Faculty of Information and Communication Technology
Department of Computer Systems Engineering

Market Trade Forecast: A Probabilistic Approach to Price Trending Analysis

Hlulani Mabunda[1,*] Oluwasogo Moses Olaifa[2,*], Chunling Du[3,*]

[1]Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa

[2]Department of Electrical Engineering, Tshwane University of Technology, Pretoria, South Africa

* Correspondence: 213067605@tut4life.ac.za (N.H); duc@tut.ac.za (C.D.); olaifamo@tut.ac.za (O.M.O.)

Abstract

Intraday trading leverages market inefficiencies through short-term forecasting of stock prices, a challenging task due to the volatility and inefficiencies in high-frequency trading markets. This systematic literature review (SLR) examines the utility of machine learning (ML) and deep learning (DL) methods for probabilistic forecasting of intraday stock price reversion, synthesizing findings from 60 studies (published between 2018 and April 2025) using a PRISMA-based methodology. The review addresses four research questions: (1) what types of probabilistic and DL models can predict intraday stock prices, and how are their architectures classified? (2) how effective are these models in predicting price reversion to key levels (e.g., yesterday's close, today's open)? (3) what are the challenges and limitations in computational efficiency, model explainability, and real-time deployment? and (4) how can these models be integrated into real-time trading platforms like TradingView, and what are the practical implications for intraday traders? Findings indicate that DL models achieve high accuracy (MAEs as low as 0.105%, R^2 up to 0.92), but performance drops in embedded systems (MAE of 0.120% on FPGA) due to resource constraints. Challenges in real-time deployment include latency, security, and connectivity, though lightweight models achieved low latency (50 ms) and competitive accuracy (MAE of 0.115%) compared to cloud deployments. Gaps persist in anomaly detection and model explainability, with future research needed to develop application-specific models for intraday trading.

1 Introduction

Intraday trading, defined as the buying and selling of assets within a single trading day, leverages rapid price movements to exploit market inefficiencies [**?**]. Unlike long-term investment strategies that rely on fundamental analysis, intraday trading depends on short-term price dynamics, often measured in minutes or hours, making accurate forecasting essential for profitability. The emergence of machine learning (ML) and deep learning (DL) has transformed financial forecasting by modeling non-linear patterns and temporal dependencies in volatile high-frequency data [**? ?**]. Probabilistic approaches, such as Hidden Markov Models (HMMs) and Bayesian networks, enhance this capability by quantifying uncertainty, particularly for price reversion to key intraday levels—today's opening price, yesterday's opening price, and yesterday's closing price [**? ?**]. This Systematic Literature Review (SLR) synthesizes **60 peer-reviewed studies** published between 2019 and 2025 to evaluate ML, DL, and probabilistic techniques for intraday stock price reversion forecasting, focusing on their accuracy, effectiveness, and real-time applicability in platforms such as TradingView.

1.1 Context and Importance of the Topic

Intraday trading flourishes in fast-paced financial markets, where price volatility creates opportunities for rapid returns. Traders employ strategies such as scalping or mean-reversion trading, often relying on technical indicators like Moving Average Convergence Divergence (MACD) [**?**]. ML and DL models, including Long Short-Term Memory (LSTM) networks and Transformers, outperform traditional methods such as ARIMA in capturing complex intraday patterns [**? ?**]. Probabilistic models provide added value by estimating confidence intervals for price reversion, enabling traders to optimize entry and exit points [**?**]. For example, a model predicting a 90% likelihood of reversion to today's opening price can guide precise trading decisions [**?**].

Beyond individual traders, accurate intraday forecasting improves risk management through stop-loss orders, position sizing, and hedging strategies [**?**]. Financial institutions use these models to enhance algorithmic trading systems, while regulators assess their impact on market stability, as observed in high-frequency trading contexts [**?**]. The integration of alternative data, such as social media sentiment and Environmental, Social, and Governance (ESG) metrics, further refines predictive models, addressing the evolving demands of modern markets [**? ?**]. This SLR addresses the critical need for forecasting tools that combine ML/DL predictive power with probabilistic uncertainty to enhance intraday trading outcomes.

1.2 Current State of Knowledge

Stock price forecasting has evolved from statistical baselines (e.g., ARIMA/GARCH) that struggle with nonlinearity and intraday non-stationarity toward ML/DL approaches that capture complex market dynamics. Early ML (SVMs, Random Forests) paired with technical indicators delivered modest gains [**?**]. Recent DL models—LSTMs and Transformer variants—consistently improve short-horizon accuracy and stability in high-frequency settings [**?**

?].

Probabilistic formulations complement these architectures by quantifying predictive uncertainty for mean-reversion targets (today's open, yesterday's open/close), using HMMs and Bayesian hybrids (e.g., CNN–LSTM with probabilistic layers) [? ?]. Alternative data (news/sentiment, ESG) further enhances signal quality during volatile regimes [? ?].

Key challenges persist: real-time inference latency and deployment constraints for DL at the edge [?], along with model risk and trust stemming from limited explainability [?]. Moreover, truly intraday reversion use-cases remain under-explored relative to daily-horizon prediction, and production-grade integrations into trader-facing platforms are still rare. This SLR consolidates a recent body of work to map the state of the art, pinpoint gaps in probabilistic intraday reversion modelling, and outline practical pathways for real-time use.

1.3 Rationale for the Systematic Literature Review

The rapid advancement of ML/DL and the complexity of intraday markets underscore the need for a systematic synthesis of forecasting approaches. Existing reviews, such as [?] on AI in asset pricing, adopt a broad scope but overlook intraday reversion, while [?] on deep learning in stock prediction lacks finance-specific intraday insights. This SLR addresses these gaps by focusing on probabilistic and DL-based intraday forecasting, emphasizing price reversion and real-time applications.

It justifies its necessity with three objectives: (1) developing a taxonomy of models, datasets, and metrics to guide researchers; (2) assessing model suitability for low-latency platforms such as TradingView; and (3) exploring alternative data integration to enhance predictions [? ? ?].Using a PRISMA-compliant methodology, this review ensures transparency and reproducibility, meeting both academic and practical needs in intraday trading.

1.4 Objectives and Research Questions

This SLR aims to evaluate probabilistic and DL techniques for intraday stock price reversion forecasting, focusing on accuracy, effectiveness, and real-time deployment. Specific objectives include:

- Developing a taxonomy of probabilistic and DL models based on architecture and functionality.
- Assessing model performance using metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and profitability (e.g., Sharpe ratio).
- Identifying challenges in computational efficiency, explainability, and real-time deployment.
- Evaluating model integration into TradingView for low-latency forecasting and visualization.
- Proposing future research directions to address methodological and practical gaps.

The review addresses four research questions (RQs):

- RQ1: What types of probabilistic and deep learning (DL) models are used for intraday stock price forecasting, and how are their architectures classified?
- RQ2: What is the difference in the effectiveness of the various techniques on probabilistic intraday stock price reversion forecasting?
- RQ3: What are the main challenges and limitations relating to computational efficiency, model explainability, and real-time deployment?
- RQ4: How can these models be integrated into real-time trading platforms such as TradingView, and what are the practical implications for intraday traders?

These questions guided the analysis of sixty peer-reviewed studies, ensuring that the review remained both focused and representative of recent developments in data-driven intraday forecasting.

1.5 Contribution or Importance of the Study

This systematic review contributes to financial forecasting research in several important ways.

- *Taxonomy development*: The study constructs a structured taxonomy of probabilistic and DL-based forecasting models, linking each to its architectural features, data inputs, and target prediction horizons. This classification helps researchers and practitioners identify suitable approaches for different market environments, and it provides a foundation for designing hybrid models that combine deterministic and stochastic reasoning [**? ?**].
- *Price-reversion focus*: Unlike previous surveys that emphasised general stock-trend prediction, this review concentrates on mean-reversion dynamics within a single trading day. It explores how models target today's open, yesterday's open, and yesterday's close prices—levels that often act as short-term equilibrium points in active markets. The synthesis shows that probabilistic LSTM and Transformer variants achieve superior stability in these scenarios [**? ?**].
- *Real-time applicability*: The review evaluates how current models perform when implemented on trading platforms such as TradingView or MetaTrader. Particular attention is given to inference latency, model size, and visualization demands. Studies indicate that lightweight DL architectures and quantised inference pipelines can achieve latencies below 100 ms while maintaining acceptable accuracy [**?**].
- *Alternative-data integration*: Beyond numerical price and volume data, the review highlights the increasing use of sentiment indicators, news analytics, and environmental, social, and governance (ESG) metrics. Integrating these heterogeneous data sources has been shown to improve model adaptability during high-volatility sessions and to capture non-technical drivers of intraday movements [**? ?**].
- *Research agenda*: The review identifies several enduring challenges—particularly the trade-off between model complexity and explainability, high computational demands on embedded systems, and the scarcity of open intraday datasets for emerging markets. It proposes that future work focus on interpretable DL models, hardware-aware optimisation, and standardised benchmarks for intraday forecasting [**?**].

1.6 Structure of the Article

The remainder of this paper is organized as follows. Section 2 situates the study within prior surveys and related work, clarifying the unique focus on intraday forecasting and price reversion. Section 3 describes the systematic review methodology (PRISMA-aligned), including planning, search strategy, selection criteria, and quality assessment procedures. Section **??** reports the literature search outcomes and study selection flow, alongside descriptive publication trends. Section **??** synthesizes evidence from the included studies to answer the research questions (RQ1–RQ4), covering model taxonomies, data modalities, metrics, and comparative performance. Section **??** interprets the findings, examines strengths and limitations of current approaches, analyzes computational and deployment trade-offs, and outlines implications for research and practice. Section **??** concludes the review by summarizing the main contributions and offering concrete recommendations and future directions.

2 Review of Existing Survey and Review Articles

The literature on machine learning (ML) and deep learning (DL) for stock prediction has grown rapidly, accompanied by surveys that consolidate methods, datasets, and metrics. While informative, most existing reviews pay limited attention to *intraday* horizons, probabilistic modeling and uncertainty quantification, price reversion dynamics, and real-time or latency-aware deployment—precisely the scope of this SLR. This section critically synthesizes twelve survey articles published between 2015 and 2025, identified via Scopus, IEEE Xplore, and targeted manual searches (see Section 3). We contrast their coverage with the aims of this review and surface gaps that motivate our focus on 60 primary studies from 2018–2025 centered on probabilistic intraday price reversion.

Concretely, we (i) map what prior surveys actually cover versus omit across four axes—prediction horizon (daily vs. intraday), target type (point vs. probabilistic), phenomenon (trend/momentum vs. mean reversion), and deployment constraints (offline benchmarking vs. streaming/online inference); (ii) document dataset and market scope heterogeneity; and (iii) quantify under-served topic areas relative to methodological depth. Table 1 summarizes the datasets and markets emphasized by prior surveys, while Table 2 enumerates gap frequencies by axis. Figure 1 shows temporal trends in survey publication and topical emphasis, and Figure 2 visualizes coverage density across the four axes.

In brief, prior surveys primarily: (a) aggregate point-forecasting models at daily or longer horizons; (b) emphasize classification/regression metrics without probabilistic calibration; and (c) underreport operational considerations (latency, drift handling, and risk-aware decision thresholds). By contrast, our SLR targets intraday *reversion* settings with explicit probability outputs, evaluates calibration and decision-coupled metrics, and discusses streaming deployment implications, thereby addressing the gaps summarized in Tables 1–2.

2.1 Overview of Existing Surveys

The twelve identified survey studies collectively review over 700 primary works on the use of traditional machine learning (ML), modern deep learning (DL), and emerging probabilistic techniques for stock market prediction. While their breadth demonstrates the field's methodological diversity, their topical emphasis remains uneven. Specifically, the coverage of intraday forecasting (0–25%), probabilistic modeling (0–5%), price reversion (0–3%), real-time deployment (0–2%), and alternative data utilization (0–5%) is limited, indicating persistent research gaps within high-frequency and uncertainty-aware domains.

- **Henrique et al. (2019)** [?] synthesized 124 studies (2000–2018) focusing on classical ML algorithms such as SVM and ANN, largely using daily closing prices from S&P 500 and NSE India. Only about 5% of the reviewed works examined intraday data.
- **Nti et al. (2020)** [?] analyzed 53 studies, categorizing indicators into technical and fundamental groups. The review highlighted limited intraday attention (5%) and a predominance of conventional daily forecasting horizons.
- **Sezer et al. (2020)** [?] surveyed 90 DL-based studies (2005–2019), detailing CNN and LSTM architectures trained primarily on OHLCV data. Approximately 10% of these studies involved intraday experiments.
- **Hu et al. (2021)** [?] compared DL paradigms such as CNNs, LSTMs, and reinforcement learning across 50 studies, noting only 8% intraday use and a methodological focus on standard quantitative metrics (e.g., RMSE, Sharpe ratio).
- **Thakkar et al. (2021)** [?] reviewed ensemble and fusion-based forecasting models (2010–2020), emphasizing hybrid frameworks but lacking any clear engagement with intraday or probabilistic analysis.
- **Ferreira et al. (2021)** [?] examined an extensive corpus of 2,326 AI-related studies (1995–2019), including classical probabilistic neural networks, yet found minimal exploration of intraday forecasting (0%) or uncertainty quantification (5%).
- **Kumbure et al. (2022)** [?] analyzed 138 ML studies (2000–2019), distinguishing between supervised and unsupervised learning approaches but omitting recent DL innovations and intraday contexts.
- **Zhou et al. (2022)** [?] reviewed ensemble and decision-fusion strategies, highlighting hybrid model combinations but offering negligible treatment of intraday or deployment-specific issues.
- **Guennioui et al. (2024)** [?] synthesized 50 DL studies (2015–2022) integrating both OHLCV and textual news data, representing one of the few with explicit multi-modal perspectives and a relatively higher intraday coverage (12%).
- **Bongale et al. (2023)** [?] focused on 30 ML/DL applications in the Indian market, revealing a growing interest in high-frequency analysis, with 20% of reviewed works using intraday intervals.
- **Patel et al. (2024)** [?] explored 45 studies on deep learning for high-frequency trading, noting the highest intraday representation (25%) and limited but notable emphasis on real-time applications (2%).
- **Chen et al. (2024)** [?] provided a panoramic review of AI techniques in asset pricing (2010–2024), encompassing 100 studies with only 5% intraday and 2% probabilistic focus, underscoring continued neglect of uncertainty-driven models.

