ChatGPT

A Critical Review of Machine Learning Methodologies for Stock Market Prediction

The stock market prediction problem sits at the intersection of finance and computer science. Classical finance theory, notably the Efficient Market Hypothesis (EMH), asserts that prices instantly reflect all available information, implying that returns follow a random walk and cannot be systematically forecasted [1]. In contrast, recent machine learning (ML) studies have uncovered persistent patterns and "pockets of predictability" (e.g. momentum, reversal, seasonal effects) that challenge pure EMH assumptions [2] [3]. For example, Murray *et al.* (2024) demonstrate that powerful ML models can predict future stock returns from price charts, calling into question EMH by showing charting and technical patterns retain predictive power [3]. These developments have been fueled by "rapid advancement of ML and DL [that] revolutionized stock market prediction" through the analysis of vast datasets [4]. This review examines how EMH frames the forecasting debate and how modern empirical evidence – including advanced ML techniques – finds exploitable patterns despite market efficiency.

Data Sources and Feature Engineering

Empirical stock-prediction models draw on heterogeneous data. **Technical indicators** derived from price and volume history (e.g. moving averages, momentum, volatility measures, and volume-based metrics) are a primary input [5] [6]. For example, Mostafavi *et al.* (2025) used 88 technical indicators (e.g. trend, momentum, volatility, volume) as features, finding that certain momentum and trend indicators are especially predictive of daily S&P 500 moves [5] [7]. **Fundamental data** (financial statements, valuation ratios, macroeconomic variables) are also widely used for longer-term forecasts. Fundamental analysis evaluates earnings, revenue, cash flows, etc., to estimate intrinsic value [8], and has been combined with technical factors to improve predictions. For instance, Xu *et al.* (2023) build models for airline stocks that include both company fundamentals and technical series (previous prices, sector index, trading volume) plus macro data like interest rates [9]. **Alternative data** such as news and social media sentiment, analyst ratings, and economic releases provide additional signals. Sentiment analysis has become common: studies show news headlines or tweets (e.g. Elon Musk's tweets affecting Tesla) can materially influence short-term prices [10]. Researchers often integrate textual features (sentiment scores, news embeddings) with price and fundamental data to enhance forecasts [11] [10].

- **Technical Indicators:** Price-derived features (moving averages, RSI, MACD, Bollinger Bands, etc.) are ubiquitous [5]. These are often computed in hundreds of variations; selecting the most relevant ones is an active research area [5] [12].
- **Fundamental Indicators:** Company metrics (earnings, cash flow, P/E, book value, etc.) and macro factors (GDP growth, interest rates) inform long-horizon models [8] [9].
- **Sentiment/Alternative Data:** News sentiment scores, social-media signals, and macro news are used to capture investor mood and unstructured information, complementing numeric features [11] [10].

Feature engineering and selection are crucial to reduce noise. Studies commonly apply filter (correlation tests), wrapper (recursive feature elimination), and embedded (LASSO, tree-based importance) methods to choose a compact feature set [13]. For example, Kumari & Swarnkar (2023) derived 83 technical indicators then fused multiple selection techniques (LASSO, information gain, forward selection) to isolate the most informative features [12]. Advanced approaches even employ dimensionality reduction (PCA) or meta-heuristic optimizers (genetic algorithms, Boruta) to construct the optimal input features for ML models [13] [14]. In general, combining diverse data sources (technical, fundamental, sentiment) and carefully curating features improves prediction accuracy, though many papers note the persistent challenge of overfitting due to high-dimensional inputs [13] [14].

