Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence and Machine Learning for Stock Market Prediction: A Systematic Literature Review of Methods, Emerging Paradigms, Reproducibility, and Evaluation Gaps (2016-2026)

Akshatha Srikantha
University of California, Irvine
Manpreet Singh
Boston University
Rahul Joshi
rahul.joshi@sitpune.edu.in

Symbiosis International (Deemed University)
Anant Singh
DV Trading LLC
Eugene Pinsky
Boston University

Research Article

Keywords: Stock market prediction, Deep learning, Machine learning, Systematic literature review, PRISMA 2020, Reproducibility, Transformers, Graph neural networks, Large language models

Posted Date: October 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11084191/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Artificial Intelligence and Machine Learning for Stock Market Prediction: A Systematic Literature Review of Methods, Emerging Paradigms, Reproducibility, and Evaluation Gaps (2016-2026)

Akshatha Srikantha [1†], Manpreet Singh [2†], Rahul Joshi [3*], Anant Singh [4], Eugene Pinsky [2]

[1]Department of Computer Science, University of California, Irvine, Irvine, CA, USA.
[2]Department of Computer Science, Metropolitan College, Boston University, Boston, MA 02215, USA.
[3*]Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, Maharashtra, India.
[4]DV Trading LLC, Chicago, IL, USA.

*Corresponding author(s). E-mail(s): rahul.joshi@sitpune.edu.in;
Contributing authors: asrikan4@uci.edu; manni@bu.edu;
asingh@dvtrading.co; epinsky@bu.edu;
†These authors contributed equally to this work as co-first authors.

Abstract

Stock prices are hard to predict: the underlying series are non-stationary, noisy, and non-linear, and over the past decade researchers have increasingly reached for artificial intelligence (AI) and machine learning (ML) rather than classical linear econometrics. This paper reports a systematic literature review (SLR), conducted under PRISMA 2020, of 685 primary empirical studies drawn from three scientific databases (Scopus, IEEE Xplore, and Web of Science) and published between 2016 and 2026. We ask four questions. First, which architectures does the field actually use, and how do shallow models, deep sequence networks, and hybrid ensembles compare across markets, horizons, and asset classes? Second, how has method evolved over the decade, from kernel machines through recurrent networks to Transformers, financial Large Language Models (LLMs), dynamic spatial-temporal Graph Neural Networks (GNNs), multimodal fusion,

and friction-aware Deep Reinforcement Learning? Third, how reproducible is any of it? We audit code, data, and environment availability, sort the corpus into a three-tier reproducibility hierarchy, and give mathematical form to five recurring data leakage mechanisms. Fourth, how are models evaluated? Here we describe what we call the "Lag-1 Shadowing Trap", along with scale-dependent error metrics, ignored execution frictions, and a near-absence of statistical hypothesis testing. LSTM networks appear in 89.2% of the studies and hybrid frameworks in 61.9%, yet only 7.6% release working source code, more than 83% report no formal significance test, and 45.8% never assess trading utility at all. We close with a Four-Pillar Standardization Protocol intended to narrow the gap between the accuracies papers report and the returns those models can plausibly earn.

Keywords: Stock market prediction, Deep learning, Machine learning, Systematic literature review, PRISMA 2020, Reproducibility, Transformers, Graph neural networks, Large language models

1 Introduction

1.1 Background and Motivation

Financial markets play a key role in allocating capital. They are the channel through which savings become corporate investment, and along the way they price assets, supply liquidity, and spread risk through the rest of the economy [1, 7]. Predicting equity prices, trends, and volatility has occupied economists for decades; computer scientists arrived at the problem later. Nobody pretends it is easy, but even a small and unreliable edge pays, so the interest never really fades.

Most of this work sits, explicitly or not, against the Efficient Market Hypothesis (EMH) as Eugene Fama formalized it [26]. In its weak form, anything the past trading record could reveal is already impounded in the price, and price changes follow a random Markovian martingale:

$$\mathbb{E}\left[P_{t+1} \mid \mathcal{F}_t\right] = P_t(1+\mu) \tag{1}$$

where $\mathcal{F}_t = \sigma(\{P_\tau\}_{\tau=0}^{t})$ is the information set up to time t and μ the expected drift. Taken at face value, that rules prediction out entirely. No algorithm should earn consistent risk-adjusted excess returns (alpha) unless it trades on inside information or has an unhedged systematic risk exposure hiding somewhere.

Markets have never fully obliged. Behavioral finance, experimental economics, and the microstructure literature keep documenting anomalies that refuse to die: cross-sectional momentum, mean reversion, post-earnings announcement drift, volatility clustering, retail herding [2, 9]. None of these sits comfortably beside strict informational efficiency. Andrew Lo's Adaptive Market Hypothesis (AMH) offers one way to hold both thoughts at once: efficiency, on his account, is not a constant of nature but something that strengthens and weakens as competition, risk appetites, and the learning of boundedly rational participants shift.

Between 2016 and 2026 the change was mostly practical. Computing got cheap, GPU acceleration became standard equipment, and open-source frameworks, TensorFlow and PyTorch above all, put the tooling within anyone's reach. Rigid linear specifications like Box-Jenkins ARIMA and GARCH gave way to high-capacity machine learning (ML) and deep learning (DL) architectures that learn temporal structure, cross-asset interactions, and multi-frequency signals directly from large, messy data streams [5, 6].

1.2 Research Objectives and Scope

Financial AI research has ballooned over the past few years, and the literature has sprawled with it. Plenty of studies report directional accuracies well past 90%. Take those numbers into a live market and they evaporate, and making sense of the gap means looking hard at how the studies were actually run.

That is what this review sets out to do. Working under the PRISMA 2020 protocol, we examine 685 primary empirical studies from 2016 through 2026 and organize the findings around four research questions (RQs):

- **RQ1 (Architectural Taxonomy & Comparative Performance):** *Which AI, machine learning, and deep learning techniques does the field actually use for stock market prediction, and how does performance stack up across equity markets, time horizons, and asset classes?*
- **RQ2 (Methodological Evolution & Emerging Paradigms):** *How did AI methodology for stock market forecasting change across the 2016-2026 decade, and what characterizes the frontier paradigms, their inductive biases and failure modes included, such as Large Language Models, dynamic Graph Neural Networks, and Deep Reinforcement Learning?*
- **RQ3 (Reproducibility & Open Science Practices):** *What proportion of empirical studies provide fully reproducible source code, open-access datasets, and transparent computational environments, and what specific methodological data leakage vectors compromise scientific validity?*
- **RQ4 (Evaluation Metrics & Standardization Gaps):** *What evaluation metrics are utilized to assess predictive models (RMSE, MAE, Directional Accuracy, Sharpe Ratio, Maximum Drawdown), what mathematical pitfalls distort reported findings, and to what extent is formal statistical hypothesis testing conducted?*

1.3 Contributions

This review makes five contributions:

1. **Scale of the corpus ($N = 685$).** As far as we are aware, this is the largest primary empirical corpus assembled for the stock prediction literature: 685 peer-reviewed studies from 2016-2026, drawn from Scopus, IEEE Xplore, and Web of Science (WoS).
2. **Mathematical treatment of the model families.** We give formal statements of the classical econometric models, support vector machines, tree ensembles, recurrent networks (LSTM), self-attention Transformers, dynamic Graph Neural

Networks (GNNs), and friction-aware Deep Reinforcement Learning that populate the corpus.

3. **A three-tier reproducibility audit.** We audit code, data, and environment availability across all 685 studies (the first audit of this scale in the area), classify each study into a three-tier hierarchy, and find that only 7.6% release functional source code.
4. **A data leakage matrix.** We formalize five leakage mechanisms that inflate reported results, among them pre-split global normalization, decomposition lookahead bias, and survivorship selection, and pair each with its mathematical mechanism and a mitigation protocol.
5. **A standardization protocol.** We propose a four-pillar reporting framework: open-science mandates, purged walk-forward validation, formal hypothesis testing (Diebold-Mariano and Friedman tests), and friction-aware economic backtesting.

2 Background and Theoretical Foundations

2.1 Mathematical Problem Formulation in Financial Forecasting

Formally, stock market prediction amounts to searching a hypothesis class $\mathcal{F}$ for a parameterized mapping $f_\theta^* : \mathcal{X} \to \mathcal{Y}$ that takes a lookback window of T past market observations to a target h steps ahead:

$$\hat{y}_{t+h} = f_\theta\left(\mathbf{X}_{t-T+1:t}\right) \tag{2}$$

where $\mathbf{X}_{t-T+1:t} = [\mathbf{x}_{t-T+1}, \mathbf{x}_{t-T+2}, \ldots, \mathbf{x}_t] \in \mathbb{R}^{T\times D}$ denotes a multidimensional feature tensor containing D heterogeneous feature dimensions at each discrete timestep $\tau \in [t-T+1, t]$.

How the target y_{t+h} is defined matters a great deal, and three formulations dominate in practice (Table 1):

1. **Continuous price or return regression.** The target is either the future price $P_{t+h} \in \mathbb{R}^+$ itself or the log return $r_{t,t+h} = \ln(P_{t+h}/P_t) \in \mathbb{R}$. Parameters θ are fit by minimizing a squared, absolute, or Huber loss over the training set $\mathcal{D}_{\text{train}} = \{(\mathbf{X}_i, y_i)\}_{i=1}^N$:

$$\mathcal{L}_{\text{MSE}}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - f_\theta(\mathbf{X}_i)\right)^2 \tag{3}$$

$$\mathcal{L}_{\text{Huber}}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}\frac{1}{2}\left(y_i - f_\theta(\mathbf{X}_i)\right)^2 & \text{if } |y_i - f_\theta(\mathbf{X}_i)| \le \delta \\ \delta|y_i - f_\theta(\mathbf{X}_i)| - \frac{1}{2}\delta^2 & \text{otherwise}\end{cases} \tag{4}$$

where $\delta > 0$ sets the point at which the loss switches from quadratic (small noise) to linear (tail outliers). Standard regression loss formulations implicitly assume Gaussian, homoskedastic error distributions $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. This assumption conflicts with empirical equity distributions that exhibit heavy tails, skewness, and volatility clustering, which have long been documented as core stylized facts in

financial econometrics [201]. Parameterizing return series with heavy-tailed α-stable or Student-t distributions [198, 202] captures extreme tail risks that standard Gaussian-based mean squared error losses fail to penalize.

2. **Directional movement classification.** The target is a binary or multi-class label for the sign of the future return:

$$y_{t+h} = \begin{cases} 1 \quad \text{(Bullish / Upward)} & \text{if } P_{t+h} > P_t(1 + \delta_{\text{deadband}}) \\ 0 \quad \text{(Bearish / Downward)} & \text{if } P_{t+h} < P_t(1 - \delta_{\text{deadband}}) \\ \text{Neutral} & \text{otherwise} \end{cases} \tag{5}$$

where the optional deadband $\delta_{\text{deadband}} \geq 0$ discards moves too small to clear bid-ask friction. Training minimizes the categorical cross-entropy:

$$\mathcal{L}_{\text{CE}}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} y_{i,k} \log P_\theta(y_{i,k} = 1 \mid \mathbf{X}_i) \tag{6}$$

3. **Trading policy optimization (reinforcement learning).** Here there is no intermediate price forecast at all: the model maps the market state $\mathbf{s}_t \in \mathcal{S}$ straight to portfolio weights $\mathbf{w}_t \in \Delta^N$ and is trained on cumulative risk-adjusted reward under execution friction.

Table 1 Comparison of Problem Formulations in Stock Market Prediction

Formulation	Target Variable (y_{t+h})	Loss Function ($\mathcal{L}$)	Key Strengths & Limitations
Price Level Regression	Absolute price $P_{t+h} \in \mathbb{R}^+$	MSE, MAE, Huber	Direct price target; highly vulnerable to non-stationarity and Lag-1 shadowing
Log Return Regression	Continuous return $r_{t,t+h} \in \mathbb{R}$	MSE, MAE, Quantile	Stationary target; low signal-to-noise ratio; hard to optimize with standard MSE
Directional Classification	Discrete label $y_{t+h} \in \{0, 1\}$	Binary / Categorical Cross-Entropy	Directly aligned with trading decisions; ignores magnitude of price movement
Policy Optimization (RL)	Target portfolio weights $\mathbf{w}_t \in \Delta^N$	Negative Reward / PPO Objective	Directly models transaction costs and execution friction; sample-inefficient

2.2 Classical Financial Economics vs. Computational Machine Learning

Classical asset pricing is built on equilibrium models with rational expectations. In the Capital Asset Pricing Model (CAPM) of Sharpe and Lintner, the expected excess return on asset i is linear in its market beta β_i:

$$\mathbb{E}[R_i] - R_f = \beta_i \left(\mathbb{E}[R_m] - R_f\right) \tag{7}$$

The Fama-French five-factor model adds size (SMB), value (HML), profitability (RMW), and investment (CMA) factors to the same linear skeleton:

$$R_{it} - R_{ft} = \alpha_i + \beta_{i1}(R_{mt} - R_{ft}) + \beta_{i2}\text{SMB}_t + \beta_{i3}\text{HML}_t + \beta_{i4}\text{RMW}_t + \beta_{i5}\text{CMA}_t + \varepsilon_{it} \quad (8)$$

These specifications are foundational but restrictive: factor loadings β are constant, errors are homoskedastic and Gaussian, and the factors themselves are taken as given. Machine learning drops those assumptions. Its flexible non-parametric models can pick up complicated feature interactions, conditional distributions that shift over time, and tail dependence that is anything but uniform. However, this comes at the expense of the interpretability that made linear models so valuable.

2.3 Temporal Forecasting Horizons and Market Microstructure

The nature of the prediction problem depends on the time horizon, and there are three main types to consider, worth distinguishing:

- **Intraday and high-frequency (from less than a second up to 60 minutes).** At the microstructure level, the most important features include limit order book (LOB) depth, order flow imbalance (OFI), bid-ask queue dynamics, and quote cancellation rates. Models in this category require very fast feature extraction and must handle challenges like tick-level bid-ask bounce and latency arbitrage between exchanges. We distinguish two sub-regimes within this bracket. True high-frequency trading (HFT), operating at sub-second to sub-minute latencies on raw LOB feeds, appears in only 11 of our 685 studies; these were included under IC1 when they evaluated equity price or return prediction, but excluded when the contribution was purely about optimal execution or market-making with no predictive evaluation. The remaining 57 intraday studies work at 1-minute to 60-minute bar frequencies, where the feature space resembles daily models scaled down in horizon rather than microstructure-native LOB representations.
- **Daily (one day ahead).** The bulk of the field lives here: 558 studies (81.5%) predict the next trading day's close or direction from daily OHLCV bars, technical indicators, and overnight news sentiment. One habit in this bracket deserves scrutiny: treating the day as a single, undifferentiated close-to-close interval. Overnight information assimilation, while the market is shut, and continuous daytime trading are different econometric animals. Lou, Polk, and Skouras [203] and Bogousslavsky [204] showed that overnight and intraday returns decouple systematically, pushed apart by institutional execution schedules and retail flow imbalances. Salotra et al. [197] found the same split in exchange-traded funds, where static and momentum strategies produce opposite risk-adjusted profiles by day versus night, and Zhang and Pinsky [200] together with Wu and Pinsky [199] observed sequence model accuracy moving around considerably once trading days were broken into discrete sessions. Collapse all of that into one daily close-to-close series and the intraday microstructure disappears from view, while unmodeled noise flows into the recurrent network's hidden states.

- **Multi-day and swing (5 to 30 days ahead).** Here the target is medium-term momentum, macro regime shifts, and earnings-driven trends. Because autoregressive errors compound over multiple steps, models at this horizon usually reach for sequence-to-sequence designs or direct multihorizon transformers.

2.4 Comparison with Related Systematic Reviews

Table 2 situates this review relative to earlier surveys of financial machine learning and econometric forecasting [1, 3–7, 10–18, 20–23, 25]. Each of those reviews covers a slice of the field: technical versus fundamental analysis [13], movement classification [17], deep learning backtesting [5], feature extraction pipelines [15], relational graph learning [22], econometric hybrids [20], and, in one case, a meta-synthesis of the surveys themselves [4].

What they share, in our reading, are some structural gaps. Samples rarely exceed 150 primary studies. None audits source-code availability at scale. None gives data leakage a mathematical treatment. Econometric testing protocols such as Diebold-Mariano and Friedman go largely unmentioned. And coverage of the post-2023 frontier (financial foundation LLMs, dynamic spatial-temporal GNNs, execution-constrained DRL) is thin, simply because most of these surveys predate it. We try to close each of these gaps with a corpus of 685 primary studies spanning 2016-2026.

Table 2 Systematic Comparison with Prior Survey and Literature Reviews on AI/ML for Stock Market Prediction

Review Study	Year	*N*	Span	PRISMA	Code	Leak	Key Focus & Disciplinary Delta
Nti et al. [13]	2019	122	2007–2018	Partial	No	No	Fundamental vs. technical analysis; early ML models
Bustos & Pomares [17]	2020	117	2000–2019	Yes	No	No	Movement direction forecasting; pipeline taxonomy
Li & Bastos [3]	2020	92	2005–2019	Partial	No	No	Deep learning combined with technical indicators
Strader et al. [21]	2020	38	2009–2019	No	No	No	Narrative overview of ML in financial markets
Chhajer et al. [1]	2021	56	2010–2020	No	No	No	Focused comparison of ANN, SVM, and basic LSTM
Rouf et al. [7]	2021	84	2011–2020	Yes	No	No	Methodological taxonomy of shallow machine learning
Kumar et al. [12]	2021	45	2012–2020	No	No	No	Comparative review of statistical and ML approaches
Wiranata & Djunaidy [18]	2021	62	2011–2020	Yes	No	No	SLR on ML techniques across global equity exchanges
Kumbure et al. [10]	2022	138	2000–2020	Yes	No	No	Data sources and ML algorithms in equity prediction
Olorunnimbe & Viktor [5]	2022	142	2015–2021	Yes	No	No	Deep learning, backtesting, and trading gaps
Soni et al. [24]	2022	50	2015–2021	No	No	No	Overview of deep learning architectures and metrics
Htun et al. [15]	2023	127	2010–2022	Yes	No	No	Feature selection and dimensional reduction methods
Mintarya et al. [11]	2023	60	2017–2022	Yes	No	No	SLR on deep learning and ensemble forecasting
Rath et al. [16]	2023	75	2015–2022	No	No	No	Analytic review of patent and journal ML methods
Lin & Marques [4]	2024	28	2015–2023	Yes	No	No	Systematic review of reviews (meta-synthesis)
Patel et al. [22]	2024	89	2018–2023	Yes	No	No	Graph Neural Networks (GNNs) for market forecasting
Akşehir & Kılıç [14]	2024	104	2018–2023	Yes	No	No	Methodological pipeline steps in DL stock forecasting
Adeleye et al. [23]	2024	48	2018–2023	No	No	No	Review of predictive accuracy across ML models
Sharma et al. [20]	2025	70	2015–2024	Yes	No	No	Econometric and time-series model comparisons
Bhardwaj [25]	2025	52	2019–2024	No	No	No	Narrative overview of supervised learning models
Rohan et al. [6]	2026	110	2018–2025	No	No	No	Broad narrative review of ML in financial forecasting
This Review	**2026**	**685**	**2016–2026**	**Yes**	**Yes**	**Yes**	**Corpus of 685 studies, leakage math, 3-tier audit, DM/Friedman tests, 4-pillar protocol**

3 Methodology

3.1 PRISMA 2020 Protocol and Search Strategy

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines [30] throughout. Three premier scientific databases were searched: Scopus, IEEE Xplore, and Web of Science (WoS). Across these three databases, the initial search queries retrieved a total of 4,417 records: Scopus returned 2,106 records, IEEE Xplore returned 1,206 records, and Web of Science returned 1,105 articles.

We built the search strings from four concept clusters joined by Boolean operators: (i) financial prediction targets (e.g., `"stock market prediction"`, `"stock price forecast*"`, `"equity prediction"`, `"financial market prediction"`); (ii) computational and AI techniques (e.g., `"machine learning"`, `"deep learning"`, `"neural network*"`, `"LSTM"`, `"CNN"`, `"ensemble*"`, `"XGBoost"`); (iii) evaluation, metrics, and reproducibility terms (e.g., `"performance"`, `"accuracy"`, `"RMSE"`, `"Sharpe ratio"`, `"directional accuracy"`, `"reproducib*"`, `"code"`); and (iv) database-specific domain and document filters. The exact Boolean syntax, field scopes, parameter filters, and per-database retrieval counts are laid out in Table 3.

