



# A Review of Statistical, Machine Learning, and Deep Learning Methods for Stock Price Prediction: Toward Explainable Hybrid Models

Kulkarni Parag

Department of Computer Science and IT,  
SBES College of Science,  
Chh. Sambhajinagar, Maharashtra, India.  
Email: [parag\\_sbeshes@outlook.com](mailto:parag_sbeshes@outlook.com)  
Tel: +91-9890053679

Sanjay Azade

Department of Computer Science and IT  
, Dr. G.Y. Pathrikar College of CS & IT, MGM University,  
Chh. Sambhajinagar, Maharashtra, India.

Corresponding author: Kulkarni Parag M.

DOI Link: <https://doi.org/10.64971/j.cph.eijtem.v13.i3.4.2026>

## Abstract:

Stock price prediction is a long-standing and challenging problem in financial data science owing to the nonlinear, non-stationary, and noisy nature of market data. Over four decades, forecasting approaches have evolved from linear statistical models such as ARIMA and GARCH, through classical machine learning methods including Support Vector Machines, Decision Trees, and Random Forests, to deep learning architectures such as LSTM, GRU, CNN, and Transformer networks. Hybrid models integrating statistical and deep learning components – for example ARIMA-LSTM, CNN-BiLSTM-Attention, SARIMA-GARCH-CNN-BiLSTM, and Transformer-GRU frameworks – now represent the state of the art, consistently outperforming standalone baselines. This paper reviews 42 peer-reviewed stock-prediction studies, supported by 10 foundational and methodological references, organised using a PRISMA-informed screening process, with the primary studies published between 2001 and 2025, with emphasis on 2023-2025. We categorise the methods through a taxonomy, synthesise their reported performance with explicit attention to dataset and metric-scale heterogeneity, present a bibliometric overview, and critically evaluate persistent gaps. Three gaps emerge as central: the absence of standardised benchmarks and economic-utility metrics, the limited use of Indian-market (Nifty 50 / NSE) data, and the near-total absence of explainability frameworks such as SHAP and LIME. We argue that the most promising direction is the development of explainable hybrid models validated on Indian equity data, and position this review as the foundation for such work.

## Keywords:

Stock price forecasting; ARIMA; GARCH; LSTM; GRU; Transformer; hybrid models; explainable AI; SHAP; Nifty 50; literature review; PRISMA-informed screening

## 1. Introduction

Forecasting stock prices accurately is of central importance to investors, portfolio managers, and policymakers. The market is shaped by the interaction of macroeconomic conditions, firm fundamentals, investor sentiment, and exogenous shocks, producing time series that are nonlinear, non-stationary, and heavy with noise [1],[2],[3]. No single modeling paradigm has proven universally superior, and the field has progressed through several distinct phases that this review traces in turn. The earliest systematic approaches were statistical. Box, Jenkins, and Reinsel formalised the ARIMA family for linear, stationary series [1], while Engle introduced the ARCH model – later generalised to GARCH – to capture the time-varying volatility characteristic of financial returns [1],[2],[4],[5]. Hyndman and Athanasopoulos provide the standard modern treatment [2], and White's early neural-network experiments on IBM returns foreshadowed the later shift to data-driven models [5]. Although

interpretable and computationally inexpensive, statistical models struggle with nonlinear dynamics, motivating their later use within hybrid frameworks.

Machine learning introduced non-parametric flexibility. Tay and Cao first applied SVMs to financial forecasting [6], Kim confirmed their effectiveness with kernel methods [7], and ensemble methods such as Random Forests reduced overfitting; Gu reported Random Forest clearly outperforming a standalone Decision Tree with  $R^2$  of 0.983 [8]. Recent comparative work continues to find Random Forest competitive with or superior to other baselines [6],[7],[8],[9],[10]. The principal limitation of these models is that they do not natively capture temporal order without engineered features.

Deep learning addressed this directly. LSTM networks learn long-range dependencies through gating; Siami-Namini et al. reported LSTM reducing forecasting error substantially relative to ARIMA [11], and Fischer and Krauss demonstrated large-scale LSTM outperformance [12]. CNNs were adapted to one-dimensional series for local pattern extraction [13], and GRU – a lighter recurrent variant – has repeatedly matched or exceeded LSTM on Indian-market data [14],[15]. Transformer architectures with self-attention now capture global temporal structure, often surpassing recurrent models [16],[17],[18]. These advances come at the cost of interpretability and computation.

The current frontier is hybridisation combined with explainability. ARIMA-LSTM models decompose a series into linear and nonlinear parts [19],[20]; CNN-BiLSTM-Attention and SARIMA-GARCH-CNN-BiLSTM models stack feature extraction, sequence modeling, and attention [21],[22],[23]; and sentiment-augmented Transformer hybrids integrate heterogeneous price and news signals [24],[25]. Yet three weaknesses persist: fragmented evaluation across incomparable datasets and metrics; under-representation of Indian equity markets; and the predominantly black-box nature of these models despite growing demand for explainability [26],[27],[28]. This review synthesises the evidence with these gaps in focus and motivates a research agenda centred on explainable hybrid models for Indian markets.

