Archives available at journals.mriindia.com

Open Access International Journal of Science and Engineering

ISSN: 2456-3293

Volume 9 Issue 05, 2026

Forecasting Financial Markets with Hybrid Deep Learning: Evidence from ARIMA–LSTM and GARCH–LSTM Models

Dhananjay N. Kalange
Associate Professor in Statistics, Arts, Commerce and Science College, Palus
Email: dnkalange@gmail.com

Peer Review Information

Submission: 10 April 2026

Revision: 26 April 2026

Acceptance: 05 May 2026

Keywords

Hybrid Forecasting, Financial Time Series, ARIMA-LSTM, GARCH-LSTM, Deep Learning, Stock Market Prediction

Abstract

Chronic Kidney Disease (CKD) poses a major global health burden due to its gradual onset and often silent progression. Traditional diagnostic methods, based on a limited set of laboratory markers, may delay detection until significant kidney damage has occurred. Machine learning (ML) offers promise for early detection by analyzing complex, multi-dimensional patient data to identify subtle patterns indicating early kidney dysfunction. In this study, we evaluate several ML classifiers — including Logistic Regression, Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GB), and k-Nearest Neighbors (KNN) — on publicly available clinical datasets. After preprocessing, feature normalization, inconsistency handling and class balancing, models are trained and evaluated. The experimental results show that ensemble-based methods outperform individual classifiers, with Random Forest achieving the highest accuracy (≈ 98.6%) and robustness to noisy clinical data. These results underscore the potential of ML-based diagnostic tools to support early CKD screening, enabling timely medical intervention and improved patient outcomes.

Introduction

Financial markets are inherently complex, characterized by volatility, uncertainty, and nonlinear dynamics. Among the numerous challenges in finance, forecasting stock prices remains one of the most critical and extensively studied problems. Stock prices are influenced by macroeconomic conditions, investor sentiment, corporate performance, and global events (Fama, 1970; Shiller, 2003). Accurate prediction of stock movements can yield significant benefits for portfolio management, risk mitigation, and policy formulation. Over the decades, researchers have developed diverse modeling approaches - ranging from classical statistical techniques to modern machine learning (ML) methods - to capture the intricate patterns of financial time series (Tsay, 2010; Sezer et al., 2020). However, despite notable progress, no single approach has proven universally effective, motivating the exploration of hybrid models that combine the strengths of both paradigms.

1. Importance of Stock Prices Forecasting

Stock price forecasting plays a pivotal role in the functioning of financial systems. For investors, precise forecasts can guide buy–sell decisions and improve returns. For institutions, they support portfolio diversification, derivative pricing, and hedging strategies. Policymakers also rely on market predictions to gauge economic conditions and detect early signals of financial instability (Campbell et al., 1997). The inherently volatile nature of stock prices - driven by both predictable cycles and unpredictable shocks - makes forecasting a challenging yet indispensable task. With the increasing availability of high-frequency trading data and advanced computational tools, the demand for robust forecasting methods has grown substantially (Atsalakis & Valavanis, 2009).

2. Statistical *versus* Machine Learning Models

Classical statistical models such as the Autoregressive Integrated Moving Average (ARIMA) (Box & Jenkins, 1970) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Engle, 1982; Bollerslev, 1986) frameworks have been foundational in financial econometrics. ARIMA is particularly effective for modeling linear dependencies and short-term trends, while GARCH excels at capturing volatility clustering and time-varying variance. Their advantages lie in interpretability, mathematical rigor, and parsimony. However, these models struggle with nonlinear structures, long memory effects, and sudden market shocks, limiting their applicability in increasingly complex financial environments (Tsay, 2010). Machine learning methods - especially deep learning models like Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) and Convolutional Neural Networks (CNNs) (LeCun et al., 1998) - offer complementary strengths. LSTMs are designed to model long-term temporal dependencies, while CNNs can detect local patterns and feature interactions. These models have demonstrated high predictive performance on nonlinear and high-dimensional datasets (Fischer & Krauss, 2018; Sezer et al., 2020). Yet, they are often criticized for being black-box models, prone to overfitting, and dependent on large amounts of training data. Moreover, their limited interpretability makes them less attractive in contexts where transparency is essential (Doshi-Velez & Kim, 2017).