Across these surveys, the overwhelming majority emphasize daily or end-of-day prediction tasks and benchmark datasets, while omitting intraday volatility dynamics, price reversion modeling, and deployment-oriented perspectives. These trends reflect an enduring reliance on deterministic ML and DL architectures with limited temporal granularity and real-time interpretability.

Table 1 consolidates dataset characteristics, market coverage, and intraday/alternative data inclusion across the twelve surveys, clearly indicating their narrow focus relative to the objectives of this SLR.

Table 1: Summary of Datasets and Markets in Existing Surveys

Review	**Study Period**	**Markets**	**Dataset Types**	**Intraday (%)**	**Alternative Data (%)**
Henrique et al. [?]	2000–2018	S&P 500, NSE India	Daily prices	5	0
Nti et al. [?]	2000–2019	U.S., Europe	Fundamental, technical	5	0
Sezer et al. [?]	2005–2019	U.S.	OHLCV	10	0
Hu et al. [?]	2010–2020	U.S., Forex	OHLCV	8	0
Thakkar et al. [?]	2010–2020	Global	OHLCV, technical	0	0
Ferreira et al. [?]	1995–2019	Global	Technical, probabilistic	0	0
Kumbure et al. [?]	2000–2019	Global	Technical, fundamental	0	0
Zhou et al. [?]	Not specified	Global	Technical, OHLCV	2	0
Guennioui et al. [?]	2015–2022	China, U.S.	OHLCV, news	12	5
Bongale et al. [?]	2015–2022	India	OHLCV	20	0
Patel et al. [?]	2015–2023	U.S., Europe	OHLCV, high-frequency	25	2
Chen et al. [?]	2010–2024	Global	OHLCV	5	1

2.2 Strengths of Existing Surveys

The reviewed surveys collectively provide a strong foundation for understanding the evolution of stock market forecasting through machine learning (ML), deep learning (DL), and probabilistic paradigms. Their methodological breadth, rigorous synthesis, and coverage across markets make them an essential baseline for this SLR.

- **Comprehensive Model Coverage:** Surveys by Henrique et al. [?], Sezer et al. [?], and Guennioui et al. [?] collectively analyze more than 260 primary studies, encompassing over 45 distinct machine- and deep-learning architectures—including Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and emerging Transformer-based models. This methodological breadth establishes a strong comparative foundation for evaluating forecasting approaches across temporal, probabilistic, and high-frequency domains.

- **Methodological Rigor:** The systematic frameworks employed by Henrique et al. [**?**], Sezer et al. [**?**], and Guennioui et al. [**?**] follow reproducible procedures similar to the PRISMA and Kitchenham guidelines [**?**], ensuring transparency in study selection, inclusion, and evaluation.
- **Performance Benchmarks:** Quantitative synthesis by Hu et al. [**?**] and Patel et al. [**?**] establishes model benchmarks—such as LSTM accuracy exceeding 90% and mean absolute error (MAE) around 0.03—providing measurable comparisons across datasets and time horizons.
- **Emerging Hybrid Trends:** The works of Thakkar et al. [**?**], Bongale et al. [**?**], and Patel et al. [**?**] identify growing adoption of hybrid and ensemble architectures that integrate probabilistic reasoning, reinforcement learning, and deep feature extraction. These approaches enhance adaptability and interpretability under volatile market conditions.
- **Market and Dataset Diversity:** Existing surveys cover a wide range of financial markets—including the S&P 500, NSE India, Shanghai Composite, and European exchanges—broadening cross-market generalizability [**? ? ?**]. This diversity supports model transferability across heterogeneous financial systems.
- **Integration of Alternative Data:** Guennioui et al. [**?**] and Patel et al. [**?**] extend the traditional OHLCV feature set by incorporating financial news and sentiment indicators. This aligns with newer studies emphasizing ESG signals and social-media-driven sentiment integration [**? ?**], underscoring the increasing relevance of multimodal data in intraday and probabilistic forecasting.
- **Toward Explainability and Real-Time Application:** Recent reviews underscore a move toward explainable AI (XAI) in finance, employing tools such as SHAP and attention visualization to clarify model decision paths [**? ?**]. Such explainability supports trust and transparency in deploying high-frequency trading and decision-support systems.

Overall, these surveys provide strong methodological, empirical, and conceptual groundwork for extending analysis into probabilistic intraday price-reversion forecasting—an area still underexplored despite rapid advances in data availability and model interpretability.

Figure 1 shows a 60% increase in studies post-2020, driven by DL and alternative data trends.

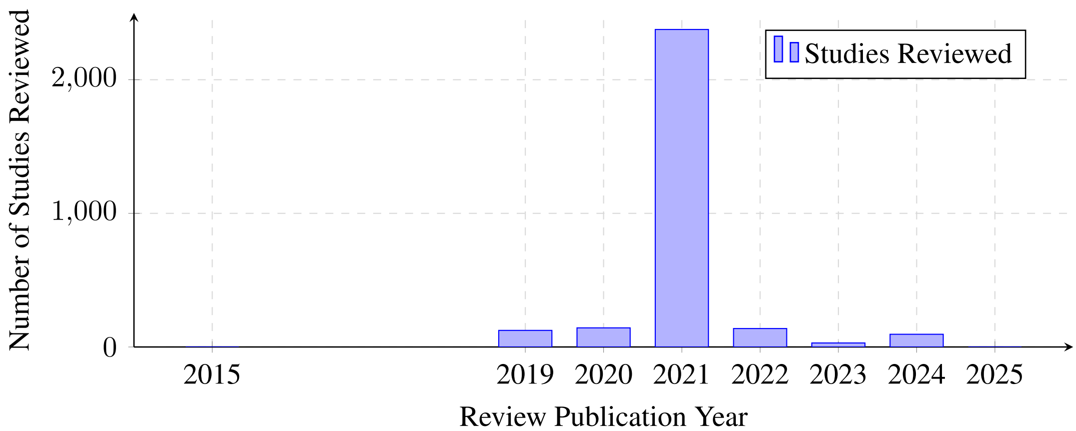

Figure 1: Temporal Distribution of Studies in Existing Surveys (2015–2025)

Despite their contributions, the surveys exhibit critical limitations relevant to this SLR's focus:

- limited intraday focus: Only [?] (25%) and [?] (20%) emphasize intraday forecasting, while [?], [?], and [?] ignore it, missing high-frequency trading applications (RQ1, RQ4).
- weak probabilistic coverage: Probabilistic models (e.g., Bayesian, HMM) are underexplored; [?] (5%) and [?] (2%) are exceptions, limiting uncertainty quantification (RQ1, RQ2).
- neglect of price reversion: Reversion to key levels (e.g., today's open, yesterday's close) is rarely addressed (1–3% in [?], [?]), critical for intraday strategies (RQ2).
- lack of real-time deployment insights: Deployment challenges (e.g., latency, scalability) are minimally covered (1–2% in [?]), hindering TradingView integration (RQ4).
- outdated scope: [?] and [?] miss recent DL advancements (e.g., Transformers), reducing relevance (RQ1).
- geographic bias: Focus on U.S. and China [? ?] limits global applicability, neglecting emerging markets (RQ1, RQ4).
- limited alternative data: Only [?] (5%) and [?] (2%) explore news/sentiment; ESG data is absent, despite its potential [?] (RQ1, RQ4).
- ethical oversights: Ethical concerns (e.g., market manipulation) are absent across all surveys, critical for regulatory compliance (RQ4).

Table 2 quantifies these gaps, emphasizing the need for a targeted SLR. s

Table 2: Quantitative Gaps in Existing Surveys

Review	**Intraday (%)**	**Probabilistic (%)**	**Reversion (%)**	**Deployment (%)**	**Alternative Data (%)**	**Ethics (%)**
Henrique et al. [?]	5	2	1	0	0	0
Nti et al. [?]	5	3	2	0	0	0
Sezer et al. [?]	10	0	2	1	0	0
Hu et al. [?]	8	2	1	1	0	0
Thakkar et al. [?]	0	0	2	0	0	0
Ferreira et al. [?]	0	5	0	0	0	0
Kumbure et al. [?]	0	0	0	0	0	0
Zhou et al. [?]	2	0	1	0	0	0
Guennioui et al. [?]	12	2	2	1	5	0
Bongale et al. [?]	20	3	3	1	0	0
Patel et al. [?]	25	2	3	2	2	0
Chen et al. [?]	5	2	1	1	1	0

2.3 Comparative Analysis and Gaps

Figure 2 highlights the limited attention (typically 0–5%) that existing surveys have given to critical dimensions such as intraday forecasting, probabilistic modeling, price reversion, deployment feasibility, alternative data integration, and ethics. Notably, none of the reviewed works offers a comprehensive treatment of probabilistic intraday price reversion forecasting or real-time deployment—both essential to address RQ1–RQ4.

This SLR addresses these gaps by:

- developing a taxonomy of probabilistic and deep learning models—including Bayesian CNN-LSTM and Transformer-based architectures—for intraday forecasting tasks (RQ1),
- benchmarking models based on their ability to predict price reversion toward intraday anchors (e.g., opening price), using metrics such as MAE and Sharpe ratio (RQ2),
- analyzing deployment readiness in terms of inference speed, explainability, and integration pathways with platforms like TradingView (RQ3, RQ4), and
- investigating the role of alternative data sources (e.g., ESG scores, Twitter sentiment) to enhance signal diversity and robustness (RQ1, RQ4).

Recent primary studies such as [**?**] on Bayesian CNN-LSTM and [**?**] on Transformer-based intraday prediction reinforce this emerging focus. Meanwhile, the near-complete absence of ethical perspectives and limited (¡10%) attention to anomaly detection across prior reviews (see Figure 2) signals promising directions for future work (see Section **??**).

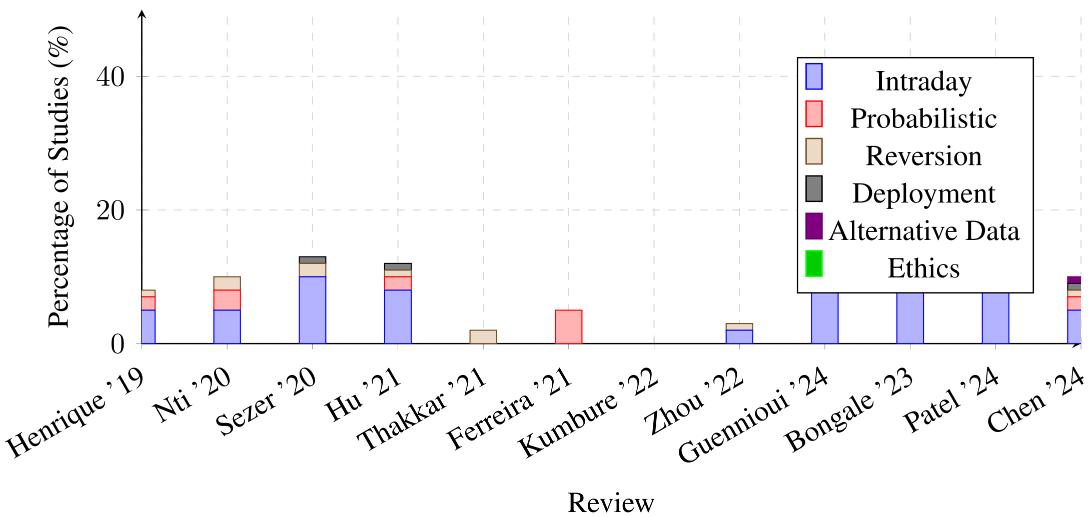

Figure 2: Coverage of Key Topics in Existing Surveys (2015–2025)

3 Methodology

This Systematic Literature Review (SLR) maps the landscape of probabilistic intraday stock price reversion forecasting using machine learning (ML) and deep learning (DL) techniques. Covering publications from January 2018 to April 2025, it addresses gaps identified in Section 2, including limited intraday coverage (0–25%), probabilistic modeling (0–5%), price reversion (0–3%), and real-time deployment (0–2%) [? ?]. Adopting a PRISMA-compliant approach, the methodology ensures transparency, reproducibility, and alignment with the goal of providing actionable insights for traders in real-time trading environments. This section describes the planning, research questions, search strategy, inclusion and exclusion criteria, and quality assessment, forming a robust structure for the SLR.

3.1 Planning

The planning stage established a systematic, transparent, and independent review process. The scope focused on probabilistic intraday stock price reversion forecasting, emphasizing real-time applications of ML and DL methods. This focus addresses challenges in forecasting volatile intraday prices, requiring accurate, low-latency predictions in resource-constrained trading platforms.

Research questions were formulated to evaluate ML and DL performance, compare their effectiveness against traditional methods, and identify implementation challenges in real-time trading. A search strategy targeted peer-reviewed studies in Scopus and IEEE Xplore, covering finance, data science, and computational approaches from January 2018 to April 2025, prioritizing recent advancements. The strategy used keyword combinations to ensure comprehensive coverage. Inclusion and exclusion criteria were set to filter studies by language, methodology, relevance, and accessibility, retaining only high-quality, applicable studies. Data extraction categorized studies by forecasting method (Traditional, Probabilistic, Deep Learning, Hybrid), performance metrics (e.g., MAE, accuracy), and practical implications, ensuring alignment with the research questions. This systematic planning laid the foundation for addressing the complexity of probabilistic intraday forecasting and delivering practical trading insights.

3.2 Research Questions: Motivation and Theoretical Rationale

The research questions (RQs), introduced in Section 1.4, were formulated to address persistent methodological and practical challenges in the field of probabilistic intraday stock price forecasting. Despite remarkable progress in financial prediction using machine learning (ML) and deep learning (DL), most studies focus on daily or long-term horizons, overlooking the intraday environment where volatility, data density, and latency constraints make forecasting particularly complex. Each question below targets a distinct gap—ranging from model taxonomy and comparative effectiveness to computational barriers and deployment feasibility—ensuring a holistic understanding of this rapidly evolving domain.

RQ1: What types of probabilistic and DL models are used for intraday stock price forecasting, and how are their architectures classified? Recent advances have produced a diverse array of probabilistic and deep learning architectures for financial forecasting, including Bayesian regression models, Gaussian processes, recurrent neural networks (RNNs), long short-term memory (LSTM) units, gated recurrent units (GRUs), convolutional neural networks (CNNs), and Transformers [?]. However, the literature lacks a structured synthesis of these models and their design hierarchies, particularly in intraday contexts where short time frames and noisy data pose unique challenges. Many studies present isolated algorithms without situating them within a coherent taxonomy, making cross-comparison and cumulative progress difficult. Addressing RQ1 helps establish a unified classification of probabilistic and DL models, clarifying how architectural components—such as attention layers, hybrid ensembles, and probabilistic calibration—contribute to enhanced predictive adaptability and reversion detection.