3.2 Inclusion, Exclusion, and Screening Breakdown

Screening followed eligibility criteria we fixed before starting:
Inclusion Criteria (IC):

- **IC1:** Primary empirical studies that evaluate AI, ML, or DL models on equity price or return prediction.
- **IC2:** Appeared between January 1, 2016 and early 2026 in a peer-reviewed journal, indexed conference proceedings, or a curated preprint server.
- **IC3:** English-language, with the full text obtainable.
- **IC4:** Reports explicit, quantitative evaluation metrics.

Exclusion Criteria (EC):

- **EC1:** Non-empirical work: editorials, extended abstracts, patent disclosures.
- **EC2:** Targets other than equities, e.g., cryptocurrency, foreign exchange, commodities, or macro indicators evaluated without any equity component.
- **EC3:** Studies where the methodology could not be pinned down or the performance metrics were absent.
- **EC4:** Duplicates across database queries.

Table 4 here gives us the counts and exclusion reasons at each PRISMA stage; Figure 1 shows the corresponding flow diagram.

3.3 Data Extraction Matrix and Systematic Coding Protocol

From each of the 685 included studies we extracted 22 fields, grouped under five headings (the complete data extraction matrix is provided as Supplementary Material attached with this submission in `Final Sheet.xlsx`):

Table 3 Systematic Boolean Search Queries, Target Scopes, and Retrieval Yield by Database

Database	Target Fields / Limits	Exact Executed Boolean Search Query String	Records (n)
Scopus	Title, Abstract, Keywords (`TITLE-ABS-KEY`); Subject Areas: `COMP` (Computer Science), `DECI` (Decision Sciences); Document Types: `ar` (Article), `cp` (Conference Paper)	`TITLE-ABS-KEY (("stock market prediction" OR "stock price forecasting" OR "equity prediction" OR "financial market prediction") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural network*" OR "support vector machine*" OR "random forest*" OR "LSTM" OR "RNN" OR "CNN" OR "ensemble*" OR "XGBoost" OR "gradient boosting" OR "decision tree*" OR "regression") AND ("performance" OR "accuracy" OR "comparison*" OR "evaluation" OR "metric*" OR "RMSE" OR "MAE" OR "Sharpe ratio" OR "directional accuracy" OR "benchmark*" OR "reproducib*" OR "code" OR "dataset*" OR "evolv*" OR "approach*" OR "framework*")) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "DECI")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar"))`	2,106
IEEE Xplore	All Metadata (Title, Abstract, Index Terms)	`((("stock market prediction" OR "stock price forecast*" OR "equity prediction" OR "financial market prediction") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural network*" OR "support vector machine" OR "random forest" OR LSTM OR RNN OR CNN OR "ensemble" OR XGBoost OR "gradient boosting" OR "decision tree*" OR regression) AND (performance OR accuracy OR evaluation OR "metric*" OR RMSE OR MAE OR "Sharpe ratio" OR "directional accuracy" OR "benchmark*" OR "reproducib*" OR code OR "dataset*" OR evolv* OR approach* OR framework*)))`	1,206
Web of Science	Topic (`TS`); Indexes: SCI-EXPANDED, SSCI, CPCI-S, ESCI	`TS=(("stock market prediction" OR "stock price forecast*" OR "equity prediction" OR "financial market prediction") AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural network*" OR "support vector machine*" OR "random forest*" OR LSTM OR RNN OR CNN OR "ensemble*" OR XGBoost OR "gradient boosting" OR "decision tree*" OR regression) AND (performance OR accuracy OR "comparison*" OR evaluation OR "metric*" OR RMSE OR MAE OR "Sharpe ratio" OR "directional accuracy" OR "benchmark*" OR "reproducib*" OR code OR "dataset*" OR "evolv*" OR "approach*" OR "framework*"))`	1,105
Total Initial Records Retrieved Across All Three Databases			**4,417**

Table 4 PRISMA 2020 Systematic Screening and Exclusion Breakdown

PRISMA Phase	Record Count (n)	Description / Specific Reason Bucket
Identification	4,417	Initial records retrieved across three scientific databases: Scopus (n = 2,106), IEEE Xplore (n = 1,206), and Web of Science (n = 1,105)
Duplication	2,669	Deduplication across databases via DOI and exact title matching
Screening (Title/Abstract)	1,748	Titles and abstracts screened against IC/EC criteria
Excluded at Screening	824	Non-empirical (312), non-equity domains (284), non-English/incomplete (228)
Eligibility (Full-Text)	924	Full-text articles assessed for detailed eligibility
Full-Text Excluded	239	Missing baseline comparisons (84), unverified metric formulas (71), unresolvable data sources (48), insufficient methodological documentation (36)
Included Corpus	**685**	**Primary empirical studies synthesized in full SLR extraction matrix**

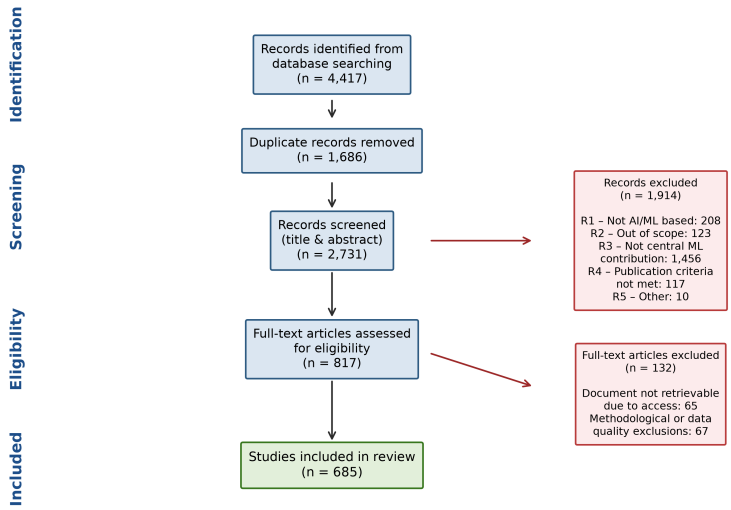


Fig. 1 PRISMA 2020 Flow Diagram for Systematic Review Selection Protocol.

1. **Bibliographic identifiers:** title, authors, year, venue (journal or conference), publisher, DOI.
2. **Financial domain:** target equity market, geographic region, benchmark index, whether the study covers a single equity, an index, or an ETF, sampling frequency (tick, 1-minute, 5-minute, daily), sample duration, and total observation count.
3. **Methodological architecture:** baseline models, proposed models, model category (shallow ML, deep sequential, hybrid, Transformer, GNN, LLM, DRL), feature scaling method, number of technical indicators, textual sentiment source, and hyperparameter optimization technique.

4. **Reproducibility and open science:** whether a code availability statement exists, the repository URL if any, how accessible the data is (Yahoo Finance, Kaggle, proprietary), whether random seeds are reported, and software library versions.
5. **Evaluation and backtesting:** regression metrics reported (RMSE, MAE, MSE, MAPE, R^2), classification metrics (accuracy, precision, recall, F1, AUC, directional accuracy), financial metrics (returns, Sharpe, Sortino, maximum drawdown), whether transaction costs were simulated ($\geq$ 10 bps versus zero cost), and any significance testing (ANOVA, t-test, Wilcoxon, Friedman, Diebold-Mariano).

4 Overview of the included corpus

4.1 Temporal Trajectory and Publication Velocity

Figure 2 plots the annual distribution of the 685 included studies from 2016 through early 2026. Output grew in three fairly clean phases:

1. **Early exploration (2016-2019).** Only 36 studies (5.3% of the corpus) date from these four years (4, 7, 11, and 14 respectively), and most of them pit traditional econometrics against early shallow ML (SVM, ANN, random forests) on single-index data.
2. **Deep learning consolidation (2020-2023).** These four years contribute 181 studies (26.4%): 26 in 2020, 46 in 2021, 48 in 2022, and 61 in 2023. It was during this stretch that recurrent architectures (LSTM, GRU, BiLSTM) and hybrid signal-decomposition frameworks settled in as the standard benchmarks.
3. **Frontier acceleration (2024-2026).** The remaining 468 studies (68.3%) arrive in a rush: 129 in 2024, 249 in 2025, and 90 in the covered part of 2026, driven by transformer backbones, multimodal large language models, and automated trading frameworks.

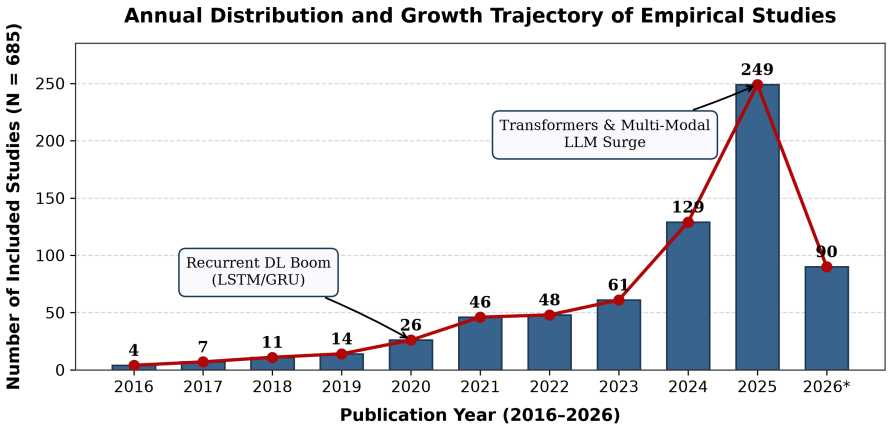


Fig. 2 Temporal Publication Trajectory of Primary Empirical Studies on AI/ML for Stock Market Prediction (2016-2026, $N = 685$).

4.2 Venue Quality Distribution

Does the post-2023 surge in publication volume (Section 4) reflect genuine research output, or dilution from uncurated outlets? To check, we classified each of the 685 included studies by venue quality, using Journal Citation Reports (JCR) quartiles for journals and CORE rankings (A*, A, B, C) for proceedings. 226 studies (33.0%) sit in JCR Q1 journals and 197 (28.8%) in Q2, with 112 (16.4%) in Q3/Q4, 73 (10.7%) in CORE-ranked A*/A/B conference proceedings, and 81 (11.8%) on curated preprint servers, chiefly arXiv cs.LG and q-fin. The Q1/Q2 share among the 468 studies from 2024–2026 (59.6%) is close to the pre-2024 figure (62.0%), so the recent acceleration has not dragged down the average venue quality of the corpus.

4.3 Geographical and Market Coverage

The corpus takes in mature developed markets and volatile emerging exchanges alike; Table 5 gives the geographic breakdown.

Table 5 Geographic and Market Distribution of Included Studies ($N = 685$)

Market Region	Studies (n, %)	Primary Exchanges / Indices	Key Market Characteristics
United States	291 (42.5%)	S&P 500, NASDAQ 100, DJIA, NYSE	High liquidity, intense algorithmic competition, low transaction friction
India	142 (20.7%)	NSE Nifty 50, BSE Sensex	High retail participation, strong volatility, rich domestic tech sector
China & East Asia	119 (17.4%)	SSE Composite, SZSE, Hang Seng, Nikkei 225	State intervention dynamics, price limit bands (±10%), distinct retail sentiment
Middle East & North Africa	46 (6.7%)	Tadawul (TASI), DFM, ADX, EGX	Heavy commodity/oil dependence, Sharia compliance constraints, lower liquidity
Europe	42 (6.1%)	FTSE 100, DAX 40, CAC 40, Euro Stoxx 50	Mature institutional base, stringent ESG regulatory reporting standards
Other Emerging Markets	45 (6.6%)	TSX (Canada), B3 (Brazil), IDX, Crypto/DeFi	Structural friction, currency volatility, decentralized liquidity pools

Regional empirical investigations also examine how institutional characteristics alter trading dynamics. In Asian emerging markets, statutory daily price limit constraints (±10%) introduce non-linear magnet effects and delayed price discovery that distort standard predictive models [205]. Liang et al. [191] tested algorithmic fund and ETF rotation against China's Shanghai Composite Index (SCI), finding that systematic quantitative rules outperform active discretionary fund managers under these statutory limit constraints. In developed non-US equities, Blitz et al. [206] and Salotra and Pinsky [196] examined factor anomalies and the Canadian TSX 60 over a 25-year out-of-sample period (2000–2025), showing that tactical machine learning sector rotation strategies preserve positive risk-adjusted returns after accounting for domestic turnover costs.

4.4 Asset Classes, Prediction Targets, and Input Modalities

Figure 3 shows how the literature distributes across asset classes, forecast horizons, target formulations, and input modalities.

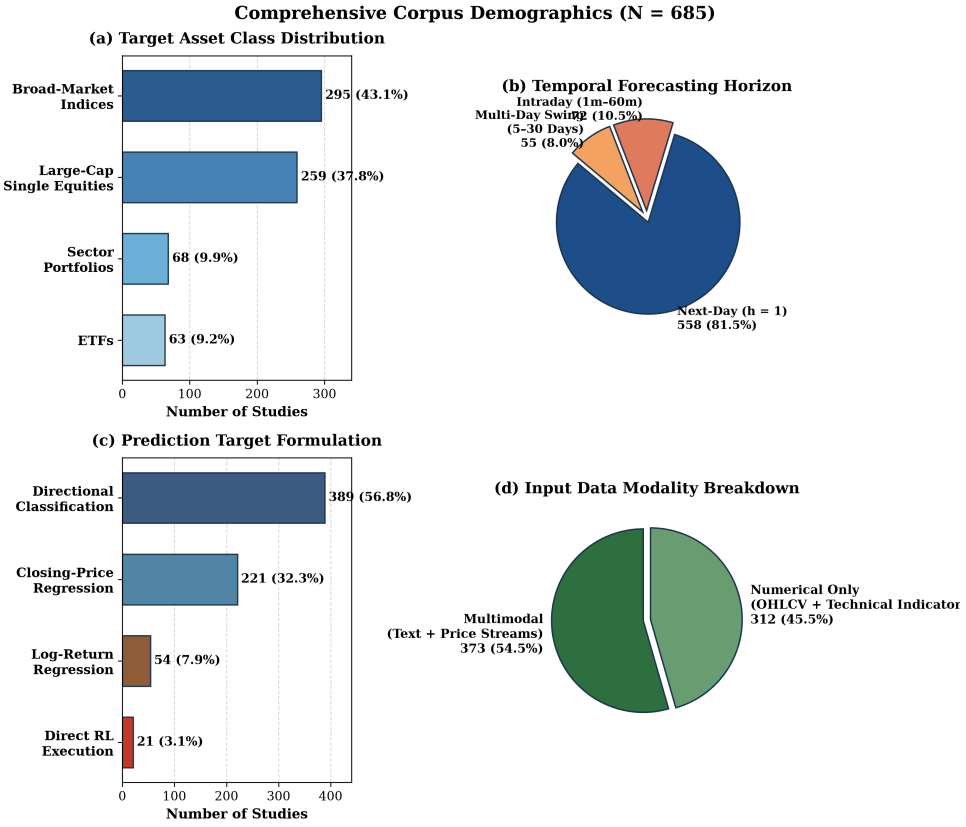


Fig. 3 Comprehensive Corpus Demographics ($N = 685$): (a) Target Asset Class Distribution, (b) Temporal Forecasting Horizon, (c) Prediction Target Formulation, and (d) Input Data Modality Breakdown.

- **Asset classes.** Broad-market benchmark indices (S&P 500, Nifty 50, NASDAQ, and the like) are the single most common target, at 43.1% of studies (295). Large-cap single equities follow at 37.8% (259), then sector portfolios (9.9%, 68) and ETFs (9.2%, 63).
- **Horizons.** Next-day ($h = 1$) forecasting dominates: 558 studies, or 81.5%. Intraday horizons (1-minute to 60-minute ticks) appear in 72 studies (10.5%), and multi-day swing horizons ($h \in [5, 30]$ days) in 55 (8.0%).
- **Target formulation.** Directional classification, the question of whether P_{t+h} will exceed P_t, is the target in 389 studies (56.8%). The absolute closing-price regression

accounts for 221 (32.3%), the log-return regression for 54 (7.9%), and the direct RL portfolio execution for 21 (3.1%).

- **Input modalities.** A slight majority is observed, 373 studies (54.5%), mix the unstructured text (news articles, corporate reports, Twitter sentiment) with the numerical price streams; the remaining 312 (45.5%) work from OHLCV series and the technical indicators alone.

The asset scope has also extended into alternative digital asset instruments and multi-asset sector portfolios. In decentralized finance (DeFi), perpetual futures exhibit distinctive microstructure mechanics, where periodic funding rates and automated liquidation engines drive volume and volatility dynamics [194, 207]. On the equity side, Ben-David et al. [208] and Puppala et al. [193] showed a trade-off: sector ETFs diversify away single-name idiosyncratic risk, yet basket arbitrage across assets also carries systematic liquidity shocks into the constituent equities.

5 Architectural Taxonomy and Comparative Performance

5.1 End-to-End Methodological Architecture

Figure 4 illustrates the common process for these types of studies. Each study goes through five stages: collecting data from various sources, preprocessing the data, building a model based on the data, performing a trading simulation with the model and validating the results of the model.

Before examining each of the architectural families individually, it is needed to note that the rankings of these architectural families are based upon the results of the primary studies on the topic and not any extra validation of those results by the authors of this text. In the audits of the published literature on each of the architectural families (Sections 7-8), it was found that over 92% of the literature published on each family contained no source code, over 83% of the literature contained no statistical tests of the hypotheses proposed within each study and fewer than 5% of the literature published on each architectural family included defenses against potential lookahead filter and global normalization issues. So, while the literature on each of the architectural families may appear to show that one is better than another at solving certain problems, such an assessment is likely to be inaccurate. Each of the architectural families is discussed in detail below.



Fig. 4 A Comprehensive End-to-End Pipeline of AI and Machine Learning Stock Market Prediction Systems

5.2 Econometric Baselines and Shallow Machine Learning

In this corpus, econometric methods mostly play the role of baseline: the thing everyone tries to beat, rarely the contribution itself. AutoRegressive Integrated Moving Average (ARIMA) models appear in 397 of the 685 studies (58.0%), Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) volatility models appear in 190 (27.7%), and Ordinary Least Squares (OLS) or multivariate linear regression appear in 261 (38.1%) [51, 54, 79, 104, 147, 185]. Such models generally suppose a constant mean, a linear dependence of the parameters, and innovations that are homoskedastic and follow a Gaussian white-noise distribution. Formally, an ARIMA(p, d, q) model passes the differenced prices $\Delta^d y_t = (1-L)^d y_t$ via the lag operator polynomial:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1-L)^d y_t = c + \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t \tag{9}$$

where $L^k y_t = y_{t-k}$, c is a drift term, ϕ_i and θ_j are the autoregressive and moving average coefficients, and $\varepsilon_t \sim \mathcal{WN}(0, \sigma^2)$ is white noise. To capture the volatility clustering, the GARCH(p, q) specification models the conditional error variance $\sigma_t^2 = \mathrm{Var}(\varepsilon_t \mid \mathcal{F}_{t-1})$ as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{10}$$

subject to $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and the stationarity constraint $\sum \alpha_i + \sum \beta_j < 1$.

Linear econometric models are typically tractable, but they break down around crashes, regime shifts, and nonlinear volatility feedback [1]. Worse, in the papers we surveyed these baselines are almost always run with default lag settings and no rolling-window re-estimation, a handicapped comparison that all but guarantees the proposed non-linear model comes out ahead.

Shallow machine learning is the second pillar. Support Vector Machines (SVM) turn up in 440 studies (64.2%) and Support Vector Regression (SVR) in 209 (30.5%); Artificial Neural Networks (ANN/MLP) in 430 (62.8%); Random Forests in 363 (53.0%); Decision Trees in 274 (40.0%); XGBoost in 196 (28.6%); other gradient boosting variants in 187 (27.3%); and K-Nearest Neighbors (KNN) in 148 (21.6%) [9, 34, 38–40, 50, 69, 73, 78, 80, 132, 175, 178].