The remainder of the paper is organised as follows. Section 2 details the PRISMA-based methodology and presents a taxonomy of methods. Section 3 synthesises the literature by model family. Section 4 presents comparative and bibliometric analysis. Section 5 offers a critical evaluation, Section 6 a discussion, Section 7 threats to validity, Section 8 the key findings, and Section 9 concludes with the proposed research direction.

## **2. Review Methodology**

To ensure transparent and repeatable study selection, this review adopts a screening process informed by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework, an approach increasingly adopted in financial-forecasting reviews [26],[29],[30]. The objective was to identify, screen, and synthesise peer-reviewed studies applying statistical, ML, DL, or hybrid models to stock price prediction.

### **2.1 Search Strategy**

Seven databases were searched: IEEE Xplore, ScienceDirect, SpringerLink, MDPI, Wiley, Scopus, and Google Scholar. The search combined the terms “stock price prediction”, “stock market forecasting”, and “financial time series” with model terms “ARIMA”, “GARCH”, “SVM”, “random forest”, “LSTM”, “GRU”, “CNN”, “transformer”, “attention”, “hybrid”, and “explainable AI”. The primary window was 2001-2025, with deliberate emphasis on 2023-2025.

### **2.2 Inclusion and Exclusion Criteria**

Studies were included if they were peer-reviewed, addressed equity price or index prediction, reported at least one quantitative metric, and were in English. Studies were excluded if they addressed non-equity assets exclusively, lacked quantitative evaluation, were non-peer-reviewed, or predated 2001 (seminal works such as [1],[4],[5] were retained).

### **2.3 Selection Process (PRISMA Flow)**

The screening funnel is summarised in Figure 1. An initial pool of approximately 260 records was reduced to 205 after duplicate removal, screened by title and abstract, assessed at full text, and finally narrowed to 42 primary stock-prediction studies retained for synthesis, together with 10 foundational and methodological works cited for context.

Figure 1. PRISMA selection flow.

|                                                                                                                                                         |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| Identification: records identified through database searching (IEEE Xplore, ScienceDirect, SpringerLink, MDPI, Wiley, Scopus, Google Scholar) (n = 260) |
| Screening: records after duplicates removed (n = 205)                                                                                                   |
| Screening: records screened by title and abstract (n = 205)                                                                                             |
| Eligibility: full-text articles assessed (excluded: non-English, no quantitative metric, non-equity assets, non-peer-reviewed) (n = 94)                 |
| Included: studies synthesised in this review (n = 42)                                                                                                   |

## 2.4 Taxonomy of Methods

Figure 2 organises the reviewed methods into five families along an increasing-complexity axis: statistical, classical ML, deep learning, attention/transformer, and hybrid-plus-explainability. This taxonomy structures the synthesis in Section 3.

Figure 2. Taxonomy of stock price prediction methods.

| STOCK PRICE PREDICTION METHODS |                                                       |                                  |                                                 |                                                     |
|--------------------------------|-------------------------------------------------------|----------------------------------|-------------------------------------------------|-----------------------------------------------------|
| Statistical                    | Classical ML                                          | Deep Learning                    | Attention / Transformer                         | Hybrid + XAI                                        |
| ARIMA, GARCH, EGARCH, SARIMA   | SVM, SVR, k-NN, Decision Tree, Random Forest, XGBoost | RNN, LSTM, BiLSTM, GRU, CNN, TCN | Transformer, TFT, Multi-head Attention, FinBERT | ARIMA-LSTM, CNN-BiLSTM-Attn, GARCH-RF, +SHAP / LIME |

## 2.5 Methodological Caveat on Cross-Study Comparison

A critical limitation governs all quantitative comparison here. The included studies use different markets, periods, normalisation strategies, horizons, and splits. Error metrics such as RMSE and MAE are scale-dependent: an RMSE of 0.14 on data normalised to [0,1] is not comparable to an RMSE of 64.8 on raw index points [21], and “accuracy %” applies only to directional classification, not price regression. Accordingly, Section 4 reports each result as stated in its source with dataset and scale made explicit, rather than as a single ranked leaderboard.

## 3. Synthesis of the Literature

### 3.1 Statistical Models (ARIMA, GARCH, and Variants)

ARIMA represents a series through autoregressive, differencing, and moving-average terms and remains a strong baseline on stationary, linear data [1],[2]. GARCH models conditional variance and is the standard volatility tool [4]. Recent work has extended these foundations: Agyarko et al. proposed a Burr-Singh GARCH that outperformed classical GARCH, EGARCH, and GJR-GARCH on NASDAQ 100 and S&P 500 by achieving the lowest RMSE, MAPE, and QLIKE [31]; and asymmetric EGARCH variants have been compared across cryptocurrencies and equities to capture leverage effects [32]. A wavelet-ARIMA-GARCH approach has been used to handle the linear-nonlinear split in volatility forecasting [33]. The consistent conclusion is that pure statistical models remain interpretable and cheap but require nonlinear augmentation, which directly motivates the hybrid models of Section 3.5.