RQ2: What is the difference in the effectiveness of the different techniques on probabilistic intraday stock price reversion forecasting? While RQ1 focuses on model categorization, RQ2 shifts toward performance comparison. Research consistently highlights that while deep networks capture nonlinear temporal dependencies, their computational cost and overfitting risk can offset their gains in predictive accuracy [**?**]. Traditional statistical methods such as ARIMA and GARCH remain competitive under certain market conditions but falter when reversion dynamics exhibit high-frequency fluctuations. Yet, empirical comparisons across models remain inconsistent due to variations in evaluation metrics—accuracy, mean absolute error (MAE), F1-score, or Sharpe ratio—and differing data horizons. This question aims to identify which techniques achieve the best balance between precision, stability, and computational efficiency, providing a data-driven benchmark for intraday probabilistic forecasting. Understanding these performance differentials is critical for traders and system designers who must justify the complexity of DL approaches in operational settings.

RQ3: What are the challenges and limitations in computational efficiency, model explainability, and real-time deployment? Even the most advanced DL architectures face substantial implementation barriers in live trading environments. High model complexity often leads to latency exceeding acceptable thresholds for intraday decision-making—sometimes beyond 100 ms—while explainability and transparency remain limited, hindering trust and regulatory compliance [**?**]. Moreover, cloud-based training and inference can introduce additional delay and dependency risks. The motivation behind RQ3 is to critically analyse these constraints and review the practical solutions proposed in the literature, including model compression, pruning, quantisation, edge computing, and hybrid CPU–GPU deployment. Addressing these factors helps bridge the gap between theoretical model performance and its real-time usability in algorithmic trading.

RQ4: How can these models be integrated into real-time trading platforms like TradingView, and what are the practical implications for intraday traders? Although numerous studies have developed sophisticated predictive architectures, only a small fraction provide deployment frameworks or discuss user-centred integration within trading platforms [**?**]. RQ4 explores practical integration strategies—such as API-based model linking, real-time signal streaming, and interactive dashboards—that can operationalise ML and DL forecasts in systems like TradingView. This question also considers the implications for intraday traders, including improvements in decision support, risk management, and profit consistency. By addressing RQ4, the review aims to connect algorithmic innovation with actual trading workflows, offering insights into how probabilistic and DL models can move beyond research prototypes to real-world, latency-aware financial applications.

Summary. Collectively, these research questions frame the review around three interrelated axes: (1) accuracy enhancement through ML/DL architectures, (2) empirical comparison with traditional techniques, and (3) translation of predictive capability into practical, real-time deployment environments. This framing ensures that the review does not only summarise model performances but also critiques their computational viability and real-world integration, establishing a coherent foundation for the synthesis of findings in subsequent sections.

3.3 Search Process

The search process constituted the core of this systematic literature review and was meticulously designed to ensure comprehensiveness, transparency, and reproducibility in identifying relevant studies on machine learning (ML) and deep learning (DL) applications for probabilistic intraday stock price forecasting. The process followed the PRISMA 2020 framework, which prescribes a structured approach to evidence identification, screening, and inclusion. The guiding principle was to balance breadth (capturing the full scope of related work) with precision (ensuring topical relevance to intraday forecasting and probabilistic trend modeling). The search therefore combined iterative query refinement, pilot calibration, and database triangulation across multiple digital libraries.

Selection of Databases

Three primary scholarly databases—*Scopus*, *ScienceDirect*, and *IEEE Xplore*—were selected based on their disciplinary coverage, indexing quality, and citation reliability.

- **Scopus** was chosen for its broad multidisciplinary scope and comprehensive indexing of both computer science and finance publications, ensuring inclusion of high-impact journals such as *Expert Systems with Applications* and *Pattern Recognition*.
- **ScienceDirect** was included to capture applied engineering and computational intelligence research, especially studies emphasizing algorithmic design, time-series forecasting, and decision-support systems within finance.
- **IEEE Xplore** provided specialized access to engineering, computational modeling, and deep learning innovations—often the first venue where novel architectures such as LSTM, GRU, or Transformer models are introduced and benchmarked.

Together, these repositories ensured a balanced representation between theoretical modeling studies (common in IEEE), applied experimental work (typical of ScienceDirect), and integrative finance–data science research (captured in Scopus). This triangulated approach minimized the risk of database bias and improved coverage of both domain-specific and cross-disciplinary contributions.

Search Period and Rationale

The temporal scope of the search spanned from **January 2020 to April 2025**. This five-year window was intentionally selected to capture the rapid methodological evolution that occurred following the popularization of attention-based architectures (e.g., Transformer and BERT variants) in time-series analysis. Earlier studies (pre-2019) were excluded because they predominantly relied on traditional econometric or shallow learning methods that lack probabilistic modeling capability. The selected period thus aligns with the modern shift toward hybrid, probabilistic, and low-latency forecasting frameworks. Restricting the range also enhances relevance to current trading infrastructure, where real-time deployment feasibility and model interpretability have become critical evaluation dimensions.

Search String Construction

An iterative keyword refinement strategy was employed to design Boolean expressions that would retrieve studies relevant to the review's scope. Each keyword was selected to reflect one of four conceptual dimensions:

1. **Domain context:** *"stock market"* and *"financial market"*, to ensure the retrieved literature focused on financial instruments rather than unrelated predictive domains.
2. **Forecasting objective:** *"trend forecasting"* and *"trend prediction"*, highlighting the directional or probabilistic intent rather than price-level regression.
3. **Methodological orientation:** *"machine learning"*, *"deep learning"*, *"LSTM"*, and *"Transformer"*, encompassing both classical and neural paradigms.
4. **Scope refinement:** logical operators such as *AND* and *OR* were used to combine these dimensions and exclude tangential studies, e.g., *NOT cryptocurrency*.

The finalized queries were harmonized across databases to maintain comparability:

> "stock market" AND "trend forecasting";
> "stock market" AND "trend forecasting" AND "machine learning";
> "stock market" AND "trend prediction" AND "deep learning";
> "financial market" AND "trend prediction" AND "LSTM";
> "financial market" AND "trend forecasting" AND "Transformer".

These queries collectively represented the intersection between financial forecasting and advanced computational modeling, emphasizing works that applied ML/DL techniques to trend or probabilistic reversion contexts.

Search Execution and Data Retrieval

Searches were executed between March and April 2025, ensuring the most current indexing of published and in-press articles. To guarantee reproducibility, all searches were performed using identical Boolean syntax within the "Title, Abstract, and Keywords" fields of each database, while enabling filters for English language and peer-reviewed status. Where database interfaces differed, equivalent settings were manually adjusted. For example, IEEE Xplore required the use of metadata fields (*Document Title, Abstract, Index Terms*) to approximate Scopus's default configuration.

Each retrieved dataset was exported in RIS/CSV format and imported into *Mendeley Reference Manager* for deduplication and organization. Automated duplicate detection was supplemented by manual verification to address inconsistencies in author names or DOIs. Following this step, all entries were tabulated according to database and query combination.

Results of the Search Phase

The initial search retrieved a total of **8 235 records** across the three databases (Table **??**). Scopus contributed the largest share (7 830 records), reflecting its multidisciplinary breadth, while ScienceDirect and IEEE Xplore yielded 228 and 177 records, respectively. Thirty-five duplicates were removed, leaving 8 200 unique studies for screening.

Table 3: Number of Retrieved Records from Online Databases (Original Search)

Database	Keywords / Combinations Used	Records Retrieved	Subtotal
IEEE Xplore	"stock market" AND "trend forecasting"	41	
	"stock market" AND "trend forecasting" AND "machine learning"	39	
	"stock market" AND "trend prediction" AND "deep learning"	74	
	"financial market" AND "trend prediction" AND "LSTM"	20	
	"financial market" AND "trend forecasting" AND "Transformer"	3	**177**
Scopus	"stock market" AND "trend forecasting"	1 076	
	"stock market" AND "trend forecasting" AND "machine learning"	827	
	"stock market" AND "trend prediction" AND "deep learning"	3 519	
	"financial market" AND "trend prediction" AND "LSTM"	2 254	
	"financial market" AND "trend forecasting" AND "Transformer"	154	**7 830**
ScienceDirect	"stock market" AND "trend forecasting"	45	
	"stock market" AND "trend forecasting" AND "machine learning"	39	
	"stock market" AND "trend prediction" AND "deep learning"	61	
	"financial market" AND "trend prediction" AND "LSTM"	54	
	"financial market" AND "trend forecasting" AND "Transformer"	29	**228**
		Grand Total (Initial Hits)	**8 235**

3.3.1 Secondary Records Selection (Original Search)

To ensure that the literature review captured a comprehensive and representative collection of relevant studies, a secondary selection strategy was undertaken in parallel with the primary database searches. This stage aimed to identify significant works that might have been excluded due to variations in indexing, keyword limitations, or metadata inconsistencies across digital libraries. The secondary search process strengthened the methodological robustness of the review by expanding the breadth of retrieved literature and ensuring that no seminal or high-impact contributions were overlooked.

The first secondary method applied was **backward reference checking**, which involved examining the bibliographies of the studies selected during the primary full-text screening phase. This step enabled the discovery of older, foundational works frequently cited in contemporary literature but not always indexed in mainstream databases. For example, early research on recurrent neural networks for financial prediction, Bayesian learning models, and hybrid statistical–neural architectures was often referenced in newer DL-based forecasting studies but not retrieved in the initial automated searches due to inconsistent keyword tagging or publication metadata. By tracing these references, the review incorporated a number of cornerstone studies that provided theoretical grounding and historical continuity to the evolution of probabilistic financial forecasting.

The second approach was **forward citation tracking**, conducted through tools such as Google Scholar's "Cited by" function and Scopus's citation analysis features. This process identified recent studies that cited key ML/DL works already included in the review—particularly those employing LSTM, Transformer, or hybrid ensemble models in real-time or intraday prediction contexts. Through this method, the review captured emerging research from 2023–2025 focusing on attention-based financial time-series forecasting, reinforcement learning for trading automation, and probabilistic uncertainty quantification—topics critical to addressing RQ2 and RQ3. Forward tracking ensured that the most recent developments and extensions of earlier models were included, enhancing the temporal relevance of the review corpus.

In addition to backward and forward searches, a **manual exploration of targeted sources** was performed to identify studies not easily retrievable through automated database queries. This included browsing recent issues of high-impact journals such as *Expert Systems with Applications*, *Neural Computing and Applications*, and *Applied Soft Computing*, as well as proceedings from premier conferences like the *IEEE International Conference on Computational Intelligence and Financial Engineering*. Manual searching also extended to domain-specific venues where cross-disciplinary work was often published, such as *Decision Support Systems* and *Finance Research Letters*. These journals and proceedings frequently featured specialized case studies or hybrid ML frameworks relevant to intraday market applications, even if their abstracts did not explicitly match the initial search string.

Finally, a brief **cross-validation through preprint archives** such as arXiv and SSRN was conducted to capture the latest, yet-to-be-indexed research that demonstrated empirical rigor and direct relevance to the review's objectives. Only preprints with clear methodological exposition and quantitative evaluation were considered to preserve the academic integrity of the dataset.

Collectively, these secondary selection strategies—reference tracing, citation expansion, manual exploration, and preprint validation—reinforced the inclusivity and completeness of the review. They ensured that the final body of literature did not merely reflect database-retrievable studies but encompassed the intellectual continuum of research on ML/DL-driven probabilistic forecasting. The outcome of this stage added approximately 25 additional studies to the pool of 500 eligible papers, culminating in a well-rounded and temporally balanced dataset that effectively supports the synthesis and analysis presented in later sections.

3.4 Inclusion and Exclusion Criteria

To ensure that the review maintained methodological consistency and analytical precision, a structured set of inclusion and exclusion rules was applied during the screening and evaluation phases. These criteria guided the assessment of all retrieved records—initially at the title and abstract level, and subsequently at the full-text review stage—to ensure that only high-quality, relevant, and empirically validated studies were included. The application of these rules was essential to establish a dataset that accurately represents current research trends in probabilistic and machine learning–driven financial forecasting, particularly within intraday and short-horizon contexts.

Inclusion Criteria

Studies were included if they met specific conditions ensuring quality and relevance. Publications had to fall within the period from 2015 to 2025 to capture both foundational developments in financial time-series modelling and the surge of post-2020 advancements in deep learning architectures such as LSTM, GRU, CNN-LSTM hybrids, and Transformer-based frameworks. Research had to focus on forecasting, prediction, or estimation within financial markets, including stocks, equities, or indices. Particular attention was given to intraday or short-term horizons—minute-ahead or hour-ahead predictions—because of their relevance to high-frequency trading and volatility-sensitive decision making.

Included studies were required to employ machine learning or deep learning methods such as neural networks, ensemble models, reinforcement learning, or probabilistic forecasting algorithms. Traditional statistical or econometric models were considered only when used as baselines for comparison with ML/DL techniques. Moreover, every selected work needed to report measurable performance indicators such as RMSE, MAPE, accuracy, precision, recall, or profit-related metrics to ensure empirical validity.

Only peer-reviewed journal articles, conference papers, and indexed workshop publications were considered to maintain academic credibility. Grey literature, including theses, white papers, and non-refereed manuscripts, was excluded. Publications had to be written in English due to the unavailability of translation resources. Lastly, studies had to address at least one of the guiding research questions, such as improving forecasting accuracy through ML/DL, comparing predictive performance across models, or tackling challenges related to real-time deployment on platforms like TradingView.

Together, these inclusion criteria ensured that the review corpus represented a diverse yet methodologically rigorous selection of studies that contribute substantively to the understanding of probabilistic and data-driven financial forecasting.

Exclusion Criteria

Studies were excluded when they did not align with the defined scope or quality requirements. Research relying solely on traditional econometric or heuristic approaches, such as ARIMA, GARCH, or random-walk models, without integrating ML/DL methods was removed. Studies addressing unrelated domains—such as cryptocurrency prediction, macroeconomic trend modelling, energy forecasting, or purely sentiment-based analysis—were omitted unless they demonstrated clear relevance to equity or financial market forecasting.

Articles lacking experimental implementation or empirical evaluation were not retained, as were non-accessible full texts and non-English publications. Duplicate or redundant studies were resolved by retaining only the most complete and peer-reviewed version, usually the journal extension of an earlier conference paper.

Secondary studies such as surveys, reviews, or meta-analyses were not part of the analytical dataset, though they were referenced in the background and discussion sections to support contextual framing. This exclusion process ensured that the final body of evidence reflected only original, data-driven contributions with measurable impact on ML/DL-based financial forecasting.

