year | Predictor |
|---|---|---|
| [43] | 2022 | LSTM |
| [19] | 2018 | Correlation FS + ANN / CNN / LSTM |
| [25] | 2023 | LSTM and GRU |
| [42] | 2019 | RNN |
| [35] | 2021 | ANN and CNN with 2D histogram |
| [5] | 2024 | GRU with exogenous variables |
| [31] | 2023 | Transformer-based sequential model |
| [29] | 2021 | Transformer with self-attention |
| [14] | 2021 | Regression on moving averages |
| [47] | 2019 | LSTM and GRU comparison |
| [62] | 2019 | GAN (LSTM generator + MLP discriminator) |
| [41] | 2023 | LSTM / GRU + Attention |
| [22] | 2023 | Correlation FS + ANN / CNN / LSTM |
| [52] | 2024 | RNN / LSTM / GRU / CNN / XGBoost |

Table 8. Statistical-ML Hybrid Models

| Statistical-ML Hybrids | | |
|---|---|---|
| **Authors** | **Year** | **Predictor** |
| [4] | 2023 | ARMA-GARCH with neural networks |
| [44] | 2019 | RIMA-LSTM And ARIMA-GRU |
| [57] | 2022 | Hybrid ARIMA-LSTM model |
| [10] | 2024 | LM-SPP multivariate learning system |

data types, [21] employs both chip-based indicators and sentiment variables, while [59] combines news sentiment compound scores with traditional stock data. This paper [49] uniquely incorporates structured research reports. This review identified several approaches for integrating alternative data with classical market data; 7 out of 10 studies treated alternative data as additional input features alongside traditional data.

In general, the magnitude of improvement ranges from 3% to 10% when compared to the classical market data approach. In some cases, the use of alternative data didn't make any improvement, as in [18], where the selected stocks are IBM and JPMorgan. Those two companies are well-established and have been listed on the

Table 9. Data-Level Hybrids models

| Authors | Year | Attribute | Type of Data |
|---|---|---|---|
| [27] | 2018 | News + OHLCV | News Data |
| [34] | 2023 | Previous closing price + Complex price + Daily amplitude + Turnover rate + OHLCV + Sentiment indicators | Sentiment Analysis |
| [58] | 2022 | Historical stock market index values + Components of each news piece: date, Headlines, Preview & Description texts | News Data |
| [13] | 2020 | Date, OHLCV, Adjusted close, TIs: MA, Increase_in_vol, increase_in_adj_close, ADX, CCI, MACD, RSI, Collected tweets about each company | Social Media Data |
| [59] | 2021 | Adjusted closing stock prices, News sentiment compound scores, Date information | News Sentiment |
| [7] | 2020 | OHLCV, Tweet attributes: Sentiment polarity, Sentiment subjectivity, Hashtags, Mentions, Capital words, Urls, Punctuations | Social Media Sentiment |
| [48] | 2024 | Sentiment scores (positive, negative, neutral) from tweets + OHLCV + Feature vectors with sentiment scores and prior day's stock price change rate | Social Media Sentiment |
| [18] | 2020 | OHLCV + 25 Headlines | Sentiment Analysis |
| [49] | 2020 | OHLCV + Research report attributes: number of research reports, research report ratings, rating changes; market sentiment indicators | Sentiment Analysis |
| [21] | 2023 | Chip-based indicators: Transaction information, Institutional investor buy/sell data, Margin loan/stock loan information, Day trading metrics, Sentiment variables | Social Media Sentiment |

stock exchange for many decades, meaning they have more stable price movements and are less reactive to daily news cycles; events that might cause dramatic price movements in smaller or newer companies often have minimal impact on these large ones.

### *3.3. Analysis of Validation and Robustness*

The 63 papers studied in this literature employ different evaluation metrics, revealing significant gaps between reported performance and practical trading viability. Classification metrics dominate the literature, as stated. In Table 10, accuracy appears in 45 papers, precision in 22, and recall in 20. However, these metrics are fundamentally detached from trading profitability. A model achieving 95% accuracy at zero transaction costs becomes unprofitable at 0.2% cost. Only [33, 39] explicitly link classification performance to realized trading returns; the rest treat accuracy as an end in itself.

Error metrics are universally reported yet inadequate. Few papers report only baseline comparisons across different datasets. Financial metrics are the most relevant for practitioners but are critically underutilized. Sharpe ratio appears in only 6 papers, despite being a crucial metric for evaluating trading strategies. Other financial metrics such as drawdown, Sortino ratio, and Calmar ratio appear in 10 papers; 97% of papers employ statistical accuracy while only 9% report the use of profitability metrics. This reflects a misalignment between academic and practical trading.

Table 10. Table of metrics used by studies

| Metrics | Total |
|---|---|
| **Classification Metrics** | |
| Accuracy | 45 |
| Precision | 22 |
| Recall | 20 |
| F1-score | 14 |
| AUC | 3 |
| Confusion Matrix | 6 |
| **Error Metrics** | |
| MSE | 34 |
| RMSE | 39 |
| MAE | 31 |
| MAPE | 17 |
| R-squared | 8 |
| **Financial metrics** | |
| Sharpe ration | 6 |
| drawdown | 1 |
| Sortino ratio | 1 |
| Calmar ratio | 1 |
| **Statistical tests** | |
| Diebold-Mariano | 3 |
| The Friedman Test | 2 |
| **Robustness Assessment Methods** | |
| Multiple timeframes testing | 19 |
| Multiple markets comparisons | 17 |

Robustness assessment reveals diverse gaps. Multiple timeframe testing is used in 19 papers, and only 10 papers compare with the buy-and-hold strategy, while only 12 papers benchmark against simple technical indicators. Statistical significance testing is rare; only 3 papers employ Diaboldo-Mariano, Friedman tests, and none reported 95% confidence intervals around reported metrics.

Deployment constraints are largely absent; only 2 papers model transaction costs, while 14% mention real-time latency without quantification. The gap between research and practice is critical, raising questions about the models' accuracy and successful economic deployment. In summary, statistical accuracy monopolizes the literature, while financial viability and profitability assessments remain scarce.