3. Motivation for Hybrid ARIMA–LSTM and GARCH–LSTM Models

To overcome the limitations of standalone models, researchers have increasingly turned to hybrid approaches that integrate statistical and ML methods. The intuition is straightforward: financial time series often exhibit both linear and nonlinear components and a single model may fail to capture this dual nature (Zhang, 2003).

- ARIMA–LSTM hybrids decompose the time series into linear and residual components, with ARIMA modeling the linear structure and LSTM capturing nonlinear dynamics (Nelson et al., 2017).
- GARCH–LSTM hybrids combine GARCH's ability to model Heteroskedastic volatility with LSTM's capacity to learn nonlinear and long-term temporal patterns (Bao et al., 2017).

By leveraging the synergy between traditional and modern techniques, hybrid models aim to improve forecasting accuracy, robustness, and reliability. They not only enhance predictive power but also offer a more comprehensive representation of stock market behavior, thereby contributing to better decision-making for investors, institutions, and policymakers.

Review of Literature

The forecasting of financial time series has progressed along two broad lines:

- Statistical/Econometric models that emphasize interpretability and explicit modeling of serial dependence and volatility (e.g., ARIMA, GARCH family), and
- Machine-learning and deep-learning methods that flexibly learn nonlinearity and complex temporal patterns (e.g., LSTM, CNN).

Hybrid approaches that combine the two aims to capture linear structure and volatility effects with the statistical component while letting the machine-learning component model residual nonlinear dynamics are often producing better empirical forecasts than either approach alone.

1. ARIMA and GARCH in finance

ARIMA (Autoregressive Integrated Moving Average) models have long been used for modeling and short-term forecasting of price and return series because of their explicit handling of autocorrelation and differencing for nonstationarity; ARIMA models are easy to estimate and interpret, making them a baseline in many studies. However, ARIMA is linear and therefore limited when the data contain nonlinear dynamics. The seminal hybrid idea of using ARIMA for linear parts and an NN for nonlinear residuals dates back to Zhang (2003), which formalized combining ARIMA and neural networks to exploit complementary strengths. Separately, volatility clustering and time-varying conditional variance in returns are modeled using ARCH/GARCH and their extensions (EGARCH, TGARCH, GJR, etc.). GARCH-family models remain standard for volatility forecasting, risk measures (VaR), and for providing heteroskedastic residuals that hybrid learners can exploit. Hybrid studies often use GARCH (or GARCH residuals) with deep learners to improve volatility/realized-variance forecasts.

2. LSTM in stock prediction

Long Short-Term Memory (LSTM) networks are specialized recurrent neural networks that handle long-range dependencies and vanishing gradients, making them well suited for sequential financial data. Empirical studies show LSTM often outperforms classical linear models (ARIMA/GARCH) for price-level or return-level forecasting when sufficient data/features are available, especially at daily or intraday frequencies. Fischer & Krauss (2018) is a high-impact demonstration of LSTM (and other deep approaches) applied to equity returns, showing strong out-of-sample performance in certain settings while highlighting issues such as overfitting and regime sensitivity.

LSTMs' principal drawbacks are: (1) they are data-hungry, (2) they are often "black-box" (low interpretability), and (3) performance can degrade when structural breaks or nonstationarity are present. These limitations motivate hybrids where a statistical model handles the stable linear/volatility structure and the LSTM models remaining nonlinear residual structure.

