International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Stock Price Prediction Using Machine Learning: A Comprehensive Review and Comparative Analysis

Saransh Singh
SRM Institute of Science and Technology,
Delhi NCR ss7238@srmist.edu.in

Abstract—Stock price prediction remains one of the most challenging tasks in financial forecasting due to the inherent volatility and nonlinearity of financial markets. This paper presents a comprehensive review of machine learning approaches for stock price prediction, comparing traditional statistical methods with modern deep learning architectures. We analyze various algorithms including Support Vector Machines (SVM), Random Forest, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks. Our comparative analysis reveals that LSTM-based deep learning models achieve superior performance with prediction accuracies ranging from 85% to 89%, significantly outperforming traditional regression methods. The study examines key technical indicators, feature engineering techniques, and model optimization strategies that enhance prediction accuracy. We also discuss the challenges of market volatility, overfitting, and computational complexity. The findings demonstrate that sophisticated neural network architectures, particularly LSTM networks, offer more reliable predictions for investors and financial analysts in volatile market conditions. This research contributes to the growing body of knowledge on intelligent financial forecasting systems and provides practical insights for implementing machine learning-based trading strategies.

Index Terms—Stock price prediction, machine learning, LSTM, deep learning, financial forecasting, technical analysis, time series prediction

I. Introduction

STOCK market prediction has long fascinated researchers, investors, and financial analysts due to its potential for generating substantial returns and managing investment risks. The stock market is a complex, dynamic system influenced by numerous factors including economic indicators, company performance, investor sentiment, political events, and global market trends[1]. Traditional statistical methods such as autoregressive integrated moving average (ARIMA) and exponential smoothing have been widely used for time series forecasting but often fail to capture the nonlinear relationships and complex patterns inherent in financial data[2].

The advent of machine learning and deep learning technologies has revolutionized stock price prediction methodologies. These advanced techniques possess the capability to process vast amounts of historical data, identify hidden patterns, and make predictions with higher accuracy than conventional approaches[3]. Machine learning algorithms can automatically extract features from raw data, handle nonlinearity, and adapt to changing market conditions, making them particularly suitable for financial market forecasting[4].

This paper presents a comprehensive analysis of various machine learning approaches for stock price prediction. We examine both traditional machine learning algorithms and state-of-the-art deep learning architectures, comparing their performance, strengths, and limitations. The primary objectives of this research are:

- To review and categorize machine learning techniques used for stock price prediction
- To analyze the performance of different algorithms based on recent empirical studies

- To identify key technical indicators and features that enhance prediction accuracy
- To discuss challenges and future directions in machine learning-based financial forecasting

The remainder of this paper is organized as follows: Section II reviews related work in stock price prediction. Section III describes various machine learning methodologies. Section IV presents comparative analysis of different approaches. Section V discusses challenges and limitations. Section VI concludes the paper with future research directions.

II. Related Work

A. Traditional Statistical Methods

Traditional time series forecasting methods have been the foundation of stock price prediction for decades. ARIMA models and their variants have been extensively used to capture temporal dependencies in financial data. However, these methods assume linearity and stationarity, which often do not hold in real-world stock markets[2]. The Efficient Market Hypothesis (EMH) suggests that stock prices reflect all available information, making prediction theoretically impossible, yet empirical evidence shows that machine learning can identify exploitable patterns[5].

B. Machine Learning Approaches

The application of machine learning to stock price prediction gained momentum in the early 2000s. Patel et al.[1] demonstrated that Artificial Neural Networks (ANN) could achieve prediction accuracy of approximately 85%, outperforming conventional statistical techniques. However, their study noted that ANN models are computationally demanding and require substantial hyperparameter tuning.

Support Vector Machines (SVM) have shown promise in handling nonlinear data and identifying complex associations in stock markets[1]. Recent research by comparative studies indicates that SVM with Radial Basis Function (RBF) kernel can achieve accuracy rates up to 88% when combined with insider trading information and technical indicators[3]. Random Forest algorithms have also demonstrated effectiveness, particularly when ensemble methods are employed to combine multiple prediction models[6].

