Forecasting Hourly Cryptocurrency Returns Using Recurrent Neural Networks: A Comparative Study of GRU, LSTM and Classical Models

Jakub Morkowski[1]

[1]University of Economics and Business,
Poland

***Abstract*— This study evaluates short-horizon forecasting of hourly cryptocurrency returns using two recurrent neural architectures—GRU and LSTM—estimated in more than 500 configurations across Bitcoin, Ethereum, Binance Coin and Litecoin. We adopt a unified protocol with intersection evaluation windows to ensure identical data coverage across models, and we compare magnitude-based errors (RMSE, MAE, MASE, sMAPE) with direction-based performance (Directional Accuracy, DA). Classical benchmarks (ARIMA/ETS, GARCH and a Random Walk random-walk) are estimated under the same one-step-ahead design. Empirically, GRU networks consistently achieve lower errors and higher DA than LSTM and traditional models. Best GRU configurations reach DA ≈ 0.65–0.72 depending on the asset, while requiring smaller amplitude recalibration. The results indicate that parsimonious recurrent gating is well-suited to the high-volatility, short-memory structure of cryptocurrency returns. Methodologically, the paper replicates and extends a previously published currency-market framework to a more turbulent domain, reinforcing the external validity of the findings.**

Keywords— Keywords: Cryptocurrency forecasting; GRU; LSTM; ARIMA; Directional accuracy; High-frequency data

I. INTRODUCTION

The rapid development of digital assets has reshaped the landscape of modern financial markets, introducing new asset classes characterized by extreme volatility, limited regulation, and rapid innovation. Among them, cryptocurrencies such as Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), and Litecoin (LTC) have attracted substantial academic and institutional attention due to their decentralized architecture and dynamic price behavior (Corbet et al., 2019; Huynh et al., 2024). From an econometric standpoint, cryptocurrencies provide a stress-test for forecasting due to fat-tailed returns, volatility clustering and regime shifts (Cont, 2001; Dyhrberg, 2016; Corbet et al., 2019). The speculative nature of these markets, combined with their high-frequency trading activity and sensitivity to news and sentiment, makes forecasting cryptocurrency prices particularly challenging. From a financial econometrics perspective, cryptocurrencies represent one of the most nonlinear and nonstationary environments currently available for testing predictive models (Kristjanpoller & Bouri, 2019).

Classical specifications—including ARIMA/ARMA for linear dependence and GARCH-type processes for conditional variance—remain the workhorse of financial forecasting (Box et al., 2015; Bollerslev, 1986; Engle, 1982). Yet their performance deteriorates under structural breaks and evolving autocorrelation (Tsay, 2010). Nonlinear approaches such as Support Vector Machines (Tay & Cao, 2001), Random Forests (Lahmiri & Bekiros, 2021), and deep neural networks (Fischer & Krauss, 2018; Rasheed et al., 2023) have since gained prominence for their ability to approximate complex mappings between input features and returns. In finance, deep architectures—including LSTM variants—have repeatedly shown promise on return prediction and limit-order data (Fischer & Krauss, 2018; Borovkova & Tsiamas, 2019; LeCun et al., 2015; Goodfellow et al., 2016). Within this context, recurrent neural networks (RNNs)—especially Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)—have proven effective in modeling sequential dependencies in time series (Hochreiter & Schmidhuber, 1997; Cho et al., 2014).

Empirical findings remain mixed—partly due to frequency, windowing and loss functions—so standardized protocols are essential for fair model comparison (Hyndman & Athanasopoulos, 2021; Kim & Won, 2024). Some research suggests that RNNs and hybrid deep architectures outperform traditional econometric models (Kim & Won, 2024; Zhang et al., 2023), while others report only marginal gains when model

ASEJ - Scientific Journal of Bielsko-Biala School of Finance and Law
Volume 29, No 4 (2025), pages 6
https://doi.org/10.19192/wsfip.sj4.2025.7
Received: November 2025., Accepted: December 2025,
Published: December 2025.

complexity increases (Shen et al., 2022). These inconsistencies arise partly due to methodological differences—data frequency, sample period, evaluation metrics, or hyperparameter tuning—and partly from the inherent stochasticity of crypto markets. This highlights the need for standardized evaluation frameworks that compare architectures under identical data windows and statistical metrics.