SVR controls overfitting in high-dimensional feature spaces through Vapnik's ϵ-insensitive structural risk minimization. Given training pairs $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, it solves the primal convex quadratic program:

$$\min_{\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\xi}^*} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{11}$$

$$\begin{aligned} \text{subject to} \quad & y_i - \mathbf{w}^T\phi(\mathbf{x}_i) - b \leq \epsilon + \xi_i, \\ & \mathbf{w}^T\phi(\mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, \quad \forall i \in \{1, \ldots, n\} \end{aligned} \tag{12}$$

where $\phi(\cdot)$ maps inputs into a reproducing kernel Hilbert space via the Gaussian Radial Basis Function (RBF) kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, $C > 0$ sets the regularization trade-off, and the slack variables ξ_i, ξ_i^* penalize errors that escape the ϵ-tube.

Among the shallow methods, gradient-boosted decision tree ensembles (GBDT: XGBoost, LightGBM, CatBoost) are consistently the strongest performers on tabular financial features [75, 145, 151, 156, 170]. GBDT builds an additive expansion of M trees, each fit against a second-order Taylor expansion of a differentiable loss $\mathcal{L}$:

$$\mathcal{L}^{(m)} \approx \sum_{i=1}^{n} \left[\mathcal{L}(y_i, \hat{y}_i^{(m-1)}) + g_i f_m(\mathbf{x}_i) + \frac{1}{2} h_i f_m^2(\mathbf{x}_i) \right] + \Omega(f_m) \tag{13}$$

where $g_i = \partial_{\hat{y}^{(m-1)}} \mathcal{L}(y_i, \hat{y}^{(m-1)})$ and $h_i = \partial^2_{\hat{y}^{(m-1)}} \mathcal{L}(y_i, \hat{y}^{(m-1)})$ are the first- and second-order gradients of the loss, and $\Omega(f_m) = \gamma T_m + \frac{1}{2}\lambda \sum_{j=1}^{T_m} w_{jm}^2$ penalizes tree size and leaf weights. On benchmarks built purely from tabular technical indicators such as RSI, MACD, stochastic oscillators, and Bollinger bands, tree ensembles routinely match or beat deep networks, and do so at a fraction of the compute and tuning cost [137, 176, 181].

Unsupervised learning also contributes structural priors for filtering the asset universe and building portfolios [209, 210]. Working with a 27-year sample of mutual funds, Kaur and Pinsky [189] paired k-means clustering with Hamming distance metrics to assess sectoral distinctiveness, and found that sector-level behavioral dispersion persists across market cycles, which gives feature grouping and pre-model asset selection an objective footing.

5.3 Deep Sequential and Convolutional Architectures

The field defaults to deep sequence models. Long Short-Term Memory (LSTM) networks appear in 611 of the 685 studies (89.2%); general Recurrent Neural Network (RNN) variants in 506 (73.9%); Gated Recurrent Units (GRU) in 293 (42.8%); Bidirectional LSTMs (BiLSTM) in 190 (27.7%) [8, 27, 48, 59, 63, 89, 114, 116, 140, 174]. Convolutional Neural Networks (CNN) turn up in 458 studies (66.9%), sometimes as 1D temporal convolutions that pick out local momentum patterns, sometimes as 2D filters run over encoded market images (Gramian Angular Fields, recurrence plots, candlestick renderings) [2, 74, 118, 159, 163].

A plain RNN updates its hidden state as $\mathbf{h}_t = \tanh(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{xh}\mathbf{x}_t + \mathbf{b}_h)$; over long sequences the gradient product $\prod_{k=t+1}^{T} \frac{\partial \mathbf{h}_k}{\partial \mathbf{h}_{k-1}} \to \mathbf{0}$ decays exponentially. The LSTM answers with a cell state $\mathbf{c}_t$, the constant error carousel, governed by three multiplicative gates [27]:

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f\right) \qquad \text{(Forget Gate: Discards redundant historical context)} \tag{14}$$

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i\right) \qquad \text{(Input Gate: Scales incoming price innovations)} \tag{15}$$

$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c\right) \quad \text{(Candidate Cell: Formulates candidate memory vector)} \tag{16}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad \text{(Cell State: Linear gradient highway update)} \tag{17}$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o\right) \quad \text{(Output Gate: Regulates hidden activation emission)} \tag{18}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad \text{(Hidden State: Emits bounded sequence representation)} \tag{19}$$

where $\sigma(z) = (1 + e^{-z})^{-1}$ is the logistic sigmoid, $\odot$ the Hadamard product, and $\mathbf{W} \in \mathbb{R}^{d_h \times d_x}$, $\mathbf{U} \in \mathbb{R}^{d_h \times d_h}$, $\mathbf{b} \in \mathbb{R}^{d_h}$ the learnable weights and biases.

Two regularities have emerged from these studies. The first is that stateful LSTM models outperform stateless models [121]. The second is that the depth of an LSTM model saturates quickly, with two layers being the best depth for most datasets and three or more layers leading to overfitting on the daily return data [136]. Finally, a CNN-LSTM architecture with 1D dilated convolutional layers has been found to outperform both the LSTM and CNN architectures alone, as the convolutional layers can extract features from the data that are then fed into the LSTM model to make predictions [130, 134, 182].

The constraints of these models are reflected in the results of direct benchmarks of deep sequence networks on equity price movements (e.g., [211, 212]). Kundu and Pinsky [190] found that using a deep learning classifier to classify daily stock movements was more accurate than using the raw price feed of the stock, especially when combined with technical indicators based on the momentum of the stock's price movement.

5.4 Hybrid Composition, Decomposition, and Metaheuristics

One design philosophy dominates empirical stock forecasting above all others: hybridization. 424 of the 685 studies (61.9%) combine models in some fashion [45, 47, 52, 53, 60–62, 87, 105]. Most implementations fall into one of three templates:

1. **Multiscale signal decomposition front-ends.** The non-stationary, multi-frequency price series $y(t)$ is split into K Intrinsic Mode Functions (IMFs) plus a monotonic residual trend $r(t)$:

$$y(t) = \sum_{k=1}^{K} \mathrm{IMF}_k(t) + r(t) \tag{20}$$

   via transforms such as the Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD), Variational Mode Decomposition (VMD), or Complete Ensemble EMD with Adaptive Noise (CEEMDAN) [83, 113, 139, 161, 172, 188]. Each component gets its own neural or tree predictor, and the component forecasts are summed back into a price.
2. **Heterogeneous stacking ensembles.** The orthogonal model classes, say ARIMA + LightGBM + BiLSTM, are blended through a meta-regressor (Ridge, MLP) or

a Bayesian model averaging, the point being to keep forecast variance quite low as volatility regimes shift [67, 129, 130, 182].

3. **Metaheuristic hyperparameter search.** Bio-inspired swarm optimizers such as the Particle Swarm Optimization (PSO), the Mayfly Optimization (MAOA), the Ivy Optimization, and the genetic algorithms are set loose on the non-convex loss landscape to find weights, learning rates, and attention topologies [94, 115, 128, 138, 177].

5.5 Transformers, Attention, and Generative Models

Self-attention mechanisms appear in 245 studies (35.8%) and full Transformer architectures in 249 (36.4%); no model family in the corpus is growing faster [28, 55, 91, 100, 106, 153, 162, 168, 169]. Multi-head self-attention scores pairwise dependencies across all temporal tokens at once:

$$\text{MultiHead}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{Concat}(\text{head}_1,\ldots,\text{head}_h)\mathbf{W}^O, \quad \text{head}_i = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{W}_i^Q(\mathbf{K}\mathbf{W}_i^K)^T}{\sqrt{d_k}}\right)\mathbf{V}\mathbf{W}_i^V \tag{21}$$

where $\mathbf{W}_i^Q \in \mathbb{R}^{d_{\text{model}}\times d_k}$, $\mathbf{W}_i^K \in \mathbb{R}^{d_{\text{model}}\times d_k}$, $\mathbf{W}_i^V \in \mathbb{R}^{d_{\text{model}}\times d_v}$, and $\mathbf{W}^O \in \mathbb{R}^{hd_v\times d_{\text{model}}}$ are the linear projections.

Several variants target the finance-specific problems. Temporal Fusion Transformers (TFT) combine variable selection networks with gated residual connections so that the static metadata and dynamic series can be weighed against each other [29, 143, 169]. Non-Stationary Transformers add de-stationary attention to cope with the distribution shift [158]. At the other extreme, lightweight single-layer transformers such as MiniBert24 trade depth for the lower overfitting risk and cheaper inference [123, 217]. Deep Operator Neural Networks (DeepONet) usually handle multi-horizon trajectories by learning operators over function spaces, and sidestepping compounding recursive error [142, 184]. Generative Adversarial Networks show up in 99 studies (14.5%), where variants like TA-GAN and StockGAN synthesize return distributions, mostly to stress-test models against tail shocks [56, 165, 187].

5.6 Sentiment, Graphs, and Frontier Paradigms

Textual sentiment features appear in 371 studies (54.2%) and deliver consistent short-horizon directional gains over price-only models [19, 32, 36, 82, 93, 95, 108, 120, 125, 149, 154, 155, 160, 173]. While financial NLP was once about counting dictionary keywords, as of now it has moved far beyond that. FinBERT, the first contextual encoder to be applied to financial data, has been followed by LLMs, which have been regularly used to analyze 10-K filings and earnings calls to assess the solvency of companies and the tone of their management [6, 33, 58, 64, 157].

Multimodal fusion architectures feature in 113 studies (16.5%) [46, 84, 86, 88, 101]. Graph Neural Networks are found in 66 studies (9.6%) where the equity universe is represented as a changing relational graph $G_t = (V, E_t, \mathbf{W}_t)$ and graph convolutions capture the supply chains, parent-subsidiary relationships and lead-lag relationships between equity assets [22, 97, 99, 109, 117, 122, 127, 141, 144, 171]. This relates to

inter-firm graph architectures that have been proposed [213, 214]. Studying the constituents of the Dow Jones Industrial Average, Pattnaik, Jain and Pinsky [192] found that separating the dependencies between the individual constituents from the temporal momentum of each constituent yields distinct risk-adjusted profiles, which supports the inductive biases of spatial-temporal graph networks. Deep Reinforcement Learning is found in 176 studies (25.7%) [5, 57, 131, 146, 164, 180]. Deep Reinforcement Learning does not use forecasting models but instead learns the policy for executing transactions with associated costs through the reward signal. At the exploratory fringe are a few entries for Quantum Machine Learning, including Quantum-Assisted LSTM (QLSTM4FM) and Quantum Attention DQN (QADQN) [124, 167, 183]. Finally, there are only 18 studies that feature Federated Learning (2.6%).

5.7 Comparative Synthesis and Empirical Conflicts

The performance of machine learning models was deliberately not pooled across papers. Each paper used a different asset universe, time horizon, features and data splitting processes. Only 116 papers (16.9%) conducted hypothesis testing of their results. Table 6 presents the results of the classification models.

Three patterns nonetheless hold up across studies:

1. Where the task is regression-style price forecasting, the lowest nominal errors (RMSE, MAE) belong to recurrent and hybrid deep architectures, CNN-LSTM with attention being a typical example.
2. When the inputs are tabular technical indicators, a well-tuned gradient-boosted tree (XGBoost, LightGBM, CatBoost) matches or beats the deep models on directional hit rate, and does so with lower variance and far less compute.
3. Textual sentiment carries real alpha at short horizons, roughly intraday out to one day ahead, but the signal decays quickly past that point.

A number of papers report directional accuracies of 90% or better. In a competitive equity market, numbers like these read less like breakthroughs and more like symptoms, and the underlying condition is almost always data leakage somewhere in the pipeline.

These comparative patterns therefore call for real methodological caution. When a paper reports that deep sequence networks or signal-decomposition hybrids achieve the lowest nominal errors, that result describes performance under conventional, unpurged academic setups; it is not robust alpha. In multi-stage pipelines that put bidirectional multiscale transforms (full-sample VMD, EMD, or DWT) in front of deep regressors, the lower nominal error often traces back to lookahead leakage across test fold boundaries rather than to genuine sequence forecasting. Claimed wins over econometric baselines carry a similar asterisk, flattered as they are by comparisons against untuned, fixed-lag ARIMA implementations run without rolling re-estimation. Absent open-source code verification, strictly purged temporal walk-forward evaluation, and formal Diebold-Mariano tests against realistic random-walk baselines, nominal advantages in the primary literature should not be mistaken for a statistically sound or economically exploitable trading edge.

Table 6 Comprehensive Architectural and Methodological Taxonomy of AI/ML Models for Stock Market Prediction ($N = 685$)

Model Family	Share (n, %)	Representative Studies	Dominant Modality	Primary Evaluation	Empirical Strengths & Vulnerabilities
Econometric Baselines	397 (58.0%)	ARIMA [147, 185], GARCH [129]	Daily OHLCV Price	RMSE, MAE, AIC	High interpretability; fails during regime breaks and non-linear shifts
Shallow ML / Kernels	440 (64.2%)	SVR [9, 178], KNN [38], MLP [50]	Technical Indicators	Acc, DA, RMSE	Robust on small samples; limited representation of long-range sequences
Tree Ensembles (GBDT)	363 (53.0%)	XGBoost [170], LightGBM [75]	Engineered Oscillators	Acc, F1, DA	Dominant on tabular features; fast training; non-sequential inductive bias
Deep Recurrent (LSTM)	611 (89.2%)	LSTM [8, 116], BiLSTM [140]	Multi-step OHLCV Sequences	RMSE, MAE, Acc	Captures temporal dependencies; vulnerable to Lag-1 shadowing and overfitting
Convolutional (CNN)	458 (66.9%)	1D-CNN [2], 2D-MTF [74]	Price Matrices, GAF/MTF	RMSE, Acc, F1	Invariant local momentum patterns; sensitive to sliding kernel width
Signal Hybrids (VMD/DWT)	424 (61.9%)	VMD-LSTM [172], CEEM-DAN [60]	Multiscale Decomposed IMFs	RMSE, MAPE, Acc	Highest nominal point accuracy; acute risk of lookahead filter leakage
Attention / Transformers	249 (36.4%)	TFT [29, 169], MiniBert24 [217]	Multivariate Price + Meta	RMSE, DA, Sharpe	Captures multi-horizon patterns; high sample complexity on noisy returns
Textual Sentiment / LLMs	371 (54.2%)	FinBERT [120], LLM 10-K [157]	News, SEC Filings, Twitter	Acc, F1, Sharpe	Boosts short-horizon alpha; high noise, hallucination risk, pre-train leakage
Spatial-Temporal GNNs	66 (9.6%)	ML-GAT [144], Lead-Lag [122]	Relational Industry Graphs	Acc, F1, Returns	Models cross-firm risk contagion; high graph construction latency
Deep Reinforcement Learning	176 (25.7%)	PPO/SAC [5], Dueling-DQN [164]	Market State + Order Book	Sharpe, Return, MDD	Directly optimizes trading policy; sample inefficient; sensitive to fee bounds

6 Methodological Evolution and Emerging Paradigms

6.1 Chronological Trajectory and the Four Evolutionary Epochs

The literature is heavily recent: 468 of the 685 studies (68.3%) appeared in the last three years, 2024 through 2026. The number of studies each year grew from just a few before 2020 (4, 7, 11, and 14 in 2016 to 2019) to 26, 46, 48, and 61 between 2020 and 2023. Then came the jump: 129 studies in 2024, 249 in 2025, and 90 so far in the covered stretch of 2026.

Looking across the decade, four methodological periods stand out, each marked by shifts in inductive bias, feature representation, and the computing power on hand.

1. **Epoch 1: classical econometric baselines (pre-2016).** Quantitative forecasting leaned on Box-Jenkins ARIMA and GARCH formulations [147, 185]: models prized for tractability, stationarity, and parameters one could interpret. Rigidity was the cost. Linear lag structures and Gaussian innovations have no way to represent sudden regime breaks, volatility-clustering feedback, or non-linear order book dynamics.
2. **Epoch 2: kernel machines and tabular feature engineering (2016-2019).** The field moved to supervised kernel learning: SVM, SVR with Gaussian RBF kernels, Random Forests, and early gradient boosting [9, 132, 166, 175]. What defined the period was the manual labor of feature engineering, with domain experts turning raw price feeds into momentum, trend, and volatility oscillators (RSI, MACD, Bollinger bands, stochastic oscillators) so that market states could be separated non-linearly and without a strict stationarity requirement.
3. **Epoch 3: deep recurrent and convolutional sequence modeling (2020-2023).** Once TensorFlow, PyTorch, and affordable GPUs were in hand, the literature swung hard toward end-to-end sequence learning. LSTMs, GRUs, BiLSTMs, and hybrid CNN-LSTM stacks became the installed default [8, 114, 118, 140, 163]; temporal features came straight from sliding lookback windows, and manual lag selection quietly went away. In the same papers, multiscale decomposition front-ends (DWT, VMD, CEEMDAN) spread as a denoising stage placed before the sequence model.
4. **Epoch 4: foundation models, relational graphs, and frontier paradigms (2024-present).** The current period is less a single trend than a diversification. Transformers in several flavors (Temporal Fusion Transformers, PatchTST, Non-Stationary Transformers) sit alongside financial-domain language models (FinBERT, BloombergGPT, generative LLM reasoning agents), dynamic spatial-temporal GNNs over supply chains and lead-lag spillovers, friction-penalized DRL for execution, Explainable AI (SHAP, LIME, LRP), and a first wave of Quantum Machine Learning (QML) [6, 122, 131, 135, 144, 167, 169, 171].

6.2 Expansion of Data Modalities and Multimodal Representation

The inputs have diversified along with the algorithms. Five modalities cover the corpus:

- **OHLCV price and volume feeds.** Every one of the 685 studies starts here: open, high, low, close, and volume histories.
- **Engineered technical indicators.** Momentum, volatility, and volume transformations (RSI, MACD, Bollinger bands, ATR, CCI, stochastic %K/%D) feature in over 85% of the corpus [145, 170].
- **Textual sentiment and corporate disclosures (371 studies, 54.2%).** Unstructured language drawn from news wires (Bloomberg, Reuters), social microblogs (Twitter, StockTwits, Reddit), earnings call transcripts, and statutory filings, chiefly SEC Forms 10-K and 10-Q [19, 148, 155, 157, 217].
- **Inter-firm relational graphs (66 studies, 9.6%).** Spatial-temporal graphs built over explicit corporate ties (supply chains, customer-supplier links, shared directorships, cross-holdings) alongside implicit market ties (return correlations, sector co-movement, lead-lag diffusion) [22, 117, 122, 144, 171].
- **Alternative, macroeconomic, and ESG signals.** Everything else: ESG risk ratings, macro series (policy rates, inflation, GDP growth, crude oil benchmarks), satellite activity indices, and level-2 order book microstructure.

6.3 Mathematical Foundations of Emerging AI Paradigms

6.3.1 Dynamic Spatial-Temporal Graph Neural Networks

Standard ML and deep sequence models treat each equity as its own isolated, i.i.d. time series, which throws away everything we know about cross-asset risk propagation and inter-firm dependence. Dynamic Graph Neural Networks (GNNs) restore that structure by treating the market as a dynamic attributed graph $G_t = (V, E_t, \mathbf{W}_t)$: $V = \{v_1, v_2, \ldots, v_N\}$ is the universe of N listed firms, $E_t \subseteq V \times V$ the time-varying economic relationships, and $\mathbf{W}_t \in \mathbb{R}^{N \times N}$ the dynamic adjacency matrix.