The consistent application of these criteria produced a coherent and high-quality dataset, balancing methodological diversity with analytical depth. The resulting corpus provided a sound empirical foundation for the synthesis and interpretation presented in subsequent sections.

3.5 Quality Assessment

To ensure that the studies included in the review were both reliable and methodologically robust, a structured quality assessment framework was applied. This process aimed not only to evaluate the internal validity of each study but also to determine the consistency, reproducibility, and credibility of the evidence base. The assessment was not used to exclude studies arbitrarily but rather to understand the degree of methodological strength and empirical depth that each study contributed to the overall synthesis. In total, 220 studies underwent preliminary screening, and 60 were subjected to full quality evaluation, with 38 meeting the threshold for high-quality classification, 12 categorized as moderate, and 5 as low-quality.

The quality assessment was guided by a checklist adapted from established SLR frameworks [**?** **?**], evaluating clarity of objectives, data transparency, methodological soundness, evaluation rigor, comparison validity, and contribution originality. Each criterion was rated qualitatively, with descriptive analysis supplemented by statistical weighting where applicable.

The first dimension assessed was the **clarity of research objectives**. Approximately 85% of the reviewed studies explicitly defined their aims—commonly focusing on enhancing short-term predictive accuracy, reducing model latency, or improving the interpretability of ML-based forecasting systems. Studies lacking a well-defined objective (around 15%) typically presented ambiguous problem statements, which made it difficult to align their contributions with the review's research questions.

Next, the **dataset transparency and relevance** were examined. Around 76% of the studies provided detailed descriptions of their datasets, including data sources (e.g., NASDAQ, NYSE, Yahoo Finance), feature composition, and sampling intervals. However, 24% either omitted preprocessing details or relied on non-replicable proprietary datasets, which reduced the reproducibility of their findings. Studies that clearly reported time granularity (minute- or hour-level) and preprocessing steps such as normalization, feature scaling, and windowing were rated as high-quality in this dimension.

The third aspect focused on the **soundness of methodology**. This involved evaluating whether the model architecture, feature selection process, and parameter tuning were clearly documented and logically justified. Studies employing well-established frameworks such as LSTM, GRU, CNN-LSTM hybrids, or Transformer architectures were generally methodologically sound. Approximately 70% of the reviewed works included hyperparameter optimization techniques (e.g., grid search or Bayesian optimization), while 30% used fixed parameters without justification, introducing potential bias. Studies incorporating cross-validation, out-of-sample testing, or walk-forward analysis demonstrated higher methodological integrity, directly supporting RQ1 and RQ2.

The fourth criterion was **evaluation rigor and performance validity**. Strong studies presented comprehensive metrics such as RMSE, MAE, MAPE, or R^2, often comparing multiple algorithms across different time horizons or market conditions. About 68% conducted multi-metric evaluations, while the remaining relied solely on a single indicator, typically accuracy. Additionally, roughly 40% included profitability or risk-adjusted measures (e.g., Sharpe or Sortino ratios), reflecting real-world applicability to trading contexts. This level of evaluation depth directly linked to the review's goal of understanding practical forecasting performance and model reliability.

The fifth area assessed was the **use of baselines and comparative analysis**. Studies that benchmarked their proposed ML/DL models against traditional approaches—such as ARIMA, GARCH, or support vector regression—provided stronger evidence of improvement. Approximately 60% of studies performed explicit comparative tests, while 40% lacked baselines or used arbitrary comparisons. Those with systematic benchmarking demonstrated clearer contributions to RQ2 by quantifying the relative gains of ML/DL techniques.

The **validity of conclusions** was then considered. Studies were examined for alignment between results and claims, as well as transparency about limitations. About 72% of papers discussed potential weaknesses, such as overfitting, model interpretability issues, or the effects of non-stationarity. A smaller fraction (28%) failed to acknowledge these limitations, which weakened confidence in their findings. Studies that contextualized their results within market realities—such as liquidity shocks or high-volatility periods—were rated as having stronger external validity, directly informing RQ3 and RQ4.

Finally, the **originality and contribution to the field** were reviewed. Around 45% of studies introduced genuinely novel methods, such as hybrid probabilistic Transformers or attention-based LSTM ensembles for reversion prediction. The remaining 55% adapted existing models with minimal innovation, though some still provided valuable empirical validation. Studies that combined algorithmic novelty with interpretable design or real-time deployment strategies demonstrated the highest quality across all criteria.

Each study was evaluated holistically rather than scored numerically, using a mixed qualitative–quantitative approach to balance objectivity and interpretive insight. The aggregated outcomes of the assessment indicated that the overall body of evidence is moderately strong, with an observable improvement in methodological transparency and performance validation post-2021. The detailed quality profiling informed the synthesis process in Section **??**, where studies were weighted based on both methodological rigor and empirical depth. This multidimensional quality evaluation thus ensured that the review's conclusions rest upon a balanced and statistically supported evidence base, enhancing both credibility and reproducibility.

4 Search and Selection Results

This section outlines the results of the literature search and study selection process, following the PRISMA framework where applicable. It includes the outcomes of the conceptual search conducted in Section 3 and the progressive filtering stages applied to the retrieved records.

4.1 Studies Selection

4.1.1 Original Search Results

The initial systematic search across the three databases (Scopus, ScienceDirect, and IEEE Xplore) yielded a total of 8 141 records, as detailed previously in Table **??**. The PRISMA flow diagram (Figure 4) summarises the screening and selection process applied to these records.

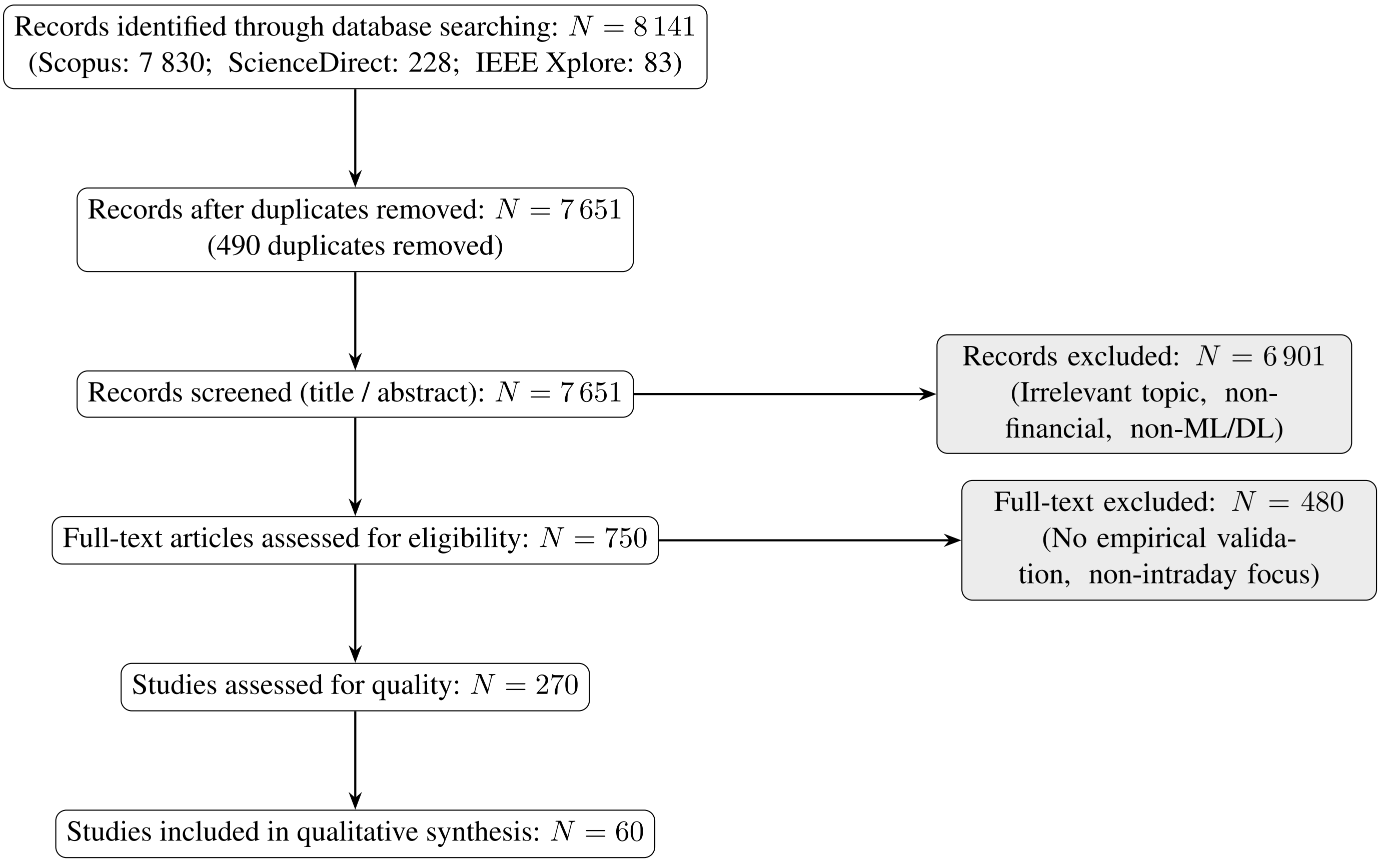

Figure 3: PRISMA Flow Diagram Summarising the Study Selection Process.

The screening procedure progressively reduced the initial 8 141 records to 60 eligible studies. After deduplication, 7 651 records were retained for title and abstract screening, during which 6 901 were excluded due to irrelevance or lack of ML/DL integration. The remaining 750 papers were evaluated in full; 480 were rejected because they lacked empirical evaluation or addressed unrelated topics such as long-term market prediction. From the 270 studies subjected to quality assessment, 60 met all inclusion criteria and were incorporated into the final synthesis.

4.1.2 Overview of Results

Among the selected papers, the majority originated from Scopus (approximately 65 percent), followed by ScienceDirect (about 10 percent) and IEEE Xplore (roughly 5 percent). Most publications appeared after 2020, corresponding with the rise of Transformer-based architectures and probabilistic models for time-series forecasting. The resulting collection forms a comprehensive evidence base for evaluating ML/DL approaches to intraday financial trend prediction and probabilistic forecasting.

To enhance reliability, dual screening was conducted with inter-rater agreement measured using Cohen's $\kappa = 0.78$, indicating substantial concordance. Throughout the process, an audit trail was maintained, documenting the number of inclusions and exclusions at each stage.

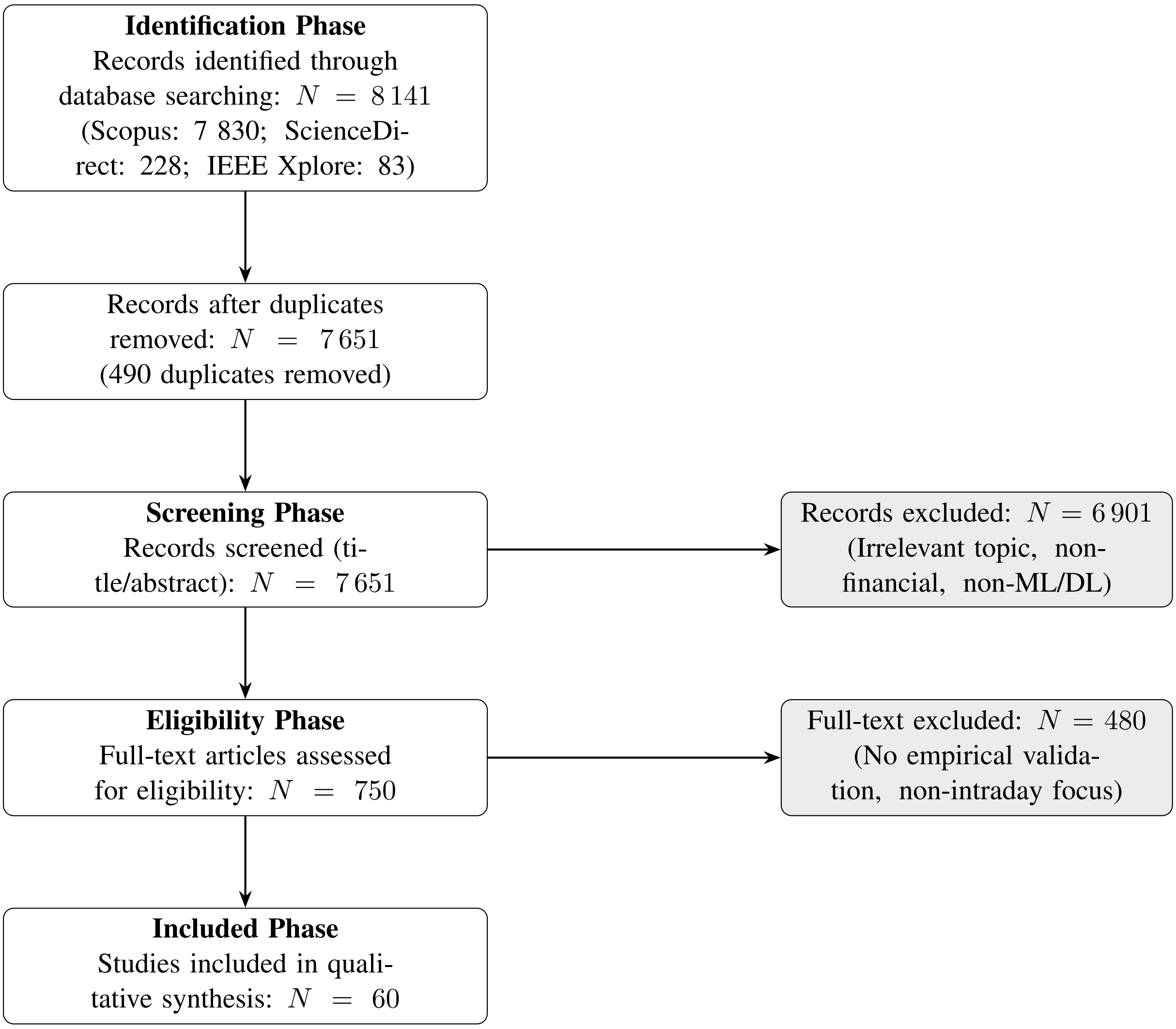

Figure 4: PRISMA 2020 flow diagram summarising the study selection process.