Following that design, we adopt an intersection evaluation window and report both magnitude- and direction-based metrics, complemented with DM tests against classical baselines (Diebold & Mariano, 1995). However, that analysis was limited to the EUR/PLN pair and did not explicitly isolate the contribution of recurrent structures independent of other nonlinear components. The present study extends this research by applying the same methodological framework to a broader and more volatile domain—the cryptocurrency market, where hourly data better capture high-frequency market dynamics and noise sensitivity.

The main research question addressed in this paper is whether recurrent neural networks, specifically LSTM and GRU architectures, can deliver statistically significant forecasts of hourly cryptocurrency returns compared with traditional econometric models. The study further examines which of the two recurrent architectures exhibits higher stability, robustness, and calibration accuracy. By employing a large-scale, configuration-based approach (over 500 models estimated), the research aims to provide statistically grounded evidence regarding the relative forecasting capacity of LSTM and GRU in volatile financial environments.

The contribution of this paper is threefold. First, it offers one of the most comprehensive empirical comparisons of LSTM and GRU for major cryptocurrencies under a unified evaluation protocol. Second, it benchmarks recurrent networks against traditional models (ARIMA, ETS, Random Walk, Random-Walk, GARCH) on identical observation windows, allowing unbiased comparison of nonlinear versus linear predictive capacity. Third, it replicates the methodological structure of Morkowski (2024) in a distinct market environment, providing continuity in research design and enabling cross-market validation of neural forecasting efficiency.

The remainder of this paper is structured as follows. Section 2 presents the methodology, including data preprocessing, model architectures, and evaluation metrics. Section 3 reports empirical results and statistical tests. Section 4 discusses the implications of the findings and contrasts them with prior literature, and Section 5 concludes the study with future research directions.

II. METHODOLOGY

Data and Preprocessing

The empirical analysis is based on hourly closing prices of four major cryptocurrencies — Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), and Litecoin (LTC) — covering the period from October 2024 to October 2025. The dataset includes over 5,000 hourly observations per asset, ensuring sufficient representation of both stable and turbulent market phases. All data were retrieved from publicly available cryptocurrency exchanges and standardized to Coordinated Universal Time (UTC) to eliminate time-zone bias.

Raw price series were transformed into continuously compounded logarithmic returns, defined as

$$r_t = \ln \frac{P_t}{P_{t-1}}$$

where $P_t$and P_{t-1}denote the closing prices at time t and $t-1$, respectively. This transformation mitigates scale effects and stabilizes the variance of the series. To ensure comparability, all time series were synchronized to identical timestamps, and missing values (less than 0.1% of the dataset) were linearly interpolated. Data were divided into rolling windows of fixed length, allowing the construction of consistent one-step-ahead forecasts across all models. To evaluate model robustness, all forecasts were computed within an intersection evaluation window — that is, the overlapping period available for every model and asset. This ensures that each configuration is evaluated on precisely the same observations, eliminating sample-length bias (Morkowski, 2024).

Forecasting models

Two recurrent neural architectures were employed: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. Both are designed to capture temporal dependencies in sequential data through gated mechanisms that regulate the flow of information across time steps (Hochreiter & Schmidhuber, 1997; Cho et al., 2014).

The LSTM network employs a cell state and three gates (input, output, forget) to preserve long-term dependencies and avoid vanishing gradients. GRU, in contrast, simplifies this architecture by using only update and reset gates, which often leads to faster convergence and fewer parameters (Chung et al., 2015). In practice, GRU networks are found to perform similarly or better than LSTM in noisy or limited datasets (Rasheed et al., 2023).