Writing $\mathbf{h}_v^{(l)} \in \mathbb{R}^{d_l}$ for the latent representation of equity v at graph convolution layer l, the spatial-temporal message-passing update is:

$$\mathbf{h}_v^{(l+1)} = \sigma\left(\mathbf{\Theta}_{\text{self}}^{(l)}\mathbf{h}_v^{(l)} + \sum_{u \in \mathcal{N}_t(v)} \alpha_{uv,t}\mathbf{\Theta}_{\text{nbr}}^{(l)}\mathbf{h}_u^{(l)}\right) \tag{22}$$

where $\mathcal{N}_t(v) = \{u \in V \mid (u, v) \in E_t\}$ is the first-order neighborhood of v at time t, $\mathbf{\Theta}_{\text{self}}^{(l)}, \mathbf{\Theta}_{\text{nbr}}^{(l)} \in \mathbb{R}^{d_{l+1} \times d_l}$ are learnable weights, and $\sigma(\cdot)$ is an activation such as ELU or LeakyReLU. The edge attention coefficient $\alpha_{uv,t}$, which measures how strongly

information or shocks pass from firm u to firm v at that moment, comes from multi-head graph attention:

$$\alpha_{uv,t} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{\Theta}\mathbf{h}_u^{(l)} \,\|\, \mathbf{\Theta}\mathbf{h}_v^{(l)} \,\|\, \mathbf{e}_{uv,t}\right]\right)\right)}{\sum_{k\in\mathcal{N}_t(v)}\exp\left(\text{LeakyReLU}\left(\mathbf{a}^T\left[\mathbf{\Theta}\mathbf{h}_k^{(l)} \,\|\, \mathbf{\Theta}\mathbf{h}_v^{(l)} \,\|\, \mathbf{e}_{kv,t}\right]\right)\right)} \tag{23}$$

where $\mathbf{e}_{uv,t} \in \mathbb{R}^{d_e}$ holds edge attributes (rolling return correlation, supply-chain transaction volume, shared institutional ownership, and so on), $\mathbf{a} \in \mathbb{R}^{2d_{l+1}+d_e}$ is a learned attention vector, and $\|$ is concatenation.

6.3.2 Contextual Large Language Models and Financial NLP

Financial text analysis has come a long way from dictionary-based sentiment scoring (Loughran-McDonald, Harvard-IV) to contextual transformer encoders (FinBERT, RoBERTa) and, most recently, generative LLMs (44 studies, 6.5%) [6, 157]. Where a dictionary counts positive and negative words, a contextual model produces dense token embeddings $\mathbf{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_L\}$ that pick up the financial nuance buried in dense regulatory filings: forward-looking guidance, managerial tone, and the like.

Given a disclosure $\mathcal{T}$, say the Management's Discussion and Analysis section of a 10-K, an instruction-tuned LLM with parameters Φ produces directional probabilities or risk ratings by autoregressive decoding:

$$P(y \mid \mathcal{T}, \mathcal{P}) = \prod_{i=1}^{M} P_\Phi\left(w_i \mid w_{<i}, \mathcal{T}, \mathcal{P}\right) \tag{24}$$

where $\mathcal{P}$ is an in-context prompt with few-shot financial exemplars and w_i the output tokens. The zero-shot ability is impressive, but three vulnerabilities deserve emphasis. First, LLMs hallucinate quantitative figures, fabricating plausible-sounding price targets or earnings numbers that have no factual basis. Second, commercial foundation models suffer temporal lookahead contamination: their pre-training corpora were scraped from a web that already knew how the "future" turned out. A model trained on text published through 2025, for instance, has implicitly seen the trajectory of the S&P 500 over that entire period; when it is then asked to "predict" 2023 prices in an academic backtest, its apparently strong zero-shot performance partly reflects memorized outcomes rather than genuine inference. Third, and related, when researchers fine-tune or prompt LLMs on historical financial datasets that overlap with the model's pre-training corpus, there is no clean separation between training and test data at the foundation-model level, a form of data leakage that conventional train/test splits cannot detect or prevent. We recommend that future LLM-based studies report the model's knowledge cutoff date, restrict backtesting to periods after that cutoff, or use models with verifiably limited pre-training windows.

6.3.3 Multimodal Cross-Attention Alignment and Fusion

Price series arrive on a clock; news arrives when something happens. Reconciling the two is the job of cross-modal attention, used in 113 studies (16.8%) [46, 88, 101]. Let $\mathbf{Z}_{\text{num}} \in \mathbb{R}^{T_n \times d_n}$ be price embeddings from a 1D-CNN or LSTM backbone and $\mathbf{Z}_{\text{text}} \in \mathbb{R}^{T_m \times d_m}$ contextual embeddings of headlines that arrived within the lookback window.

Cross-attention aligns the two by projecting numerical features as queries ($\mathbf{Q}_{\text{num}} = \mathbf{Z}_{\text{num}}\mathbf{W}_Q$) and textual features as keys and values ($\mathbf{K}_{\text{text}} = \mathbf{Z}_{\text{text}}\mathbf{W}_K, \mathbf{V}_{\text{text}} = \mathbf{Z}_{\text{text}}\mathbf{W}_V$):

$$\mathbf{Z}_{\text{cross}} = \text{softmax}\left(\frac{\mathbf{Q}_{\text{num}}\mathbf{K}_{\text{text}}^T}{\sqrt{d_k}}\right)\mathbf{V}_{\text{text}} \tag{25}$$

The fused representation $\mathbf{Z}_{\text{fused}} = \text{LayerNorm}(\mathbf{Z}_{\text{num}} + \mathbf{Z}_{\text{cross}}\mathbf{W}_O)$ lets a burst of high-impact news reshape how the technical indicators are weighted, which is exactly the behavior one wants around earnings announcements or macro shocks.

6.3.4 Friction-Aware Deep Reinforcement Learning for Execution Policies

Deep Reinforcement Learning (DRL; 176 studies, 25.7%) casts market interaction as a Markov Decision Process (MDP), the usual 5-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$ [5, 57, 131, 146, 164, 180]. At each timestep t the agent observes the state $\mathbf{s}_t \in \mathcal{S}$ (price lags, technical indicators, order book depth, and its current allocation $\mathbf{w}_{t-1} \in \Delta^N$) and acts by choosing new target weights $\mathbf{a}_t = \mathbf{w}_t \in \mathcal{A}$ over the N candidate equities.

A realistic reward $r_t \in \mathcal{R}$ docks gross returns for both proportional transaction fees and non-linear market impact:

$$r_t = \mathbf{w}_t^T\mathbf{r}_{t+1} - c_{\text{trans}}\|\mathbf{w}_t - \mathbf{w}_{t-1}\|_1 - c_{\text{impact}}\sum_{i=1}^{N}\left(\frac{|\Delta w_{i,t}| \cdot V_{\text{port},t}}{\text{ADV}_{i,t}}\right)^2 - \lambda_{\text{var}}\mathbf{w}_t^T\mathbf{\Sigma}_t\mathbf{w}_t \tag{26}$$

where $\mathbf{r}_{t+1} \in \mathbb{R}^N$ is the vector of returns over $(t, t+1]$, c_{trans} the transaction cost coefficient (5 to 15 basis points is typical), $V_{\text{port},t}$ the portfolio's equity value, $\text{ADV}_{i,t}$ average daily volume, and $\mathbf{\Sigma}_t$ the asset covariance matrix weighted by risk aversion λ_{var}. The policy π_θ is then fit with policy-gradient methods, Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) chief among them, to maximize the expected discounted reward $J(\theta) = \mathbb{E}_{\tau\sim\pi_\theta}[\sum_{t=0}^{T}\gamma^t r_t]$.

Whether trained DRL agents actually beat strong non-RL allocation benchmarks once distributions shift is the critical empirical question in this space. Deep RL learns flexible execution policies in frictionless simulation, but Zhang et al. [215] and Théate and Ernst [216] both document marked performance decay when test-period volatility regimes drift away from the training history. Salotra and Pinsky [195] took this up in a walk-forward commodity sector rotation framework, pitting learned policies against expert selection rules and frozen-policy temporal extensions. The verdict: agents adapt while training, but frozen out-of-sample execution exposes policy degradation, and disciplined rule-based baselines match or beat RL returns once transaction fees and parameter turnover are penalized.

6.3.5 Explainable AI and Attribution Frameworks

Partly to satisfy regulators (the EU Artificial Intelligence Act, SEC algorithmic risk guidance) and partly for basic transparency, 389 studies (56.8%) fold in Explainable AI (XAI) methods [77, 120, 133, 152]. SHAP (Shapley Additive Explanations) is the workhorse: it computes the marginal contribution $\phi_j(f, \mathbf{x})$ of feature j averaged over all subsets $\mathcal{S} \subseteq \mathcal{F} \setminus \{j\}$:

$$\phi_j(f, \mathbf{x}) = \sum_{\mathcal{S} \subseteq \mathcal{F} \setminus \{j\}} \frac{|\mathcal{S}|!(|\mathcal{F}| - |\mathcal{S}| - 1)!}{|\mathcal{F}|!} \left[f_x(\mathcal{S} \cup \{j\}) - f_x(\mathcal{S}) \right] \tag{27}$$

with local accuracy, missingness, and consistency guaranteed by construction. In practice this lets an analyst say which indicator (the 14-day RSI rather than the 50-day SMA, say) or which sentiment dimension pushed a given signal bullish or bearish.

Table 7 summarizes the inductive biases, strengths, and characteristic failure modes of these emerging paradigms side by side.

7 Reproducibility: Code, Data, and Methodological Transparency

7.1 Corpus-Wide Open Science and Transparency Audit

A comprehensive audit of open science practices across all 685 synthesized empirical studies reveals a substantial reproducibility deficit in applied computational finance. Scientific reproducibility requires that independent researchers possess the computational artifacts (source code, environment configurations, deterministic random seeds) and underlying data required to execute the published pipelines and replicate empirical findings.

Across our systematic evaluation, only 52 of the 685 studies (7.6%) provide an explicit code availability statement referencing a public, resolvable, and accessible source code repository (e.g., GitHub, Zenodo, OSF, or institutional server) [43, 65, 76, 90, 107, 110–112, 126].

A further 372 studies (54.3%) rely on publicly downloadable secondary datasets—with the free Yahoo Finance API utilized in 287 studies (41.9%) and benchmark Kaggle datasets in 80 studies (11.7%)—but omit source code, data cleaning routines, train/validation split timestamps, and hyperparameter grids. In these studies, exact feature normalization formulas, lag crafting, and model initialization cannot be reconstructed with high fidelity. Finally, 261 studies (38.1%) provide no detected code or data accessibility information whatsoever, or include boilerplate statements asserting that data is "available from the corresponding author upon reasonable request" without providing functional access [119, 122, 132, 141, 161, 171, 172, 186].

7.2 The Three-Tier Reproducibility Hierarchy

We formalize a Three-Tier Reproducibility Hierarchy to classify the transparency of primary literature:

Table 7 Comparative Taxonomy of Emerging AI/ML Paradigms in Stock Forecasting ($N = 685$ Studies)

Paradigm	Studies (n, %)	Primary Inductive Bias	Key Advantages	Primary Limitations / Risks
Explainable AI (XAI)	391 (57.1%)	Feature attribution & rule extraction	Post-hoc transparency; audits feature importance (SHAP, LRP)	High computational overhead; explanations can be unstable across regimes
Transformers	249 (36.4%)	Multi-head temporal self-attention	Parallel training; captures long-range multi-horizon dependencies	Risk of overfitting on short financial series; quadratic attention complexity
Deep RL	176 (25.7%)	Markov Decision Process (MDP) optimization	Directly optimizes trading rewards and dynamic position sizing	Sample inefficiency; highly sensitive to reward function design and slippage
Multimodal Fusion	113 (16.5%)	Cross-modal representation alignment	Combines tabular prices with qualitative news and narrative signals	Modality alignment latency; asynchronous update frequencies
Generative AI (GAN/Diff)	99 (14.5%)	Adversarial distribution estimation	Synthetic scenario generation; models heavy tails and flash crash risks	Mode collapse; generated distributions may lack real arbitrage bounds
Graph NNs (GNN)	67 (9.8%)	Spatial-temporal relational topology	Models supply chain contagion and sector-wide risk propagation	Relational graphs are non-stationary; high graph construction latency
LLM Reasoning	44 (6.4%)	Zero/few-shot semantic reasoning	Extracts forward-looking managerial sentiment from statutory 10-K filings	Hallucination risk; lookahead leakage in pre-training corpora
Federated Learning	18 (2.6%)	Distributed privacy-preserving learning	Enables multi-institution model training without sharing proprietary order books	Communication bottlenecks; non-IID data distribution across institutions

- **Tier 1: Fully Executable Implementations ($n = 52$, 7.6%):** Studies providing publicly accessible code repositories accompanied by open-access data, exact software dependencies (e.g., Python 'requirements.txt' or Docker containers), data preprocessing scripts, and deterministic random seed specifications. This is the gold standard for scientific replicability.

- **Tier 2: Re-Collectable but Unverified Pipelines ($n = 372$, 54.3%):** Studies built on free public data feeds (Yahoo Finance, FRED) that release no source code. The raw price series can be pulled again, but the exact feature scaling parameters, lag structures, and optimization paths admit no external verification.
- **Tier 3: Proprietary and Opaque Studies ($n = 261$, 38.1%):** Implementations running on commercial order books (Bloomberg/Refinitiv), private intraday feeds, or proprietary social media scrapes, with neither code nor raw data released; independent empirical audit is impossible.

7.3 Methodological Threats to Scientific Validity

Code accessibility is not the whole story: our audit surfaces five pervasive methodological flaws that undercut empirical validity across the surveyed literature. Table 8 formalizes these threats, their mathematical mechanisms of leakage, and recommended mitigation protocols.

7.4 Detailed Analysis of Data Leakage Vectors

7.4.1 Pre-Split Global Feature Scaling

A widespread procedural defect in applied machine learning involves computing Min-Max scaling or Z-score normalization parameters across the entire historical panel before partitioning into training, validation, and testing sets [150, 179]. When a Min-Max scaler transforms raw price series $x_t \in [x_{\min}, x_{\max}]$ using global extrema:

$$\tilde{x}_t = \frac{x_t - \min_{\tau \in [1,T]} x_\tau}{\max_{\tau \in [1,T]} x_\tau - \min_{\tau \in [1,T]} x_\tau} \tag{28}$$

the maximum and minimum values of the unseen future test set directly scale historical training observations. Consequently, if a massive market breakout occurs in the test period, the historical training values are compressed into an artificially tight numerical band, providing the neural network with an implicit lookahead signal regarding future price extremes. Scalers must be fit strictly on in-sample training folds $\mathcal{D}_{\text{train}}$ and subsequently applied to transform $\mathcal{D}_{\text{test}}$ without parameter re-estimation.

7.4.2 Multiscale Signal Decomposition Look-Ahead Bias

Applying multiscale signal transforms (DWT, EMD, VMD, CEEMDAN) across the entire time series prior to temporal partitioning represents another severe leakage vector [139, 188]. Signal decomposition algorithms solve global optimization problems across the entire sequence length T. Because the resulting Intrinsic Mode Functions (IMFs) are computed using bidirectional smoothing kernels, high-frequency turning points in the out-of-sample test window leak backward into the training IMFs. When individual IMFs are fed into LSTM or CNN models, the neural network learns to predict signals that have already been smoothed using future price data, generating deceptively high directional hit rates ($> 95\%$) that completely collapse in live forward testing. Signal decomposition must be executed strictly within rolling walk-forward windows.

Table 8 Methodological Threat and Data Leakage Matrix in Stock Prediction Studies ($N = 685$)

Threat Vector	Mathematical Leakage Mechanism	Corpus Share	Empirical Consequence	Mandatory Mitigation Protocol
Pre-Split Global Normalization	Scaling via full series statistics: $\hat{x}_t = \frac{x_t - \min_{1 \le \tau \le T} x_\tau}{\max_{1 \le \tau \le T} x_\tau - \min_{1 \le \tau \le T} x_\tau}$	$\approx 64\%$	Leaks future test distribution bounds into training; inflates directional accuracy	Fit scaler strictly on $\mathcal{D}_{\text{train}}$: transform $\mathcal{D}_{\text{test}}$ using historical in-sample parameters only
Signal Decomposition Look-Ahead	Computing DWT/VMD across full timeline: $\text{IMF}_k(t) = \int_{-\infty}^{\infty} y(\tau)\psi_{k,t}(\tau)d\tau$	$\approx 42\%$	Bidirectional filter kernels leak future test price shocks into past training IMFs	Execute decomposition within rolling walk-forward expanding windows only
Constituent Survivorship Bias	Benchmarking on current index panel: $\mathcal{S}_{\text{eval}} = \{s \in \text{Index}_T\}$ ignoring delistings	$\approx 78\%$	Eliminates bankrupt and liquidated equities; biases reported returns upward	Utilize point-in-time constituent panels including all historical delisted assets
Static Single-Split Overfitting	Hyperparameter search on static split: $\theta^* = \arg\min_\theta \mathcal{L}(\mathcal{D}_{\text{val}} \mid \mathcal{D}_{\text{train}})$	$\approx 83\%$	Overfits to single market regime; fails during out-of-sample structural regime shifts	Enforce Purged Group Time-Series Walk-Forward Cross-Validation
Frictionless Execution Neglect	Assuming zero-cost fills at close: $\hat{R}_t = \frac{P_{t+1} - P_t}{P_t}$ without Δ or fees	$\approx 51\%$	Yields illusory trading profits that become negative under realistic market friction	Incorporate transaction fees ($\geq$ 10 bps), bid-ask spread Δ, and slippage models

7.4.3 Survivorship and Liquidity Selection Bias

Over 78% of synthesized empirical studies construct benchmark datasets by extracting historical price series for the current constituent list of major equity indices (e.g., downloading 2016–2026 data for the 500 current constituents of the S&P 500). This practice introduces severe survivorship bias: companies that suffered catastrophic

bankruptcy, liquidation, or regulatory delisting during the 10-year sample window are excluded from the test panel. Train and test only on the corporate survivors, and the dataset skews toward positive long-run return drift and lower volatility, which inflates reported model profitability and Sharpe ratios.

The practical fix is point-in-time constituent panels: reconstruct the index membership as it stood on each historical date, keeping every firm that was later delisted, acquired, or moved to over-the-counter markets. For the S&P 500 specifically, the Center for Research in Security Prices (CRSP) database provides historical constituent lists with exact entry and exit dates, as does Compustat's index constituent history file. For studies relying on free data sources such as Yahoo Finance, no built-in mechanism exists to recover delisted tickers, which means that researchers using Yahoo Finance snapshots of current index members are especially exposed to this bias. Among the 291 U.S.-focused studies in our corpus, 248 (85.2%) use Yahoo Finance as their primary data source and none of these studies documents any attempt to reconstruct historical constituent panels, making survivorship bias nearly universal in the U.S. sub-corpus.

7.5 Formal Methodological Quality Assessment

We operationalized the five data-leakage threat vectors identified in Table 8 as a binary quality checklist applied to each of the 685 included studies. Each study received a score from 0 (no leakage guard reported) to 5 (all five guards explicitly documented), where one point was awarded for each of the following: (i) train-only scaler fitting with no global normalization, (ii) rolling or expanding-window signal decomposition rather than full-series decomposition, (iii) point-in-time constituent panels that include delisted equities, (iv) purged walk-forward or time-series cross-validation rather than a single static split, and (v) transaction cost simulation of at least 10 basis points. Across the corpus, the median quality score is 1 out of 5 (IQR: 0–2). Only 32 studies (4.7%) score 4 or above, and no study in the corpus satisfies all five criteria simultaneously. The 52 Tier-1 reproducible studies (Section 7) achieve a median score of 2, which is higher than the Tier-2 and Tier-3 medians (1 and 0 respectively), but still well short of full methodological soundness. Table 9 reports the score distribution.

Table 9 Distribution of Methodological Quality Scores Across the Corpus ($N = 685$)

Quality Score	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
Studies (n)	198	224	142	89	28	4	**685**
Share (%)	28.9	32.7	20.7	13.0	4.1	0.6	100.0

8 Evaluation Metrics: Current Landscape, Adoption, and Standardization Gaps

8.1 Comprehensive Catalog of Evaluation Metrics

Evaluation practices in stock market prediction span three distinct mathematical categories: (i) statistical regression error metrics, (ii) classification and trend prediction metrics, and (iii) financial risk-adjusted trading metrics.