### *3.4. Analysis based on real-world challenges*

During this study, we identified several real-world challenges affecting the implementation and performance of machine learning models for stock market prediction.

30% of studies acknowledge using narrow datasets, either focusing on a single market such as the US or China or limited assets such as ETFs like S&P500, NASDAQ, or Shanghai Composite Index or famous stocks such as Apple, Amazon, Microsoft, and Oracle in the US market or other famous firms in the Chinese markets as well, such as BYD or Shanghai Airport, or short time periods such as intraday trading. This study reveals a significant bias in this field: many markets are rarely studied or not mentioned as highlighted above. For instance, Latin American, African, and South Asian countries are understudied or not mentioned in the literature. The field focuses on developed and large emerging markets such as the US, UK, and Chinese markets. This phenomenon can be explained by many factors, such as the quality and abundance of data, which helps researchers produce statistically significant results. On the other hand, the scale of the US or Chinese markets is incomparable to the other markets, where the capitalization of the developed markets stands at trillions, while others stand at only billions. Finally, the high trading volumes in the US and Chinese markets allow institutional investors to easily enter or exit the market. These characteristics have made these markets attractive to investors and academics for decades, resulting in a large number of high-quality papers being produced and cited.

We noticed that 25% of the papers exhibit geographical limitations: while models are trained on specific markets, they fail in predicting stocks from other markets, highlighting a lack of generalizability. In these works [27, 34, 58, 59, 48, 38, 26, 49, 11, 21], which represent 19% of the studies, researchers encountered difficulties incorporating alternative data sources like news, sentiment, and social media information. Technical limitations, such as computational complexity, are highlighted in these papers, [37, 6, 36, 5, 39, 29, 60, 20, 11, 61, 28]. 17.2% of studies reported high computational limitations when working with deep learning models, and 15% especially mentioned long times limitation for LSTM-based models compared to alternatives. Another technical limitation identified in this study is overfitting. In [45, 37, 53, 51, 60, 23], especially with complex models, this issue is identified in 10% of the selected papers. We also noticed practical trading implementation challenges in these papers [56, 33, 63, 54]. 6.3% acknowledged of studies acknowledged that transaction costs significantly impact profitability.

Profitability is crucial in this field, yet it is mentioned in only 7 papers. Only 7.8% of selected papers [56, 37, 33, 39, 60], identify slippage when implementing trading strategies, and 17% noted difficulties translating prediction accuracy into profitable trading strategies. While accuracy is mentioned extensively for assessing the model's performance, profitability remains absent, which suggests a bias between statistical and economic significance because a mathematically accurate model can still be unprofitable. To calculate profitability, researchers should include brokerage fees, exchange fees, taxes, slippage, and profitability metrics such as the Sharpe ratio, maximum drawdown, and PnL. These calculations may involve numerous assumptions that vary by broker and country, and disagreement over them may lead to the paper's rejection. Furthermore, the studies are oriented toward a data science perspective in which the ultimate goal is to minimize the loss function. This is noticed by the calculation of diverse error metrics rather than maximizing utility through risk-adjusted returns.

In addition, practical deployment is crucial; 10 articles underline deployment challenges. In [46], we noticed that forecasting delay, the time between data retrieval and data prediction, reveals a serious concern about real-time trading. The lack of detailed exploration of real-world trading strategy integration is evident in [39] because it involves real risk and real money. Another important challenge is interpretability and trust. We can not trust a model even if it has high accuracy and high prediction potential if we can not understand how predictions are generated. This is called the black box problem. [5, 31, 12, 53, 51, 39, 44, 3, 18, 60, 20, 11, 10, 28] These are the types of challenges encountered when using deep learning models for prediction; 14 papers out of the dataset highlighted this issue. Additionally, these studies [37, 24, 48, 5, 31, 12, 39, 44, 29, 60, 62, 52] representing around one-fifth of the papers we studied, which revealed reliability issues—namely, inconsistent performance across different market conditions and fundamental limitations in prediction accuracy and confidence. In [48], we noticed that the error margin remains too high for a practical real-world application, while [52] proposes a model that struggles with rapid price fluctuation, particularly in periods of higher volatility.This model's [63] performance varied across different time periods.

### *3.5. Analysis based on hybrid models*

Since this study focuses on machine learning for predicting the financial market, it is essential to examine what drives superior market prediction. Many researchers have developed research using hybrid models for market prediction, whereas the majority of them are linked to neural networks. However, the term "hybrid" is used differently across studies, mixing together very different types of approaches. To bring clarity to this confusion, we propose a four-tiered taxonomy to provide conceptual clarity.

- data-level hybrids
- model-level hybrids
- parallel-ensemble methods
- architectural hybrids

This distinction is crucial because performance claims cannot be compared without understanding how these models are actually built .

**Data-Level Hybrids:** These models combine multiple sources, such as sentiment, technical indicators, fundamental data, and news, at the preprocessing stage and work with a single prediction architecture in these papers [19, 17, 48, 5]. The assumption behind this approach is that diverse data types contain complementary signals when integrated early, allowing the model to learn about signal interactions. However, we notice that the literature reports modest improvements. This paper [58] shows 3.39% accuracy gain after adding news data.

**Model-Level Hybrids** This type of hybridization sequentially combines two or more distinct algorithms, typically for denoising or feature extraction, followed by prediction. The DAELM architecture that incorporates DWT-AE-ELM layers in [56] is an illustration of this: the discrete wavelet transform removes noise, an autoencoder extracts features, and extreme learning machines generate predictions. Similarly, [3] uses ICEEMDAN decomposition followed by ensemble deep learning, achieving RMSE values of 0.031 to 0.244 across indices. The critical insight here is that the increase in performance comes mainly from the denoising stage rather than the model combination. This is evident when comparing results such as $R^2 = 0.905$ to 0.998 against traditional deep learning approaches; this result suggests it is not the hybrid architecture but the decomposition-based preprocessing that drives the performance. Yet the literature conflates preprocessing rigor with hybrid synergy, which leads to confusion.