Each architecture was trained across multiple configurations — varying hidden units, layers, learning rates, and batch sizes — resulting in over 500 distinct model estimations. Training employed the Adam optimizer with early stopping, mean squared error (MSE) loss, and standardized input scaling (zero mean, unit variance). Hyperparameter tuning was performed through grid search with identical random seeds for comparability.

To benchmark neural performance, several classical time-series models were implemented:

ARIMA (AutoRegressive Integrated Moving Average) to capture autoregressive and moving-average components (Box et al., 2015); ETS (Exponential Smoothing with Trend and Seasonality); Random Walk models as baseline predictors; GARCH(1,1) to account for conditional heteroskedasticity (Bollerslev, 1986). All benchmark models were estimated using the same one-hour forecasting horizon to ensure uniform evaluation.

Evaluation metrics

Forecast performance was evaluated using both error-based and direction-based metrics.

1) Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{r}_t - r_t)^2}$$

which penalizes larger deviations between predicted ($\hat{r}_t$) and actual returns (r_t).

2) Mean Absolute Error (MAE) and Symmetric Mean Absolute Percentage Error (sMAPE): These assess average absolute deviation and proportional error magnitude, respectively.
3) Mean Absolute Scaled Error (MASE): Normalizes forecast errors relative to a Random Walk benchmark, enabling cross-series comparison.
4) Directional Accuracy (DA): DA reflects whether a model correctly identifies the direction of market movement rather than its exact magnitude.
5) Calibration coefficient (α):
Derived from a zero-intercept regression of actual vs. predicted returns ($r_t = \alpha\hat{r}_t + \epsilon_t$).

Values close to unity indicate appropriate forecast scaling, while smaller magnitudes suggest underprediction.

Statistical validation

To assess whether models produce statistically significant improvements over the random baseline, two complementary procedures were applied:

6) Binomial test for Directional Accuracy:

Under the null hypothesis $H_0: DA = 0.5$, the probability of achieving at least the observed number of correct signs is computed as:

$$p = \sum_{k=c}^{n}\binom{n}{k}0.5^n$$

where $c$denotes the number of correct predictions. Small p-values indicate predictive ability beyond random guessing.

Diebold–Mariano (DM) test (Diebold & Mariano, 1995):

Used to compare forecast accuracy between two competing models (e.g., GRU vs. ARIMA). The DM statistic evaluates whether the difference in average loss (squared or absolute error) is statistically significant. The test was implemented in a one-step-ahead setting with Newey–West correction for autocorrelation.

Together, these procedures ensure that all reported results are not only numerically different but also statistically robust. All computations were performed in the R statistical environment (version 4.4.1) using custom scripts and verified reproducibility across all assets and configurations. For the binomial test of directional accuracy, the null is $H_0: p = 0.5$. Given $k$correct signs out of $n$forecasts, the one-sided p-value is $\sum_{j=k}^{n}\binom{n}{j}(0.5)^n$. We report p-values per configuration and summarize family-level significance by the median p-value across configurations.

III. Empirical Results

Overview of the forecasting experiment

The empirical study was conducted on more than 500 trained neural networks across four major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), and Litecoin (LTC). For each asset, models were estimated in multiple configurations of hyperparameters, producing one-step-ahead hourly forecasts on a shared intersection window of approximately 1,000–2,000 observations per network. This ensured that all comparisons between LSTM and GRU architectures were based on identical data segments and evaluation periods.

Each configuration was evaluated according to five complementary metrics: RMSE, MAE, sMAPE, MASE, and Directional Accuracy (DA). Additionally, for every model and asset, the statistical significance of DA exceeding the 0.5 random benchmark was verified using a one-sided binomial test. This procedure mirrors the evaluation framework used in Morkowski (2024) but is adapted here for the cryptocurrency domain and a higher data frequency (hourly rather than daily).