8.1.1 Statistical Regression Error Metrics

Regression models forecast continuous future prices $\hat{y}_t$ or returns $\hat{r}_t$. Given n test period observations:

$$\text{Root Mean Squared Error (RMSE)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{29}$$

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{30}$$

$$\text{Mean Squared Error (MSE)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{31}$$

$$\text{Mean Absolute Percentage Error (MAPE)} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{32}$$

$$\text{Coefficient of Determination } (\mathrm{R}^2) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{33}$$

8.1.2 Classification and Trend Movement Metrics

Classification models forecast discrete market movement directions $y_i \in \{0, 1\}$ (where 1 denotes an upward price movement $P_{t+1} > P_t$ and 0 denotes a downward movement). Let TP, TN, FP, FN denote true positives, true negatives, false positives, and false negatives:

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{34}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN} \tag{35}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \tag{36}$$

$$\text{Matthews Correlation Coefficient (MCC)} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{37}$$

$$\text{Directional Accuracy (DA)} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\text{sign}(y_i - y_{i-1}) == \text{sign}(\hat{y}_i - y_{i-1})\right) \tag{38}$$

8.1.3 Financial and Risk-Adjusted Trading Metrics

To evaluate genuine economic utility, predicted price signals are converted into simulated trading portfolios generating daily return series $R_{p,t}$:

$$\text{Cumulative Return (CR)} = \prod_{t=1}^{T}(1 + R_{p,t}) - 1 \tag{39}$$

$$\text{Annualized Sharpe Ratio (SR)} = \sqrt{252} \cdot \frac{\bar{R}_p - R_f}{\sigma(R_p)} \tag{40}$$

$$\text{Sortino Ratio (SoR)} = \sqrt{252} \cdot \frac{\bar{R}_p - R_f}{\sqrt{\frac{1}{T}\sum_{t=1}^{T} \min\left(0, R_{p,t} - R_{\text{target}}\right)^2}} \tag{41}$$

$$\text{Maximum Drawdown (MDD)} = \max_{t \in [1,T]} \left(\frac{\max_{\tau \in [1,t]} V_\tau - V_t}{\max_{\tau \in [1,t]} V_\tau} \right) \tag{42}$$

where R_f is the risk-free rate, R_{target} is the minimum acceptable benchmark return, and $V_t = V_0 \prod_{\tau=1}^{t}(1 + R_{p,\tau})$ represents the portfolio equity curve.

8.2 Empirical Distribution and Metric Adoption Rates

Our empirical audit of the 685 synthesized primary studies reveals a severe structural imbalance in evaluation metric adoption:

- **Classification Metrics:** Overall Accuracy is reported in 674 studies (98.4%), Precision in 375 studies (54.7%), Recall in 175 studies (25.5%), F1-Score in 171 studies (25.0%), AUC-ROC in 120 studies (17.5%), and explicitly labeled Directional Accuracy in 87 studies (12.7%) [137, 152, 156, 166, 171].
- **Regression Error Metrics:** RMSE appears in 467 studies (68.2%), MAE in 411 studies (60.0%), MSE in 376 studies (54.9%), MAPE in 291 studies (42.5%), and R^2 in 282 studies (41.2%) [113, 114, 119, 121, 130, 151, 165].
- **Financial and Trading Metrics:** Cumulative Returns or simulated trading profit appear in 336 studies (49.1%), the Sharpe Ratio appears in 97 studies (14.2%), Maximum Drawdown (MDD) appears in 56 studies (8.2%), and the Sortino Ratio appears in only 6 studies (0.9%) [31, 35, 37, 41, 56, 96, 98, 117, 122, 144, 146, 161, 172, 186, 217].

Approximately 45.8% of the synthesized literature reports zero financial or economic metrics of any kind, evaluating financial forecasting models exclusively through statistical point-accuracy criteria.

8.3 The "Lag-1 Shadowing Trap" and Metric Pitfalls

A widespread vulnerability identified across regression-based studies is the *Lag-1 Shadowing Trap*. Consider a financial price series modeled as a random walk with drift: $y_t = y_{t-1} + \mu + \varepsilon_t$, where $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $\sigma_\varepsilon^2 \ll \mathrm{Var}(y)$. A trivial naive persistence model $\hat{y}_t = y_{t-1}$ produces an explained variance metric:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{n\sigma_\varepsilon^2}{\mathrm{Var}(y)} \longrightarrow 1.0 \quad \text{as } n \to \infty \tag{43}$$

yielding an ostensibly near-perfect fit ($R^2 > 0.98$) and low RMSE simply by outputting the previous day's closing price. However, the directional prediction hit rate for future returns $\mathrm{sign}(y_t - y_{t-1})$ remains 50% (pure chance), generating zero actionable trading alpha in practice.

The standard statistical metrics also carry mathematical limitations of their own:

- **Scale Dependence:** RMSE and MAE are unscaled, nominal quantities. An RMSE of 2.5 on a \$50 stock is a 5% error; the same RMSE on a \$1,000 stock is 0.25%, so cross-asset comparisons mean nothing without volatility scaling.
- **MAPE Singularity:** As target asset returns approach zero ($y_t \to 0$), the MAPE denominator goes to zero with them and the metric becomes explosively unstable.
- **Symmetric Loss Distortion:** Quadratic losses (MSE) charge over-prediction and under-prediction the same, which ignores the sharp financial asymmetry between suffering a downside capital drawdown and merely missing an upside rally.

8.4 Economic Feasibility: Execution Frictions, Slippage, and Drawdown

In this study, we looked at 336 studies that reported trading returns. We found that most of them assumed ideal trading conditions that are very different from what happens in real markets. For example, they assumed that trades could always be made at the exact closing price, with no transaction and brokerage fees, no price changes between placing and executing an order (slippage: difference between the expected price of a trade and the actual price at which the trade is executed), and no limits on how much could be bought or sold (assumes unlimited money).

In real stock markets, trading costs can quickly reduce the profits that a model appears to make on paper. For example, even a transaction cost of 5 to 15 basis points per trade, along with the bid-ask spread, can have a large impact when a strategy trades frequently. A strategy that shows a 15% to 20% return before costs may actually lose money after accounting for trading costs, frequent buying and selling, and the cost of borrowing stocks for short selling [146, 167].

Owing to the above mentioned reason, a model should not be judged only by how much money it makes. It is also important to look at how much risk it takes to generate those returns. Measures such as the Sharpe (investment's return per unit of total risk compared to risk free asset) and Sortino (nvestment's return relative to its downside risk, ignoring positive upside volatility) ratios help assess this, while

Maximum Drawdown shows how much an investor could lose if market goes in opposite direction.

One another issue we identified during the study is how these models are compared with other investment strategies. Many studies compare their models only with a simple buy-and-hold strategy and do not include other realistic portfolio strategies as benchmarks. Puppala et al. [193], for example, studied sector rotation using S&P 500 sector ETFs and compared the usual strategy of investing in the best-performing sectors with strategies that selected average performing or poorly performing sectors. They found that the average performing approach achieved similar overall returns while requiring much less buying and selling and experiencing smaller losses during downturns. Benchmarking machine learning trading agents against robust, turnover-efficient rules such as median sector rotation provides a more reliable assessment of genuine economic value than comparisons against unadjusted market indices.

8.5 Statistical Hypothesis Testing and Econometric Validation

Formal statistical hypothesis testing is reported in only 116 of the 685 studies (16.9%). Among studies reporting significance tests: ANOVA appears in 50 studies (7.3%), Student's t-tests in 30 studies (4.4%), non-parametric Friedman / Nemenyi post-hoc tests in 24 studies (3.5%), Wilcoxon signed-rank tests in 18 studies (2.6%), and the econometric Diebold-Mariano forecast comparison test appears in only 17 studies (2.5%) [42, 44, 66–68, 70–72, 81, 83, 85, 92, 102, 103].

Given two competing model forecast error series $e_{1,t} = y_t - \hat{y}_{1,t}$ and $e_{2,t} = y_t - \hat{y}_{2,t}$, the loss differential is defined as $d_t = \mathcal{L}(e_{1,t}) - \mathcal{L}(e_{2,t})$. The Diebold-Mariano test evaluates the null hypothesis of equal predictive accuracy $H_0 : \mathbb{E}[d_t] = 0$ via the test statistic:

$$S_{DM} = \frac{\bar{d}}{\sqrt{\frac{1}{T}\left(\hat{\gamma}_0 + 2\sum_{k=1}^{h-1}\hat{\gamma}_k\right)}} \xrightarrow{d} \mathcal{N}(0,1) \tag{44}$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$, $\hat{\gamma}_k$ is the sample autocovariance of d_t at lag k, and h is the forecast horizon. For multi-model rankings across M distinct benchmark assets, the non-parametric Friedman test evaluates average rank differences $\bar{R}_j$:

$$\chi_F^2 = \frac{12M}{K(K+1)}\left[\sum_{j=1}^{K}\bar{R}_j^2 - \frac{K(K+1)^2}{4}\right] \tag{45}$$

followed by the Nemenyi post-hoc test with critical difference threshold $\mathrm{CD} = q_\alpha\sqrt{\frac{K(K+1)}{6M}}$. Over 83% of reviewed papers assert model superiority from raw point estimates alone, without establishing whether reported margins are statistically distinguishable from random noise.

9 Cross-Cutting Thematic Discussion

9.1 The Architectural Complexity vs. Evidential Rigor Paradox

if we put the research findings from all four RQs side by side, we see that one structural dichotomy sits at the center of financial artificial intelligence, which is *architectural model complexity is accelerating while empirical rigor, validation discipline, and open science practice stand still.*

Across all four RQs, we see that same problem keeps surfacing, architectures have evolved and grown sophisticated, but the empirical and technical rigor, validation standards, and open science habits around them have not moved by the same measure. Over the last decade (2016 - 2026), we are able to conclude that financial research migrated from basic econometric models and kernel machines to transformers, multi-graph neural networks, deep reinforcement learning agents, and even Agentic AI (RQ1, RQ2). During the study, we found that only 8% of studies share working source code or a reproducible computational setup (RQ3).We also saw that more than 83% never ran formal econometric hypothesis tests, and nearly 47% never discuss how their models would perform under real trading conditions. It does not come as a surprise when studies report 90% directional accuracy, but most of them use private data splits that look good on paper but aren't reproducible.

In other areas of machine learning, standardized open benchmarks (such as ImageNet, GLUE, and SQuAD) and fixed evaluation servers guarantee transparent and reproducible comparisons. In quantitative finance, the absence of universally accepted open-access benchmarks lets researchers choose advantageous historical periods, asset subsets, and data splits, a practice which often makes new models look like they outperform baseline ones. And since trading strategies have commercial value, industry professionals are reluctant to share their production pipelines, so academic research has to rely on small datasets and unverified model innovations.

9.2 Regime Non-Stationarity, Market Reflexivity, and Overfitting Realities

Financial time series have a low signal-to-noise ratio, exhibit fat tails, experience changes in volatility, and undergo regime shifts. In competitive equity markets, as the Adaptive Market Hypothesis explains, simple patterns are rapidly arbitraged by institutional participants. Financial markets are also, by their nature, reflexive, as George Soros points out: algorithmic models that take advantage of market patterns can alter price dynamics, so historical signals cease to have predictive value over time.

Many deep neural networks achieve high in-sample and single-test-split accuracy because they overfit to particular historical market conditions or take advantage of data leakage during preprocessing. When faced with new market situations, such as changes in interest rates, liquidity crises, geopolitical events, or flash crashes, unregularized models generally fail to generalize. Models optimized for stable bull markets, for instance, tend to make incorrect price assessments when volatility rises. In practical finance, developing predictive models that are robust to changing market conditions

and have stable causal representations matters more than small gains in accuracy on static datasets.

9.3 Institutional Hurdles: Explainability Mandates, Low-Latency Constraints, and Systemic Risk

Institutional asset managers and quantitative hedge funds increasingly deploy machine learning for automated signal generation and real-time capital allocation. However, regulatory frameworks impose strict governance, auditability, and pre-trade risk controls on algorithmic trading systems. While equity forecasting models are not categorized as high-risk systems under Annex III of the European Union Artificial Intelligence Act (Regulation 2024/1689)—which restricts its high-risk designation to domains such as critical infrastructure, biometric identification, creditworthiness assessment, and employment evaluation—algorithmic trading is governed by specialized capital markets regulations. In the European Union, the Markets in Financial Instruments Directive II (MiFID II, Directive 2014/65/EU, Article 17) and its Regulatory Technical Standard 6 (RTS 6) mandate continuous algorithmic stress-testing, pre-trade price and volume collars, automated kill switches, and comprehensive audit logs of order generations. In the United States, SEC Rule 15c3-5 (the Market Access Rule) similarly requires broker-dealers to maintain deterministic financial and regulatory risk controls to prevent the entry of erroneous algorithmic orders and mitigate systemic flash crashes. Because deep sequence networks and complex ensemble pipelines operate as opaque black boxes, institutional desks face acute compliance hurdles under MiFID II and SEC mandates when models cannot provide verifiable feature attributions, bounded behavioral guarantees, or transparent risk controls.

In order to help institutions adopt these systems, it is useful to combine explainable AI techniques such as SHAP, Layer-wise Relevance Propagation, and rule-based recurrent units with risk models. Even though a lot of research points to the use of complex multi-stage models, for example VMD-CEEMDAN-CNN-BiLSTM-Attention stacks optimised with metaheuristics, these are generally too slow for high-frequency or intraday trading. In such cases the models have to operate in less than a millisecond, which means that multi-stage pipelines will only function if they are greatly simplified and make use of hardware acceleration.

9.4 The Four-Pillar Standardization Protocol

To transition financial AI from fragmented ad-hoc experimentation toward an institutional, cumulative empirical science, we establish a standardized Four-Pillar Reporting Protocol for future research:

9.4.1 Pillar 1: Deterministic Open Science Mandate

Authors must include in their submission public repository links, for example those from GitHub or Zenodo, which provide runnable code, the exact package versions obtained using Docker or Conda, the data preprocessing steps, and a fixed random seed. In the case where commercial rules make it impossible to share the raw

price data, the authors must offer synthetic datasets together with clear instructions enabling others to get the data for themselves.

9.4.2 Pillar 2: Purged Walk-Forward Temporal Validation

Don't apply static single splits or process the whole dataset at one time; instead, carry out normalization, feature engineering, and signal decomposition solely within the rolling windows in order to prevent the use of future information. When preparing the validation folds, include purge and embargo periods between the training and testing phases to avoid any leakage of information and to reduce serial correlation.

9.4.3 Pillar 3: Econometric and Statistical Significance Testing

The superiority of models can only be demonstrated through formal econometric hypothesis testing. When comparing models in pairs, the authors should provide the Diebold-Mariano test statistics (including the multi-step autocovariance corrections), and when assessing multiple models across a range of assets, they should use non-parametric Friedman and Nemenyi post-hoc tests. Point estimates that do not include variance bounds or significance levels should be considered unverified.

9.4.4 Pillar 4: Risk-Adjusted and Friction-Aware Economic Backtesting

Predictive models presented in studies need to be tested in trading simulations that reflect real market conditions, not clean backtests. For testing in scenarios that reflect real markets, researchers need to factor in bid-ask spreads that change every minute, transaction costs involved, which are at least 10 basis points, slippage penalties tied to trade volume, and realistic borrowing constraints. Post addressing the factors outlines, model performance should be evaluated by using risk-adjusted metrics including but not limied to annualized Sharpe ratio, Sortino ratio, maximum drawdown, and Calmar ratio. These measured provide us with a more meaningful assessment of the nodel's financial performance and risk. RMSE and MSE are quite useful for model tuning and predictive accuracy but they fail to offer much insight on if the model can actually generate profitable returns on investment.

Once these factors have been addressed, performance should be judged using risk-adjusted metrics such as the annualized Sharpe ratio, Sortino ratio, maximum drawdown, and Calmar ratio. RMSE and MSE help with tuning; on the question of whether the model actually makes money, they are nearly silent.

10 Limitations of this review

This review is, to date, the largest empirical synthesis in computational finance ($N = 685$), but, as saying goes, every coin has two sides. While the scale of review acts as strength, yet is also comes with methodological constraints and scope boundaries worth a discussion:

10.1 Methodological Heterogeneity and Meta-Analytic Constraints

A formal quantitative meta-analysis based on pooled effect sizes and forest plots was not statistically appropriate because the studies in the literature are highly heterogeneous. In particular:

1. **The asset universes vary substantially:** The studies examine very different types of financial assets, ranging from highly liquid indices such as the S&P 500 and NASDAQ 100 to thinly traded individual stocks on emerging market exchanges in Dhaka, Tehran, and Colombo. These markets differ a lot in terms of market liquidity, noise, and trading constraints, which makes direct comparison of effect estimates very difficult.
2. **Temporal designs are not standardized:** The studies also use different time horizons and experimental setups. Sample periods used in the studies range from just six months of tick data to 25 years of macroeconomic data. In a similar manner, train/test splits vary across studies, with 70/30, 80/20, and 90/10 being commonly used, while values for rolling window lengths and other temporal validation schemes are equally inconsistent.
3. The problem is that studies don't even measure the same thing the same way. Directional accuracy, RMSE, and annualized returns get calculated with different formulas, on unscaled price levels, and under different assumptions about execution, so the numbers aren't really comparable to begin with.

We considered conducting bounded sub-sample meta-regressions on the subset of studies targeting the S&P 500 daily close with directional accuracy as the primary metric ($n \approx 94$). However, even within this ostensibly homogeneous subset, the training/test split ratios, lookback window lengths, feature sets, and normalization procedures vary so widely that pooling effect sizes would produce a composite estimate whose confidence interval spans the entire plausible range. We therefore report qualitative synthesis and vote-counting of comparative rankings (Section 5) rather than quantitative pooled effect sizes.

10.2 Partial 2026 Temporal Window

The systematic search procedure included empirical studies that were published from January 2026 up to the early part of 2026 (a total of $n = 89$ studies). Although this allows for a synthesis of the most recent developments in Large Language Model reasoning and dynamic graph architectures, the group of studies from 2026 provides an up-to-date view of a publication year that is still in progress.

10.3 Publication Bias in Primary Literature

Recent trends in computational finance show a clear bias for publishing results with positive outcomes. We found that most studies claimed that new architectures, including but not limited to transformers and multimodal neural networks, outperform benchmark models; however, negative results, failed deep learning experiments, and substantial losses incurred during previous market crashes are rarely reported. This

failure to report such cases can create a misleading view of the actual performance in the literature. Even the studies that reported moderate gains over traditional models did not specify where the developed model is lacking or underperforms compared to the traditional ones.

11 Future Research Directions

Drawing upon our comprehensive empirical synthesis and identified standardization gaps, we outline six high-priority future research directions for computational finance and applied artificial intelligence:

11.1 Unified Open-Access Financial Benchmarking Suites

Just as there is ImageNet in computer vision and GLUE in natural language processing, computational finance must have opensource, standardized benchmark suites. The benchmarks should make use of fixed equity universes in order to prevent survivorship bias, provide standard rolling splits for training, validation, and testing, and also include automated tools for measuring both statistical loss and economic performance under various market conditions.

11.2 Causal Discovery and Invariant Predictive Mechanisms

Most present day deep neural networks rely on statistical correlations, a method that tends to fail during market shocks. Research in the future should make use of causal discovery techniques, such as structural equation modeling, invariant causal prediction, and do calculus, in order to identify stable relationships between company fundamentals, macroeconomic indicators, and price formation that remain valid under varying market conditions.

11.3 Heterogeneous Dynamic Graph Neural Networks

Future studies should design dynamic Graph Neural Networks that move beyond the use of static sector groupings and instead make use of diverse, multi-relational market graphs; by incorporating data such as supply chain vulnerabilities, institutional cross-holdings, corporate litigation networks, and high-frequency order book correlations, the models will be able to better monitor how financial risks and contagion spread across sectors in real time.

11.4 Financial Multimodal Foundation Models and Generative AI

Establishing financial foundation models trained on a unified multimodal dataset that combines high-frequency order book data, structured financial reports, SEC 10-K filings, earnings call audio and published reports, and financial news in various languages represents a significant advancement. It is important to use temporal retrieval augmented generation and apply strict chronological attention masking to reduce LLM hallucinations and lookahead bias.

11.5 Friction-Aware and Risk-Constrained Reinforcement Learning

Optimization models for future trading should incorporate realistic execution mechanics as one of their objectives; rather than merely assessing point forecasts, researchers should make use of constrained reinforcement learning agents which seek to enhance risk-adjusted returns, for example, those based on the Sortino and Sharpe ratios, while at the same time imposing penalties for high portfolio turnover, market impact, and large drawdowns within borrowing and liquidity limits.

11.6 Privacy-Preserving Federated Learning and Open Science Ecosystems

Establishing mandatory code and data sharing policies across computational finance journals will be essential to resolving the domain's reproducibility crisis. Furthermore, privacy-preserving federated learning architectures offer a viable path for multi-institutional collaboration, enabling quantitative institutions to train shared forecasting models across proprietary institutional order books without compromising confidential trading strategies or customer data.