C. Deep Learning Revolution

The introduction of deep learning architectures has significantly advanced stock price prediction capabilities. Long Short-Term Memory (LSTM) networks, a specialized form of Recurrent Neural Networks (RNN), have emerged as particularly effective for financial time series forecasting[7].
LSTM networks can capture long-term dependencies in sequential data, making them ideal for modeling temporal patterns in stock prices.

Recent studies demonstrate that LSTM models substantially enhance prediction accuracy compared to other machine learning methodologies and traditional investment strategies[2]. Research by various authors shows LSTM achieving accuracies of 89.34%, with precision of 90.21% and recall of 88.76%, delivering 15.78% better results than Linear Regression and 6.91% better than Random Forest[7]. The superior performance of LSTM is attributed to its ability to handle temporal dependencies and maintain information across long sequences.

Deep learning's nonlinear modeling capabilities and automatic feature extraction make it better suited to handle complex nonlinear relationships and large-scale data in financial markets[4]. Studies have shown that combining deep learning with big data analytics can significantly improve prediction accuracy and stability in financial market forecasting[4].

D. Technical Analysis Integration

Many studies have explored the integration of technical analysis indicators with machine learning models. Common indicators include Bollinger Bands, Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Moving Averages (MA), and Stochastic Momentum[8]. These indicators help capture market volatility patterns and momentum, providing valuable features for prediction models.

III. Methodology

A. Data Collection and Preprocessing

Stock price prediction requires comprehensive historical data including:

- Opening price, closing price, high, and low prices

- Trading volume and market capitalization
- Technical indicators (MACD, RSI, MA, Bollinger Bands)
- Economic indicators (interest rates, inflation, GDP growth)
- Sentiment data from news and social media

Data preprocessing is crucial for model performance. This includes handling missing values, normalizing data to a common scale (typically 0-1 range using Min-Max scaling), removing outliers, and creating time-lagged features to capture temporal dependencies.

B. Support Vector Regression (SVR)

Support Vector Regression is a powerful machine learning technique renowned for its capacity to handle nonlinear data. Unlike conventional regression models, SVR seeks to identify the optimal hyperplane that minimizes prediction errors within a specified margin[1]. The algorithm maps stock market data into higher-dimensional spaces using kernel functions:

- Linear kernel: Suitable for linearly separable data
- Polynomial kernel: Captures polynomial relationships
- Radial Basis Function (RBF): Most commonly used for stock prediction, handles complex nonlinear patterns

The SVR optimization problem can be formulated as:

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

subject to:

$$y_i - (w^T\phi(x_i) + b) \leq \epsilon + \xi_i$$

$$(w^T\phi(x_i) + b) - y_i \leq \epsilon + \xi_i^*$$

where C is the regularization parameter, ϵ is the insensitive loss parameter, and ξ_i, ξ^* are slack variables.

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction. For stock price prediction, Random Forest offers several advantages:

- Handles high-dimensional feature spaces effectively
- Provides feature importance rankings
- Reduces overfitting through ensemble averaging
- Robust to outliers and noise

The algorithm's prediction is given by:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} f_t(x)$$

where T is the number of trees and $f_t(x)$ is the prediction of the t-th tree.

D. Artificial Neural Networks (ANN)

ANNs are computational models inspired by biological neural networks. A typical feedforward ANN for stock prediction consists of:

- Input layer: Receives normalized features (technical indicators, price data)
- Hidden layers: Extract complex patterns through nonlinear transformations
- Output layer: Produces predicted stock price or price movement direction

The activation function commonly used is the Rectified Linear Unit (ReLU) for hidden layers and linear activation for regression output:

$$h = g(Wx + b)$$

$$\hat{y} = W_o h + b_o$$

C. Random Forest

where g is the activation function, W and W_o are weight matrices, and b and b_o are bias vectors.

E. Long Short-Term Memory (LSTM) Networks

LSTM networks represent the state-of-the-art approach for sequential data prediction. Unlike traditional RNNs, LSTM networks can learn long-term dependencies through specialized gating mechanisms:

- Forget gate: Determines what information to discard from cell state
- Input gate: Decides which new information to store in cell state
- Output gate: Controls what information to output based on cell state

The LSTM cell update equations are:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

where f_t, i_t, and o_t are the forget, input, and output gates respectively, C_t is the cell state, h_t is the hidden state, and σ is the sigmoid function.