**Parallel-Ensemble Hybrids:** Parallel ensembles train independent models and combine predictions at a decision layer via voting or stacking mechanisms. This paper [59] uses blending ensemble learning combining RNNs with a fully connected network, which helped to reduce the MSE from 438.94 to 186.32, which means a 57.55% reduction. Similarly, [50] implements a Random Forest + XGBoost + LSTM ensemble, and [20] uses stacking policies that improve accuracy by 2-7% and reduce error by 0.01-0.03 compared to single models. Parallel-ensemble approaches offer theoretical advantages through their straightforward design; these models are capable of reducing prediction variance without increasing bias when base learners maintain adequate diversity and independence. However, in the current research, the process of examining metrics for evaluating base learner diversity is absent. This analytical gap raises concerns about whether reported ensemble improvements genuinely result from sound methodological principles or from selective reporting of favorable base learner combinations

**Architectural Hybrids:** Architectural hybrids combine different neural network modules within a single, end-to-end differentiable system. For instance, CNN-LSTM [1, 8] architectures stack convolutional layers to extract spatial features with LSTM layers, which learn temporal dependencies as a unified structure. Similarly, attention-augmented models like Transformer-LSTM-Attention [38, 14] and the proposed CNN-BiLSTM-ATT [31] employ self-attention to selectively weight temporal features. Here, the hybrid nature is intrinsic because models are jointly optimized via backpropagation rather than the additive approach of combining separate algorithms after training. The distinction is important: architectural hybrids may face challenges with gradient flow when dealing with deep sequences of different modules. Whereas ensemble methods avoid this issue by design. A specific paper [53] reports that the attention CNN_BiLSTM with LGBM regressor achieves better performance than the first experimental set that contains several baseline models. Yet determining whether the attention mechanism or the regression model substitution contributed more to these improvements remains difficult. This analytical limitation is common in architectural hybrid research, as identifying the performance contributions of individual modules typically requires ablation studies, which the literature fails to include.

**Serial Statistical-Machine Learning Hybrids:** Serial statistical-machine learning hybrid models, for instance, ARIMA + LSTM/GRU, argue that linear models capture trend while neural networks capture nonlinearity. This paper [30] reports that ARIMA-LSTM and ARIMA-GRU outperform standalone LSTM/GRU when tested on the Shanghai Composite opening spread. Specifically, ARIMA-GRU achieves the best performance. Additional study [44] confirms that hybrid approaches integrate complementary strengths, stating that the ARIMA-LSTM model performs better in prediction and outperforms other benchmark methods. The dependence on dataset size is underexplored. Small datasets with fewer than 500 samples may favor classical statistical methods, which tend to favor low variance, whereas deep learning tends to favor high variance.

## 4. Discussion

From the analyses in previous subsections and the information in Tables 5, 6, 7, 8, 9, we can visualize the most used tools and techniques and the gaps and challenges over the years. In this systematic research, we observed a peak in 2023, showing that machine learning techniques have been widely used in the field of stock predictions, but the number declined in 2024.

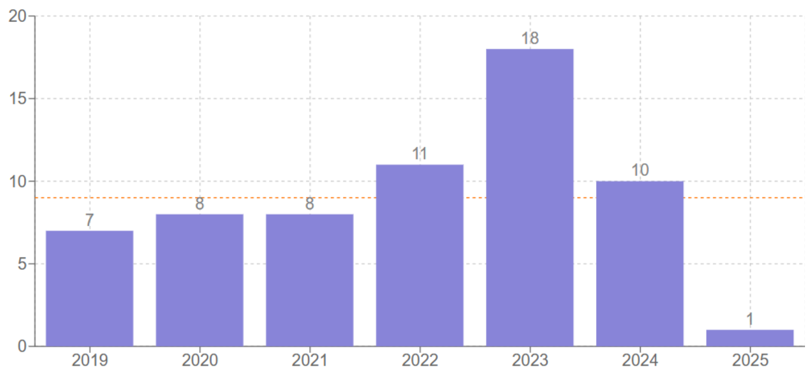


Figure 2. Included publications by year

On the other hand, Figure 3 shows an increase in the adoption of Transformer models and attention mechanisms over the last three years. Transformers excel in capturing high-lag dependencies, which can be beneficial in the stock market, where patterns can exist across different time scales. A distant temporal relationship can be of paramount importance, and the pre-training capability of transformers allows them to leverage knowledge from broader financial contexts.

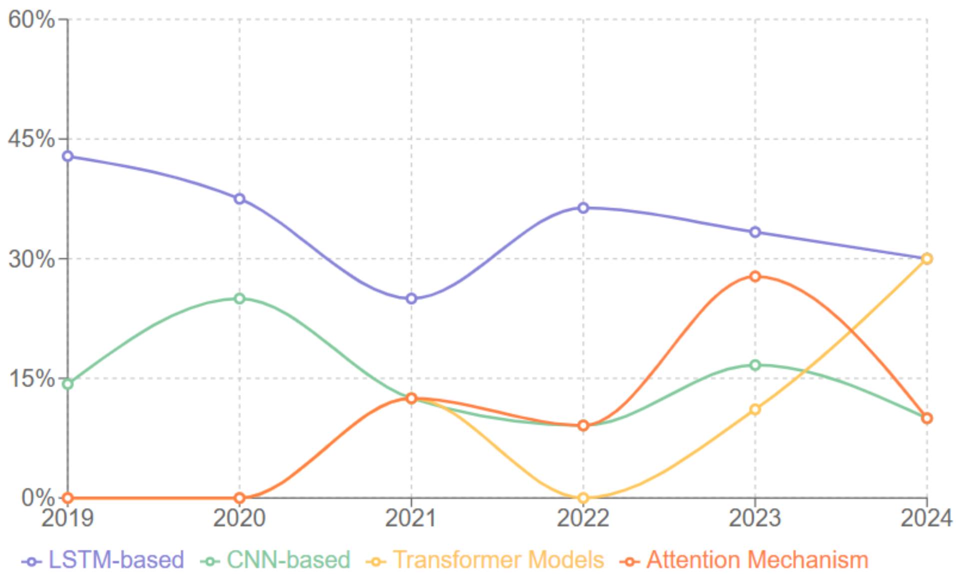


Figure 3. Relative adoption of ML architectures by year.