3. Existing hybrid models

The hybrid modeling literature is now mature and diverse. The canonical line (ARIMA + NN residual modeling) originates with Zhang (2003), who combined ARIMA with feed-forward ANNs and found gains in forecast accuracy on several series. Later work replaced generic ANNs with RNN/LSTM modules (ARIMA+LSTM) to better capture sequential nonlinear dependencies; many empirical studies (Choi 2018) report ARIMA-LSTM hybrids achieving lower RMSE/MAPE than pure ARIMA or pure LSTM on stock/index datasets. For volatility forecasting and risk (VaR), hybrids integrating GARCH with LSTM (or GRU) have been proposed: the GARCH component models conditional Heteroskedasticity while the LSTM captures nonlinear/time-varying patterns in the volatility or in external covariates (e.g., sentiment). Kim & Won (2018) and subsequent studies (Hu 2020; various 2023–2025 papers; see summary Table no. 1 below) document performance improvements in volatility and VaR tasks from GARCH-LSTM hybrids.

Table 1: Summary of Recent Hybrid models

Study	**Hybrid model**	**Data**	**Key finding**
Zhang (2003)	ARIMA + ANN	Various economic & exchange series	Proposed ARIMA+ANN framework — improved accuracy vs ARIMA/ANN alone.
Fischer & Krauss (2018)	LSTM & other DL models (benchmarking)	US equities, daily	LSTM models can beat classic baselines in certain settings; careful feature selection & regularization required.
Choi (2018) (arXiv)	ARIMA + LSTM	Multiple stock indices	ARIMA-LSTM shows lower RMSE than standalones on tested indices.
Kim & Won (2018)	LSTM + multiple GARCH variants	Stock index volatilities	Hybrid GARCH + LSTM improved volatility forecasts vs single GARCH.
Hu et al. (2020)	LSTM + GARCH (or LSTM+ANN+GARCH)	Copper price volatility	Integrating GARCH with LSTM/ANN produced better volatility forecasts.
Shah (2022)	Survey of ARIMA, LSTM, CNN hybrids	Multiple studies	Comprehensive review: hybrid deep-learning approaches (ARIMA+LSTM, CNN+LSTM, wavelet + hybrids) show consistent empirical gains; calls for standardized benchmarks.
Recent 2023–2025 studies	Wavelet + ARIMA + LSTM; sentiment-augmented GARCH-LSTM	Indices, equities with sentiment features	Newer hybrids add pre-filtering (wavelet), external features (sentiment) and still report improved accuracy — but results depend on sample and tuning.

Recent variants add wavelet transforms, external features (macro variables, sentiment), or use multistage pipelines (e.g., wavelet → ARIMA → LSTM). Finally, systematic reviews highlight that hybrid architectures (ARIMA/LSTM, Wavelet + ARIMA + LSTM, GARCH + LSTM, CNN+LSTM, etc.) tend to outperform single-method baselines in most reported experiments - although study designs (different datasets, horizons, evaluation metrics, and hyperparameters tuning) vary widely, so careful benchmarking on common datasets is still needed.

Methodology

This section outlines the dataset, preprocessing steps, proposed hybrid models, and evaluation criteria. The focus is on developing hybrid ARIMA-LSTM and GARCH-LSTM frameworks for forecasting financial time series.

1. Data

The dataset consists of daily closing prices from four financial markets: Bombay Stock Exchange (BSE), National Stock Exchange of India (NSE), Standard & Poor's 500 Index (S&P 500), and Bitcoin, spanning the period January 2000 to January 2025. These markets are chosen to capture a mix of emerging (BSE), developed (S\&P 500), and digital asset (Bitcoin) financial environments, thereby testing the generalizability of the proposed models.

2. Preprocessing

To stabilize variance and reduce nonstationarity, logarithmic returns are computed as:

$$r_t = log\left(\frac{P_t}{P_{t-1}}\right)$$

where P_t is the daily closing price at time t.

Stationarity of the return series is assessed using Augmented Dickey--Fuller (ADF) and Kwiatkowski--Phillips--Schmidt--Shin (KPSS) tests. Differencing is applied if necessary to ensure weak stationarity before model estimation.

3. Mathematical Formulation of ARIMA–LSTM Hybrid Model

Let r_t denote a univariate financial time series representing log-returns. The ARIMA component captures the linear structure in the return series.