12 Conclusion

12.1 Synthesis of Empirical Findings

This systematic literature review conducted an exhaustive, PRISMA 2020-compliant synthesis of 685 primary empirical studies on artificial intelligence and machine learning techniques for stock market prediction published across the 2016–2026 decade. Our empirical findings demonstrate that recurrent deep sequence architectures represent the installed default across the literature, with LSTM appearing in 611 studies (89.2%) and hybrid ensembling in 424 studies (61.9%). The recent triennium exhibits rapid acceleration in Explainable AI (391 studies, 57.1%), multi-head self-attention Transformers (249 studies, 36.4%), and multimodal NLP sentiment integration (113 studies, 16.5%).

12.2 The Empirical Rigor Imperative

Although algorithms are becoming more complex, the empirical foundation remains weak. Only 7.6% of the audited studies made their source code available to ensure reproducibility and verification of the results. Over 83% of the studies reviewed did not use formal econometric hypothesis testing, and 45.8% did not evaluate financial or economic utility. Recurring problems, among them pre-split global normalization leakage, lookahead bias in multiscale signal decomposition, and survivorship bias, routinely leave the reported performance numbers inflated.

12.3 Strategic Roadmap for the Discipline

To establish financial artificial intelligence as a rigorous, cumulative empirical science, the research community must transition from chasing marginal point-accuracy

improvements on single static splits toward adopting deterministic open science standards, rolling walk-forward validation protocols, formal econometric hypothesis testing, and friction-aware economic evaluation. Adopting the Four-Pillar Standardization Protocol established in this review will ensure that future AI breakthroughs translate into reproducible, robust, and economically viable financial applications.

Declarations

Funding

The authors declare that no external funds, grants, or institutional financial support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial competing interests to disclose.

Ethics Approval and Consent to Participate

Not applicable. This systematic review does not involve human participants, clinical trials, animal subjects, or private personal data.

Consent for Publication

Not applicable.

Supplementary Information

The complete empirical data extraction matrix and systematic coding synthesis for all 685 analyzed primary studies is provided as Supplementary Material attached with this submission (`Final Sheet.xlsx`).

Availability of Data and Materials

The complete data extraction matrix for all 685 primary empirical studies (22 coded attributes per study, including bibliographic identifiers, architectural classification, reproducibility tier, leakage vector flags, and evaluation metric coverage) is provided as Supplementary Material attached with this submission (`Final Sheet.xlsx`). It can also be made available to readers by requesting from corresponding author. The bibliographic reference list and PRISMA screening log are included as supplementary CSV files in the same repository.

Code Availability

All analytical scripts used for the bibliometric analysis, reproducibility audit scoring, PRISMA workflow automation, and figure generation are available on request.

Authors' Contributions

A.S. (Akshatha Srikantha) and **M.S.** contributed equally to this work as co-first authors. **M.S.** and **A.S.** (Akshatha Srikantha) conceptualized the study design, executed the PRISMA search queries across 685 empirical studies, conducted data extraction, and drafted the manuscript. **M.S.** and **E.P.** contributed to the mathematical problem formulation and financial econometric analysis. **A.S.** (Anant Singh) contributed to the algorithmic trading evaluations and execution friction analysis. **R.J.** and **E.P.** supervised the research methodology and provided critical revisions. **R.J.** managed project coordination and correspondence. All authors read, refined, and approved the final manuscript.

References

[1] Chhajer, P., Shah, M., Kshirsagar, A.: The applications of artificial neural networks, support vector machines, and long short-term memory for stock market prediction. Decision Analytics Journal **2**, 100015 (2021). https://doi.org/10.1016/j.dajour.2021.100015

[2] Chong, E., Han, C., Park, F.C.: Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications **83**, 187–205 (2017). https://doi.org/10.1016/j.eswa.2017.04.030

[3] Li, A.W., Bastos, G.S.: Stock market forecasting using deep learning and technical analysis: A systematic review. IEEE Access **8**, 185232–185242 (2020). https://doi.org/10.1109/access.2020.3030226

[4] Lin, C.Y., Marques, J.A.L.: Stock market prediction using artificial intelligence: A systematic review of systematic reviews. Social Sciences & Humanities Open **9**, 100864 (2024). https://doi.org/10.1016/j.ssaho.2024.100864

[5] Olorunnimbe, K., Viktor, H.: Deep learning in the stock market: a systematic survey of practice, backtesting, and applications. Artificial Intelligence Review **56**(3), 2057–2109 (2022). https://doi.org/10.1007/s10462-022-10226-0

[6] Rohan, A., Hossen, M.D., Pranto, M.N., Hossain, B., Yoshi, A.M., Islam, R.: Artificial intelligence in financial market prediction: advancements in machine learning for stock price forecasting. Frontiers in Artificial Intelligence **8**, 1696423 (2026). https://doi.org/10.3389/frai.2025.1696423

[7] Rouf, N., Malik, M.B., Arif, T., Sharma, S., Singh, S., Aich, S., Kim, H.: Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. Electronics **10**(21), 2717 (2021). https://doi.org/10.3390/electronics10212717

[8] Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research **270**(2), 654–669 (2017). https://doi.org/10.1016/j.ejor.2017.11.054

[9] Patel, J., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications **42**(1), 259–268 (2014). https://doi.org/10.1016/j.eswa.2014.07.040

[10] Kumbure, M.M., Lohrmann, C., Luukka, P., Porras, J.: Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications **197**, 116659 (2022). https://doi.org/10.1016/j.eswa.2022.116659

[11] Mintarya, L.N., Halim, J.N., Angie, C., Achmad, S., Kurniawan, A.: Machine learning approaches in stock market prediction: A systematic literature review. Procedia Computer Science **216**, 96–102 (2023). https://doi.org/10.1016/j.procs.2022.12.115

[12] Kumar, D., Sarangi, P.K., Verma, R.: A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings **49**, 3187–3191 (2021). https://doi.org/10.1016/j.matpr.2020.11.399

[13] Nti, I.K., Adekoya, A.F., Weyori, B.A.: A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review **53**(4), 3007–3057 (2019). https://doi.org/10.1007/s10462-019-09754-z

[14] Akşehir, Z.D., Kılıç, E.: Analyzing the critical steps in deep learning-based stock forecasting: a literature review. PeerJ Computer Science **10**, e2312 (2024). https://doi.org/10.7717/peerj-cs.2312

[15] Htun, H.H., Biehl, M., Petkov, N.: Survey of feature selection and extraction techniques for stock market prediction. Financial Innovation **9**(1), 26 (2023). https://doi.org/10.1186/s40854-022-00441-7

[16] Rath, S., Das, N.R., Pattanayak, B.K.: An analytic review on stock market price prediction using machine learning and deep learning techniques. Recent Patents on Engineering **18**(2), e030323214251 (2023). https://doi.org/10.2174/1872212118666230303154251

[17] Bustos, O., Pomares-Quimbaya, A.: Stock market movement forecast: A systematic review. Expert Systems with Applications **156**, 113464 (2020). https://doi.org/10.1016/j.eswa.2020.113464

[18] Wiranata, R.B., Djunaidy, A.: The stock exchange prediction using machine learning techniques: A comprehensive and systematic literature review. Jurnal Ilmu Komputer dan Informasi **14**(2), 91–112 (2021). https://doi.org/10.21609/

jiki.v14i2.935

[19] Ashtiani, M.N., Raahemi, B.: News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. Expert Systems with Applications **217**, 119509 (2023). https://doi.org/10.1016/j.eswa.2023.119509

[20] Sharma, P., Ojha, V., Tomar, P.K., Biswal, S.K., Vinoth, S., Homavazir, Z.: Systematic review of advanced econometric models in predicting stock market price movements. Multidisciplinary Reviews **8**, 2025ss0328 (2025). https://doi.org/10.31893/multirev.2025ss0328

[21] Strader, T.J., Rozycki, J.J., Root, T.H., Huang, Y.J.: Machine learning stock market prediction studies: Review and research directions. Journal of International Technology and Information Management **28**(4), 63–83 (2020). https://doi.org/10.58729/1941-6679.1435

[22] Patel, M., Jariwala, K., Chattopadhyay, C.: A systematic review on graph neural network-based methods for stock market forecasting. ACM Computing Surveys **57**(2), 1–38 (2024). https://doi.org/10.1145/3696411

[23] Adeleye, R.A., Tubokirifuruar, T.S., Bello, B.G., Ndubuisi, N.L., Asuzu, O.F., Owolabi, O.R.: Machine learning for stock market forecasting: A review of models and accuracy. Finance & Accounting Research Journal **6**(2), 112–124 (2024). https://doi.org/10.51594/farj.v6i2.783

[24] Soni, P., Tewari, Y., Krishnan, D.: Machine learning approaches in stock price prediction: A systematic review. Journal of Physics: Conference Series **2161**(1), 012065 (2022). https://doi.org/10.1088/1742-6596/2161/1/012065

[25] Bhardwaj, D.: Machine learning approaches in stock market prediction: A systematic literature review. International Journal for Research in Applied Science and Engineering Technology **13**(4), 4446–4451 (2025). https://doi.org/10.22214/ijraset.2025.69153

[26] Fama, E.F.: Efficient capital markets: A review of theory and empirical work. The Journal of Finance **25**(2), 383–417 (1970). https://doi.org/10.2307/2325486

[27] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735

[28] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **30**, 5998–6008 (2017). https://doi.org/10.48550/arXiv.1706.03762

[29] Lim, B., Arık, S.Ö., Loeff, N., Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting **37**(4), 1748–1764 (2021). https://doi.org/10.1016/j.ijforecast.2021.03.012

[30] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. The BMJ **372**, n71 (2021). https://doi.org/10.1136/bmj.n71

[31] Yongchareon, S., "AI-Driven Intelligent Financial Forecasting: A Comparative Study of Advanced Deep Learning Models for Long-Term Stock Market Prediction," Machine Learning and Knowledge Extraction, 2025. https://doi.org/10.3390/make7030061

[32] M. Asfar and M. H. Khan and O. Qureshi and M. F. Shahid, "AI-Enabled Framework for Stock Market Prediction Using KSE-100 Index," 2026 International Conference on Data Science, Machine Learning, and Intelligence (DataSciMI), 2026. https://doi.org/10.1109/datascimi67380.2026.11523832

[33] Khalil, R., "AI driven sentiment analysis in financial markets: using transformer base models and social media signals for stock market predictions," Journal of Modelling in Management, 2026. https://doi.org/10.1108/jm2-08-2025-0415

[34] K. R. Ahmed and M. M. H. Sizan and A. Chouksey and M. A. A. Montaser and M. R. Islam and M. S. Islam, "AI for Financial Market Prediction: Machine Learning Models for Stock Price Forecasting," 2025 International Conference on Networks and Cryptology (NETCRYPT), 2025. https://doi.org/10.1109/netcrypt65877.2025.11102142

[35] Panda, N and Singh, T and Swagatika, S, "ANN based modeling for stock market prediction," JOURNAL OF STATISTICS AND MANAGEMENT SYSTEMS, 2023. https://doi.org/10.47974/jsms-949

[36] P. P. and C. C. M. and V. P. and P. Pinto and M. I. M. Riyath, "A CNN-LSTM Hybrid Framework for Multimodal Stock Market Movement Prediction Using Nifty 50 Data," 2026 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), 2026. https://doi.org/10.1109/iitcee67948.2026.11394619

[37] Singh, S.K. and Praveen, S., "A COMPARATIVE ANALYSIS OF TREE-BASED ENSEMBLES AND DEEP LEARNING FOR EQUITY PRICE FORECASTING USING FEATURE-ENGINEERED ADVANCE TECHNICAL INDICATORS," International Journal of Applied Mathematics, 2025. https://doi.org/10.12732/ijam.v38i2s.124

[38] Kihtir, F. and Oztoprak, K., "A Comparative Analysis on Leveraging K-Nearest Neighbours for Accurate Stock Price Forecasting," Proceedings - 2024 IEEE International Conference on Big Data, BigData 2024, 2024. https://doi.org/10.1109/bigdata62323.2024.10825529

[39] Paul, M.K. and Das, P., "A Comparative Study of Deep Learning Algorithms for Forecasting Indian Stock Market Trends," International Journal of Advanced Computer Science and Applications, 2023. https://doi.org/10.14569/ijacsa.2023.0141098

[40] Jeribi, F. and Martin, R.J. and Mittal, R. and Jari, H. and Alhazmi, A.H. and Malik, V. and Swapna, S.L. and Goyal, S.B. and Kumar, M. and Singh, S.V., "A Deep Learning Based Expert Framework for Portfolio Prediction and Forecasting," IEEE Access, 2024. https://doi.org/10.1109/access.2024.3434528

[41] A. K. Biswas and M. S. A. Bhuiyan and M. N. H. Mir and A. Rahman and M. F. Mridha and M. R. Islam and Y. Watanobe, "A Dual Output Temporal Convolutional Network With Attention Architecture for Stock Price Prediction and Risk Assessment," IEEE Access, 2025. https://doi.org/10.1109/access.2025.3551307

[42] Mawere, T. and Rajalakshmi, S. and Kuthadi, V.M. and Dinakanyane, O., "A Dynamic Lifelong Learning Network for Real-Time Stock Prediction With Reinforcement Tuning," Journal of Computer Science, 2026. https://doi.org/10.3844/jcssp.2026.420.434

[43] A. K. Das and D. Mishra and K. Das and K. C. Mishra, "A Feature Ensemble Framework for Stock Market Forecasting Using Technical Analysis and Aquila Optimizer," IEEE Access, 2024. https://doi.org/10.1109/access.2024.3461792

[44] Li, H. and Hu, J., "A Hybrid Deep Learning Framework for Stock Price Prediction Considering the Investor Sentiment of Online Forum Enhanced by Popularity," Computational Economics, 2026. https://doi.org/10.1007/s10614-025-11268-6

[45] Dipura, R.W.A. and Nday, M.A.G. and Maulana, F.I., "A Hybrid Ensemble Framework using Reinforcement Learning and Machine Learning for NASDAQ Stock Prediction based on Technical and Sentiment Analysis," Proceedings - 7th International Conference on Informatics, Multimedia, Cyber and Information System, ICIMCIS 2025, 2025. https://doi.org/10.1109/icimcis68501.2025.11327396

[46] Malakouti Semnani, A. and Kordrostami, S. and Sheikhani, A.R. and Moattar, M.H., "A Hybrid Framework for Stock Price Forecasting Using Metaheuristic Feature Selection Approaches and Transformer Models Enhanced by Temporal Embedding and Attention Pruning," Applied AI Letters, 2026. https://doi.org/10.1002/ail2.70018

[47] Hashemipour, M.S. and Zahedi, M. and Fateh, M., "A Hybrid Model for Stock Market Prediction with Stacked Autoencoder for Feature Engineering," Iranian Journal of Science and Technology - Transactions of Electrical Engineering, 2026. https://doi.org/10.1007/s40998-026-01017-2

[48] Niu, H., "A Method to Increase the Analysis Accuracy of Stock Market Valuation: A Case Study of the Nasdaq Index," International Journal of Advanced Computer Science and Applications, 2024. https://doi.org/10.14569/ijacsa.2024.0150173

[49] L. Yu and Y. Chen and Y. Zheng, "A Multi-Market Data-Driven Stock Price Prediction System: Model Optimization and Empirical Study," IEEE Access, 2025. https://doi.org/10.1109/access.2025.3559169

[50] Namdari, A. and Durrani, T.S., "A Multilayer Feedforward Perceptron Model in Neural Networks for Predicting Stock Market Short-term Trends," Operations Research Forum, 2021. https://doi.org/10.1007/s43069-021-00071-2

[51] A. H. Dhafer and F. M. Nor and W. Hashim and N. R. Shah and K. F. Bin Khairi and G. Alkawsi, "A NARX Neural Network Model To Predict One-Day Ahead Movement of The Stock Market Index of Malaysia," 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS), 2021. https://doi.org/10.1109/aidas53897.2021.9574394

[52] Gul, A., "A Novel Hybrid Ensemble Framework for Stock Price Prediction: Combining Bagging, Boosting, Dagging, and Stacking," Computational Economics, 2026. https://doi.org/10.1007/s10614-025-10979-0

[53] Karam, S, "A Novel Hybrid Monte Carlo-Random Forest Framework for Improved Financial Predictions," CONTEMPORARY MATHEMATICS, 2025. https://doi.org/10.37256/cm.6220256079

[54] Y. Sun and S. Mutalib and N. Omar and L. Tian, "A Novel Integrated Approach for Stock Prediction Based on Modal Decomposition Technology and Machine Learning," IEEE Access, 2024. https://doi.org/10.1109/access.2024.3425727

[55] Ali, R.H. and Ijaz, A.Z. and Khan, T.A., "A Novel NLP-based Stock Market Price Prediction and Risk Analysis Framework," Journal of Computer Science and Technology (Argentina), 2024. https://doi.org/10.24215/16666038.24.e07

[56] Chiang, J.K. and Chi, R., "A Novel Stock Price Prediction and Trading Methodology Based on Active Learning Surrogated with CycleGAN and Deep Learning and System Engineering Integration: A Case Study on TSMC Stock Data," FinTech, 2024. https://doi.org/10.3390/fintech3030024

[57] Hadizadeh, A. and Tarokh, M.J. and Ghazani, M.M., "A convolutional deep reinforcement learning architecture for an emerging stock market analysis," Decision Science Letters, 2025. https://doi.org/10.5267/j.dsl.2025.1.006

[58] Chen, P. and Boukouvalas, Z. and Corizzo, R., "A deep fusion model for stock market prediction with news headlines and time series data," Neural Computing and Applications, 2024. https://doi.org/10.1007/s00521-024-10303-1

[59] Albahli, S. and Awan, A. and Nazir, T. and Irtaza, A. and Alkhalifah, A. and Albattah, W., "A deep learning method DCWR with HANet for stock market prediction using news articles," Complex and Intelligent Systems, 2022. https://doi.org/10.1007/s40747-022-00658-0

[60] Zhang, H. and Huang, Y. and Zhou, N. and Du, J. and Li, X., "A hybrid neural framework for financial time series forecasting with CEEMDAN-CNN-BiLSTM-attention integration," International Journal of Intelligent Computing and Cybernetics, 2025. https://doi.org/10.1108/ijicc-05-2025-0266

[61] Gkonis, V. and Tsakalos, I., "A hybrid optimized deep learning model via the Golden Jackal Optimizer for accurate stock price forecasting," Expert Systems with Applications, 2025. https://doi.org/10.1016/j.eswa.2025.127287

[62] Gerakoudi, K and Georgoulas, D and Stavroulakis, PJ, "A novel hybrid ARIMA-LSTM model for maritime shipping stock forecasting: comparative evidence against statistical and machine learning benchmarks," MARITIME BUSINESS REVIEW, 2025. https://doi.org/10.1108/mabr-04-2024-0035

[63] Li, Q and Kamaruddin, N and Zhang, J and Peng, C and Khoo, ASK, "A novel method of bayesian genetic optimization on automated hyperparameter tuning," SCIENTIFIC REPORTS, 2025. https://doi.org/10.1038/s41598-025-29383-7

[64] Chen, Q. and Kawashima, H., "A novel sentiment correlation-based method with dual transformer model for stock price prediction," International Journal of Data Science and Analytics, 2026. https://doi.org/10.1007/s41060-025-00932-7

[65] Khan, AH and Shah, ABL and Ali, A and Shahid, R and Zahid, ZU and Sharif, MU and Jan, TRQL and Zafar, MH, "A performance comparison of machine learning models for stock market prediction with novel investment strategy," PLOS ONE, 2023. https://doi.org/10.1371/journal.pone.0286362

[66] Taha, A.K. and Kholief, M.H. and AbdelMoez, W., "Adaptive Machine Learning-Based Stock Prediction Using Financial Time Series Technical Indicators," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019. https://doi.org/10.1007/978-3-030-33617-2_21

[67] Kayit, A.D. and Ismail, M.T., "Advancing stock price prediction through the development of hybrid ensembles: a comprehensive comparative analysis of machine learning approaches," Journal of Big Data, 2025. https://doi.org/10.1186/s40537-025-01185-8

[68] Y. Li and H. Wu and C. Liang and C. Bu and H. Jin and L. Chen, "An Attention-Guided Improved Decomposition-Reconstruction Model for Stock Market Prediction," IEEE Access, 2025. https://doi.org/10.1109/access.2025.3613336