### 3.2 Classical Machine Learning (SVM, k-NN, Decision Trees, Random Forests)

SVMs map inputs to high-dimensional spaces and perform well on directional classification [6],[7]; Kurani et al. provide a comprehensive ANN-versus-SVM comparison for stock forecasting [34]. k-NN is simple but scales poorly with noise. Decision Trees are interpretable but overfit; Random Forests mitigate this through ensemble averaging. Gu reported RF achieving  $R^2$  of 0.983 against a weaker single tree [8], and a comparative study found RF (MAE 17.766, RMSE 22.376,  $R^2$  0.997) outperforming SVR on five years of bank-stock data [10]. Khan et al. compared nine ML models and showed that with an improved investment strategy, Random Forest reached 91.27% directional accuracy, up from an 85.51% logistic-regression baseline [9]. San embedded Random Forest within an ARIMA-GARCH ensemble for forecasting and portfolio optimisation [35]. These models require explicit feature engineering to encode temporal structure.

### **3.3 Deep Learning (LSTM, GRU, CNN, TCN)**

Early work established the viability of recurrent and convolutional models; Selvin et al. compared LSTM, RNN, and CNN sliding-window models for stock prediction [36]. LSTM dominates the recurrent literature owing to gating mechanisms that mitigate vanishing gradients and capture long-memory effects [11],[12],[37]. GRU offers comparable accuracy with fewer parameters and emerged as the top performer in several Indian-market studies [14],[15]. CNNs extract localised temporal features through one-dimensional convolution [13]. Indian-market studies have optimised LSTM for the NSE [38] and applied it to the Nifty 50 index specifically [39]. Comparative studies on the Nifty 50 index find that CNN, GRU, and Temporal Convolutional Network (TCN) variants each lead under different conditions, underscoring that no single architecture dominates universally [14],[18],[40]. An efficient time-series RNN variant (TRNN) has also been proposed to reduce computational overhead while preserving accuracy [41].

### **3.4 Transformer and Attention-Based Models**

Transformer architectures replace recurrence with self-attention, weighting all time steps globally and capturing long-range dependencies more effectively than recurrent models in several studies [16],[17]. Incremental-learning and multi-perspective Transformer variants address non-stationarity and online updating [42],[43]; an improved BiLSTM-Transformer hybrid has reported strong accuracy on benchmark indices [44]; and Transformer-GRU and FinBERT-augmented hybrids fuse price and textual sentiment [24],[25]. The principal drawbacks are high computational cost and reduced transparency, which the explainability methods of Section 3.6 seek to address.

### **3.5 Hybrid Models**

Hybrid architectures combine complementary strengths. ARIMA-LSTM separates linear and nonlinear components [19],[20]. CNN-BiLSTM-Attention integrates feature extraction, bidirectional modeling, and attention, achieving lower RMSE than simpler variants on Chinese indices [21], while the closely related CNN-BiLSTM-AM reported MAE 21.952 and RMSE 31.694 as the best of eight methods on the Shanghai Composite [22]. Araya proposed a four-stage SARIMA-GARCH-CNN-BiLSTM model that achieved the lowest MAE and RMSE against GARCH-family baselines on Apple data [23]. On Indian data, a Nifty 50 volatility study found GARCH-family models (notably the asymmetric EGARCH and TARCh) slightly outperforming an RNN-based LSTM, underscoring that deep models do not automatically dominate well-specified econometric baselines [45]. Bi-LSTM-GRU and decomposition-based hybrids further improve robustness on Nifty 50 data [40],[46]. Bao et al. combined wavelet transforms, stacked autoencoders, and LSTM [47], and Zhang et al. fused events, sentiment, and quantitative data via multiple-instance learning [48]. The consistent finding is that hybrids outperform their constituents, at the cost of complexity and interpretability.

### **3.6 Explainable AI in Stock Prediction**

As DL models grow more accurate, their opacity has become a barrier to adoption in regulated finance. SHapley Additive exPlanations (SHAP) [27] and Local Interpretable Model-agnostic Explanations (LIME) [49] are the dominant post-hoc attribution methods. Systematic reviews of explainable AI in finance identify stock prediction as one of the most common application areas, yet note that explainability is applied far less often than predictive modeling [26],[50]. Muhammad et al. demonstrated an explainable DL approach for trend prediction using feature attribution [28], and comparative XAI studies confirm that attention visualisation, SHAP, and LIME can reveal the influence of features such as volume and sentiment [51]. Mehtab and Sen integrated deep learning with natural-language processing of financial news for Indian equities, an early step toward feature-aware prediction [52]. However, the integration of explainability into hybrid forecasting pipelines – especially for Indian equities – remains rare, defining the central gap this review identifies.