An ARIMA(p, d, q) process is given by:

$$\varphi(B)(1-B)^d\, r_t = \theta(B)\varepsilon_t,\ \ \varepsilon_t \sim WN(0,\sigma^2) \quad \text{---------} \ (1)$$

The fitted ARIMA model produces linear forecasts $\hat{r}_t^{ARIMA}$ and residuals

$$e_t = r_t - \hat{r}_t^{ARIMA}$$

These residuals capture nonlinear dynamics and are modeled using an LSTM network.

An LSTM cell is defined as:

$$f_t = \sigma(W_f\,[h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i\,[h_{t-1}, x_t] + b_i)$$
$$\tilde{c}_t = tanh(W_c\,[h_{t-1}, x_t] + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$o_t = \sigma(W_o\,[h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \odot tanh(c_t)$$

where $f_t, i_t and\ o_t$ are forget, input, and output gates, c_t is the cell state, h_t the hidden state, and $\sigma(.)$ the sigmoid activation.

The final ARIMA-LSTM forecast is given by:

$$\hat{y}_t = \hat{y}_t^{ARIMA} + \hat{e}_t^{LSTM} \quad \text{--------} \ (2)$$

4. Mathematical Formulation of GARCH--LSTM Hybrid Model

The GARCH component models volatility clustering in returns. A GARCH(p, q) process is defined as:

$$r_t = \mu + \epsilon_t,\ \ \epsilon_t = \sigma_t z_t,\qquad z_t \sim iid\ N(0,1)$$

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

Where σ_t^2 is the conditional variance.

The conditional variance series σ_t^2 is then passed to the LSTM, which captures nonlinear patterns in volatility. The hybrid forecast is:

$$\hat{\sigma}_t^2 = \hat{\sigma}_t^{GARCH} + \hat{u}_t^{LSTM} \ \text{-----------}(3)$$

where $\hat{\sigma}_t^{GARCH}$ is the GARCH-based volatility forecast and $\hat{u}_t^{LSTM}$ is the nonlinear adjustment from the LSTM.

5. Evaluation Metrics

Model performance is evaluated using multiple error metrics:

Mean Absolute Error (MAE): MAE is the average of the absolute differences between actual and predicted values.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i|$$

Mean Squared Error (MSE): MSE squares the prediction error, penalizing larger errors more heavily.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Root Mean Squared Error (RMSE): RMSE is the square root of MSE and provides error magnitude in the original units.

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

Mean Absolute Percentage Error (MAPE): MAPE expresses accuracy as a percentage and is scale-independent:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100$$

These metrics are commonly used in financial forecasting studies (Fischer & Krauss, 2018; Shah et al., 2019).

6. Comparison Strategy

To assess generalizability, models are trained on an in-sample dataset (e.g., 2000 - 2020) and evaluated on an out-of-sample test set (2021 - 2025). Forecast accuracy between ARIMA-LSTM, and GARCH-LSTM is compared based on out-of-sample performance, with statistical tests (e.g., Diebold - Mariano test) used to evaluate the significance of performance differences by using a Python code written explicitly for the analysis of results of this study.

Results and Discussion

This section evaluates the out-of-sample forecasting performance of the proposed ARIMA–LSTM and GARCH–LSTM hybrid models across three heterogeneous financial markets, namely Bitcoin (2009–2025), BSE (2000–2025), and S&P 500 (2000–2025). Model accuracy is assessed using MAE, MSE, RMSE, and MAPE, consistent with the evaluation framework described in the methodology.