Furthermore, we analyze ML techniques used in each architecture type, thus answering the first research question.

(**RQ1: What ML techniques are most commonly used for stock market prediction models, and how do they differ from traditional technical indicators?**) By far, the LSTM network is the most widely used, in 53% of the papers, as it is a powerful algorithm for time series forecasting, particularly effective for processing sequential financial data, and capturing long-term dependencies. The fundamental differences between the ML approaches

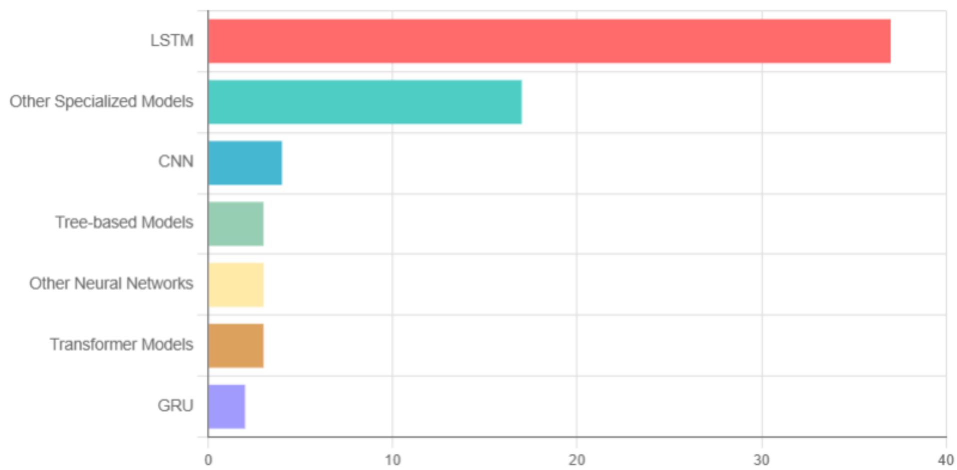


Figure 4. Frequency of ML techniques across reviewed studies.

and traditional technical indicators lie in adaptive learning. AI models learn patterns directly from data, while traditional indicators rely on fixed formulas. They can automatically discover relevant features and incorporate non-traditional data sources like sentiment and news. They can also identify complex non-linear patterns that simple moving averages or oscillators cannot detect.

Answering the second question (**RQ2 How do AI-based models incorporate alternative data?**) Only 9 papers out of 63 incorporate alternative data; the vast majority rely solely on direct input data. Although a limited number of papers incorporate alternative data, the findings are worth examining critically. Among the reviewed papers, three studies use social media sentiment, two papers use news data combined with sentiment analysis, and only one paper each uses news sentiment and raw social media data.

Besides the fact that few papers incorporated alternative data, the findings still hold relevant insights. First, data quality and potential bias remain significant issues, as noted by papers [51, 38], where sentiment dictionaries and context-dependent interpretation may introduce subjectivity. Second, even among papers that combine alternative and numerical data, the architectural design consistently privileges price-based inputs; for instance, in models such as LSTM and CNN, layers are predominantly trained on OHLCV prices rather than textual or sentiment features. These so-called alternative features are used as secondary signals and are not treated as structurally equivalent. This raises questions of whether true multimodal fusion is achieved or merely feature concatenation. Third, the adoption of alternative data appears constrained by reproducibility and accessibility barriers: several papers operate with non-English or emerging market contexts, suggesting that language specificity and data availability limit the generalizability when using alternative data approaches across markets. Finally, none of the reviewed papers use industry alternative data such as satellite imagery, earnings call transcripts, or macroeconomic indices, pointing to a meaningful gap between academic research practice and real-world deployment.

Figure 6 illustrates the number of studies using error metrics, financial metrics, classification metrics, and statistical tests. The figure shows that error metrics are the most used to compare models, thus answering the third question (**RQ3 - What evaluation frameworks are used to validate AI-driven indicators, and how do they address overfitting or survivorship bias?**) The current literature exhibits systematic methodological limitations. First, financial prediction literature lacks standardized evaluation protocols, which makes the comparisons across architectures difficult. Second, ablation studies are absent. Complex architecture using hybrid or ensemble methods