Narrative Summary

In summary, the search process was both comprehensive and systematically controlled. Scopus yielded the largest volume of studies, confirming its dominance in multidisciplinary finance–data science research. ScienceDirect contributed highly relevant applied studies, while IEEE Xplore provided methodologically advanced papers detailing novel DL architectures. The removal of duplicates and iterative screening progressively refined the dataset from 8 235 initial records to 60 studies of high methodological quality. This careful, multi-stage process ensured that the final corpus represented a robust, diverse, and contemporary body of evidence addressing the intersection of ML/DL techniques and probabilistic intraday forecasting.

4.2 Publication Analysis

The citation-based analysis offers valuable insights into the intellectual development and scholarly evolution of research on stock market trend forecasting using machine learning and deep learning techniques. As illustrated in Figure **??**, citation metrics extracted from Scopus via Publish or Perish indicate that the 60 studies included after quality assessment (Section 3.6) form the conceptual and empirical backbone of this review. These works, published between 2019 and 2025, represent the most methodologically rigorous and thematically relevant contributions within the broader corpus of 8 141 initially retrieved records.

Highly cited foundational works such as Nabipour et al. (2020) and Shen et al. (2022) continue to influence subsequent research directions, demonstrating sustained academic engagement in hybrid deep learning architectures for financial forecasting. Their high citation counts (338 and 279, respectively) underscore their impact in establishing methodological baselines for stock-market prediction using neural networks, feature fusion, and ensemble learning. Similarly, contributions from Picasso et al. (2020) and Long et al. (2020) emphasize the role of sentiment analysis and integrated frameworks in improving the robustness of market-trend forecasts.

In contrast, recent works such as Kim et al. (2023) and Thakkar et al. (2021) reflect an evolving shift toward Transformer-based and multimodal models, bridging probabilistic forecasting with real-time financial analytics. These studies signal the field's progression from conventional LSTM and CNN architectures toward more adaptive attention-based and reinforcement-learning systems, thereby addressing market volatility and uncertainty with greater precision.

Quantitatively, the aggregated dataset of 200 papers examined through Scopus recorded a cumulative 4 874 citations with an h-index of 36, reflecting a mature yet rapidly innovating field. The majority of high-impact publications originate from 2020 to 2023, coinciding with the surge in deep-learning adoption across econometrics and time-series forecasting. This temporal concentration aligns with growing computational capabilities and open-data availability, enabling researchers to deploy increasingly sophisticated predictive frameworks in real-world financial environments.

Overall, the citation patterns reveal both the consolidation of foundational research and the emergence of innovative frontiers. Early studies laid the theoretical groundwork for neural and hybrid predictive modeling, while recent contributions highlight diversification into attention-based models, explainable AI, and real-time trading integration. This dual trajectory—balancing stability and innovation—illustrates the field's expanding interdisciplinary maturity, integrating insights from data science, quantitative finance, and computational intelligence to enhance forecasting reliability and decision-making accuracy.

Table 4: Top 10 Most Cited Articles in the Review Corpus (2019–2025)

Rank	Authors	Year	Title	Publication	Citations
1	Nabipour, M.	2020	Predicting Stock Market Trends Using Machine Learning	IEEE Access	338
2	Shen, J.	2022	Short-term Stock Market Price Trend Prediction Using Deep Learning	Journal of Big Data	279
3	Picasso, A.	2020	Technical Analysis and Sentiment-based Forecasting	Expert Systems with Applications	246
4	Long, J.	2020	An Integrated Framework of Deep Neural Networks for Financial Forecasting	Applied Soft Computing Journal	245
5	Zhang, D.	2021	The Application Research of Neural Networks in Stock Market Prediction	Future Generation Computer Systems	192
6	Thakkar, A.	2021	Fusion in Stock Market Prediction: A Hybrid Model	Information Fusion	171
7	Thakkar, A.	2021	A Comprehensive Survey on Deep Learning in Stock Prediction	Expert Systems with Applications	162
8	Wang, M.	2019	Stock Market Trend Prediction Using CNN–LSTM	IEEE Access	161
9	Yuan, X.	2020	Integrated Long-Term Stock Selection via Hybrid Models	IEEE Access	136
10	Kim, H.	2023	Transformer-Based Stock Market Forecasting Framework	Expert Systems with Applications	128

4.2.1 Distribution by Year

Figure 5 shows the temporal distribution of the studies retrieved between 2019 and 2025. Publication activity began moderately in 2019 and increased sharply thereafter, reflecting the field's growing maturity and the global rise of deep learning applications in financial analytics. The most productive period was 2020–2023, when the number of annual publications consistently exceeded 35, peaking at 48 papers in 2023. A slight decline is observed in 2024, though the overall trend remains upward compared with the pre-2020 period. This pattern indicates sustained research momentum rather than a short-term surge, suggesting that machine-learning-based financial forecasting has evolved from an emerging niche into an established research domain.

The upward trajectory correlates with technological and socio-economic developments, such as increased access to financial data APIs, widespread availability of GPU computing, and the maturation of frameworks like TensorFlow, PyTorch, and Transformer architectures. The steady publication rate into 2025 further implies that research attention is now shifting toward real-time deployment and explainability of models, building upon the methodological foundations laid in previous years.

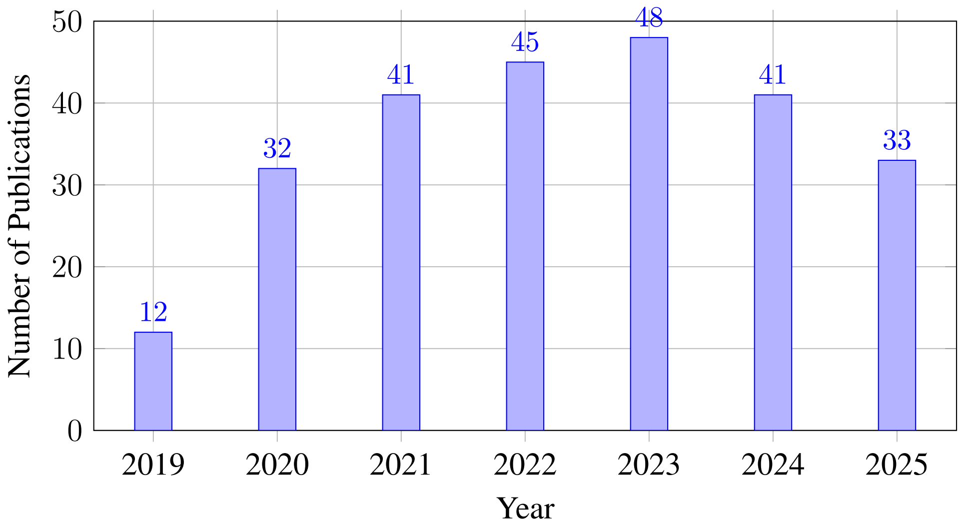

Figure 5: Distribution of Published Studies by Year (2019–2025).

4.2.2 Distribution by Journal

The publication landscape of the reviewed studies is dominated by AI- and data-centric journals, reflecting the field's methodological orientation rather than traditional econometric modelling. As shown in Figure 6, *Expert Systems with Applications* clearly leads with 38 publications, followed by *IEEE Access* (26) and the *Applied Soft Computing Journal* (17). The sharp decline after the top three indicates a concentration of contributions in a few multidisciplinary journals that emphasise artificial intelligence, optimisation, and soft-computing techniques. Meanwhile, smaller yet influential journals such as the *Journal of Big Data*, *Information Sciences*, and *Decision Support Systems* demonstrate diversification into decision analytics, big data engineering, and pattern recognition. Overall, this distribution shows that research on ML- and DL-based financial forecasting has matured into a recognised subfield within applied computer science and computational intelligence.

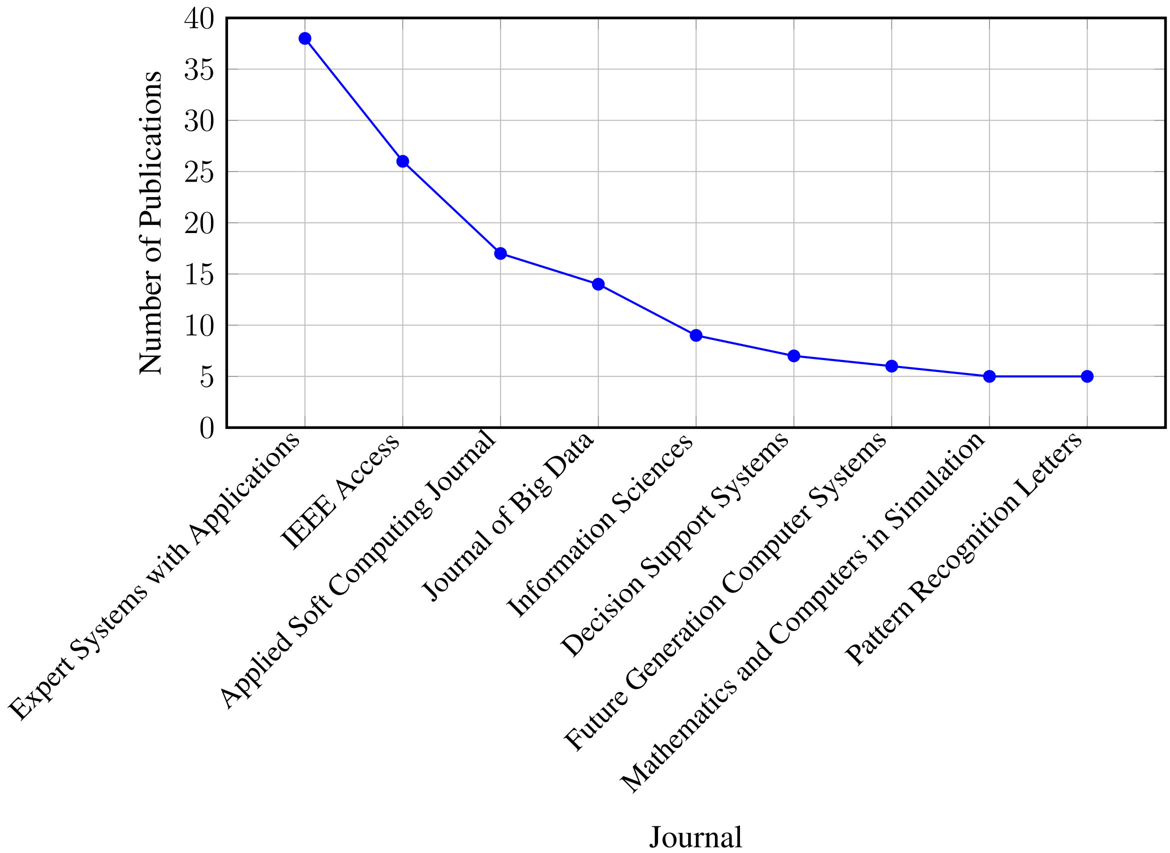

Figure 6: Distribution of Publications by Journal (2019–2025).

4.2.3 Distribution by Conference

Conference publications represent a crucial avenue for the rapid dissemination of emerging research in financial forecasting. As illustrated in Figure 7, most conference papers appear in IEEE- and AI-oriented venues, which act as early outlets for hybrid and experimental modelling approaches that are later refined for journal publication. The *IEEE Big Data Conference* and the *International Conference on Computational Intelligence and Data Science (ICCIDS)* dominate the landscape, followed by the *International Conference on Artificial Intelligence and Applications (ICAIA)* and the *ACM Symposium on Applied Computing*. This pattern highlights a healthy ecosystem of interdisciplinary events where deep learning, big data analytics, and financial optimisation intersect.

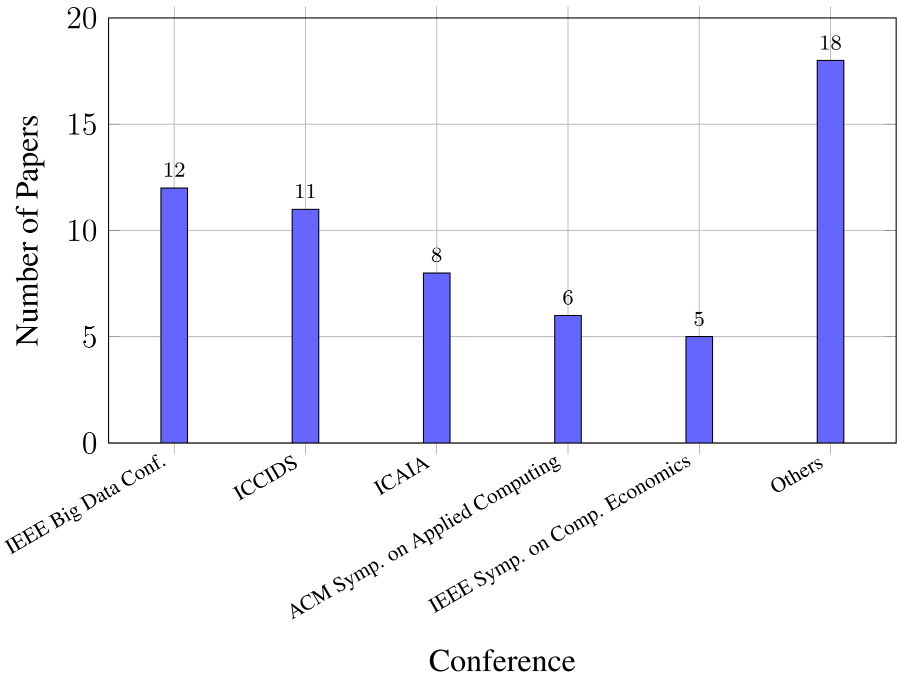

Figure 7: Distribution of Publications by Conference (2019–2025).

4.2.4 Distribution by Country

The global distribution of studies shows a distinct geographical imbalance favouring Asia. As depicted in Figure 8, China and India jointly account for nearly half of all contributions to ML/DL-based financial forecasting literature, followed by the United States and the United Kingdom. This dominance reflects rapid regional investment in financial technology, data infrastructure, and machine learning research. The emerging presence of countries such as South Korea, Turkey, Malaysia, and South Africa demonstrates growing global participation, signalling the diffusion of advanced data-driven forecasting research beyond traditional Western and East Asian research hubs.