Table 2: Model Evaluation metrics

Sr. No.	Data set	Model	MAE	MSE	RMSE	MAPE
1	Bitcoin_2009_2025	ARIMA-LSTM	0.017	0.001	0.024	225.20
2	Bitcoin_2009_2025	GARCH-LSTM	0.031	0.001	0.037	2114.92
3	BSE_2000_2025	ARIMA-LSTM	0.006	0.000	0.008	304.93
4	BSE_2000_2025	GARCH-LSTM	0.008	0.000	0.010	963.64
5	SP500_2000_2025	ARIMA-LSTM	0.008	0.000	0.011	234.66
6	SP500_2000_2025	GARCH-LSTM	0.011	0.000	0.015	998.86

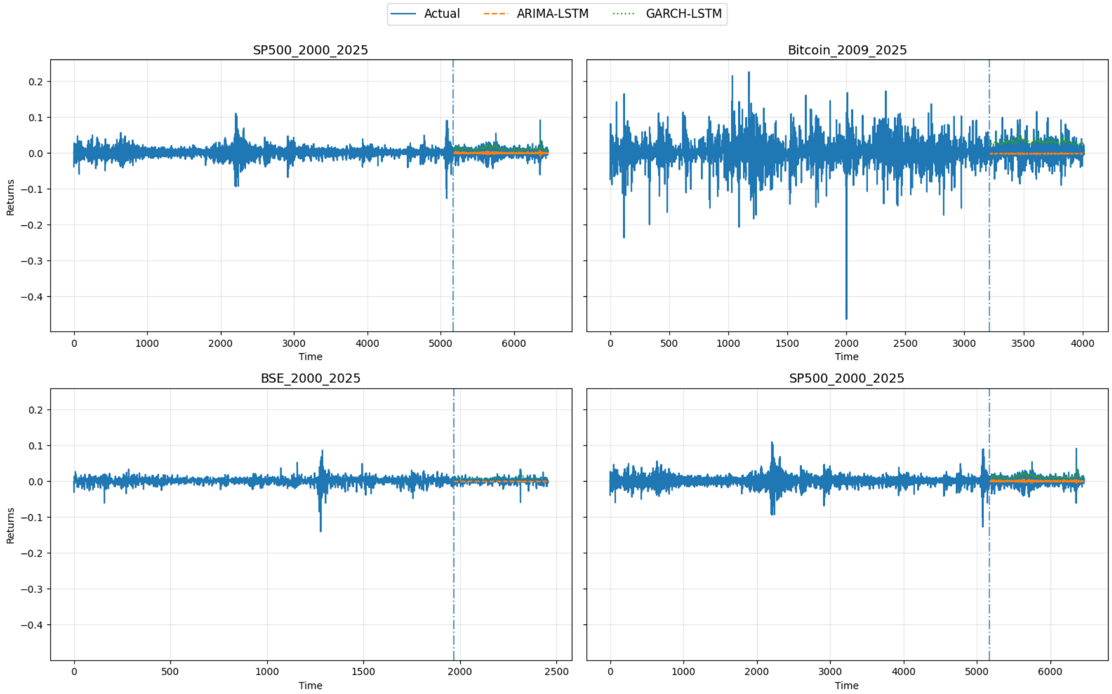

Figure 1: Actual and Predicted Returns for Test Period for Hybrid models

1. Overall comparative performance

Across all three datasets, the ARIMA–LSTM hybrid consistently outperforms the GARCH–LSTM hybrid on every error metric (MAE, MSE, RMSE, and MAPE). This indicates that modeling the linear structure first (ARIMA) and learning nonlinear residuals through LSTM provides more accurate forecasts than combining conditional volatility modeling with LSTM. This consistent dominance suggests that:

- return dynamics contain strong linear dependencies that ARIMA captures effectively,
- remaining nonlinearities are well learned by LSTM,
- whereas volatility modeling alone (GARCH) does not sufficiently explain price/return movements for direct prediction.

2. Dataset-wise interpretation

(a) Bitcoin (Cryptocurrency market)

Model	MAE	RMSE
ARIMA–LSTM	0.0171	0.0244
GARCH–LSTM	0.0313	0.0371

It is observed that ARIMA–LSTM reduces MAE by ≈ 45%, RMSE decreases by ≈ 34% and GARCH–LSTM shows substantially larger forecast dispersion. Thus, Bitcoin data exhibits extreme volatility, nonlinear jumps, and regime shifts. Although GARCH captures volatility clustering, it does not model the directional structure of returns effectively. The ARIMA–LSTM hybrid benefits from ARIMA → short-term linear memory and LSTM → nonlinear shocks. Hence, ARIMA–LSTM adapts better to cryptocurrency dynamics. The very large MAPE values are expected because returns are often close to zero, and percentage errors become inflated. Thus, RMSE/MAE are more reliable metrics for returns.