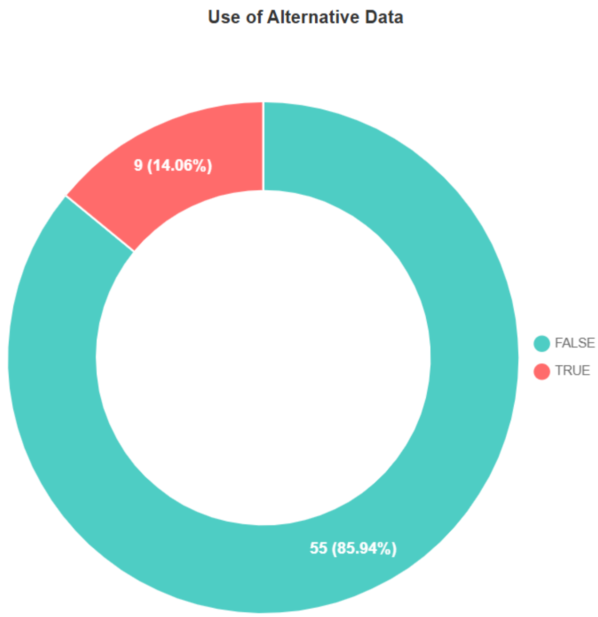


Figure 5. Proportion of studies using alternative data

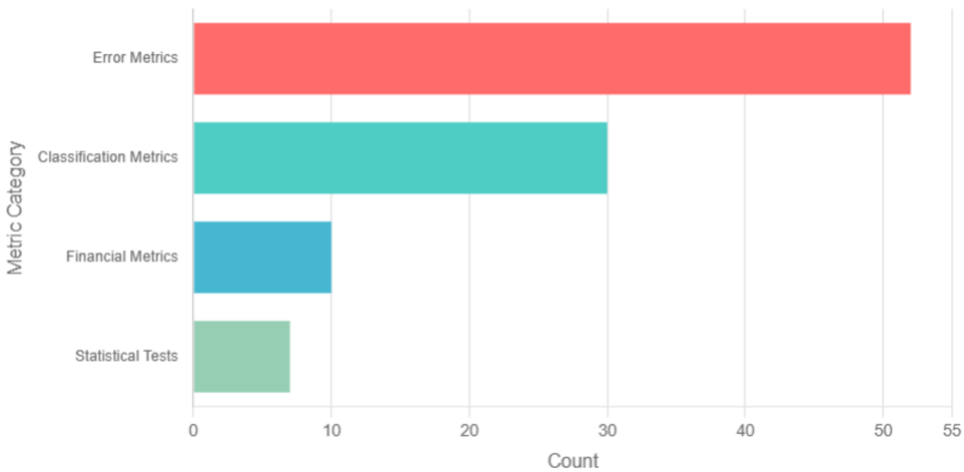


Figure 6. Number of studies by metrics

should report the contribution of each component to enable the identification of which component is the source of improvement. Few papers adopt this approach, while it remains absent in the other studies. Third, statistical significance is absent; for example, a 2% improvement in accuracy may be within the noise of cross-validation variability without confidence intervals or hypothesis tests to quantify uncertainty. Fourth, profitability metrics such as transaction costs, slippage, Sharpe ratio, and P&L are rarely incorporated; the literature employs metrics that measure prediction fidelity, not trading profitability, because a model achieving 95% RMSE still can achieve negative Sharpe ratios after calculating transaction costs. Only 4 papers apply backtesting with realistic costs; the remaining papers conflate accuracy with alpha generation. Fifth, 85% of papers report only test-set performance, making the overfitting magnitude invisible. To prevent this problem, many papers such as [19, 8, 50] use expanding-window cross-validation, which reduces but does not eliminate look-ahead bias. Sixth, survivorship bias is acknowledged but rarely addressed; only 3 papers mention this problem. This paper [5] mentions that bankrupt stocks are delisted from analysis, while the remaining papers conduct analysis on established indices such as KSE300, DSE30, and HS300, which exclude failed companies from inception. The results are reconstructed based on a selection bias, which will lead to an overestimate of future performance on new or failing companies.

Answering the fourth question **RQ4: What are the practical challenges in deploying AI-driven indicators in live trading?** Based on the analysis of 63 papers, we identified several key challenges.

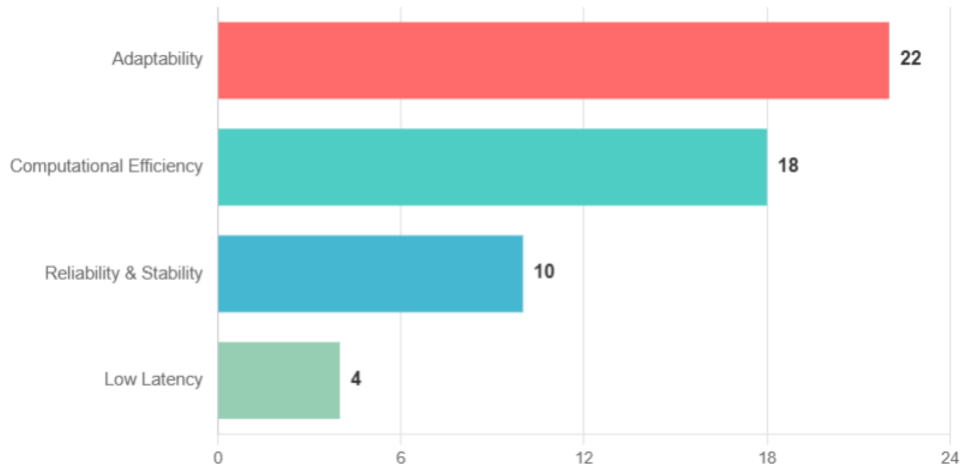


Figure 7. Deployment challenges by frequency

The deployment of AI-driven indicators needs low-latency processing because of the need to generate signals with minimal delay, processing incoming market data in real-time, and delivering actionable insights before market opportunities disappear. As stated in these articles, [36], [48], [42], and [52], the proposed models demonstrated a delay in capturing rapid price changes and movements.

For an indicator to be high-performing, it should operate within the computational constraints of trading platforms and must adapt during high-volume market periods. The authors of these articles, [58, 40, 6, 23, 39] highlight that the use of complex methodologies requires significant computational resources, making deployment very challenging and costly.

Furthermore, reliability and stability are crucial because the system must function continuously without crashes or unexpected behavior and needs to maintain consistent performance during market volatility. These studies [40, 1, 8, 19, 29] identify that observed models may not account for unexpected market events or external factors, revealing a fundamental vulnerability, and that no single model consistently outperformed all others across all datasets, suggesting that reliance on any single model is risky.

Finally, adaptability is a cornerstone of performance because the indicator should continuously learn from new market data, must detect and adjust to regime changes, and should incorporate feedback loops from actual trading results, as demonstrated in these articles [46, 62, 39, 1, 14]. Fixed parameters selected during initial training gradually lose effectiveness, and there is a constant retraining requirement, indicating that models need regular monitoring and updates to maintain accuracy.

Figure 7 shows that although LSTMs offer theoretical advantages in sequence modeling, they entail the highest implementation costs.

Finally, answering the fifth question **(RQ5—How do hybrid approaches enhance the robustness of trading signals?)** To answer this question, we analyze the success rates across sub-categories
Quantitative synthesis of performance across taxonomy tiers reveals striking disparities.