Model Architectures

A wide range of ML models have been applied. Classical statistical models like ARIMA serve as baselines for time-series forecasting. In ML, **support vector machines (SVM)** and **kernel methods** were popular among earlier studies due to their ability to handle nonlinearity [15]. **Tree ensembles** (Random Forests, gradient boosting such as XGBoost/LightGBM) are now commonplace, prized for handling heterogenous features and providing feature-importance insight [15]. In fact, Saberironaghi *et al.* (2025) note that ensemble models (like gradient-boosted trees) and deep neural networks are "widely employed in the literature" for stock prediction [15]. **Deep learning** introduces neural architectures that learn complex patterns directly. Feed-forward and convolutional neural networks (CNNs) have been used for pattern recognition on price "charts" or multi-day image representations. Recurrent models, especially **Long Short-Term Memory (LSTM)** and **Gated Recurrent Unit (GRU)** networks, are extensively applied for sequence forecasting due to their ability to capture time dependencies. Indeed, Darwish *et al.* (2025) report that CNNs, RNNs (LSTM/BiLSTM/GRU) are "demonstrated to be significant" in forecasting stock trends [16].

Comparative performance: The literature lacks consensus on the single best model. Sophisticated deep/RL models can capture richer dynamics, but simpler methods often perform comparably. Some studies find that relatively simpler models (SVM or linear regressions) can outperform large neural nets on limited data [15]. For example, ensemble SVM+RF+XGBoost models have shown strong performance in stock-selection tasks [17], while LSTM excels at capturing long-term patterns but is sensitive to noise. Hybrid models (e.g. combining ARIMA with LSTM) and stacked/hybrid networks are frequently proposed to leverage both linear and nonlinear dynamics [18].

Emerging architectures: Transformers and large language models (LLMs) are recent entrants. These attention-based networks excel at capturing long-range dependencies and have shown promise when applied to financial time series or textual features. For instance, transformer networks have been used in deep RL trading models to model long-term price dependencies [19]. Darwish *et al.* (2025) observe that "the use of transformers and LLMs in finance has gained popularity" in recent years [20]. Newer hybrid models incorporate attention layers on top of LSTMs or integrate text embeddings from BERT-like models for sentiment. Nonetheless, empirical gains from Transformers are still an active research topic. Many authors urge caution: even LSTM/Transformer models "remain vulnerable" to rare events and require careful validation to avoid spurious patterns [21].

Overall, the architecture landscape spans **classical time-series methods (ARIMA)** to **shallow ML (SVM, decision trees)**, **ensemble learners (RF, XGBoost)**, **deep nets (CNN/RNN/LSTM/GRU)**, and **attention-based models (Transformers/LLMs)**. Reviews emphasize that ensemble and deep approaches are prevalent, but simpler SVMs/ensemble trees sometimes match or beat them depending on the scenario [15] [16]. The choice often depends on data availability, with deep models requiring large samples, and on the prediction target (direction vs return).

Evaluation and Backtesting Methodologies

Evaluating stock prediction models demands metrics and validation protocols beyond simple accuracy. Common academic metrics include the classification accuracy or RMSE, but rigorous studies also report financial performance measures. Important metrics are: cumulative or annualized return, **Sharpe ratio** (risk-adjusted return), **maximum drawdown**, win rate, and other risk metrics. For example, Li *et al.* (2019) use cumulative return, excess return, win rate, maximum drawdown, and Sharpe ratio to assess model-based portfolios [22]. Similarly, recent work benchmarks models by cumulative return, annualized Sharpe, and maximum drawdown against baseline strategies [23]. These financial metrics capture both profitability and stability of the strategy, which accuracy alone cannot measure.

Validation methodology is equally critical. Because stock data are autocorrelated and non-stationary, **time-series cross-validation** (such as rolling or expanding window "walk-forward" validation) is recommended [21]. Unlike random CV, walk-forward tests models on forward-moving segments of time to mimic live trading. Studies emphasize **out-of-sample testing** and even paper-trading stages to ensure robustness [24] [21]. For example, recent reviews note that models require "thorough backtesting, staged paper trading, and ongoing out-of-sample monitoring" for deployment [24].