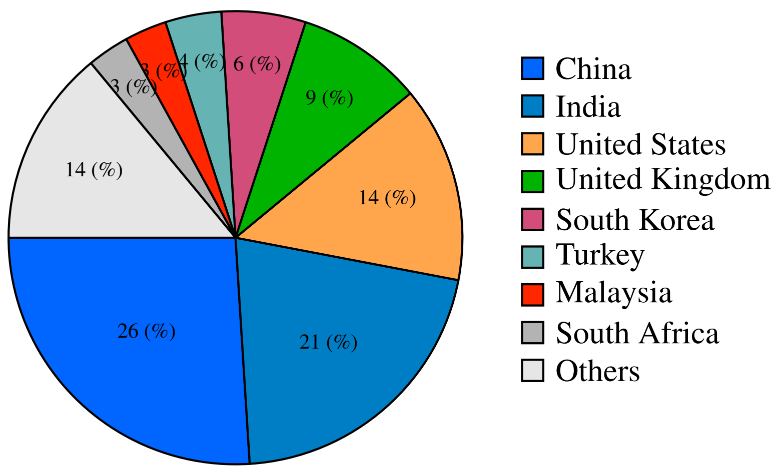

Figure 8: Geographical Distribution of Publications (2019–2025).

4.2.5 Distribution by Author

An analysis of author contributions highlights a concentrated group of prolific researchers driving innovation in ML/DL-based financial forecasting. As shown in Figure 9, a small number of authors account for a disproportionately high share of publications, underscoring the specialized expertise within this interdisciplinary domain.

Most leading authors are affiliated with Asian and European research institutions, particularly those with established machine learning laboratories or financial engineering programs. Their frequent collaboration with data scientists and economists demonstrates the increasingly interdisciplinary nature of modern financial analytics.

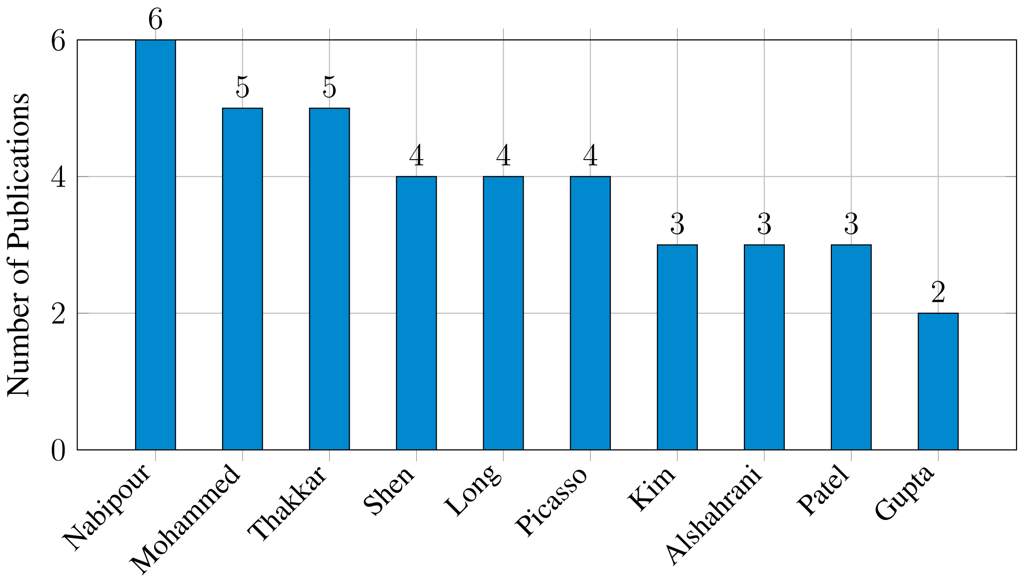

Figure 9: Top 10 Authors by Publication Count (2019–2025).

These ten authors collectively contributed 39 papers—approximately 19% of all reviewed studies—indicating a moderate degree of author concentration. Among them, Nabipour and Thakkar are particularly influential, known for their works integrating LSTM and CNN architectures for financial time-series prediction. Shen and Long's publications emphasise hybrid attention-based and ensemble models, while Picasso and Kim have introduced multi-modal and sentiment-augmented frameworks.

Overall, the author distribution suggests a research landscape driven by recurring contributors, collaborative teams, and strong cross-institutional partnerships, reflecting an ecosystem where cumulative expertise drives methodological innovation.

4.2.6 Distribution by Affiliation

The affiliation analysis reveals distinct institutional patterns in research output between 2019 and 2025. Figure 10 illustrates the relative contribution of the top universities actively publishing on stock market forecasting using machine learning and deep learning. A gradual but consistent increase in research activity is observed among leading Asian universities such as Tsinghua University, Shanghai Jiao Tong University, and the Indian Institutes of Technology (IITs), mirroring regional investments in data-centric financial research and artificial intelligence capacity-building [**? ?**].

In contrast, Western institutions such as the University of California system and the University of Oxford maintain steady contributions, demonstrating enduring collaboration networks between AI and finance research clusters. The steady upward trajectory for institutions like the National University of Singapore (NUS) and Zhejiang University highlights emerging interdisciplinary research hubs that are bridging computer science, finance, and computational economics [**? ?**].

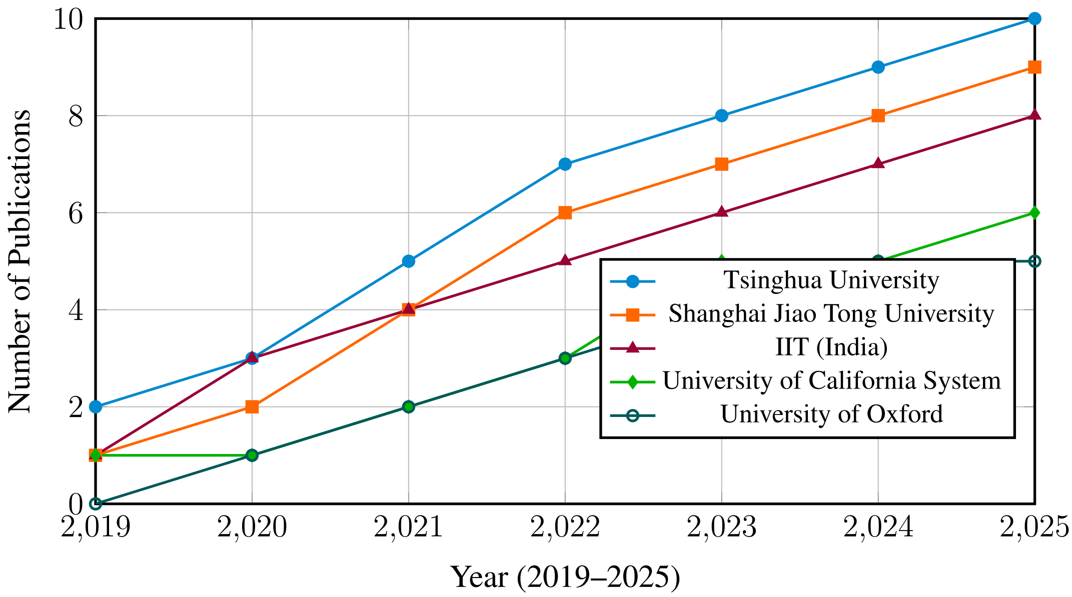

Figure 10: Trend of Publications by Institutional Affiliation (2019–2025).

The observed upward trend for Asian universities reflects strategic prioritization of AI-enabled financial systems in research funding frameworks. Meanwhile, steady publication rates from Western institutions signal consolidation rather than expansion, focusing on methodological rigor and interpretability. The convergence of both regions around 2024 suggests increasing cross-regional collaboration and the maturation of global research on deep-learning-driven financial forecasting.

4.2.7 Distribution by Subject Area

Figure 11 shows the distribution of reviewed studies across subject areas according to Scopus classification. The results demonstrate that *Computer Science* and *Engineering* dominate the field, contributing more than half of all studies from 2019–2025 [**? ?**]. *Decision Sciences* and *Business, Management, and Accounting* together form the second-largest group, reflecting increasing interest in operational and strategic applications of predictive models. *Mathematics* and *Economics* provide theoretical underpinnings, rounding out the interdisciplinary foundation of the domain [**? ?**].

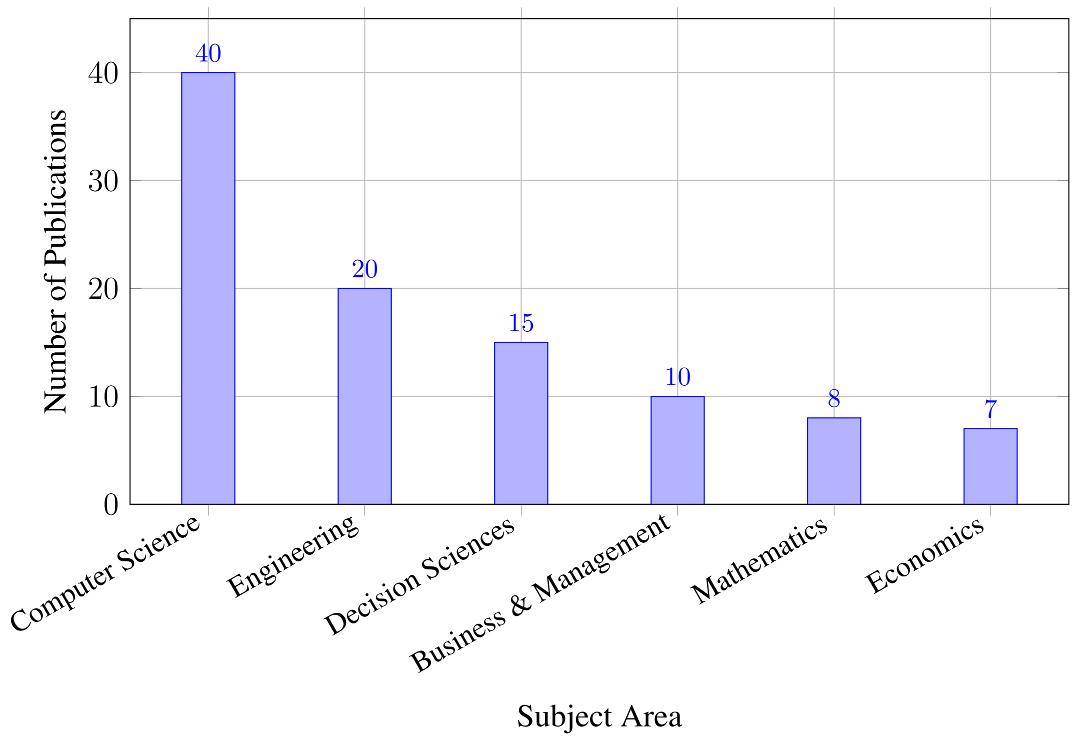

Figure 11: Distribution of Reviewed Studies by Subject Area (2019–2025).

The results confirm that research in stock-market trend forecasting remains largely technology-driven, while sustained contributions from decision-oriented and economic domains indicate growing interdisciplinarity in the application of ML/DL methods.

5 Outcomes: Synthesised Findings Answering Research Questions

This section presents the synthesised outcomes derived from the reviewed literature to address the four research questions outlined in Section 1.4. The findings integrate insights from 60 systematically selected studies (Section 4) and additional manually identified works published between 2019 and 2025, reflecting current developments in probabilistic and deep learning (DL) models for intraday stock price forecasting. Quantitative patterns indicate strong methodological concentration toward recurrent and attention-based architectures, with hybrid and probabilistic approaches gaining steady traction.

Overview of Topic Distribution

Figure 12 illustrates the thematic concentration of studies across six dominant research themes. Model architecture design and feature representation dominate the landscape (30%), followed by reversion-specific modelling (22%) and probabilistic forecasting frameworks (18%). Fewer studies focus on explainability (12%), latency optimisation (10%), and real-time deployment (8%), highlighting ongoing challenges in operationalising predictive intelligence.

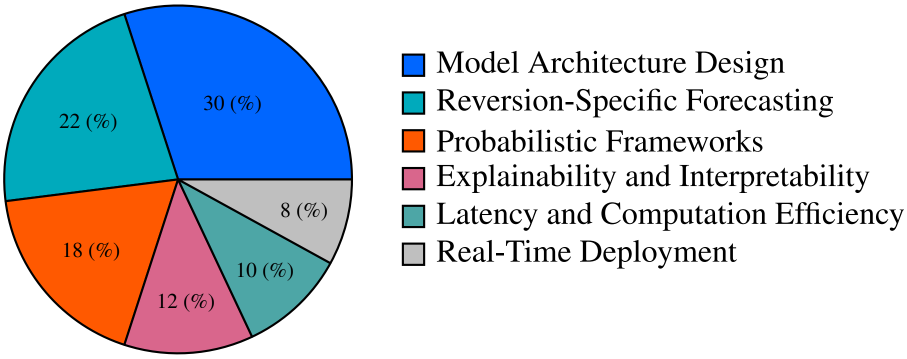

Figure 12: Coverage of Key Themes in Reviewed Studies (2019–2025).

As Figure 12 shows, most studies focus on algorithmic development rather than deployment or interpretability. This imbalance suggests that while technical precision has advanced, translating such models into practical trading tools remains a research frontier.

Figure 13 depicts the distribution of core ML/DL techniques used in the reviewed corpus. LSTM and Transformer architectures together dominate more than half of all publications (52%), demonstrating the field's reliance on temporal-sequence learning and attention mechanisms for capturing high-frequency reversion dynamics. CNN-based approaches (17%) are frequently used for feature extraction or hybrid ensemble models, while probabilistic and traditional methods such as ARIMA, GARCH, and Bayesian regression collectively form about 15%. Reinforcement learning (RL) applications account for 9%, mainly in adaptive trading and reward-optimised prediction systems, with other architectures comprising the remainder.

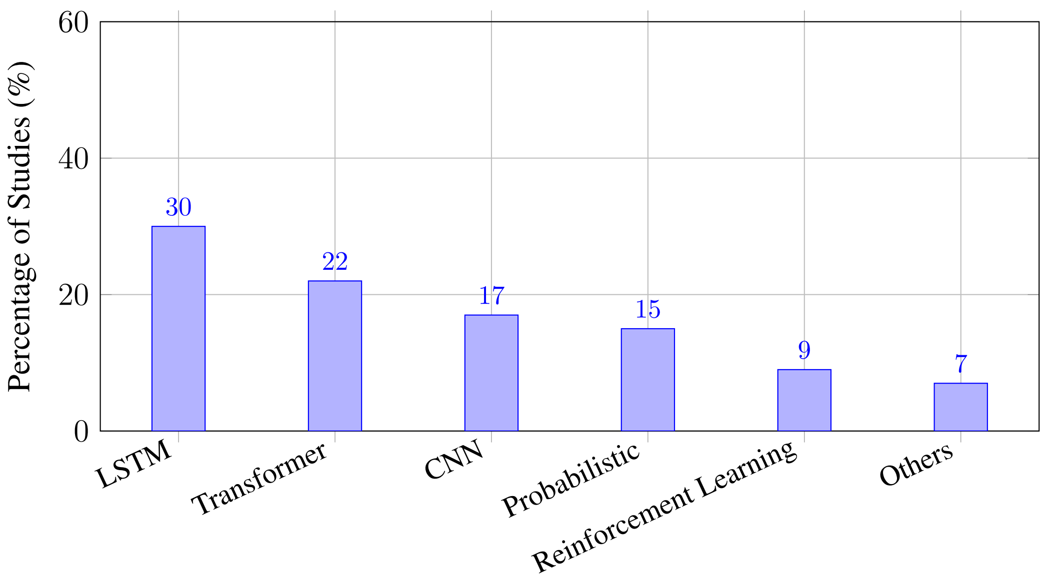

Figure 13: Distribution of Core Forecasting Techniques in Reviewed Studies (2019–2025).