F. Model Evaluation Metrics

Several metrics are used to evaluate prediction performance:

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Accuracy, Precision, Recall, and F1-Score for classification-based approaches that predict price movement direction.

A. Performance Comparison

Algorithm	Accuracy	Precision	Recall
Linear Regression	68-73%	70.5%	69.2%
Random Forest	82-85%	84.3%	81.5%
SVM (RBF Kernel)	85-88%	87.1%	85.8%
ANN	77-85%	82.6%	80.3%
LSTM	85-89.34%	90.21%	88.76 %

Table 1: Comparison of machine learning algorithms for stock price prediction based on recent studies[1][3][7]

The comparative analysis reveals several key findings:

1. **LSTM Superiority**: LSTM networks demonstrate the highest prediction accuracy (89.34%) and precision (90.21%), significantly outperforming traditional methods[7]
2. **SVM Effectiveness**: SVM with RBF kernel achieves competitive accuracy (88%), particularly effective when combined with insider trading information[3]
3. **Random Forest Reliability**: Random Forest provides robust performance (82-85%) with good interpretability through feature importance analysis[6]
4. **ANN Performance Variance**: ANN accuracy varies widely (77-85%) depending on architecture and hyperparameter tuning[1]
5. **Linear Regression Limitations**: Traditional regression methods show the lowest

IV. Comparative Analysis

accuracy (68-73%), confirming their inadequacy for capturing nonlinear market dynamics[7]

B. Computational Complexity

Algorithm	Training Time	Inference Speed
Linear Regression	Low	Very Fast
Random Forest	Medium	Fast
SVM	High	Medium
ANN	High	Fast
LSTM	Very High	Medium

Table 2: Computational complexity comparison of different algorithms

While LSTM networks offer superior accuracy, they require substantial computational resources for training. The trade-off between accuracy and computational efficiency must be considered based on application requirements and available resources.

C. Feature Importance Analysis

Studies indicate that certain features contribute more significantly to prediction accuracy:

- Historical closing prices (highest importance)
- Trading volume
- Moving averages (50-day, 200-day)
- MACD and RSI indicators
- Market volatility indices
- Sector-specific economic indicators

Random Forest models provide explicit feature importance rankings, while deep learning models learn feature representations automatically through hidden layers[6].

V. Challenges and Limitations

A. Market Volatility and Noise

Stock markets are inherently volatile and influenced by unpredictable events such as political instability, natural disasters, and sudden economic shifts. Machine learning models trained on historical data may fail to predict prices during unprecedented market conditions or "black swan" events[4].

B. Overfitting

Complex models, particularly deep neural networks, are prone to overfitting when trained on limited data. Regularization techniques such as dropout, L1/L2 regularization, and early stopping are essential to prevent models from memorizing training data rather than learning generalizable patterns[7].

C. Non-Stationarity

Financial time series are non-stationary, meaning their statistical properties change over time. This violates assumptions of many traditional statistical methods and requires models to continuously adapt to evolving market conditions[8].

D. Data Quality and Availability

Accurate predictions require high-quality, comprehensive data. Missing values, data errors, and delayed information can significantly impact model performance. Additionally, access to alternative data sources (sentiment analysis, news feeds) may be limited or expensive[9].

E. Computational Resources

Deep learning models, particularly LSTM networks, require substantial computational power for training and hyperparameter optimization. This may limit their accessibility for individual investors or small firms without adequate infrastructure[1].

F. Interpretability

While deep learning models achieve high accuracy, they function as "black boxes," making it difficult to understand the reasoning behind specific predictions. This lack of interpretability poses challenges for regulatory compliance and investor trust[4].

VI. Future Directions

A. Ensemble Methods

Combining multiple models through ensemble techniques can leverage the strengths of different algorithms while mitigating individual weaknesses. Hybrid approaches that integrate LSTM with attention mechanisms or transformer architectures show promising results[6].

B. Alternative Data Integration

Incorporating alternative data sources such as social media sentiment, news sentiment analysis, satellite imagery, and web traffic can provide additional signals for prediction models. Natural Language Processing (NLP) techniques can extract valuable insights from textual data[8].

C. Explainable AI

Developing interpretable machine learning models that can explain their predictions is crucial for building trust and meeting regulatory requirements. Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) can enhance model transparency[9].