(b) BSE (Emerging market)

Model	**MAE**	**RMSE**
ARIMA–LSTM	0.00599	0.00802
GARCH–LSTM	0.00799	0.01024

The key observations are approximately 25–30% improvement in all metrics and Lower absolute errors compared to Bitcoin. BSE data shows smoother trends, moderate volatility, and stronger linear components. Hence, ARIMA efficiently captures most predictable behavior, leaving only small nonlinear residuals for LSTM. The hybrid therefore achieves high accuracy with minimal error. GARCH–LSTM is less competitive because volatility clustering is weaker relative to directional effects.

(c) S&P 500 (Developed market)

Model	**MAE**	**RMSE**
ARIMA–LSTM	0.00799	0.01097
GARCH–LSTM	0.01125	0.01495

It is observed that approximately 30–35% reduction in errors and similar pattern to BSE. The S&P 500 data is relatively efficient, less noisy, and exhibits stable autocorrelation. Linear dependencies dominate, favoring ARIMA-based modeling. Consequently, ARIMA–LSTM again provides superior forecasts, while volatility-based modeling adds limited incremental information.

3. Cross-market comparison

A few interesting patterns emerge: **Volatility versus accuracy**

Market	**RMSE (best model)**	**Volatility level**
Bitcoin	highest	very high
S&P 500	medium	moderate
BSE	lowest	lower

Higher volatility implies higher prediction error. This confirms forecasting difficulty increases with market instability. The findings suggest that, for investors, ARIMA–LSTM forecasts provide better entry/exit timing and improved risk-adjusted trading strategies. For risk managers, hybrid models enhance short-term return prediction and useful for VaR and portfolio optimization and for researchers, combining statistical and deep learning approaches is beneficial and ARIMA-based hybrids are preferable to volatility-only hybrids

Conclusion

The empirical results demonstrate that the ARIMA–LSTM hybrid model consistently outperforms the GARCH–LSTM hybrid across all three markets and all evaluation metrics. The improvement is particularly pronounced for highly volatile assets such as Bitcoin, where the ARIMA–LSTM model achieves nearly 40–50% reductions in forecasting errors. This indicates that separating linear structure through ARIMA and modeling nonlinear residual dynamics via LSTM provides a more effective representation of financial return processes than volatility-based hybridization. While GARCH–LSTM successfully captures conditional Heteroskedasticity, volatility information alone is insufficient for accurate directional forecasting. Overall, the results confirm that ARIMA–LSTM offers superior robustness, generalizability, and predictive accuracy across emerging, developed, and cryptocurrency markets.

References

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. Expert Systems with Applications, 36(3), 5932–5941. https://doi.org/10.1016/j.eswa.2008.07.006

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long short-term memory. PLOS ONE, 12(7), e0180944. https://doi.org/10.1371/journal.pone.0180944

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: Forecasting and control (5th ed.). Wiley.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). The econometrics of financial markets. Princeton University Press.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

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

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

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

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Systems with Applications, 103, 25–37. https://doi.org/10.1016/j.eswa.2018.03.002

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Nelson, D. M. Q., Pereira, A. C. M., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. International Joint Conference on Neural Networks (IJCNN), 1419–1426.

Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review. Applied Soft Computing, 90, 106181. https://doi.org/10.1016/j.asoc.2020.106181

Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), 26. https://doi.org/10.3390/ijfs7020026

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. Journal of Economic Perspectives, 17(1), 83–104. https://doi.org/10.1257/089533003321164967

Tsay, R. S. (2010). Analysis of financial time series (3rd ed.). Wiley.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0