## **4. Comparative and Bibliometric Analysis**

Table 1 reports verified results exactly as stated in their source studies, with dataset and metric-scale annotations. Table 2 characterises the evaluation metrics used across the corpus. Table 3 summarises the datasets. Table 4 provides a qualitative capability matrix, and Figure 3 shows the temporal distribution of the reviewed studies.

**Table 1. Reported performance of representative models (results as stated in each source; scales not directly comparable).**

| Model (Study, Year)                             | Market / Dataset                     | Reported Result (as stated in source)                          | Metric & Scale Note                   | Ref       |
|-------------------------------------------------|--------------------------------------|----------------------------------------------------------------|---------------------------------------|-----------|
| ARIMA-LSTM (Kulshreshtha & Vijayalakshmi, 2020) | S&P 500 (live API)                   | RMSE 1.74; MAPE 0.009; beat Prophet (RMSE 27.59)               | Absolute price (USD); not normalised  | [19]      |
| LSTM-ARIMA (Kalange, 2025)                      | 5 Indian bank stocks (NSE)           | Hybrid lowest RMSE/MAE vs standalone ARIMA & LSTM              | Per-stock price (INR)                 | [20]      |
| RF vs Decision Tree (Gu, 2024)                  | Single equity, CV                    | RF: MSE 2.605, MAE 1.152, $R^2$ 0.983 > DT                     | Regression; price scale               | [8]       |
| RF vs SVR (2024)                                | BNI stock, 2019-24                   | RF: MAE 17.766, RMSE 22.376, $R^2$ 0.997 > SVR                 | Price scale (IDR)                     | [10]      |
| 9 ML models + strategy (Khan et al., 2023)      | Directional, 1-day                   | RF best 91.27% with proposed strategy; LR 85.51% baseline      | Directional accuracy (%)              | [9]       |
| CNN, GRU, LSTM, Attn-LSTM (2024)                | HDFC, TCS, ICICI, Reliance, Nifty 50 | CNN & GRU generally best (stock-dependent)                     | Normalised [0,1]; MAE/MSE/ $R^2$      | [14]      |
| GRU vs LSTM variants (2024)                     | 5 sectors, NSE India                 | GRU top performer across sectors                               | Normalised; sector-wise               | [15]      |
| CNN-BiLSTM-Attention (Zhang et al., 2023)       | CSI 300, SSE (China)                 | RMSE 64.84 on CSI 300; below LSTM/CNN-LSTM/CNN-LSTM-Attn       | Index points; not normalised          | [21]      |
| CNN-BiLSTM-AM (Lu et al., 2021)                 | Shanghai Composite                   | MAE 21.952; RMSE 31.694 (best of 8)                            | Index points                          | [22]      |
| SARIMA-GARCH-CNN-BiLSTM (Araya, 2024)           | Apple Inc., 2018-23                  | Hybrid lowest MAE/RMSE vs GARCH-family baselines               | Volatility; price scale (USD)         | [23]      |
| GARCH family vs LSTM (Mahajan et al., 2022)     | Nifty 50 / India VIX                 | GARCH models slightly outperformed RNN-LSTM; EGARCH/TARCH best | Volatility (India VIX); MAPE/RMSE/MDA | [45]      |
| Transformer vs LSTM/Prophet (2022-25)           | Multiple indices                     | Transformer superior on long-range dependencies                | Varies by study                       | [16],[17] |

**Table 2. Evaluation metrics used in the reviewed literature.**

| Metric            | Type           | Scale-Dependent? | Frequency of Use | Note                                               |
|-------------------|----------------|------------------|------------------|----------------------------------------------------|
| RMSE              | Regression     | Yes              | Very high        | Penalises large errors; not cross-study comparable |
| MAE               | Regression     | Yes              | High             | Robust to outliers; scale-bound                    |
| MAPE              | Regression     | No (relative)    | High             | Comparable across scales; unstable near zero       |
| R <sup>2</sup>    | Regression     | No               | Moderate         | Variance explained; good for cross-study           |
| Accuracy / F1     | Classification | No               | Moderate         | Only valid for directional prediction              |
| Sharpe / Drawdown | Economic       | No               | Very low         | Captures real trading value; rarely reported       |

**Table 3. Characteristics of datasets used across the reviewed studies.**

| Dataset / Market        | Region | Frequency      | Typical Span | Used by                  |
|-------------------------|--------|----------------|--------------|--------------------------|
| S&P 500 / DJIA          | USA    | Daily/Intraday | 5-20 yrs     | [12],[19]                |
| CSI 300 / SSE           | China  | Daily          | ~10 yrs      | [21],[22]                |
| Nifty 50 / NSE          | India  | Daily          | 5-16 yrs     | [20],[14],[15],[45],[46] |
| BSE Sensex              | India  | Daily          | ~15 yrs      | [8]                      |
| NASDAQ / NEPSE / others | Mixed  | Daily/Intraday | Varies       | [17],[42]                |