D. Transfer Learning

Transfer learning approaches that leverage pre-trained models on large financial datasets can improve prediction accuracy, especially for stocks with limited historical data or newly listed companies.

E. Real-Time Prediction Systems

Developing real-time prediction systems that can process streaming data and adapt to market changes dynamically represents an important frontier. This requires efficient algorithms and scalable infrastructure.

F. Risk Management Integration

Future research should focus on integrating prediction models with comprehensive risk management frameworks, incorporating portfolio optimization, risk assessment, and automated trading strategies.

VII. Conclusion

This paper has presented a comprehensive review of machine learning approaches for stock price prediction, comparing traditional statistical methods with modern deep learning architectures. Our analysis demonstrates that LSTM-based deep learning models achieve superior performance, with prediction accuracies reaching 89.34%, significantly outperforming traditional regression methods and other machine learning algorithms[7].

The key findings of this research include:

- LSTM networks excel at capturing temporal dependencies and long-term patterns in financial time series, making them the most effective approach for stock price prediction
- SVM with RBF kernel and Random Forest provide competitive accuracy with lower computational requirements, offering practical alternatives for resource-constrained applications
- Integration of technical indicators (MACD, RSI, Moving Averages) substantially enhances prediction accuracy across all algorithms
- Despite high accuracy, challenges remain in handling market volatility, preventing overfitting, and ensuring model interpretability

The findings underscore the potential of sophisticated neural network architectures in yielding more dependable predictions within the inherently volatile realm of stock market forecasting[2]. However, successful implementation requires careful consideration of data quality, feature engineering, model selection, and computational resources.

Machine learning-based stock price prediction offers valuable decision-making assistance for investors and financial analysts by improving risk management and supporting informed investment decisions[7]. As financial markets continue to evolve and alternative data sources become more accessible, the integration of advanced machine learning techniques with domain expertise will play an increasingly important role in financial forecasting.

Future research should focus on developing explainable AI models, integrating alternative data sources, and creating real-time adaptive systems that can respond to rapidly changing market conditions. The convergence of big data, deep learning, and financial analytics promises continued advancement in the accuracy and reliability of stock price prediction systems.

References

[1] Patel et al., "Stock Market Price Prediction using Machine Learning," *International Journal of Engineering Research & Technology*, May 2025. [Online]. Available: https://www.ijert.org/stock-market-price-prediction-using-machine-learning

[2] "LSTM Price Movement Prediction for Stock Market," *Granthaalayah*, vol. 13, no. 4, Apr. 2025. [Online]. Available: https://www.granthaalayahpublication.org/journals/granthaalayah/article/view/6058

[3] "A Comparative Study of Machine Learning Algorithms for Stock Price Prediction," *arXiv preprint*, Feb. 2025. [Online]. Available: https://arxiv.org/html/2502.08728v1

[4] "Boosting Financial Market Prediction Accuracy With Deep Learning and Big Data," *IGI Global*, Dec. 2023. [Online]. Available: https://www.igi-global.com/article/boosting-financial-market-prediction-accuracy-with-deep-learning-and-big-data/358454

[5] J. Malkiel, *A Random Walk Down Wall Street*. Norton & Company, 2019.

[6] "Stock Market Forecasting with Machine Learning," *SCITEPRESS Digital Library*, 2025. [Online]. Available: https://www.scitepress.org/Papers/2025/139424/139424.pdf

[7] "Prediction of Stock Prices using LSTM Model," *International Journal for Research in Applied Science and Engineering Technology*, 2025. [Online]. Available: https://www.ijraset.com/research-paper/prediction-of-stock-prices-using-lstm-model

[8] T. Phuoc et al., "Applying machine learning algorithms to predict the stock price trend in the stock market in an emerging economy," *Scientific Reports*, vol. 14, Mar. 2024. [Online]. Available: https://www.nature.com/articles/s41599-024-02807-x

[9] Y. Sun et al., "Deep Learning for Intelligent Assessment of Financial Market Volatility," *PMC*, Oct. 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9578860/

[10] "Market Stock Price Prediction Using Machine Learning with PSO and LS-SVM," *Scribd*, Feb. 2026. [Online]. Available: https://www.scribd.com/document/828102152/d-2203032024.