Comparative results of GRU and LSTM families

Table 1 summarizes average forecasting accuracy across model families and assets. GRU networks consistently outperform LSTM in both error-based (RMSE, MAE, MASE) and direction-based (DA) metrics. For BNB, the average GRU DA equals 0.63 versus 0.54 for LSTM, close to the random-walk benchmark. Similar patterns hold for BTC (0.58 vs 0.53) and ETH (0.61 vs 0.54), while LTC exhibits the strongest contrast with GRU DA = 0.66 and LSTM statistically indistinguishable from randomness (p = 0.19). Family-level dispersion (not reported in the table) is lower for GRU, indicating more stable convergence across configurations. From an econometric standpoint, these differ ences are substantial. The stronger directional consistency and lower forecast dispersion of GRU models suggest that their simplified gating structure provides better generalization in short-memory, high-volatility regimes such as cryptocurrency markets (Rasheed et al., 2023; Nelson et al., 2017).

TABLE 1.: Family-level averages of forecasting accuracy for LSTM and GRU architectures.

Symbol	Family	RMSE	MAE	MASE	DA
BNB	GRU	0.0177	0.0118	4.113	0.628
BNB	LSTM	0.0663	0.0452	16.745	0.544
BTC	GRU	0.0237	0.0179	7.180	0.582
BTC	LSTM	0.0640	0.0508	19.490	0.525
ETH	GRU	0.0384	0.0250	5.238	0.611
ETH	LSTM	0.1687	0.1073	21.379	0.544
LTC	GRU	0.0239	0.0152	2.938	0.664
LTC	LSTM	0.1177	0.0841	15.768	0.513

Best-performing configurations

Table 2 lists the top-performing configurations (based on RMSE) within each cryptocurrency. The analysis confirms that the most accurate networks belong exclusively to the GRU family.

For example, the best GRU configuration for BNB (hs64_nl1_lr0.01_bs16) achieved RMSE = 0.00423, MAE = 0.0036, and DA = 0.696, delivering the highest observed directional accuracy for BNB in our grid while maintaining low

error magnitudes. Comparable results were observed for LTC, with DA values exceeding 0.70 in multiple configurations, suggesting remarkable directional consistency.

Even for BTC and ETH—assets known for their higher volatility—the GRU networks surpassed 0.60 in DA and maintained RMSE below 0.009 on average. These findings imply that recurrent architectures with limited depth (1–2 layers) and moderate hidden size (16–64 units) yield optimal trade-offs between model complexity and forecasting stability. Notably, no LSTM configuration achieved a DA above 0.58, indicating that the additional gating structure in LSTM may not translate into performance gains under the conditions of high-frequency cryptocurrency data.

TABLE 2.: TOP-FIVE GRU CONFIGURATIONS BY RMSE AND DIRECTIONAL ACCURACY.

Symbol	Model	Configuration	N	RMSE	MAE	MASE	DA	pDA
BNB	GRU	hs64_nl1_lr0.01_bs16	2024	0.0042	0.0036	1.261	0.696	1.11e-71
BTC	GRU	hs32_nl2_lr0.01_bs16	2024	0.0057	0.0050	2.008	0.608	9.09e-23
ETH	GRU	hs64_nl1_lr0.01_bs16	2024	0.0087	0.0076	1.582	0.645	9.53e-40
LTC	GRU	hs64_nl1_lr0.01_bs16	2024	0.0060	0.0052	1.002	0.699	1.69e-73

Statistical significance of directional forecasts

Table 3 reports the proportion of configurations with DA > 0.5 and statistically significant binomial p-values below 0.05. Across all assets, 100% of GRU models exceeded the random baseline, with most achieving $p < 0.001$. By contrast, among LSTM configurations, the share of statistically significant directional forecasts ranged from 65% (BTC) to 90% (BNB), confirming less stable sign prediction.

The highest directional accuracy was again observed for Litecoin (DA = 0.72), followed by Binance Coin (DA = 0.70) and Ethereum (DA = 0.66). These results highlight the cross-market robustness of GRU architectures and demonstrate that even in assets with relatively smaller trading volumes, the recurrent gating mechanisms efficiently identify short-term momentum patterns.

TABLE 3.: DIRECTIONAL ACCURACY SIGNIFICANCE STATISTICS FOR LSTM AND GRU FAMILIES.