**Table 4. Qualitative capability matrix by model family.**

| Model Family            | Nonlinearity  | Temporal Modeling   | Interpretability | Compute Cost   |
|-------------------------|---------------|---------------------|------------------|----------------|
| ARIMA / GARCH           | Low           | Linear / volatility | High             | Very low       |
| SVM / k-NN              | High (kernel) | Low                 | Moderate         | Low-moderate   |
| Decision Tree / RF      | High          | Low                 | Moderate-high    | Moderate       |
| LSTM / GRU / BiLSTM     | High          | High                | Low              | High           |
| CNN                     | High          | Moderate (local)    | Low              | Moderate       |
| Transformer / Attention | High          | High (global)       | Low-moderate     | Very high      |
| Hybrid + XAI            | High          | High                | Moderate-high    | High-very high |

#### 4.1 Bibliometric Overview

Figure 3 shows the distribution of the 52 references by publication year (42 reviewed studies together with 10 foundational and methodological works). The corpus is deliberately weighted toward recent work: more than two-thirds of the included studies were published in 2023-2025, reflecting the rapid acceleration of transformer, GRU, hybrid, and explainability research. The most frequently used datasets are US indices (S&P 500, DJIA) and Chinese indices (CSI 300, SSE), with Indian markets (Nifty

50, NSE, BSE Sensex) appearing primarily in 2024-2025 studies – quantitative evidence of the Indian-market gap this review highlights.

**Figure 3. Distribution of reviewed studies by publication period.**

| Year    | Publications in this review (relative)                                            | n  |
|---------|-----------------------------------------------------------------------------------|----|
| ≤2018   |  | 6  |
| 2019-21 |  | 6  |
| 2022    |  | 3  |
| 2023    |  | 12 |
| 2024    |  | 13 |
| 2025    |  | 12 |

## 5. Critical Evaluation

Three observations follow from the synthesis. First, hybrid and deep architectures consistently outperform standalone statistical and classical ML models within their respective studies [19],[20],[21],[22],[23], confirming the field's trajectory toward integrated nonlinear modeling. Second, the evidence base is fragmented: results are reported on incomparable datasets (S&P 500, CSI 300, Shanghai Composite, Nifty 50) using metrics on different scales, with directional accuracy and regression error often conflated [8],[9],[21]. This makes any global ranking of models unreliable and is the single greatest obstacle to cumulative progress. Third, economic-utility metrics – Sharpe ratio, maximum drawdown, backtested profitability – are rarely reported (Table 2), so statistical accuracy is not connected to real trading value [2],[29].

Two further gaps are especially relevant to a doctoral research agenda. The Indian equity market (Nifty 50 / NSE) is comparatively under-studied: although several recent works address it [14],[15],[40],[45],[46], the majority of high-profile hybrid and transformer studies use US or Chinese data, as Table 3 and Figure 3 make explicit. And explainability, despite mature tooling in SHAP [27] and LIME [49] and repeated calls in the literature [26],[28],[50], is almost never embedded in hybrid forecasting models. The combination of these gaps defines a clear and defensible research opportunity.

## 6. Discussion

The literature describes a clear evolution: from interpretable but rigid statistical models, through flexible but feature-dependent ML, to powerful but opaque DL and hybrid architectures. The dominant trend is hybridisation, and the dominant unsolved problem is the tension between accuracy and interpretability. Transformer and attention mechanisms partially address long-range dependency capture [16],[43], and GRU offers an efficient recurrent alternative [15], but neither resolves the transparency deficit.

A second theme is the neglect of domain knowledge and economic realism. Few models incorporate macroeconomic indicators or news sentiment systematically [24],[25],[48], and fewer still evaluate predictions through backtesting in realistic trading conditions [29]. Closing the gap between academic metrics and deployable systems requires both explainability and economically meaningful evaluation. The combination of an explainable hybrid model and economic-utility evaluation, validated on under-studied Indian markets, therefore represents a coherent and high-value research direction.

## 7. Threats to Validity

As with any literature review, several threats to validity apply. Selection bias: despite searching seven databases, relevant studies in non-indexed venues or other languages may have been missed; the inclusion window from 2001 also excludes some early work. Publication bias: studies reporting positive results are more likely to be published, potentially inflating the apparent superiority of hybrid models. Comparability bias: as emphasised throughout, heterogeneous datasets, scales, and metrics prevent precise quantitative meta-analysis, so the synthesis is necessarily narrative and qualitative for cross-study claims. Extraction bias: performance values were taken as reported by original authors and were not independently reproduced. These threats were mitigated by following PRISMA, applying explicit inclusion criteria, annotating every reported metric with its dataset and scale, and refraining from constructing a single ranked leaderboard. Readers should interpret all cross-study comparisons as indicative rather than definitive.