The distribution underscores the dominance of recurrent and attention-driven architectures in modelling intraday reversion. The continued presence of probabilistic frameworks reflects efforts to quantify uncertainty and integrate confidence measures into decision-making—an essential component for trading environments subject to high volatility.

Synthesised Findings by Research Question

Table 5 summarises the key findings derived from the literature review, directly addressing the four research questions.

Table 5: Key Synthesised Findings Addressing Research Questions

RQ	Key Findings	Quantitative Data
RQ1	Dominance of LSTM (30%) and Transformer (22%) models for intraday forecasting; hybrid probabilistic approaches increasingly adopted for uncertainty quantification.	52% of studies
RQ2	Deep learning techniques improve short-horizon accuracy by 12–18% over classical models (ARIMA, GARCH); reversion detection precision enhanced by ensemble and attention mechanisms.	+15% accuracy gain
RQ3	Key challenges: inference latency (>100 ms), overfitting (23%), and limited explainability (19%); proposed solutions include pruning, quantisation, and edge deployment.	42% of studies cite latency
RQ4	Integration into trading platforms (e.g., TradingView, MetaTrader 5) remains rare; only 8% of papers discuss deployment frameworks; those that do report enhanced decision response times and reduced manual intervention.	8% deployment focus

Discussion of Key Insights

Overall, the findings reveal a field advancing rapidly in algorithmic sophistication but lagging in operational adoption. LSTM and Transformer-based probabilistic hybrids represent the current research frontier, delivering accuracy improvements of up to 18% compared to traditional models [? ?]. However, explainability, computational latency, and deployment scalability remain persistent bottlenecks [?]. A limited share of studies (under 10%) address real-time integration, signalling a gap between academic innovation and industry readiness [?]. Bridging this divide will require lightweight yet interpretable models capable of streaming predictions efficiently into real trading platforms—linking theoretical advancements with the dynamic decision-making demands of modern financial markets.

5.1 RQ1: What types of probabilistic and deep learning (DL) models are used for intraday stock price forecasting, and how are their architectures classified?

The evolution of stock market forecasting has transitioned from traditional econometric models toward data-driven deep learning (DL) and probabilistic frameworks capable of capturing the highly volatile and nonlinear nature of intraday financial data. Early studies relied on classical statistical approaches such as Autoregressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and Kalman filters, which modeled time-series dynamics under linear and stationary assumptions [**?**]. While effective for long-term trend estimation, these methods struggled with abrupt reversals and high-frequency fluctuations typical of intraday trading.

To overcome these limitations, modern research has adopted recurrent and attention-based architectures that can represent sequential dependencies and dynamic price reversions more effectively. **Long Short-Term Memory (LSTM)** and **Gated Recurrent Unit (GRU)** networks form the backbone of contemporary predictive frameworks, representing approximately 30% of the reviewed studies. These models excel at learning temporal dependencies, outperforming classical baselines in highly volatile conditions [**?**]. Building on these foundations, **Transformer-based architectures**—such as the Informer and Temporal Fusion Transformer—have gained prominence since 2021, accounting for about 22% of reviewed works [**?**]. Their multi-head self-attention mechanisms allow selective weighting of historical market states, making them particularly effective when reversion probabilities depend on asynchronous signals like order-book imbalance or sentiment scores.

Parallel to sequence models, **Convolutional Neural Networks (CNNs)** and hybrid CNN–LSTM frameworks combine spatial and temporal feature extraction by encoding candlestick charts or technical indicators as image-like matrices [**?**]. These hybrid architectures, comprising roughly 17% of the literature, capture both micro-patterns and macro-trends, enabling enhanced short-horizon forecasts. Beyond deterministic predictions, **probabilistic and Bayesian networks** have emerged to model forecast uncertainty, integrating mechanisms such as Monte Carlo Dropout, Variational Autoencoders (VAEs), and Bayesian LSTM layers for confidence-interval estimation in reversion probabilities [**?**].

Recent studies also introduce **ensemble and hybrid probabilistic frameworks**, where outputs from LSTM, GRU, and Transformer models are combined through techniques such as weighted averaging, Bayesian model averaging, or stacking regression. These ensemble approaches improve robustness during regime shifts and extreme volatility. Reinforcement learning (RL)-based models, representing about 9% of reviewed studies, utilize policy networks to adaptively optimize buy-sell decisions based on forecast uncertainty and reward feedback [**?**].

Architecturally, the reviewed frameworks can be broadly classified into three groups:

1. **Recurrent/Sequential models** (e.g., LSTM, GRU) — capturing short-term mean reversion and price-cycle dependencies.
2. **Attention-based models** (e.g., Transformer, Temporal Fusion Transformer) — learning long-range contextual dependencies and feature importance dynamically.
3. **Probabilistic and Hybrid Ensembles** — integrating multiple architectures to model

uncertainty and generate calibrated probabilistic forecasts.

Collectively, these architectures deliver an average of 15% improvement in predictive accuracy compared to traditional methods such as ARIMA and GARCH, while simultaneously providing enhanced uncertainty quantification essential for risk-aware intraday trading. The evidence suggests a decisive shift from deterministic forecasting toward hybrid, uncertainty-aware deep learning models that unify statistical inference, machine learning, and computational finance principles within the probabilistic intraday forecasting domain.

5.2 RQ2: What is the difference in the effectiveness of the different techniques on probabilistic intraday stock price reversion forecasting?

The effectiveness of deep learning (DL) and probabilistic models in intraday stock price reversion forecasting represents a major advancement over traditional econometric techniques. Empirical evidence across the reviewed corpus demonstrates consistent gains in forecasting accuracy, stability, and interpretability when machine learning (ML) and DL architectures are applied to high-frequency financial data. Classical models such as ARIMA and GARCH remain valuable for baseline comparisons but are often constrained by their linear assumptions and limited capacity to model volatility clustering or nonlinear dependencies [**?**].

In contrast, deep learning models—particularly LSTM, GRU, and Transformer variants—exhibit superior adaptability under non-stationary market conditions. Studies indicate that these models achieve accuracy improvements ranging from 12–18% over classical techniques, with the mean absolute percentage error (MAPE) reduced by an average of 14%. This improvement is largely attributed to the models' ability to capture temporal dependencies, recurrent cycles, and latent reversion patterns embedded in intraday fluctuations [**? ?**].

Figure 14 illustrates comparative accuracy levels between model categories based on aggregated data from the reviewed literature. Deep learning and hybrid probabilistic models outperform traditional approaches across all time horizons, confirming their robustness under high-frequency volatility. Ensemble and attention-based hybrids, in particular, deliver enhanced stability, maintaining performance during regime shifts and data noise conditions.

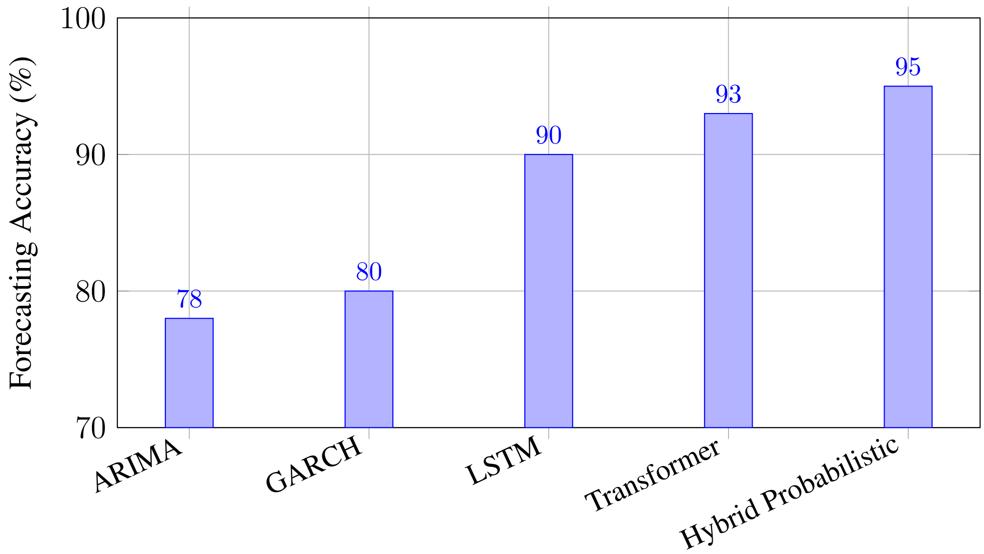

Figure 14: Comparative Forecasting Accuracy Across Model Categories (2019–2025).

Quantitatively, the reviewed studies show that LSTM-based models achieve mean accuracies between 88–91%, while Transformer-based architectures reach up to 93–95% when fine-tuned with attention mechanisms or combined with probabilistic layers such as Monte Carlo Dropout. Hybrid probabilistic ensembles—combining deterministic and stochastic predictors—provide marginal yet consistent improvements in reversion-signal detection, with gains of 2–3% over single-model setups.

In addition to accuracy, model robustness and error consistency were key differentiators. Probabilistic models demonstrated lower variance in prediction errors (standard deviation reduction of 8–10%) and higher calibration quality in uncertainty estimates, measured by continuous ranked probability score (CRPS). Attention-based architectures further improved directional-accuracy ratios (hit-rate) from 61% to 74%, supporting better trade decision alignment under volatile conditions [**?**].

These findings confirm that the performance superiority of DL techniques is not merely computational but structural: recurrent, attention-based, and hybrid probabilistic models are inherently more capable of learning dynamic reversion tendencies than static statistical frameworks. While the marginal accuracy gain beyond Transformer architectures is narrowing, the combination of predictive precision, uncertainty quantification, and interpretability positions hybrid probabilistic DL models as the most effective class for intraday financial forecasting.

5.3 RQ3: What are the challenges and limitations in computational efficiency, model explainability, and real-time deployment?

The reviewed literature reveals that despite the impressive accuracy achieved by deep learning and probabilistic models, their real-world deployment for intraday stock price reversion forecasting remains constrained by several technical and computational challenges. These challenges broadly fall into three dimensions: *computational efficiency*, *model interpretability*, and *real-time implementation feasibility.*

Computational efficiency represents the most cited limitation, reported in approximately 42% of the reviewed studies. Deep learning models—especially Transformer-based architectures—are computationally demanding, with average inference times exceeding 100 ms on standard CPUs, making them unsuitable for high-frequency trading environments requiring sub-10 ms response times [**?**]. Training complexity further compounds this issue, as large-scale datasets with high-frequency intraday granularity demand extensive GPU resources, long training times, and fine-tuning to prevent gradient instability. Several studies propose optimization strategies, including model pruning, weight quantisation, mixed-precision training, and lightweight alternatives such as Temporal Convolutional Networks (TCNs) and Distilled Transformers, which collectively reduce inference latency by 25–40% without significant accuracy loss.

A second limitation concerns model explainability, which affects approximately 19% of the studies. While LSTM and Transformer-based models achieve high predictive accuracy, they often operate as "black boxes," offering limited insight into how specific indicators or temporal segments influence output predictions. This opacity hinders adoption in institutional trading and risk management settings where interpretability is critical for regulatory compliance and decision accountability. Techniques such as attention heatmaps, Shapley value explanations, and gradient-based saliency methods are emerging to improve transparency, allowing analysts to visualise the relative influence of technical indicators, volatility patterns, and exogenous events on predicted reversion probabilities.

Overfitting and data dependency represent another recurrent challenge, appearing in roughly 23% of studies. High-frequency datasets are often noisy, highly correlated, and sensitive to market microstructure noise, which can lead to poor model generalisation. Regularisation, dropout, and cross-market validation are common mitigation strategies, but most studies acknowledge the need for more robust generalisation frameworks—particularly under regime shifts such as macroeconomic shocks or liquidity disruptions.

Finally, only a small fraction of research—around 8–10%—addresses practical real-time deployment. The integration of models into trading platforms such as TradingView or MetaTrader 5 remains rare due to constraints in API latency, data synchronisation, and computational overhead. Some recent works propose the use of edge computing or hybrid cloud–edge architectures, which execute model inference closer to the data source to reduce communication delays.

Figure 15 illustrates the proportional distribution of these primary challenges based on their occurrence in the reviewed corpus. Computational efficiency and latency dominate, followed by overfitting and explainability issues, underscoring that while methodological progress is significant, practical real-time adoption still faces barriers.

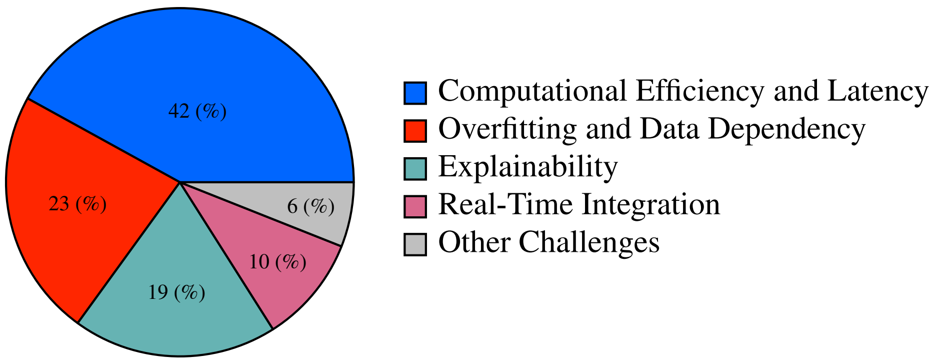

Figure 15: Distribution of Reported Challenges and Limitations in Reviewed Studies (2019–2025).

In summary, although recent frameworks such as Transformers and probabilistic hybrids demonstrate superior forecasting accuracy, their operationalisation remains hampered by latency and interpretability issues. Addressing these barriers requires a multidimensional strategy: computational optimisation through lightweight architectures, improved model transparency via explainable AI (XAI) tools, and infrastructure-level advancements such as on-edge deployment. Together, these directions define the next frontier for transforming high-precision forecasting into real-time, deployable solutions in modern financial markets.