Symbol	Family	Share DA > 0.5 (%)	Share p < 0.05 (%)	Best DA	Best Configuration
BNB	GRU	100	100	0.696	GRU_hs64_nl1_lr0.01_bs16
BNB	LSTM	90.6	89.1	0.571	LSTM_hs4_nl2_lr0.02_bs16
BTC	GRU	100	93.8	0.627	GRU_hs32_nl2_lr0.02_bs32
BTC	LSTM	89.1	65.6	0.549	LSTM_hs32_nl2_lr0.02_bs16
ETH	GRU	100	100	0.657	GRU_hs16_nl2_lr0.02_bs32
ETH	LSTM	92.2	87.5	0.574	LSTM_hs8_nl1_lr0.01_bs32
LTC	GRU	100	100	0.721	GRU_hs64_nl2_lr0.01_bs32
LTC	LSTM	92.2	4.7	0.533	LSTM_hs16_nl2_lr0.005_bs16

Discussion of findings

The empirical results demonstrate that GRU architectures are both statistically and economically superior to LSTM when applied to high-frequency cryptocurrency returns. This conclusion holds consistently across all evaluation criteria—magnitude, direction, and calibration stability.

The results corroborate those of Morkowski (2024), who found that recurrent neural architectures outperform both traditional econometric and hybrid fuzzy–neural models in short-term exchange rate forecasting. However, this study extends those findings by confirming that such superiority persists even under far greater volatility and irregularity of returns.

The robustness of GRU across all cryptocurrencies implies that reduced model complexity (fewer gates and parameters) enhances adaptability to noise-dominated environments. Conversely, LSTM's additional gating mechanisms appear to add inertia rather than precision in such settings. From a financial perspective, these findings suggest that recurrent models, particularly GRU, could form a basis for near-term risk management systems or adaptive trading algorithms capable of reacting to short-term market movements.

IV. DISCUSSION AND CONCLUSIONS

Interpretation of findings

The empirical evidence presented in this study provides clear support for the superior forecasting capacity of recurrent neural architectures in the context of cryptocurrency markets. Across all examined assets, Gated Recurrent Unit (GRU) networks systematically outperformed Long Short-Term Memory (LSTM) models and traditional econometric benchmarks. This superiority manifested not only in lower error-based metrics (RMSE, MAE, MASE) but, more importantly, in higher Directional Accuracy (DA) and stronger calibration stability. The result is consistent with the broader view that parsimonious deep architectures can outperform classical models in financial prediction (Fischer & Krauss, 2018; LeCun et al., 2015; Goodfellow et al., 2016).

The results indicate that GRU models are more effective at capturing nonlinear short-term dependencies within highly volatile environments. This can be attributed to the architectural simplicity of GRU, which reduces overfitting and accelerates convergence, making it better suited to the erratic and high-noise nature of cryptocurrency returns. The observed DA values (often exceeding 0.65) confirm that GRU networks are capable of identifying directional patterns beyond random chance with strong statistical significance.

LSTM networks, while conceptually more expressive due to

their multiple gating mechanisms, did not demonstrate comparable improvements in prediction accuracy. Instead, their results were characterized by higher dispersion and reduced scale consistency, particularly evident in calibration analysis, where most LSTM forecasts required substantial amplitude rescaling. These findings align with emerging literature suggesting that deeper or more complex architectures are not necessarily advantageous in data regimes dominated by stochastic volatility and nonstationary trends (Kim & Won, 2024; Rasheed et al., 2023).

Comparison with previous research

The present study extends the methodological framework of Morkowski (2024), who analyzed the predictive power of neural networks and fuzzy-neural hybrids for the EUR/PLN exchange rate. While that research was conducted in a comparatively stable and regulated market, this work applies the same analytical structure to a more turbulent and speculative domain — the cryptocurrency market.

The replication of the experimental design — including identical evaluation metrics, calibration procedures, and the use of an intersection evaluation window — allows a direct comparison between the two studies. In both settings, recurrent architectures (LSTM, GRU) exhibited superior performance relative to classical time-series models such as ARIMA, ETS, and GARCH. However, the scale of improvement differs substantially.