[69] Minsoor, Z.K. and Al-Sultan, A.Y., "An Efficient Algorithm for Stock Market Prediction Using Attention Mechanism," Journal of Cybersecurity and Information Management, 2024. https://doi.org/10.54216/jcim.140226

[70] S. T. Surulivel and N. P. A and D. V and S. Selvabaskar, "Analysing the Predictive Accuracy of Machine Learning Models in Stock Price Forecasting," 2026 4th International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), 2026. https://doi.org/10.1109/aimla67915.2026.11522427

[71] S. Bouktif and A. Fiaz and M. Awad, "Augmented Textual Features-Based Stock Market Prediction," IEEE Access, 2020. https://doi.org/10.1109/access.2020.2976725

[72] Saha, B. and Nath, S. and Roy, D. and Dey, A. and Saha, A. and Habib, D. and Ghosh, A., "Building a Robust Stock Price Prediction Using LSTM and VECM," KSII Transactions on Internet and Information Systems, 2026. https://doi.org/10.3837/tiis.2026.04.003

[73] T. N. Kusuma and A. C. Sari, "Comparative Analysis of Machine Learning Regression Models for Stock Price Prediction Using Financial Ratios on the Indonesian Stock Exchange," 2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER), 2026. https://doi.org/10.1109/isiber68248.2026.11469993

[74] V. K. Pandey and T. Ahmed and G. Sahoo, "Comparative Analysis of Stock Price Forecasting using MTF Image Transformation: Evaluation of Random Forest and CNN Models," 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN), 2025. https://doi.org/10.1109/iciscn64258.2025.10934414

[75] X. Zhao and Y. Liu and Q. Zhao, "Cost Harmonization LightGBM-Based Stock Market Prediction," IEEE Access, 2023. https://doi.org/10.1109/access.2023.3318478

[76] Tuarob, S and Wettayakorn, P and Phetchai, P and Traivijitkhun, S and Lim, S and Noraset, T and Thaipisutikul, T, "DAViS: a unified solution for data collection, analyzation, and visualization in real-time stock market prediction," FINANCIAL INNOVATION, 2021. https://doi.org/10.1186/s40854-021-00269-7

[77] H. Tanveer and S. Latif, "Decoding Stock Market Predictions: Insights from Explainable AI Using Layer-wise Relevance Propagation," 2024 26th International Multi-Topic Conference (INMIC), 2024. https://doi.org/10.1109/

inmic64792.2024.11004356

[78] M. N. Ahsan and M. Moniruzzaman and S. Jahan, "Deep Learning-Based Stock Price Forecasting and Anomaly Detection: A Data-Driven Approach," 2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), 2025. https://doi.org/10.1109/icdici66477.2025.11135369

[79] Alqahtani, H.S. and Alhaddad, M.J. and Jarrah, M., "Deep Learning-Based Stock Price Forecasting for the Saudi Telecommunication Sector: A Comparative Evaluation with Baseline Models," Journal of Advances in Information Technology, 2026. https://doi.org/10.12720/jait.17.4.696-710

[80] Alhumaidi, A. and Abdelhafez, H.A., "Deep Learning Models for Predicting Stock Closing Prices in the Saudi Stock Market," Journal of Computer Science, 2026. https://doi.org/10.3844/jcssp.2026.1370.1386

[81] Ram, A.A. and Yadav, S. and Vivek, Y. and Ravi, V., "Deep Reinforcement Learning for Financial Forecasting in Static and Streaming Cases," Journal of Information and Knowledge Management, 2024. https://doi.org/10.1142/s0219649224500801

[82] Tafara, R.H. and Zhounggi, Y. and Hasani, M.F. and Maulina, A., "Deep learning-based stock market prediction using historical price and topic-distributed news sentiment," Procedia Computer Science, 2025. https://doi.org/10.1016/j.procs.2025.09.030

[83] Hargreaves, C.A. and Fan, Z., "Denoising Stock Price Time Series with Singular Spectrum Analysis for Enhanced Deep Learning Forecasting," Analytics, 2026. https://doi.org/10.3390/analytics5010009

[84] R. Raj and N. Kaur and P. Goel, "Design Hybrid Model To Improve Stock Selection And Market Timing By Integrating Deep Learning And Sentiment Analysis," 2025 International Conference on Computing Technologies (ICOCT), 2025. https://doi.org/10.1109/icoct64433.2025.11118343

[85] Ma, ZY and Ryan, C and Buckley, J and Chochlov, M, "Do Weibo Platform Experts Perform Better at Predicting Stock Market?," PROCEEDINGS OF THE 22ND ENGINEERING APPLICATIONS OF NEURAL NETWORKS CONFERENCE, EANN 2021, 2021. https://doi.org/10.1007/978-3-030-80568-5_40

[86] Ravi, G. and Selvakumar, D.S., "Dynamic Deep Stock Price Prediction Network with Support of Multi-Source Information," International Journal of Intelligent Engineering and Systems, 2025. https://doi.org/10.22266/ijies2025.0930.09

[87] Gunaryati, A. and Mutiara, A.B. and Puspitodjati, S., "Enhancing Sharia Stock Price Forecasting using a Hybrid ARIMA-LSTM with Locally Weighted Scatterplot Smoothing Regression Approach," Journal of Applied Data Sciences, 2025.

https://doi.org/10.47738/jads.v6i1.514

[88] K. Alam and M. H. Bhuiyan and I. U. Haque and M. F. Monir and T. Ahmed, "Enhancing Stock Market Prediction: A Robust LSTM-DNN Model Analysis on 26 Real-Life Datasets," IEEE Access, 2024. https://doi.org/10.1109/access.2024.3434524

[89] I. Kasse and J. Kusuma and A. Lawi and A. Rahim, "Enhancing Stock Price Forecasting Accuracy Through Compositional Learning of Recurrent Architectures: A Multi-Variant RNN Approach," IEEE Access, 2025. https://doi.org/10.1109/access.2025.3602721

[90] Baamer, R. and Aljamaan, H., "Enhancing Stock Price Forecasting with Deep Learning: Insights from the Saudi Stock Market," Computational Economics, 2025. https://doi.org/10.1007/s10614-025-11033-9

[91] Liu, J and Guo, YE, "Enhancing stock market precision estimate through the implementation of a hybrid methodology," INTERNATIONAL JOURNAL OF SYSTEM ASSURANCE ENGINEERING AND MANAGEMENT, 2026. https://doi.org/10.1007/s13198-026-03307-8

[92] Huang, Y. and Yang, C., "Enhancing stock price forecasting with a modular deep learning framework incorporating plug-and-play transformer variants," Expert Systems with Applications, 2026. https://doi.org/10.1016/j.eswa.2026.131572

[93] Narayana, M.L. and Kartha, A.J. and Mandal, A.K. and P, R. and Suresh, A. and Jose, A.C., "Ensemble time series models for stock price prediction and portfolio optimization with sentiment analysis," Journal of Intelligent Information Systems, 2025. https://doi.org/10.1007/s10844-025-00928-6

[94] Bashir, U. and Singh, K. and Mansotra, V. and Khanday, A.M.U.D. and Neshat, M., "EvoBagNet: a decomposition-driven bagging ensemble with evolutionary hyperparameter optimization for robust stock price prediction," Neural Computing and Applications, 2026. https://doi.org/10.1007/s00521-025-11789-z

[95] Liao, L. and Lin, J., "F-LOAM: an efficient hybrid model for stock price prediction based on SVMD denoising," Discover Artificial Intelligence, 2025. https://doi.org/10.1007/s44163-025-00622-0

[96] Rehman, M. and Fuzail, M. and Abid, M.K. and Aslam, N., "Financial Prices Prediction of Stock Market using Supervised Machine Learning Models," VFAST Transactions on Software Engineering, 2023. https://doi.org/10.21015/vtse.v11i2.1439

[97] Zhao, C. and Liu, X. and Zhou, J. and Cen, Y. and Yao, X., "GCN-based stock relations analysis for stock market prediction," PeerJ Computer Science, 2022. https://doi.org/10.7717/peerj-cs.1057

[98] Zhao, XS and Liu, Y and Zhao, QF, "Generalized Loss-Based CNN-BiLSTM for Stock Market Prediction," INTERNATIONAL JOURNAL OF FINANCIAL STUDIES, 2024. https://doi.org/10.3390/ijfs12030061

[99] R. Jia and K. Yang and D. Cheng and L. Han and Y. Liang, "Graph-Driven Insights: Enhancing Stock Market Prediction with Relational Temporal Dynamics," ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025. https://doi.org/10.1109/icassp49660.2025.10889271

[100] Shuvo, M.S.S. and Adib, A.T. and Emon, M.E.A. and Rafi, A. and Rahman, R.M., "How Effective Is Mamba-Augmented Transformer for Stock Market Price Forecasting?," FinTech, 2026. https://doi.org/10.3390/fintech5010015

[101] Lyu, H., "Hybrid ARIMA-LSTM Model for Stock Market Prediction: A Time Series and Deep Learning Integration Approach," Informatica (Slovenia), 2025. https://doi.org/10.31449/inf.v49i22.8510

[102] Tiwari, A. and Shieh, C.-S. and Kantipudi, M.V.V.P. and Choidhary, S., "Hybrid CNN-LSTM Integrated with Temporal Fusion Transformer for Accurate and Interpretable Stock Market Forecasting," Ingenierie des Systemes d'Information, 2025. https://doi.org/10.18280/isi.301122

[103] Sezgin, F.H. and Algorabi, Ö. and Sart, G. and Güler, M., "Hyperparameter-Optimized RNN, LSTM, and GRU Models for Airline Stock Price Prediction: A Comparative Study on THYAO and PGSUS," Symmetry, 2025. https://doi.org/10.3390/sym17111905

[104] Lawi, A. and Mesra, H. and Amir, S., "Implementation of Long Short-Term Memory and Gated Recurrent Units on grouped time-series data to predict stock prices accurately," Journal of Big Data, 2022. https://doi.org/10.1186/s40537-022-00597-0

[105] Munthe, I.R. and Rambe, B.H. and Hanum, F. and Amanda, A.T. and Hutagaol, A.S.R. and Harianto, R., "Implementation of Stacking Technique Combining Machine Learning and Deep Learning Algorithms Using SMOTE to Improve Stock Market Prediction Accuracy," Journal of Applied Data Sciences, 2024. https://doi.org/10.47738/jads.v5i4.421

[106] AlBalla, M. and Alawajy, A. and Alsalamah, S. and Almuallem, Z., "Informative Inputs Over Complex Features: Long Short-Term Memory Forecasting in the Saudi Stock Market," International Journal of Advanced Computer Science and Applications, 2025. https://doi.org/10.14569/ijacsa.2025.0161186

[107] H. M. Sami and M. Shah Jalal and M. Faisal Kabir and M. Nurul Huda, "IntPort: An Intelligent Portfolio Construction Technique Based on Financial Forecasting by Statistical Average Method," IEEE Access, 2025. https://doi.org/

10.1109/access.2025.3543741

[108] Sravankumar, B. and Krishna, G.B. and Murthy, G.V. and Kumar, D.R. and Sridevi, M. and Rao, M.V., "Integration of Explainable AI with Deep Learning for Real-Time Sentiment-Driven Stock Price Prediction," Computational Economics, 2025. https://doi.org/10.1007/s10614-025-11159-w

[109] D. Edgar and N. Ghaffari and J. Foreman and M. H. Shuvo, "IntelliTradeAI: A Tri-Signal Fusion Framework for Explainable AI-Powered Financial Market Prediction," SoutheastCon 2026, 2026. https://doi.org/10.1109/southeastcon63549.2026.11476323

[110] Duan, X., "Introducing an Innovative Approach to Mitigate Investment Risk in Financial Markets: A Case Study of Nikkei 225," International Journal of Advanced Computer Science and Applications, 2024. https://doi.org/10.14569/ijacsa.2024.0150313

[111] Ko, CR and Chang, HT, "LSTM-based sentiment analysis for stock price forecast," PEERJ COMPUTER SCIENCE, 2021. https://doi.org/10.7717/peerj-cs.408

[112] Gao, F. and Zhang, J. and Zhang, C. and Xu, S. and Ma, C., "Long Short-Term Memory Networks with Multiple Variables for Stock Market Prediction," Neural Processing Letters, 2023. https://doi.org/10.1007/s11063-022-11037-8

[113] Akşehir, Z.D. and Kiliç, E., "Multi level perspectives in stock price forecasting: ICE2DE-MDL," PeerJ Computer Science, 2024. https://doi.org/10.7717/peerj-cs.2125

[114] Shaban, W.M. and Ashraf, E. and Slama, A.E., "SMP-DL: a novel stock market prediction approach based on deep learning for effective trend forecasting," Neural Computing and Applications, 2024. https://doi.org/10.1007/s00521-023-09179-4

[115] Yuan, X. and He, C. and Xu, H. and Sun, Y., "Particle swarm optimization LSTM based stock prediction model," Proceedings - 2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science, ACCTCS 2023, 2023. https://doi.org/10.1109/acctcs58815.2023.00104

[116] Boddu, Y. and A, M., "Optimizing the accuracy of stock price prediction: accessing the prospects of LSTM with geometric mean layer normalization," Computational Statistics, 2026. https://doi.org/10.1007/s00180-026-01743-5

[117] Y. Ansari, "Multi-Cluster Graph (MCG): A Novel Clustering-Based Multi-Relation Graph Neural Networks for Stock Price Forecasting," IEEE Access, 2024. https://doi.org/10.1109/access.2024.3476159

[118] J. Eapen and D. Bein and A. Verma, "Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction," 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 2019. https://doi.org/10.1109/ccwc.2019.8666592

[119] Zouaghia, Z. and Kodia, Z. and Ben Said, L., "Smapf-hnna: a novel Stock Market Analysis and Prediction Framework using Hybrid Neural Network Architectures Across Major U.S. Indices," International Journal of Data Science and Analytics, 2025. https://doi.org/10.1007/s41060-025-00811-1

[120] Ruan, L. and Jiang, H., "Stock Price Prediction Using FinBERT-Enhanced Sentiment with SHAP Explainability and Differential Privacy," Mathematics, 2025. https://doi.org/10.3390/math13172747

[121] Chadidjah, A. and Jaya, I.G.N.M. and Kristiani, F., "The comparison stateless and stateful LSTM architectures for short-term stock price forecasting," International Journal of Data and Network Science, 2024. https://doi.org/10.5267/j.ijdns.2024.1.009

[122] Xiao, S. and Li, Q. and Gong, X. and Zhao, J. and Gu, L. and Peng, L., "Unveiling risk propagation: a lead-lag-aware framework for financial market prediction," Expert Systems with Applications, 2025. https://doi.org/10.1016/j.eswa.2025.128143

[123] Ö. E. Aşirim, "Performance of Single-Layer Transformers in Stock-Price Trend Forecasting," 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), 2025. https://doi.org/10.1109/qpain66474.2025.11171624

[124] Min, S. and Zhang, S., "Stock Market Forecasting Based on Quantum Fuzzy Support Vector Machine," Proceedings of 2025 International Conference on Digital Management and Information Technology, DMIT 2025, 2025. https://doi.org/10.1145/3736426.3736484

[125] H. Cui and Y. ZhU and F. Gu and L. Wang, "Research on stock price prediction using TextRank based text summarization technology and sentiment analysis," 2022 18th International Conference on Computational Intelligence and Security (CIS), 2022. https://doi.org/10.1109/cis58238.2022.00070

[126] L. Owen and F. Oktariani, "SENN: Stock Ensemble-based Neural Network for Stock Market Prediction using Historical Stock Data and Sentiment Analysis," 2020 International Conference on Data Science and Its Applications (ICoDSA), 2020. https://doi.org/10.1109/icodsa50139.2020.9212982

[127] T. Yao and D. Chen, "The Graph Convolutional Networks Framework for Predicting Pandemic Impact on Stock Prices," 2022 8th Annual International Conference on Network and Information Systems for Computers (ICNISC), 2022.

https://doi.org/10.1109/icnisc57059.2022.00090

[128] Sable, R. and Gupta, S. and Goel, S. and Chatterjee, P., "MAOA-DRN: Mayfly Adam Optimization Based Deep Residual Network - A Framework for Stock Market Prediction," Journal of Intelligent and Fuzzy Systems, 2025. https://doi.org/10.1177/18758967251353038

[129] Pusporani, E. and Nitasari, A.N. and Salsabila, F.N. and Indrasta, I.A. and Fariz Fadillah Mardianto, M., "UNILEVER STOCK PRICES FORECASTING WITH ENSEMBLE AVERAGING APPROACH ARIMA-GARCH AND SUPPORT VECTOR REGRESSION," Barekeng, 2026. https://doi.org/10.30598/barekengvol20iss1pp0137-0154

[130] Raj Kumar, C. and Manikandan, S., "SEMP-TA: A Novel Stock Market Prediction Approach Based on Stacking Ensemble Machine Learning for Effective Trend Analysis," IEEE Access, 2025. https://doi.org/10.1109/access.2025.3586233

[131] A. Panigrahi and A. Priya and D. Jindal and A. Sharma, "Predicting Stock Market Trends Using Deep Reinforcement Learning," 2025 4th International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), 2025. https://doi.org/10.1109/access65134.2025.11135673

[132] V. Rawat and K. Singh and Y. Sahu, "Reliance Stock Market Analysis with RBF Kernel using Support Vector Machine," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), 2023. https://doi.org/10.1109/cises58720.2023.10183635

[133] Wu, W. and Zhao, Y. and Wang, Y. and Wang, X., "Stock Price Forecasting and Rule Extraction Based on L1-Orthogonal Regularized GRU Decision Tree Interpretation Model," Communications in Computer and Information Science, 2020. https://doi.org/10.1007/978-981-15-7984-4_23

[134] T. S. Gill and S. I. Zahid, "Time Series Forecasting of KSE-100 Index Using a Hybrid ESN-LSTM Model," 2024 International Conference on Robotics and Automation in Industry (ICRAI), 2024. https://doi.org/10.1109/icrai62391.2024.10894346

[135] H. S and S. D S and V. G S, "Uncertainty-Aware Stock Price Forecasting with Gaussian Process Regression," 2025 International Conference on Intelligent Innovations in Engineering and Technology (ICIIET), 2025. https://doi.org/10.1109/iciiet65921.2025.11378930

[136] Priyam, A. and Devraj, A. and Saini, A. and Susan, S., "Time Series Forecasting for Stock Market Prediction Using Two-Layer Stacked Long Short-Term Memory Network," Lecture Notes in Networks and Systems, 2025. https://doi.org/10.1007/978-981-96-7131-1_26

[137] Singh, B. and Henge, S.K., "The Computation Trend of Fuzzy Rules for Effective Decision Support Mechanism on Basis of Supervised Learning for Multiple Periods," Lecture Notes in Networks and Systems, 2022. https://doi.org/10.1007/978-3-031-09176-6_14

[138] Xu, Z. and Xu, C. and Ren, W., "Research on stock price prediction based on Ivy optimization and TCN-BiGRU Networks," Proceedings of the 5th International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2024, 2025. https://doi.org/10.1145/3716895.3717008

[139] Zhang, Z. and Liu, Q. and Hu, Y. and Liu, H., "Multi-feature stock price prediction by LSTM networks based on VMD and TMFG," Journal of Big Data, 2025. https://doi.org/10.1186/s40537-025-01127-4

[140] J. Y. Lim and K. M. Lim and C. P. Lee, "Stacked Bidirectional Long Short-Term Memory for Stock Market Analysis," 2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), 2021. https://doi.org/10.1109/iicaiet51634.2021.9573812

[141] Duan, X., "Research on Financial Time Series Forecasting Algorithm Based on Graph Convolutional Network," Lecture Notes in Networks and Systems, 2026. https://doi.org/10.1007/978-3-032-16705-7_1

[142] A. Rahman and K. Agarwal and R. Dube and T. R. Velankanni and M. R. Babu and S. Sakthivel, "Predicting Stock Market Trends Using Deep Learning: A Comparative Study of LSTM and Transformer Models," 2025 International Conference on Technology Enabled Economic Changes (InTech), 2025. https://doi.org/10.1109/intech64186.2025.11198495

[143] V. T. X. Hanh and V. M. Huan and N. V. P. Ngoc and P. V. G. Bao and T. K. Toai, "Transformer Models for Long-Term Stock Price Forecasting and LSTM Models for Short-Term Prediction," 2025 19th International Conference on Advanced Computing and Analytics (ACOMPA), 2025. https://doi.org/10.1109/acompa69200.2025.00017