## 8. Key Findings

- (a) Statistical models (ARIMA, GARCH, EGARCH, SARIMA) remain valuable interpretable baselines but are insufficient alone for nonlinear markets [1],[2],[4],[31].
- (b) Classical ML (SVM, Random Forest) improves nonlinear handling, with RF frequently the strongest classical model, but depends on feature engineering [6],[7],[8],[9],[10].
- (c) Deep learning (LSTM, GRU, CNN, TCN) and Transformer/attention models capture temporal structure most effectively, with GRU competitive on Indian-market data [11],[12],[14],[15],[16].
- (d) Hybrid models (ARIMA-LSTM, CNN-BiLSTM-Attention, SARIMA-GARCH-CNN-BiLSTM) consistently outperform their components within their studies and represent the current state of the art [19],[20],[21],[22],[23].
- (e) The absence of standardised datasets, scales, and metrics prevents reliable cross-study comparison and is the field's central methodological weakness [21],[26],[29].
- (f) Economic-utility metrics (Sharpe ratio, drawdown, backtested return) are rarely reported, disconnecting accuracy from trading value [2],[29].
- (g) Explainability frameworks (SHAP, LIME) are mature but almost never integrated into hybrid stock-prediction models, and Indian equity markets remain under-studied – together defining the principal research gap [26],[27],[28],[49].

## 9. Conclusion and Research Direction

This review of 42 stock-prediction studies traced the evolution of stock price prediction from ARIMA and GARCH to deep, attention-based, and hybrid architectures. The evidence supports hybrid models as the state of the art while exposing three persistent gaps: fragmented and scale-inconsistent evaluation, the neglect of economic-utility metrics, and the near-absence of explainability – compounded by limited attention to Indian equity markets.

These gaps motivate a concrete research direction, which the present authors intend to pursue: the development of an explainable hybrid forecasting framework – integrating a linear statistical component (ARIMA) with a nonlinear deep component (LSTM or a Temporal Fusion Transformer) and a SHAP-based post-hoc explanation layer – validated on Nifty 50 / NSE equity data across multiple market regimes, and evaluated not only on RMSE and MAPE but also on directional accuracy and backtested economic utility. Such a framework would simultaneously address the interpretability deficit, the Indian-market gap, and the disconnect between statistical accuracy and practical trading value identified throughout this review. By combining state-of-the-art predictive performance with transparency, this direction aims to bridge the persistent gap between academic forecasting research and deployable, trustworthy financial decision-support systems.

## References

- [1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. Hoboken, NJ: Wiley, 2008.
- [2] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021. [Online]. Available: <https://otexts.com/fpp3/>
- [3] P. H. Vuong, L. H. Phu, T. H. V. Nguyen, L. N. Duy, P. T. Bao, and T. D. Trinh, "A bibliometric literature review of stock price forecasting: From statistical model to deep learning approach," *Sci. Prog.*, vol. 107, no. 1, 2024, doi: 10.1177/00368504241236557.
- [4] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987-1007, 1982, doi: 10.2307/1912773.
- [5] H. White, "Economic prediction using neural networks: The case of IBM daily stock returns," in *Proc. IEEE Int. Conf. Neural Networks*, vol. 2, San Diego, CA, 1988, pp. 451-458, doi: 10.1109/ICNN.1988.23959.
- [6] F. E. H. Tay and L. J. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, vol. 29, no. 4, pp. 309-317, 2001, doi: 10.1016/S0305-0483(01)00026-3.
- [7] K. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1-2, pp. 307-319, 2003, doi: 10.1016/S0925-2312(03)00372-2.
- [8] Y. Gu, "A comparative analysis study of stock prediction based on Random Forest and Decision Tree," in *Proc. 2024 Int. Conf. Electronics and Devices, Computational Science (ICEDCS)*, 2024, pp. 96-101, doi: 10.1109/ICEDCS64328.2024.00025.
- [9] A. H. Khan, A. Shah, A. Ali, R. Shahid, Z. U. Zahid, M. U. Sharif, et al., "A performance comparison of machine learning models for stock market prediction with a novel investment strategy," *PLOS ONE*, vol. 18, no. 9, art. e0286362, 2023, doi: 10.1371/journal.pone.0286362.
- [10] R. Panggabean et al., "Performance comparison of Random Forest Regression and SVR models in stock price prediction," *Jurnal Pilar Nusa Mandiri*, vol. 20, no. 2, 2024, doi: 10.33480/pilar.v20i2.6072. [VERIFY author/volume]
- [11] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *Proc. 17th IEEE Int. Conf. Machine Learning and Applications (ICMLA)*, Orlando, FL, 2018, pp. 1394-1401, doi: 10.1109/ICMLA.2018.00227.
- [12] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654-669, 2018, doi: 10.1016/j.ejor.2017.11.054.
- [13] M. Agrawal, P. K. Shukla, R. Nair, A. Nayyar, and M. Masud, "Stock prediction based on technical indicators using deep learning model," *Comput. Mater. Continua*, vol. 70, no. 1, pp. 287-304, 2022, doi: 10.32604/cmc.2022.014683.