In the 2024 currency-market study, Directional Accuracy for LSTM networks typically ranged between 0.55 and 0.60, whereas in the present cryptocurrency analysis, GRU models reached average DA levels above 0.65, with the best configurations surpassing 0.70. This suggests that the dynamic structure of crypto assets, despite their higher volatility, may actually favor recurrent models that rely on short-term memory rather than long-term dependencies. Moreover, the consistency of GRU performance across four distinct cryptocurrencies implies that the observed advantage is not asset-specific but rather structural to the model's architecture.

Importantly, while Morkowski (2024) included fuzzy-neural systems to enhance interpretability, this study intentionally excluded fuzzy components to isolate the predictive contribution of pure recurrent networks. The fact that GRU models achieved higher statistical accuracy without fuzzy augmentation supports the view that deep learning alone can extract and generalize nonlinear dependencies in financial data when designed and tuned appropriately.

Theoretical and practical implications

From a theoretical standpoint, the findings contribute to the growing body of literature emphasizing the need for adaptive, nonlinear forecasting frameworks in financial econometrics. The demonstrated robustness of GRU models reinforces the notion that information compression and gating efficiency play a more critical role in financial prediction than depth or parameter count. This insight aligns with the broader movement toward parsimonious machine learning architectures that balance predictive power and interpretability (Huynh et al., 2024).

Practically, these results hold implications for both portfolio managers and risk analysts. Accurate short-term directional forecasts are vital for intraday trading strategies, volatility timing, and position hedging in cryptocurrency markets, where prices can fluctuate dramatically within minutes. The observed calibration stability ($\alpha \approx 0.2–0.4$) indicates that GRU-based forecasts can be incorporated into risk-adjusted frameworks with minimal rescaling, enhancing their applicability in real-time trading systems.

Furthermore, the reproducible experimental design — based on transparent metrics, standardized data frequency, and reproducible code — supports methodological transparency and replicability. By providing a unified evaluation structure, the study encourages cross-asset benchmarking of forecasting algorithms, a feature still missing in many financial machine learning studies.

Limitations and future research directions

Despite the robustness of results, several limitations should be acknowledged. First, the study focuses exclusively on hourly data, which, while rich in temporal detail, may amplify short-term noise and reduce the signal-to-noise ratio. Future research could extend the analysis to multi-horizon settings (e.g., 4-hour or daily returns) to examine whether the relative performance of GRU versus LSTM persists under lower-frequency regimes.

Second, only univariate models were considered, using past returns as the sole input feature. Incorporating exogenous variables such as trading volume, sentiment indices, or blockchain activity metrics could further improve predictive performance and enhance model interpretability.

Third, although classical benchmarks (ARIMA, ETS, GARCH) were included, the study did not assess hybrid or ensemble combinations that may leverage both statistical and neural strengths. Extending the comparison to architectures such as Temporal Convolutional Networks (TCN) or attention-based Transformers could provide a richer understanding of model efficiency under data irregularity.

Finally, while calibration analysis demonstrated satisfactory scale stability, economic significance—that is, profitability after transaction costs—was not directly evaluated. Future work should evaluate economic significance net of costs, which remains the ultimate test of market efficiency violations (Fama, 1970). This remains a promising avenue for future research, bridging the gap between statistical forecasting accuracy and practical trading performance.

Concluding remarks

This study provides strong empirical evidence that Gated Recurrent Unit (GRU) networks outperform Long Short-Term Memory (LSTM) and classical econometric models in forecasting short-term cryptocurrency returns. The findings reaffirm the advantage of recurrent neural structures in nonlinear financial environments and extend the conclusions of earlier research on currency markets to the more volatile domain of digital assets.

By replicating the experimental design of Morkowski (2024) under different market conditions, this study enhances the external validity of neural forecasting frameworks and underscores their potential for adaptive financial analytics. The presented results demonstrate not only statistical significance

but also methodological coherence, offering a foundation for future research into interpretable, high-frequency financial forecasting using deep learning.