Pitfalls abound if evaluation is naïve. **Overfitting** on historical data is a serious issue. Walk-forward validation and nested cross-validation are advised to mitigate look-ahead bias [21]. Another common flaw is **ignoring transaction costs and market impact**: many ML papers report unrealistic returns by assuming zero commissions. Analysts warn that costs must be included in backtests, as suggested in studies that explicitly incorporate transaction costs into model assessment [25]. Additionally, slippage, illiquidity, and changing market regimes can degrade real-world performance relative to in-sample results. Academically accepted evaluation, therefore, goes beyond accuracy: it focuses on realistic trading simulations, risk-adjusted performance, and statistically sound validation protocols (e.g. walk-forward with fixed training window, and multiple market scenarios) [24] [21].

Common Pitfalls and Future Directions

Across the literature, several recurring issues emerge. A major pitfall is **data snooping**: many studies test numerous models and select the best ex post, leading to over-optimistic claims. Similarly, using overlapping time windows or improperly scaled features can leak future information. Another concern is **short study horizons**: many papers evaluate on only a few years of data, which may not capture different market cycles [15]. Research highlights that training on a single index or short span often produces overfit models.

Moreover, the **failure to account for trading frictions** (costs, slippage, liquidity) means reported Sharpe ratios and returns are typically overstated [25]. Rigorous studies explicitly simulate transaction costs and adjust for multiple comparisons. Feature stability is also an issue: models often deteriorate when reused in out-of-sample periods, underscoring the market's adaptiveness.

Future research is steering toward more robust frameworks. Multi-modal models that fuse market data with text (via news or social media) are advancing. Also, explainability methods (SHAP values, attention visualization) are being explored to make ML models more transparent. Some suggest **causal learning** to distinguish real signals from spurious correlations. Finally, continued development of evaluation standards – e.g. public benchmarks and walk-forward contests – is needed to judge progress.

In summary, the machine-learning frontier has pushed the boundaries of stock prediction beyond the strict EMH paradigm. Modern studies leverage rich feature sets and powerful algorithms (from ensembles to deep nets) [15] [16]. Results indicate that while no "silver bullet" exists, certain patterns and models do capture exploitable structure [3] [2]. The most successful approaches carefully engineer features from multiple data domains, use architectures suited to time series (e.g. LSTM/Transformer), and evaluate strategies using proper backtesting with risk metrics [22] [24]. Yet the literature also repeatedly cautions that overly optimistic findings are common, and that robust out-of-sample tests – including accounting for costs – are essential for credible results [21] [25].

Figure: Taxonomy of Stock-Forecasting Approaches. Stock prediction methods can be grouped into traditional numeric ML/DL approaches (e.g. feature engineering, ensemble learning, RL/ MARL, sentiment analysis) versus emerging LLM-based techniques [26].

Sources: Authoritative reviews and empirical studies are cited throughout: market efficiency and predictability [1] [2] [3], data and features [5] [6] [13], model classes [15] [16] [20], and evaluation methods [22] [24] [25]. These sources represent recent high-impact journal and conference publications (2015–2025) in quantitative finance and machine learning.

---

[1] [15] Machine learning, stock market forecasting, and market efficiency: a comparative study | International Journal of Data Science and Analytics
https://link.springer.com/article/10.1007/s41060-025-00854-4

[2] [20] [21] [23] [24] Increase Alpha: Performance and Risk of an AI-Driven Trading Framework
https://arxiv.org/html/2509.16707v1

[3] Charting by machines
https://ideas.repec.org/a/eee/jfinec/v153y2024ics0304405x2400014x.html

[4] Stock Market Prediction Using Machine Learning and Deep Learning Techniques: A Review
https://www.mdpi.com/2673-9909/5/3/76

[5] [7] Key technical indicators for stock market prediction
https://spiral.imperial.ac.uk/entities/publication/84e28dee-de1a-452d-94b1-4a070c5ec1ed

[6] [8] [9] [10] [11] [12] [13] [14] [16] [18] [19] [25] [26] Stock Market Forecasting: From Traditional Predictive Models to Large Language Models | Computational Economics
https://link.springer.com/article/10.1007/s10614-025-11024-w

[17] [22] Stock Prediction Based on Machine Learning Models
https://www.atlantis-press.com/article/126015271.pdf