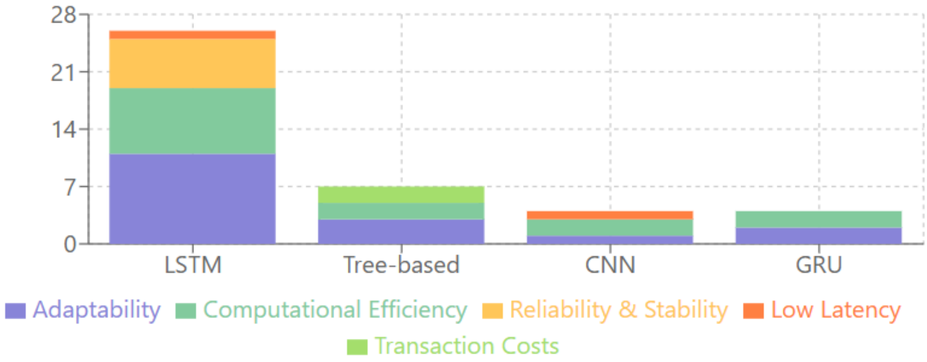


Figure 8. Deployment challenges by primary model type

**Data-level hybrid models** show inconsistent gains. This study [27] reports that an attention-based combination of news and numerical data achieves superiority, but it does not compare results with baseline models. Conversely, [34] integrates sentiment indices and LSTM, achieving a 10.74% $R^2$ improvement, yet this paper [17] provides a similar approach fusing sentiment with CNNBiLSTM-ATT, showing sentiment index improve the accuracy without specifying the metric of the baseline. We notice the absence of standardized evaluation protocols across papers. This highlights a methodological crisis. The lack of employing consistent metrics is a gap that makes the comparison of data-level hybrid performance difficult.

**Model-level hybrids** demonstrate more consistent, quantifiable improvements. For instance, this paper [56] reports that DAELM outperforms previous single-model approaches on 400 stocks. Additionally, this paper [40]reports that EMD-TI-LSTM achieves improvements of 39.56%, 36.86%, and 39.90% in MAPE, RMSE, and MAE, respectively, over baseline LSTM, while [3] reported that ICE2D2-MDL achieves $R^2$ values exceeding 0.90 across multiple datasets, substantially outperforming other baseline models. Yet a critical confound emerges: these gains conflate preprocessing quality, such as EMD or ICEEMDAN, with hybrid architectures. When [40] states the EMD-TI-LSTM outperforms LSTM, it is unclear whether the baseline LSTM received equivalent preprocessing. Otherwise, the improvement may be due to the noise-reduction efficacy. Few papers report this ablation; uniquely this study [3] explicitly evaluates LSTM-BN, GRU, and SVR on individual decomposed components before combining them in an ensemble. This approach lends credibility to the claims of the architectural advantage.

**Parallel-ensemble methods** show consistent but modest gains relative to single models. These papers [50, 63, 28] report 27% accuracy improvements. This paper [63] states that the ensemble combination strategy matters more than diversity. However, interpreting a 2% accuracy improvement in terms of trading profitability or risk-adjusted returns is rarely considered important. A 2% improvement in directional accuracy might translate to 0% improvement in the Sharpe ratio if transaction costs consume the edge. The literature treats machine learning accuracy as an end in itself without connecting the statistical metrics to the field reality.

**Architectural hybrids** show the highest variability. CNN-LSTM models and their variants show promise and report gains in many studies [1, 8, 53], yet [8] states clearly that no model shows consistent dominance across various stocks and conditions. This suggests that architectural hybrids exhibit high variance across assets, different market regimes, and diverse training windows. [31] proposed a CNN-BiLSTM-ATT model with sentiment, which achieves improved accuracy relative to the baseline, but the improvement magnitude is not quantified. Furthermore, [38] states that the transformer model achieves an accuracy above 90% in predicting the closing price; this metric lacks context to be adequately interpreted. This paper didn't define the prediction task exactly, which leads to a blurred interpretation.

**Serial Statistical-Machine Learning Hybrids:** Appears to outperform traditional statistical and traditional machine learning methods on small datasets, but we have not found clear evidence that they perform better on large datasets. Additionally, to conclude the superiority of this type of model, we need to inspect whether differencing a dataset and preprocessing the features before feeding them to deep learning models would give the same results.

This will help to understand if the superiority is due to the preprocessing quality or to the architecture itself. Finally, experiments over varying dataset sizes are absent from the literature and need careful examination.

## 5. Conclusion

This article aims to review the academic literature on stock market forecasting using machine learning models to assess the potential of transforming models into an AI-driven indicator for stock trading. Through the application of a structured research protocol and well-defined inclusion and exclusion criteria, 63 articles were selected for this study. Thus, analysis and discussion are structured around four main perspectives: predictor techniques, key findings, limitations, and profitability metrics. We acknowledge the following limitations of this study: the collected papers are limited to English-language research, we relied solely on Scopus and Web of Science databases, limiting this study to peer-reviewed papers only and missing proprietary research, gray literature, and non-English publications. Furthermore, the collected papers use different datasets, making the qualitative assessment difficult.

This systematic literature review reveals a persistent crisis of interpretability. The literature overwhelmingly favors deep learning precisely and other complex architectures, yet provides minimal insight into the mechanism of prediction. This gap creates multiple systemic risks, such as regulatory risk and model accountability, data drift, and model failure, demanding immediate XAI adoption to support the explanation of the decision mechanisms, diagnose causes and likelihoods of failure, and ensure continuous monitoring and rapid diagnosis of model degradation. This critical review establishes a taxonomy of hybrid learning, compares empirical success rates across subcategories, and reveals critical gaps, including the absence of controlled experiments isolating the effects of dataset size and preprocessing quality, as well as ablation studies and profitability assessments. Until such experiments are conducted, claims about hybrid superiority remain speculative.

The field would benefit from standardized benchmarks, rigorous ablation, realistic backtesting, and statistical hypothesis testing. These foundational practices remain rare in financial prediction literature.