- [14] M. Barua, T. Kumar, K. Raj, and A. M. Roy, "Comparative analysis of deep learning models for stock price prediction in the Indian market," *FinTech*, vol. 3, no. 4, pp. 551-568, 2024, doi: 10.3390/fintech3040029.
- [15] M. Kolambe and S. Arora, "Comparative analysis of LSTM variants for stock price forecasting on NSE India: GRU's dominance and enhancements," *Int. J. Inf. Technol. Comput. Sci.*, vol. 16, no. 6, pp. 43-60, 2024, doi: 10.5815/ijitcs.2024.06.04.
- [16] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep Transformer model," *Expert Syst. Appl.*, vol. 208, art. 118128, 2022, doi: 10.1016/j.eswa.2022.118128.
- [17] T. Muhammad, A. B. Aftab, M. M. Ahsan, M. M. Muhu, M. Ibrahim, S. I. Khan, et al., "Transformer-based deep learning model for stock price prediction: A case study on Bangladesh stock market," *Int. J. Comput. Intell. Appl.*, vol. 22, no. 3, art. 2350013, 2023, doi: 10.1142/S146902682350013X.
- [18] S. P. Kallimath, N. Darapaneni, and A. R. Paduri, "Deep learning approaches for stock price prediction: A comparative study on Nifty 50 dataset," *EAI Endorsed Trans. Intell. Syst. Mach. Learn. Appl.*, vol. 1, 2025, doi: 10.4108/eai.ismla.7481.
- [19] S. Kulshreshtha and A. Vijayalakshmi, "An ARIMA-LSTM hybrid model for stock market prediction using live data," *J. Eng. Sci. Technol. Rev.*, vol. 13, no. 4, pp. 117-123, 2020, doi: 10.25103/jestr.134.11.
- [20] D. N. Kalange, "Prediction of stock prices using LSTM-ARIMA hybrid deep learning model," *Asian J. Probab. Stat.*, vol. 27, no. 8, pp. 28-39, 2025, doi: 10.9734/ajpas/2025/v27i8791.
- [21] J. Zhang, L. Ye, and Y. Lai, "Stock price prediction using CNN-BiLSTM-Attention model," *Mathematics*, vol. 11, no. 9, art. 1985, 2023, doi: 10.3390/math11091985.
- [22] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 33, pp. 4741-4753, 2021, doi: 10.1007/s00521-020-05532-z.
- [23] E. Araya et al., "A hybrid GARCH and deep learning method for volatility prediction," *J. Appl. Math.*, vol. 2024, art. 6305525, 2024, doi: 10.1155/2024/6305525.
- [24] A. T. Haryono, R. Sarno, and K. R. Sungkono, "Transformer-gated recurrent unit method for predicting stock price based on news sentiments and technical indicators," *IEEE Access*, vol. 11, pp. 77132-77146, 2023, doi: 10.1109/ACCESS.2023.3298445.
- [25] P. Hajek and J. Novotny, "Beyond sentiment in stock price prediction: Integrating news sentiment and investor attention with Temporal Fusion Transformer," in *Artificial Intelligence Applications and Innovations (AIAI 2024)*, IFIP AICT, vol. 713, Springer, 2024, pp. 30-43, doi: 10.1007/978-3-031-63219-8\_3.
- [26] M. Cerneviene and A. Kabasinskas, "Explainable artificial intelligence (XAI) in finance: A systematic literature review," *Artif. Intell. Rev.*, vol. 57, art. 216, 2024, doi: 10.1007/s10462-024-10854-8.
- [27] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 4765-4774.
- [28] D. Muhammad, I. Ahmed, K. Naveed, and M. Bendeache, "An explainable deep learning approach for stock market trend prediction," *Heliyon*, vol. 10, no. 21, art. e40095, 2024, doi: 10.1016/j.heliyon.2024.e40095.
- [29] M. N. Ashtiani and B. Raahemi, "News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review," *Expert Syst. Appl.*, vol. 217, art. 119509, 2023, doi: 10.1016/j.eswa.2023.119509.
- [30] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: A systematic literature review," *Procedia Comput. Sci.*, vol. 216, pp. 96-102, 2023, doi: 10.1016/j.procs.2022.12.115.
- [31] P. Agyarko et al., "Hybrid model for stock market volatility," *J. Probab. Stat.*, vol. 2023, art. 6124649, 2023, doi: 10.1155/2023/6124649.
- [32] I. Dharma et al., "Volatility forecasting using GARCH versus EGARCH models for cryptocurrencies, Indonesian stocks, and U.S. stocks," *Integr. J. Bus. Econ.*, vol. 9, no. 2, 2025.
- [33] L. Rubio, A. Palacio Pinedo, A. Mejía Castaño, and F. Ramos, "Forecasting volatility by using wavelet transform, ARIMA and GARCH models," *Eurasian Econ. Rev.*, vol. 13, no. 3, pp. 803-830, 2023, doi: 10.1007/s40822-023-00243-x.
- [34] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," *Ann. Data Sci.*, vol. 10, no. 1, pp. 183-208, 2023, doi: 10.1007/s40745-021-00344-x.
- [35] J. San, "Stock forecasting and portfolio optimization based on ARIMA-GARCH, Random Forest and Monte Carlo models," in *Proc. ICDEBA 2024, Adv. Econ. Bus. Manag. Res.*, 2025, pp. 582-591.
- [36] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *Proc. 2017 ICACCI*, 2017, pp. 1643-1647, doi: 10.1109/ICACCI.2017.8126078.
- [37] V. Vincent, B. J. Saputra, S. Y. Izzatunnisa, H. Lucky, and I. A. Iswanto, "Stock market prediction system using LSTM with technical indicators as voters," in *Proc. 2023 4th Int. Conf. Artificial Intelligence and Data Sciences (AiDAS)*, 2023, pp. 229-234, doi: 10.1109/AiDAS60501.2023.10284674.
- [38] A. Yadav, C. K. Jha, and A. Sharan, "Optimizing LSTM for time series prediction in Indian stock market," *Procedia Comput. Sci.*, vol. 167, pp. 2091-2100, 2020, doi: 10.1016/j.procs.2020.03.257.
- [39] P. S. Sisodia, A. Gupta, Y. Kumar, and G. K. Ameta, "Stock market analysis and prediction for Nifty 50 using LSTM deep learning approach," in *Proc. 2022 2nd ICIPTM*, vol. 2, 2022, pp. 156-161, doi: 10.1109/ICIPTM54933.2022.9753956.
- [40] M. E. Karim, M. Foysal, and S. Das, "Stock price prediction using Bi-LSTM and GRU-based hybrid deep learning approach," in *Proc. ICDSA, Lecture Notes in Networks and Systems*, Springer, 2023, pp. 701-711, doi: 10.1007/978-981-19-3148-2\_60.
- [41] M. Lu and X. Xu, "TRNN: An efficient time-series recurrent neural network for stock price prediction," *Inf. Sci.*, vol. 657, art. 119951, 2024, doi: 10.1016/j.ins.2023.119951.
- [42] Y. Qian, "An enhanced Transformer framework with incremental learning for online stock price prediction," *PLOS ONE*, vol. 20, no. 1, art. e0316955, 2025, doi: 10.1371/journal.pone.0316955.
- [43] X. Li, S. Chen, X. Qiao, and F. Zhao, "Multi-perspective learning based on Transformer for stock price trend," *Int. J. Comput. Intell. Syst.*, vol. 18, art. 44, 2025, doi: 10.1007/s44196-025-00768-w.
- [44] S. Wang, "A stock price prediction method based on BiLSTM and improved Transformer," *IEEE Access*, vol. 11, pp. 104211-104223, 2023, doi: 10.1109/ACCESS.2023.3296537.
- [45] V. Mahajan, S. Thakan, and A. Malik, "Modeling and forecasting the volatility of NIFTY 50 using GARCH and RNN models," *Economies*, vol. 10, no. 5, art. 102, 2022, doi: 10.3390/economies10050102.