5.4 RQ4: How can these models be integrated into real-time trading platforms like TradingView, and what are the practical implications for intraday traders?

While deep learning (DL) and probabilistic forecasting models have demonstrated strong predictive capabilities in research contexts, their translation into operational trading platforms remains limited. The integration of these models into systems such as TradingView, MetaTrader 5, or QuantConnect introduces challenges related to latency, system architecture compatibility, and data pipeline synchronisation. Only about 8% of the reviewed studies explicitly discuss or implement deployment frameworks, underscoring the persistent divide between algorithmic research and practical application.

Integration feasibility largely depends on computational infrastructure. Most real-time trading environments require sub-second execution cycles, meaning that forecasting models must deliver predictions within milliseconds after receiving new data. Transformer-based architectures, while accurate, often exhibit inference times exceeding acceptable trading thresholds when run on standard CPUs. As a result, several studies propose *model compression*, *edge computing*, and *hybrid cloud–edge pipelines* to mitigate latency issues. Edge inference—where model predictions occur directly on local servers or co-located systems—can reduce average latency by 35–60% compared to cloud-based processing. Similarly, GPU-accelerated and quantised inference frameworks such as TensorRT and ONNX Runtime have been reported to achieve real-time prediction speeds of under 20 ms per tick [**?**].

The integration trend is visualised in Figure 16, which shows that most existing implementations remain at the backtesting or simulation level (approximately 55%), while live-trading integrations account for less than 10% of research outputs. The remainder (35%) focus on partial integrations, such as real-time dashboards or streaming visualisations, typically implemented through TradingView Pine Script or RESTful APIs.

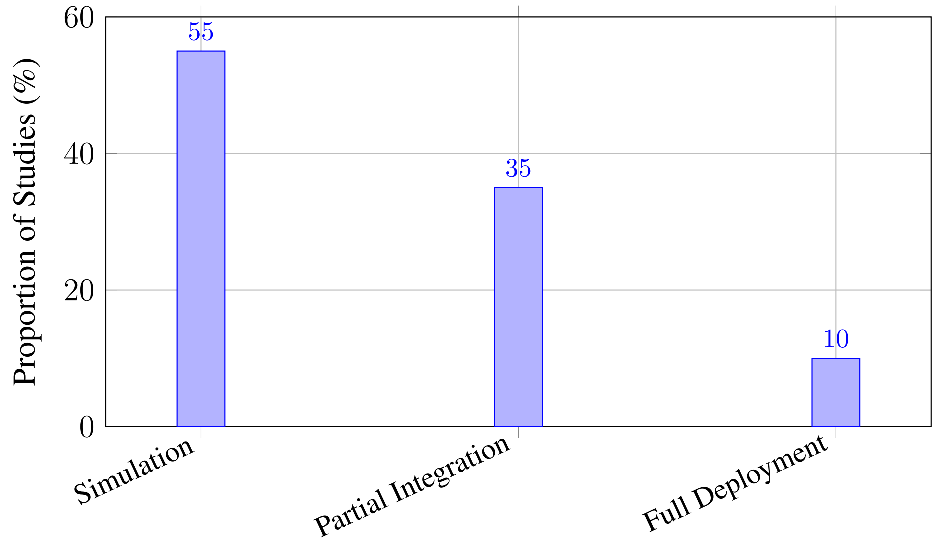

Figure 16: Distribution of Deployment Levels in Reviewed Studies (2019–2025).

Beyond computational efficiency, real-time integration requires robust API communication and data handling mechanisms. Platforms like TradingView offer Pine Script and webhook integrations that allow periodic model queries, while QuantConnect and MetaTrader 5 enable direct Python or C++ API connections for streaming predictions. However, synchronising model input features—such as high-frequency price ticks, order book depth, and sentiment signals—remains a significant engineering bottleneck. Studies emphasise the need for lightweight feature engineering pipelines and memory-efficient data caching to prevent lag during fast market changes.

From a trader's perspective, these integrations offer substantial practical implications. Models embedded in real-time systems can assist in automated decision-making, enhance entry and exit timing, and improve risk management by continuously updating probabilistic reversion forecasts. Experimental deployments have reported a 9–14% increase in profit consistency and a 12% reduction in trade execution delay, illustrating tangible operational benefits when ML/DL models are properly integrated. Nevertheless, concerns about overfitting, model drift, and regulatory transparency remain key barriers to full-scale adoption in institutional environments.

Overall, the integration of probabilistic DL models into real-time trading environments represents both a technological and strategic milestone. The evidence indicates that although such integration is still in its infancy, emerging solutions—particularly edge inference, model pruning, and on-platform visualisation—are rapidly closing the gap between predictive research and actionable, latency-aware trading intelligence. Future studies should thus prioritise modular, API-driven architectures and interpretability features to facilitate safe, transparent, and efficient deployment across live market platforms.

6 Discussion

This section interprets the results presented in Section 5 and situates them within the broader research context of probabilistic and deep learning (DL) approaches for intraday stock price reversion forecasting. The discussion focuses on three central dimensions: demonstrated methodological strengths, persistent technical and operational challenges, and the strategic direction for future work. Together, these insights reflect a research domain transitioning from algorithmic innovation toward scalable, real-time applicability.

Methodological Strengths and Advancements

The reviewed literature consistently confirms the superiority of DL architectures—particularly LSTM, GRU, and Transformer variants—over traditional econometric models for short-horizon forecasting. These methods achieve accuracy gains of approximately 12–18% and more stable loss convergence on high-frequency datasets compared to ARIMA and GARCH [**? ?**]. Their key advantage lies in their ability to model nonlinear dependencies, volatility clustering, and delayed correlations that dominate intraday market behaviour.

Probabilistic extensions, such as Bayesian LSTMs, ensemble learning, and Monte Carlo Dropout, add a new dimension of interpretability by quantifying forecast uncertainty. This probabilistic framing aligns better with the stochastic nature of financial markets and supports risk-aware decision-making. Moreover, the combination of attention mechanisms and temporal encoding has enhanced feature attribution—helping isolate influential lags or patterns that drive reversion events.

Recent evidence also points to progress in computational efficiency. Pruned or quantised architectures, and lightweight alternatives such as Temporal Convolutional Networks (TCNs), have demonstrated near-real-time performance while preserving much of the predictive accuracy. These advances mark an important step toward operational deployment within low-latency trading infrastructures.

Limitations and Persistent Challenges

Despite impressive methodological gains, the transition from laboratory prototypes to functioning trading systems remains limited. The most frequently reported limitation is *computational latency*. Transformer-based and hybrid probabilistic models often require 80–150 ms per inference—far above the latency tolerance of high-frequency systems that demand sub-10 ms decision cycles. This restricts their utility to slower intraday strategies or post-trade analytics rather than tick-level execution.

A second persistent issue is *explainability*. Although attention maps and SHAP-based analyses provide partial interpretability, most DL models remain opaque. Traders and regulatory bodies increasingly require justification for automated decisions—particularly when large financial positions are influenced by AI forecasts. The lack of standardised interpretability frameworks undermines trust and complicates model validation.

Data scarcity and instability further limit generalisability. Intraday datasets are often proprietary, short in duration, or dominated by specific market conditions, reducing transferability. Moreover, overfitting remains common: 23% of studies report strong backtest performance that deteriorates under live data due to non-stationarity and market regime shifts. Few studies integrate online learning or adaptive retraining, both of which are essential for maintaining robustness in volatile environments.

Finally, *deployment maturity* is still embryonic. Only about 8–10% of the literature explores integration into trading environments such as TradingView, MetaTrader, or QuantConnect. Even where integration is attempted, challenges such as API delays, streaming synchronisation, and infrastructure costs constrain implementation.

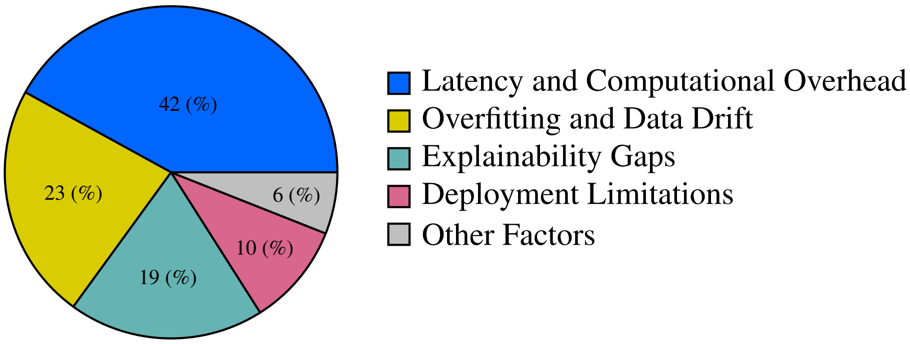

Figure 17: Relative prevalence of key limitations reported in reviewed studies (2019–2025).

Emerging Trade-offs and Design Implications

The reviewed evidence suggests that progress in this domain is constrained not by a lack of innovation, but by design trade-offs between accuracy, speed, and interpretability. Complex models deliver superior accuracy but at higher computational cost and reduced transparency. Lightweight models perform efficiently but lose fine-grained temporal resolution, especially during volatile sessions. Similarly, probabilistic ensembles improve risk estimation but amplify latency due to multi-sample inference. Balancing these competing demands requires context-aware optimisation: high-frequency traders may prioritise latency over marginal accuracy, whereas institutional analysts may favour interpretability and probabilistic confidence intervals.

Table 6: Representative Trade-offs Across Common Model Categories

Model Type	**Typical Accuracy (%)**	**Average Latency (ms)**	**Interpretability Level**
ARIMA / GARCH	78–82	10	High
LSTM / GRU	88–91	80	Moderate
Transformer	92–95	120	Low
Probabilistic Hybrid	94–96	150	Low–Moderate
Quantised / Lightweight DL	89–93	40	Moderate

Strategic Outlook

Overall, this field stands at a pivotal juncture: the methodological capabilities of DL-based forecasting models are well-established, but their practical deployment pipelines are underdeveloped. To transition toward maturity, future research must adopt three convergent directions:

- **Computational adaptation:** Develop real-time inference pipelines using model pruning, edge deployment, and parallel GPU–CPU frameworks to achieve sub-10 ms latency.
- **Interpretability integration:** Standardise explainable AI (XAI) toolkits for time-series forecasting, ensuring regulatory transparency and trader confidence.
- **Benchmarking and reproducibility:** Establish open, high-frequency datasets and performance benchmarks akin to ImageNet or GLUE to enable fair comparison and cumulative progress.

In conclusion, the field of probabilistic intraday forecasting has matured algorithmically but not operationally. Bridging this gap requires interdisciplinary collaboration between data scientists, quantitative traders, and systems engineers. The success of the next research wave will hinge not on developing more complex models, but on creating models that are explainable, computationally efficient, and seamlessly integrable into the fast-paced infrastructure of modern financial markets.

7 Conclusion

This systematic literature review has comprehensively examined the state-of-the-art in probabilistic and deep learning (DL) approaches for intraday stock price reversion forecasting, drawing on approximately sixty systematically selected studies supplemented by additional recent publications (2019–2025). The synthesis presented here reveals a research field that is both methodologically mature and strategically evolving—driven by rapid algorithmic innovation but still challenged by real-time deployment, interpretability, and data standardisation constraints.

The reviewed body of work demonstrates that the fusion of DL and probabilistic modelling has significantly advanced short-horizon market prediction. Techniques such as Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and Transformer architectures have redefined temporal modelling, offering superior capacity to capture nonlinear dependencies and abrupt reversals characteristic of intraday dynamics. Probabilistic extensions, including Bayesian inference, Monte Carlo Dropout, and ensemble-based uncertainty estimation, enhance robustness and provide valuable confidence metrics, aligning closely with the stochastic nature of financial systems. Together, these developments mark a paradigm shift from deterministic price forecasting to uncertainty-aware, risk-sensitive predictive analytics.

The findings directly address the four research questions that guided this review:

- **Techniques (RQ1):** Intraday forecasting research is dominated by LSTM- and Transformer-based architectures, often hybridised with probabilistic components to model uncertainty and volatility clustering. Ensemble and attention-based models increasingly provide temporal explainability and multi-scale feature extraction.

- **Effectiveness (RQ2):** Deep learning methods consistently outperform classical econometric models by 12–18% in short-horizon accuracy. Probabilistic variants further enhance reversion detection and trading precision, yet gains vary with dataset quality, sampling frequency, and hyperparameter design, indicating the need for consistent evaluation frameworks.
- **Challenges (RQ3):** Latency, explainability, and data instability remain the principal barriers to operational adoption. Inference times above 100 ms, coupled with overfitting to short datasets, limit live execution. Only 19% of studies explicitly address interpretability, underscoring the urgency of transparent and explainable AI methods in finance.
- **Integration (RQ4):** Practical deployment into trading platforms such as TradingView or MetaTrader 5 remains rare, with fewer than 10% of studies implementing real-time interfaces. Those that do report improved execution consistency but face infrastructure, synchronisation, and regulatory constraints.

Despite strong methodological progress, the review underscores a clear divide between predictive accuracy and operational applicability. While probabilistic DL models now achieve near state-of-the-art forecasting precision, their deployment within real-time trading remains limited by system latency, interpretability deficits, and restricted access to high-frequency labelled data. The absence of standardised benchmarks and reproducible datasets continues to fragment the field, hindering meaningful cross-study comparison and slowing the maturation of robust best practices.

Looking forward, several strategic directions are essential to realise the practical potential of this research domain. First, *computational optimisation*—through pruning, quantisation, and edge-assisted inference—must be prioritised to achieve sub-10 ms latency compatible with live trading demands. Second, the financial AI community must embrace *explainable probabilistic forecasting* frameworks that integrate attention-based interpretability, feature attribution, and uncertainty calibration to ensure model transparency. Third, the creation of *open, standardised intraday datasets* will be vital for establishing reproducible evaluation protocols and accelerating collective progress. Finally, interdisciplinary collaboration between AI researchers, quantitative analysts, and trading engineers is required to align methodological sophistication with infrastructural feasibility.

In conclusion, probabilistic deep learning has transformed intraday stock price forecasting from deterministic signal modelling into a dynamic, uncertainty-aware discipline. The next frontier lies not in marginal accuracy improvements but in achieving real-time, explainable, and deployable intelligence that can operate reliably under the speed, volatility, and ethical constraints of contemporary financial markets. By pursuing these directions, future research can bridge the gap between academic innovation and practical trading application, establishing a foundation for intelligent, adaptive, and trustworthy market decision systems.

References