REFERENCES

1. W. Abdullah and A. Salah. A novel hybrid deep learning model for price prediction. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(3):3420, June 2023.
2. H. Agarwal, G. Mahajan, A. Shrotriya, and D. Shekhawat. Predictive data analysis: Leveraging rnn and lstm techniques for time series dataset. *Procedia Computer Science*, 235:979–989, 2024.
3. Z. D. Akşehir and E. Kılıç. Multi level perspectives in stock price forecasting: Ice2de-mdl. *PeerJ Computer Science*, 10:e2125, 2024.
4. D. Alaminos, M. B. Salas, and A. Partal-Ureña. Hybrid arma-garch-neural networks for intraday strategy exploration in high-frequency trading. *Pattern Recognition*, 148:110139, Apr. 2024.
5. F. M. M. Alsheebah and B. A. Al-Fuhaidi. Emerging stock market prediction using gru algorithm: Incorporating endogenous and exogenous variables. *IEEE Access*, 12(while):132964–132971, 2024.
6. A. Bimo Gumelar, S. Nilowardono, A. Teguh, W. Fakultas, I. Komputer, H. Setyorini, M. Sulistyono, D. Pramono, A. Widodo, and E. Christine. 2020 international seminar on application for technology of information and communication (is emantic) boosting the accuracy of stock market prediction using xgboost and long short-term memory. IEEE, 2020.
7. S. Bouktif, A. Fiaz, and M. Awad. Augmented textual features-based stock market prediction. *IEEE Access*, 8:40269–40282, 2020.
8. W. Buachuen and P. Kantavat. Automated stock trading system using technical analysis and deep learning models. In *Proceedings of the 13th International Conference on Advances in Information Technology*, number practices in IAIT 2023, pages 1–9. ACM, Dec. 2023.
9. T. Bui, Khoa, D. Phuoc Vinh, T. Hiep, N. Quang, and T. Huynh. Predict stock price movement using long short-term memory recurrent neural network and support vector machine. *Quality - Access to Success*, 24(194), Jan. 2023.
10. Z. Chang and Z. Zhang. *The Composite Multi-Level Prediction Model for Stock Market Time Series*. IOS Press, Mar. 2024.
11. S. Chen and C. Zhou. Stock prediction based on genetic algorithm feature selection and long short-term memory neural network. *IEEE Access*, 9:9066–9072, 2021.
12. J. Choi, S. Yoo, X. Zhou, and Y. Kim. Hybrid information mixing module for stock movement prediction. *IEEE Access*, 11(using):28781–28790, 2023.
13. N. Deepika and M. Nirupamabhat. An optimized machine learning model for stock trend anticipation. *Ingénierie des systèmes d information*, 25(6):783–792, Dec. 2020.
14. S. Dinesh, N. R. R, S. P. Anusha, and S. R. Prediction of trends in stock market using moving averages and machine learning. In *2021 6th International Conference for Convergence in Technology (I2CT)*, number experiment, pages 1–5. IEEE, Apr. 2021.

15. J. Faritha Banu, S. B. Rajeshwari, J. S. Kallimani, S. Vasanthi, A. Mateen Buttar, M. Sangeetha, and S. Bhargava. Modeling of hyperparameter tuned hybrid cnn and lstm for prediction model. *Intelligent Automation amp; Soft Computing*, 33(3):1393–1405, 2022.
16. D. A. Fattah, A. A. Naim, A. S. Desuky, and M. S. Zaki. Autokeras and particle swarm optimization to predict the price trend of stock exchange. *Bulletin of Electrical Engineering and Informatics*, 11(2):1100–1109, Apr. 2022.
17. S. Gite, H. Khatavkar, K. Kotecha, S. Srivastava, P. Maheshwari, and N. Pandey. Explainable stock prices prediction from financial news articles using sentiment analysis. *PeerJ Computer Science*, 7:e340, 2021.
18. J. Guo and B. Tuckfield. News-based machine learning and deep learning methods for stock prediction. *Journal of Physics: Conference Series*, 1642(1):012014, Sept. 2020.
19. M. Gurav. Adaptive stock forecasting model using modified backpropagation neural network (mbnn). IEEE, 2018.
20. L. Hang, D. Liu, and F. Xie. A hybrid model using pca and bp neural network for time series prediction in chinese stock market with topsis analysis. *Scientific Programming*, 2023:1–15, June 2023.
21. J.-Y. Huang, C.-L. Tung, and W.-Z. Lin. Using social network sentiment analysis and genetic algorithm to improve the stock prediction accuracy of the deep learning-based approach. *International Journal of Computational Intelligence Systems*, 16(1), May 2023.
22. A. Ifleh and M. El Kabbouri. Stock price indices prediction combining deep learning algorithms and selected technical indicators based on correlation. *Arab Gulf Journal of Scientific Research*, 42(4):1237–1256, Oct. 2023.
23. I. Javid, R. Ghazali, I. Syed, M. Zulqarnain, and N. A. Husaini. Study on the pakistan stock market using a new stock crisis prediction method. *PLOS ONE*, 17(10):e0275022, Oct. 2022.
24. Y. Jia, A. Anaissi, and B. Suleiman. Resnls: An improved model for stock price forecasting. *Computational Intelligence*, 40(1), Nov. 2023.
25. S. Joseph, N. Mduma, and D. Nyambo. A deep learning model for predicting stock prices in tanzania. 2023.
26. P. Q. Khang, M. Hernes, K. Kuziak, A. Rot, and W. Gryncewicz. Liquidity prediction on vietnamese stock market using deep learning. *Procedia Computer Science*, 176:2050–2058, 2020.
27. d. Liu and X. Wang. A numerical-based attention method for stock market prediction with dual information. IEEE, 2018.
28. H. Lu and X. Peng. Better effectiveness of multi-integrated neural networks: Take stock big data as an example. *Wireless Communications and Mobile Computing*, 2021(1), Jan. 2021.
29. N. Malibari, I. Katib, and R. Mehmood. Predicting stock closing prices in emerging markets with transformer neural networks: The saudi stock exchange case. *International Journal of Advanced Computer Science and Applications*, 12(12), 2021.
30. P. Mehta, S. Pandya, and K. Kotecha. Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*, 7:e476, Apr. 2021.
31. T. S. Mian. Evaluation of stock closing prices using transformer learning. *Engineering, Technology amp; Applied Science Research*, 13(5):11635–11642, Oct. 2023.
32. S. T. Mndawe, B. S. Paul, and W. Doorsamy. Development of a stock price prediction framework for intelligent media and technical analysis. *Applied Sciences*, 12(2):719, 2022.
33. S. A. Mohammed. Day trading strategy based on transformer model, technical indicators and multiresolution analysis. *International Journal of Advanced Computer Science & Applications*, 15(4), 2024.
34. G. Mu, N. Gao, Y. Wang, and L. Dai. A stock price prediction model based on investor sentiment and optimized deep learning. *IEEE Access*, 11:51353–51367, 2023.
35. S. Mukherjee, B. Sadhukhan, N. Sarkar, D. Roy, and S. De. Stock market prediction using deep learning algorithms. *CAAI Transactions on Intelligence Technology*, 8(1):82–94, Aug. 2021.
36. M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. S. Deep learning for stock market prediction. *Entropy*, 22(8):840, July 2020.
37. V. Naidoo and S. Du. A deep learning method for the detection and compensation of outlier events in stock data. *Electronics*, 11(21):3465, Oct. 2022.
38. H. Niu, Q. Pan, and K. Xu. Hybrid deep learning models with multi-classification investor sentiment to forecast the prices of china's leading stocks. *PLOS ONE*, 18(11):e0294460, Nov. 2023.
39. K. Noor and U. Fatima. Meta learning strategies for comparative and efficient adaptation to financial datasets. *IEEE Access*, 13:24158–24170, 2025.
40. O. Ozupek, R. Yilmaz, B. Ghasemkhani, D. Birant, and R. A. Kut. A novel hybrid model (emd-ti-lstm) for enhanced financial forecasting with machine learning. *Mathematics*, 12(17), 2024.
41. A. Saboor, A. Hussain, B. Lord Y. Agbley, A. ul Haq, J. Ping Li, and R. Kumar. Stock market index prediction using machine learning and deep learning techniques. *Intelligent Automation amp; Soft Computing*, 37(2):1325–1344, 6 2023.
42. A. Sachdeva, B. Geet, V. Adapalli, C. Krishna, and Manjunath. An effective time series analysis for equity market prediction using deep learning model. IEEE, 2019.
43. T. Sanboon, K. Keatruangkamala, and S. Jaiyen. A deep learning model for predicting buy and sell recommendations in stock exchange of thailand using long short-term memory. In *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, pages 757–760, 2019.
44. Y. Si, S. Nadarajah, Z. Zhang, and C. Xu. Modeling opening price spread of shanghai composite index based on arima-gru/lstm hybrid model. *PLOS ONE*, 19(3):e0299164, Mar. 2024.
45. T. Sidogi, W. T. Mongwe, R. Mbuvha, P. Olukanmi, and T. Marwala. A signature transform of limit order book data for stock price prediction. *IEEE Access*, 11:70598–70609, 2023.
46. T. Singh, R. Kalra, S. Mishra, Satakshi, and M. Kumar. An efficient real-time stock prediction exploiting incremental learning and deep learning. *Evolving Systems*, 14(6):919–937, Dec. 2022.
47. A. Site, D. Birant, and Z. Işık. Stock market forecasting using machine learning models. In *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pages 1–6, 2019.