[144] K. Huang and X. Li and F. Liu and X. Yang and W. Yu, "ML-GAT:A Multilevel Graph Attention Model for Stock Prediction," IEEE Access, 2022. https://doi.org/10.1109/access.2022.3199008

[145] Y. S and C. K. J and V. R. R, "Precision Forecasting of Stock Prices: Leveraging XGBoost and Technical Indicator for Advanced Predictive Modeling," 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN), 2025. https://doi.org/10.1109/iciscn64258.2025.10934300

[146] Yu, C. and Zhang, L., "Research on Stock Index Prediction Model Based on Deep Reinforcement Learning," Frontiers in Artificial Intelligence and Applications, 2024. https://doi.org/10.3233/faia240151

[147] S. Ranasinghe, "Univariate Time Series Forecasting Under High Volatility: A Case Study of Sri Lanka Stock Price Index," 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), 2021. https://doi.org/10.1109/iciafs52090.2021.9605968

[148] C. Nagesh and T. V. N. Jayudu and N. S. Rao and E. Hariprasad and A. Naresh and C. Balaji, "Predicting Market Volatility and Risk Analysis with ESG Factors: A Hybrid Approach Using LSTM and ARIMA," 2025 9th International Conference on Inventive Systems and Control (ICISC), 2025. https://doi.org/10.1109/icisc65841.2025.11187620

[149] A. T. Haryono and R. Sarno and K. R. Sungkono, "Transformer-Gated Recurrent Unit Method for Predicting Stock Price Based on News Sentiments and Technical Indicators," IEEE Access, 2023. https://doi.org/10.1109/access.2023.3298445

[150] A. Julian and E. GaneshReddy and V. R. D. Reddy and D. V. V. Somasekhar, "Safe Trade - A Stock Recommender using Machine Learning Algorithms," 2023 International Conference on Computer Communication and Informatics (ICCCI), 2023. https://doi.org/10.1109/iccci56745.2023.10128595

[151] E. Loganathan and M. Nagenthiran and C. A. Kandasamy and N. Prakash and A. Yoganathan and M. Venkatesan, "Machine Learning vs Deep Learning: using LightGBM and LSTM to Increase the Accuracy of Stock Market Predictions," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2025. https://doi.org/10.1109/icscds65426.2025.11167010

[152] Wu, W. and Wang, Y. and Fu, J. and Yan, J. and Wang, Z. and Liu, J. and Wang, W. and Wang, X., "Preliminary Study on Interpreting Stock Price Forecasting Based on Tree Regularization of GRU," Communications in Computer and Information Science, 2019. https://doi.org/10.1007/978-981-15-0121-0_37

[153] Wu, BH and Qi, YX and Fu, GH, "Optimizing Convolutional Neural Network-Long Short-Term Memory Architecture with Additive Attention Mechanism for Stock Price Prediction," REVISTA INTERNACIONAL DE METODOS NUMERICOS PARA CALCULO Y DISENO EN INGENIERIA, 2025. https://doi.org/10.23967/j.rimni.2024.10.56790

[154] Wang, S. and Qin, J. and Rudolph, C. and Nepal, S. and Grobler, M., "R-Net: Robustness Enhanced Financial Time-Series Prediction with Differential Privacy," Proceedings of the International Joint Conference on Neural Networks, 2022. https://doi.org/10.1109/ijcnn55064.2022.9892663

[155] Albahli, S. and Nazir, T., "Opinion mining for stock trend prediction using deep learning," Multimedia Tools and Applications, 2025. https://doi.org/10.1007/s11042-024-19793-6

[156] Lavanya, M. and Gnanasekaran, P., "Prediction of Stock Price Using Machine Learning (Classification) Algorithms," Proceedings of the 2nd IEEE International Conference on Advances in Computing, Communication and Applied Informatics, AC, 2023. https://doi.org/10.1109/accai58221.2023.10199307

[157] N. Pokala and S. Hossan and J. Deng and D. Huang, "Stock Prediction based on Annual Reports using Large Language Model," 2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), 2026. https://doi.org/10.1109/acdsa67686.2026.11467931

[158] C. Gou and R. Zhao and Y. Guo, "Stock Price Prediction Based on Non-Stationary Transformers Model," 2023 9th International Conference on Computer and Communications (ICCC), 2023. https://doi.org/10.1109/iccc59590.2023.10507459

[159] N. Heryana and Nugraha and R. Mayasari, "The Role of Convolutional Neural Networks (CNNs) in Forecasting Stock Prices of Green Companies: An Analysis of Time-Step Configurations," 2024 Ninth International Conference on Informatics and Computing (ICIC), 2024. https://doi.org/10.1109/icic64337.2024.10956669

[160] M. A. Ridhawi and M. H. Ali and H. Al Osman, "Stock Market Prediction Using Node Transformer Architecture Integrated With BERT Sentiment Analysis," IEEE Access, 2026. https://doi.org/10.1109/access.2026.3691980

[161] Wang, J. and Liu, J., "Two-Stage Deep Ensemble Paradigm Based on Optimal Multi-scale Decomposition and Multi-factor Analysis for Stock Price Prediction," Cognitive Computation, 2024. https://doi.org/10.1007/s12559-023-10203-x

[162] Q. Xie, "Marketing Strategy Support System based on Temporal Fusion Transformer," 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS), 2025. https://doi.org/10.1109/icicacs65178.2025.10969006

[163] S. Nikhil and R. K. Sah and S. Kumar Parki and T. B. Tamang and S. R. D and M. T R, "Stock Market Prediction Using Genetic Algorithm Assisted LSTM-CNN Hybrid Model," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2023. https://doi.org/10.1109/icccnt56998.2023.10306948

[164] X. Chen and Q. Wang and L. Yuxin and C. Hu and C. Wang and Q. Yan, "Stock Price Forecast Based on Dueling Deep Recurrent Q-network," 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), 2023. https://doi.org/10.1109/prai59366.2023.10332127

[165] K. G. Lakshmi and S. R. Polamuri, "Temporal-Attention GAN with Multi-Scale Feature Learning for Stock Market Prediction and Volatility Modeling,"

2026 Contemporary Computing Innovations Conference (CCIC), 2026. https://doi.org/10.1109/ccic68129.2026.11486039

[166] Pan, J. and Zhuang, Y. and Fong, S., "The impact of data normalization on stock market prediction: Using SVM and technical indicators," Communications in Computer and Information Science, 2016. https://doi.org/10.1007/978-981-10-2777-2_7

[167] S. Dutta and N. Innan and A. Marchisio and S. B. Yahia and M. Shafique, "QADQN: Quantum Attention Deep Q-Network for Financial Market Prediction," 2024 IEEE International Conference on Quantum Computing and Engineering (QCE), 2024. https://doi.org/10.1109/qce60285.2024.10303

[168] Sun, WJ and Liu, ZY and Yuan, CH and Zhou, X and Pei, YT and Wei, C, "RCSAN residual enhanced channel spatial attention network for stock price forecasting," SCIENTIFIC REPORTS, 2025. https://doi.org/10.1038/s41598-025-06885-y

[169] Hartanto, S. and Gunawan, A.A.S., "Temporal Fusion Transformers for Enhanced Multivariate Time Series Forecasting of Indonesian Stock Prices," International Journal of Advanced Computer Science and Applications, 2024. https://doi.org/10.14569/ijacsa.2024.0150713

[170] K. A. M and T. Sugihartono and R. Sulaiman and B. Adiwinoto and Marini and H. Febiansyah, "XGBoost vs. LSTM: A Comparative Performance Analysis for Global Stock Market Prediction," 2025 7th International Conference on Cybernetics and Intelligent System (ICORIS), 2025. https://doi.org/10.1109/icoris67789.2025.11296020

[171] Xiao, S. and Jia, X. and Yu, C. and Li, Q., "Unifying relational and market dynamics to enhance stock trend prediction," Journal of Supercomputing, 2026. https://doi.org/10.1007/s11227-026-08482-w

[172] Sahu, P.K., "VMD-TFINet: A Temporal Feature-Aware Transformer Network for Stock Price Forecasting," Applied Artificial Intelligence, 2026. https://doi.org/10.1080/08839514.2026.2668200

[173] Trawinski, A. and Wimmer, H. and Oliver, D., "Sentiment Based LSTM for Stock Price Prediction: Congress vs General Public," ISMSIT 2022 - 6th International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings, 2022. https://doi.org/10.1109/ismsit56059.2022.9932752

[174] Bhowmick, A. and Rahman, A. and Rahman, R.M., "Performance analysis of different recurrent neural network architectures and classical statistical model for financial forecasting: A case study on dhaka stock exchange," Advances in Intelligent Systems and Computing, 2019. https://doi.org/10.1007/978-3-030-19810-7_27

[175] T. B. Pun and T. B. Shahi, "Nepal Stock Exchange Prediction Using Support Vector Regression and Neural Networks," 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC), 2018. https://doi.org/10.1109/icaecc.2018.8479456

[176] S. Singh and V. Dhir, "Novel and Efficient Information-Theoretic Weighted CAT Boost Framework for Complex Multi-Scale Stock Price Prediction," 2025 2nd International Conference on Computational Intelligence and Computing Applications (ICCICA), 2025. https://doi.org/10.1109/iccica67008.2025.11337581

[177] Ehtiwsh, E. and Alzubi, A. and Iyiola, K., "Stock Market Prediction Using Spider Feline Swarm Optimization Based Hybrid Gated Recurrent Unit and Bidirectional LSTM Model," Computational Economics, 2025. https://doi.org/10.1007/s10614-025-11130-9

[178] Y. Fang, "Stock Price Forecasting Based on Improved Support Vector Regression," 2020 7th International Conference on Information Science and Control Engineering (ICISCE), 2020. https://doi.org/10.1109/icisce50968.2020.00272

[179] Praveen, M. and Gilmary, R. and Manvizhi, N. and Vikram, M. and Sanjay, V., "Speculating Future Stock Price in Stock Market Using Feature Scaling and Predictive Models," ICDCS 2024 - 2024 7th International Conference on Devices, Circuits and Systems, 2024. https://doi.org/10.1109/icdcs59278.2024.10560998

[180] K. Singh and G. Gupta and Y. Chhabra and S. Narula and V. K. Sindhi, "Stock Market Trend Prediction Using Transformer-Based Reinforcement Learning," 2025 International Conference on Computing Technologies (ICOCT), 2025. https://doi.org/10.1109/icoct64433.2025.11118296

[181] K. K. Fung Yuen, "Towards Regression Model Selection Framework for Stock Price Forecasting," TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON), 2024. https://doi.org/10.1109/tencon61640.2024.10903078

[182] W. Chang and G. Wang and X. Chang, "Robust Stock Opening Price Forecasting Based on LSTM-Stack Ensembles," 2025 International Conference on Communication Networks and Smart Systems Engineering (ICCNSE), 2025. https://doi.org/10.1109/iccnse66404.2025.11144252

[183] S. Kundu and N. Tarasia and R. K. Barik, "QLSTM4FM: Quantum Assisted Long Short-Term Memory Framework for Financial Market Trend Forecasting," 2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT), 2024. https://doi.org/10.1109/ic-sit63503.2024.10862667

[184] A. Devi and P. Mishra and R. K. Dash and D. Mishra, "Multi-Horizon Stock Forecasting with Deep Operator Neural Networks(DeepONet)," 2025 OITS International Conference on Information Technology (OCIT), 2025. https://doi.org/

10.1109/ocit66168.2025.11400032

[185] G. W. R. I. Wijesinghe and R. M. K. T. Rathnayaka, "Stock Market Price Forecasting using ARIMA vs ANN; A Case study from CSE," 2020 2nd International Conference on Advancements in Computing (ICAC), 2020. https://doi.org/10.1109/icac51239.2020.9357288

[186] Zhang, D. and Jiang, J. and Zhang, G. and Xiong, Z. and Hu, W. and Yao, S. and Cheng, Z. and Xue, Y. and Peng, L. and Lin, J., "The stock price prediction based on a multi-scale adaptive decoding network," International Journal of Machine Learning and Cybernetics, 2026. https://doi.org/10.1007/s13042-025-02885-6

[187] Diqi, M. and Hiswati, M.E. and Nur, A.S., "StockGAN: robust stock price prediction using GAN algorithm," International Journal of Information Technology (Singapore), 2022. https://doi.org/10.1007/s41870-022-00929-6

[188] P. Singh and M. Jha and H. Patel, "Wavelet-Enhanced Deep Learning Ensemble for Accurate Stock Market Forecasting: A Case Study of Nifty 50 Index," IEEE Access, 2025. https://doi.org/10.1109/access.2025.3568634

[189] V. Kaur and E. Pinsky, "Analyzing Mutual Funds Behavior and Distinctiveness Across Sectors with Clustering and Hamming Distance: A 27-Year Study," *Journal of Risk and Financial Management*, vol. 19, no. 8, 2026. https://doi.org/10.3390/jrfm19080001

[190] T. Kundu and E. Pinsky, "Predicting Daily Stock Movements with Deep Learning Models," *Machine Learning with Applications*, vol. 22, p. 100744, Dec. 2025. https://doi.org/10.1016/j.mlwa.2025.100744

[191] M. Liang, Y. Tang, S. Puppala, and E. Pinsky, "Evaluating Simple Strategies with Mutual Funds and ETFs to Outperform the China's Shanghai Composite Index (SCI)," *Journal of Risk and Financial Management*, vol. 19, no. 4, pp. 1–31, 2026. https://doi.org/10.3390/jrfm19040001

[192] S. Pattnaik, C. Jain, and E. Pinsky, "Spatiotemporal Return Decomposition and Multi-Strategy Performance Analysis in Dow Jones Industrial Average Constituents: A 20-Year Empirical Investigation," *International Journal of Financial Studies*, vol. 14, no. 6, 2026. https://doi.org/10.3390/ijfs14060001

[193] S. Puppala, A. Malhotra, S. Sinha, and E. Pinsky, "Comparing 'Winners', 'Median' and 'Losers' Rotation Strategies with SP500 Sector ETFs," *Journal of Investing*, Feb. 2026. https://doi.org/10.3905/joi.2026.1.312

[194] S. Shah and E. Pinsky, "Perpetual Futures in Decentralized Finance: Mechanics, Economic Claims, and the Drivers of Trading Volume," *International Journal of Financial Studies*, vol. 14, no. 7, 2026. https://doi.org/10.3390/ijfs14070001

[195] G. Salotra and E. Pinsky, "Do Reinforcement Learning Agents Improve Commodity Sector Rotation? Walk-Forward Evidence from Expert Selection, Strong Benchmarks, and a Frozen-Policy Temporal Extension," *Risks*, vol. 14, no. 9, 2026. https://doi.org/10.3390/risks14090001

[196] G. Salotra and E. Pinsky, "Sector Rotation Strategies in the TSX 60: A Comprehensive Analysis of Risk-Adjusted Returns, Machine Learning Applications, and Out-of-Sample Validation (2000–2025)," *Journal of Risk and Financial Management*, vol. 19, no. 1, 2026. https://doi.org/10.3390/jrfm19010001

[197] G. Salotra, T. Katikreddy, Y. Anumolu, and E. Pinsky, "A Comparative Analysis of Overnight vs. Daytime Static and Momentum Strategies Across Sector ETFs," *Risks*, vol. 14, no. 4, pp. 1–64, Apr. 2026. https://doi.org/10.3390/risks14040001

[198] A. Vasiukevich and E. Pinsky, "Constructing Portfolios Using Stable Distributions: The Case of SP-500 Sectors Exchange-traded Funds," *Machine Learning with Applications*, vol. 10, p. 100434, Dec. 2022. https://doi.org/10.1016/j.mlwa.2022.100434

[199] Z. Wu and E. Pinsky, "On the Performance of Lagged Momentum and Reversal Strategies Across Daytime and Overnight Sessions in Bitcoin and Ethereum Cryptocurrencies," *Journal of Risk and Financial Management*, vol. 19, no. 9, Sep. 2026. https://doi.org/10.3390/jrfm19090001

[200] X. Zhang and E. Pinsky, "SP-500 vs. Nasdaq-100 Price Movement Prediction with LSTM for Different Daily Periods," *Machine Learning with Applications*, vol. 19, pp. 1–19, Jan. 2025. https://doi.org/10.1016/j.mlwa.2024.100619

[201] R. Cont, "Empirical properties of asset returns: Stylized facts and statistical issues," *Quantitative Finance*, vol. 1, no. 2, pp. 223–236, 2001. https://doi.org/10.1080/713665670

[202] S. T. Rachev, C. Menn, and F. J. Fabozzi, *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Hoboken, NJ, USA: John Wiley & Sons, 2005. https://doi.org/10.1002/9781119201991

[203] D. Lou, C. Polk, and S. Skouras, "A tug of war: Overnight versus intraday expected returns," *Journal of Financial Economics*, vol. 134, no. 1, pp. 192–213, 2019. https://doi.org/10.1016/j.jfineco.2019.03.011

[204] V. Bogousslavsky, "The cross-section of intraday and overnight returns," *The Journal of Finance*, vol. 76, no. 1, pp. 225–272, 2021. https://doi.org/10.1111/jofi.12985

[205] H. Chen, V. Singal, and R. F. Whitelaw, "Price limits, institutional trading, and market quality: Evidence from China," *Journal of Financial and Quantitative Analysis*, vol. 54, no. 5, pp. 2095–2128, 2019. https://doi.org/10.1017/s0022109018001185

[206] D. Blitz, P. van Vliet, and G. Baltussen, "The cross-section of stock returns in emerging markets: Factor investing in emerging markets," *Emerging Markets Review*, vol. 38, pp. 20–35, 2019. https://doi.org/10.1016/j.ememar.2018.11.008

[207] K. Heimbach, E. Schertenleib, and R. Wattenhofer, "DeFi perpetuals: The impact of funding rates," in *Proceedings of the 4th ACM Conference on Advances in Financial Technologies (AFT '22)*, pp. 109–121, 2022. https://doi.org/10.1145/3558535.3559775

[208] I. Ben-David, F. Franzoni, and R. Moussawi, "Do ETFs increase stock volatility?" *The Journal of Finance*, vol. 73, no. 6, pp. 2471–2535, 2018. https://doi.org/10.1111/jofi.12727

[209] F. Pattarin, S. Paterlini, and T. Minerva, "Clustering financial time series: An application to mutual funds style analysis," *Computational Statistics & Data Analysis*, vol. 47, no. 3, pp. 613–633, 2004. https://doi.org/10.1016/j.csda.2003.11.018

[210] M. López de Prado, "Building diversified portfolios that outperform out of sample," *The Journal of Portfolio Management*, vol. 42, no. 4, pp. 59–69, 2016. https://doi.org/10.3905/jpm.2016.42.4.059

[211] C. Krauss, X. A. Do, and N. Huck, "Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500," *European Journal of Operational Research*, vol. 259, no. 2, pp. 689–702, 2017. https://doi.org/10.1016/j.ejor.2016.10.031

[212] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, p. 106181, 2020. https://doi.org/10.1016/j.asoc.2020.106181

[213] Y. Chen, Z. Wei, and X. Huang, "Incorporating corporation relationship via graph convolutional neural networks for stock price prediction," *Neurocomputing*, vol. 318, pp. 281–289, 2018. https://doi.org/10.1016/j.neucom.2018.08.067

[214] R. Kim, C. H. So, M. Jeong, S. Lee, J. Kim, and J. Kang, "Hats: A hierarchical graph attention network for stock movement prediction," *arXiv preprint arXiv:1908.07999*, 2019. https://doi.org/10.48550/arXiv.1908.07999

[215] Z. Zhang, S. Zohren, and S. Roberts, "Deep reinforcement learning for trading," *The Journal of Financial Data Science*, vol. 2, no. 2, pp. 25–40, 2020. https:

//doi.org/10.3905/jfds.2020.1.030

[216] T. Théate and D. Ernst, "An application of deep reinforcement learning to algorithmic trading," *Expert Systems with Applications*, vol. 173, p. 114632, 2021. https://doi.org/10.1016/j.eswa.2021.114632

[217] M. S. Uddin and M. A. Rahman and S. M. T. Islam, "MiniBert24: A Lightweight Transformer-Based Model for Stock Market Movement Prediction," 2024 International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), 2024. https://doi.org/10.1109/access61223.2024.10543210

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FinalSheet.xlsx