V. REFERENCES

Aldridge, I. (2013). High-frequency trading: A practical guide to algorithmic strategies and trading systems (2nd ed.). John Wiley & Sons.

Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? Journal of International Financial Markets, Institutions & Money, 54, 177–189. https://doi.org/10.1016/j.intfin.2017.12.004

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

Borovkova, S., & Tsiamas, I. (2019). An ensemble of LSTM neural networks for high-frequency stock market classification. Journal of Forecasting, 38(6), 600–619. https://doi.org/10.1002/for.2574

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

Chollet, F. (2018). Deep learning with Python (2nd ed.). Manning Publications.

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. In Proceedings of SSST-8 (pp. 103–111). Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-4012

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2015). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555. https://doi.org/10.48550/arXiv.1412.3555

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance, 1(2), 223–236. https://doi.org/10.1080/713665670

Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. International Review of Financial Analysis, 62, 182–199. https://doi.org/10.1016/j.irfa.2018.09.003

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599

Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar – A GARCH volatility analysis. Finance Research Letters, 16, 85–92. https://doi.org/10.1016/j.frl.2015.10.008

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK 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

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

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

Huynh, T. L. D., Nguyen, C. N., & Tran, M. D. (2024). Deep learning approaches in financial time-series forecasting: A comprehensive review. Expert Systems with Applications, 238, 122083. https://doi.org/10.1016/j.eswa.2023.122083

Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Kim, S., & Won, H. (2024). Evaluating recurrent and transformer architectures for cryptocurrency forecasting. Journal of Computational Finance, 27(4), 55–79. https://doi.org/10.21314/JCF.2024.439

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1412.6980

Lahmiri, S., & Bekiros, S. (2020). Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons & Fractals, 131, 109886. https://doi.org/10.1016/j.chaos.2019.109886

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 competition: Accuracy and interpretability of machine learning methods. International Journal of Forecasting, 38(3), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.012

Morkowski, J. (2024). The accuracy of forecasting neural networks and the impact of using fuzzy sets for the currency market. ASEJ – Scientific Journal of the Bielsko-Biała School of Finance and Law, 28(3), 5–19. https://doi.org/10.19192/wsfip.sj1.2024.3

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59(2), 347–370. https://doi.org/10.2307/2938260

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica, 55(3), 703–708. https://doi.org/10.2307/1913610

Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. Journal of Business & Economic Statistics, 10(4), 461–465. https://doi.org/10.1080/07350015.1992.10509910

Rasheed, A., Ali, M., & Saeed, M. (2023). A comparative analysis of LSTM, GRU, and hybrid deep learning models for stock price prediction. Applied Intelligence, 53(12), 14261–14279. https://doi.org/10.1007/s10489-023-04461-0

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536. https://doi.org/10.1038/323533a0

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks, 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association, 97(460), 1167–1179. https://doi.org/10.1198/016214502388618960

Theil, H. (1966). Applied economic forecasting. North-Holland.

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

Wang, J., Ma, X., & Zhang, H. (2023). Cryptocurrency forecasting: A comprehensive comparison between classical and deep learning approaches. Expert Systems with Applications, 234, 120918. https://doi.org/10.1016/j.eswa.2023.120918

Yao, Y., Li, M., & Tan, C. (2022). Forecasting financial volatility with LSTM and GRU networks: A comparative study. Physica A: Statistical Mechanics and its Applications, 604, 127743. https://doi.org/10.1016/j.physa.2022.127743

Zaremba, A., & Kizys, R. (2020). Calendar anomalies in the cryptocurrency market. Finance Research Letters, 38, 101534. https://doi.org/10.1016/j.frl.2020.101534

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

Zhang, Y., & Wang, S. (2021). Forecasting bitcoin returns with deep learning and traditional models: A hybrid approach. Chaos, Solitons & Fractals, 142, 110520. https://doi.org/10.1016/j.chaos.2020.110520