48. N. Smatov, R. Kalashnikov, and A. Kartbayev. Development of context-based sentiment classification for intelligent stock market prediction. *Big Data and Cognitive Computing*, 8(6):51, May 2024.
49. H. Song, D. Peng, and X. Huang. Incorporating research reports and market sentiment for stock excess return prediction: A case of mainland china. *Scientific Programming*, 2020:1–7, May 2020.
50. G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, 11(3):94, July 2023.
51. M. Tao, S. Gao, D. Mao, and H. Huang. Knowledge graph and deep learning combined with a stock price prediction network focusing on related stocks and mutation points. *Journal of King Saud University - Computer and Information Sciences*, 34(7):4322–4334, July 2022.
52. D. M. Teixeira and R. S. Barbosa. Stock price prediction in the financial market using machine learning models. *Computation*, 13(1):3, Dec. 2024.
53. I. Valova, N. Gueorguieva, T. Aayushi, P. Nikitha, and H. Mohamed. Hybrid deep learning architectures for stock market prediction. In *Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Science*, number various in EECSS 2023. Avestia Publishing, Aug. 2023.
54. S. Wei, S. Wang, S. Sun, and Y. Xu. Stock ranking prediction based on an adversarial game neural network. *IEEE Access*, 10:65028–65036, 2022.
55. M. Wen, P. Li, L. Zhang, and Y. Chen. Stock market trend prediction using high-order information of time series. *IEEE Access*, 7:28299–28308, 2019.
56. D. Wu, X. Wang, and S. Wu. A hybrid framework based on extreme learning machine, discrete wavelet transform, and autoencoder with feature penalty for stock prediction. *Expert Systems with Applications*, 207:118006, Nov. 2022.
57. D. Xiao and J. Su. Research on stock price time series prediction based on deep learning and autoregressive integrated moving average. *Scientific Programming*, 2022(strategies.):1–12, Mar. 2022.
58. D. Xu, Z. Xu, S. Chen, and H. Fujita. A multi-channel cross-residual deep learning framework for news-oriented stock movement prediction. *Economic Research-Ekonomska Istraživanja*, 36(2), Aug. 2022.
59. W. Yang, R. Wang, S. Pinghe, S. Shanghai, and P. School. Detection of anomaly stock price based on time series deep learning models. In *2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID)*. IEEE, 2020.
60. T. Yin, X. Du, W. Zhang, Y. Zhao, B. Han, and J. Yan. Real-trading-oriented price prediction with explainable multiobjective optimization in quantitative trading. *IEEE Access*, 10:57685–57695, 2022.
61. Z. Zhan and S.-K. Kim. Versatile time-window sliding machine learning techniques for stock market forecasting. *Artificial Intelligence Review*, 57(8), July 2024.
62. K. Zhang, G. Zhong, J. Dong, S. Wang, and Y. Wang. Stock market prediction based on generative adversarial network. *Procedia Computer Science*, 147(technical):400–406, 2019.
63. P. Zhang, J.-Y. Yang, H. Zhu, Y.-J. Hou, Y. Liu, and C.-C. Zhou. Failure in stock price prediction: A comparison bettwen the curve-shape-feature and non-curve-shape-feature modes of existing machine learning algorithms. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS amp; CONTROL*, 16(6), Dec. 2021.