- [46] D. S. Kotha, T. K. Gundu, and U. M. G. Krishna, "Diving into financial forecasting: Analyzing stock price prediction with LSTM, GRU, and GAN for Nifty50 NSE index," in *Bio-Inspired Computing (IBICA 2023)*, LNNS, vol. 1229, Springer, 2025, doi: 10.1007/978-3-031-78940-3\_25.
- [47] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLOS ONE*, vol. 12, no. 7, art. e0180944, 2017, doi: 10.1371/journal.pone.0180944.
- [48] X. Zhang, S. Qu, J. Huang, B. Fang, and P. S. Yu, "Stock market prediction via multi-source multiple instance learning," *IEEE Access*, vol. 6, pp. 50720-50728, 2018, doi: 10.1109/ACCESS.2018.2869735.
- [49] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, 2016, pp. 1135-1144, doi: 10.1145/2939672.2939778.
- [50] W. J. Yeo, W. Van Der Heever, R. Mao, E. Cambria, R. Satapathy, and G. Mengaldo, "A comprehensive review on financial explainable AI," *Artif. Intell. Rev.*, vol. 58, art. 189, 2025, doi: 10.1007/s10462-024-11077-7.
- [51] L. Dowson, X. Song, and X. Li, "Explainable AI for stock price prediction: A comparative study of deep learning and feature attribution methods," in *Proc. 2025 11th Int. Conf. Communication and Information Processing (ICCIP)*, Lingshui, China, 2025, pp. 488-494, doi: 10.1145/3784833.3784870.
- [52] S. Mehtab and J. Sen, "A robust predictive model for stock price prediction using deep learning and natural language processing," in *Proc. BAICONF*, 2019, doi: 10.36227/techrxiv.15023361.

#### How do I cite this article?

Kulkarni Parag et.al, A Review of Statistical, Machine Learning, and Deep Learning Methods for Stock Price Prediction: Toward Explainable Hybrid Models, *Excel International Journal of Technology, Engineering and Management*, 2026; Volume -13, Issue-3\_Page\_28-37. DOI Link: <https://doi.org/10.64971/j.cph.ejtem.v13.i3.4.2026>



This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)