ISSN: 1672 - 6553

JOURNAL OF DYNAMICS AND CONTROL

VOLUME 9 ISSUE 11: P60-75

VOLATILITY DYNAMICS IN CRYPTOCURRENCIES: COMPARING TRADITIONAL AND HYBRID DEEP LEARNING PREDICTIVE TECHNIQUES

Tam Phan Huy, Khuong Huu Pham, Linh Nguyen Thi Thuy, Nam Nguyen Nhat, Bach Le Danh

University of Economics and Law and Vietnam National University, Ho Chi Minh City, Vietnam

https://doi.org/10.71058/jodac.v9i11005

VOLATILITY DYNAMICS IN CRYPTOCURRENCIES: COMPARING TRADITIONAL AND HYBRID DEEP LEARNING PREDICTIVE TECHNIQUES

Tam Phan Huy*, Khuong Huu Pham, Linh Nguyen Thi Thuy, Nam Nguyen Nhat, Bach Le Danh
University of Economics and Law and Vietnam National University, Ho Chi Minh City, Vietnam

Corresponding Author: tamph@uel.edu.vn
The research is funded by the University of Economics and Law, Vietnam National University, Ho Chi Minh City, Vietnam

Abstract: *This study investigates the predictive performance of traditional econometric and deep learning models in forecasting the return and volatility of Ethereum, a leading cryptocurrency known for its high price fluctuation and market sensitivity. Using historical daily price data from January 2018 to March 2025 obtained from Yahoo Finance, we construct a comprehensive set of features including log returns, multiple volatility estimators (Rolling, Parkinson, Garman-Klass, and Rogers-Satchell), and technical indicators such as moving averages, momentum oscillators, and volume-based metrics. Six predictive models—ARIMA, GARCH, VAR, Random Forest, LSTM, and GRU—are evaluated based on their ability to predict log returns and volatility using multiple performance metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2), along with cross-validation to assess overfitting. The results reveal that machine learning and deep learning models significantly outperform traditional econometric methods. Random Forest yields the highest accuracy in volatility prediction, while GRU demonstrates the most stable and consistent performance across both return and volatility forecasts. Traditional models, particularly GARCH, show poor generalization and limited predictive capability under high market uncertainty. The findings highlight the importance of model selection and volatility measure choice in cryptocurrency forecasting. This study contributes to the growing literature on crypto analytics by providing a direct comparative evaluation of predictive models and offering practical insights into model suitability for high-frequency, volatile digital asset markets.*

Keywords: *Ethereum, volatility forecasting, deep learning, machine learning.*

JEL codes: *C22, C45, G17*

1. Introduction

The unprecedented growth and volatility of cryptocurrencies have captured significant attention from investors, policymakers, and researchers alike. Among them, Ethereum has emerged as one of the leading digital assets, not only due to its role as a financial instrument but also because of its unique technological utility in decentralized applications and smart contracts. However, the rapid price fluctuations and frequent market swings observed in cryptocurrencies pose serious challenges for risk management, investment decisions, and regulatory oversight. Traditional financial assets such as stocks and commodities generally follow more predictable patterns, but cryptocurrencies defy these norms with high-frequency volatility, speculative bubbles, and limited underlying fundamentals (Urquhart 2016, Liu and Tsyvinski 2021). Consequently, forecasting the return and volatility of cryptocurrencies, particularly Ethereum, remains a complex yet critical research problem.

Existing studies have explored various approaches to model and predict cryptocurrency behavior, ranging from traditional econometric models like ARIMA and GARCH to modern machine learning and deep learning frameworks such as Random Forests, Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) (Bouri, Azzi and Dyhrberg 2017, Katsiampa 2017, Lahmiri and Bekiros 2019). While GARCH models have been widely applied to account for volatility clustering and time-varying variance, they often fall short in capturing nonlinear dynamics and regime shifts that are inherent in crypto markets.

https://doi.org/10.71058/jodac.v9i11005

Deep learning models, on the other hand, have demonstrated strong performance in time series forecasting due to their ability to capture sequential dependencies and nonlinear interactions (Patel, Tanwar et al. 2020, Vaddi, Neelisetty et al. 2020, Goutte, Le et al. 2023, Seabe, Moutsinga and Pindza 2023). Nonetheless, comparative evaluations of these models within the same framework, particularly using multiple volatility estimators, remain limited in the existing literature. Furthermore, many studies have focused solely on Bitcoin, leaving Ethereum's volatility dynamics relatively underexplored.

The main objective of this study is to evaluate and compare the predictive performance of traditional and hybrid deep learning models in forecasting Ethereum's daily return and volatility. The analysis focuses on assessing model accuracy using multiple error metrics and includes overfitting diagnostics through cross-validation and train-test performance comparisons. Specifically, the research aims to (1) assess the strengths and weaknesses of each algorithm, (2) determine the most effective models for volatility forecasting under different volatility estimation techniques, and (3) provide empirical insights into model suitability for cryptocurrency data.

This research is significant as it provides a comprehensive comparison between traditional econometric models and advanced machine learning algorithms for predicting both the return and volatility of Ethereum. By employing a unified modeling framework that incorporates ARIMA, GARCH, VAR, Random Forest, LSTM, and GRU models, the study bridges a critical gap in empirical research and offers insights into the relative performance and applicability of each approach. The inclusion of multiple volatility estimators' further value, allowing for a more nuanced understanding of how different models respond to alternative volatility definitions. The findings can benefit investors in developing more robust trading strategies, inform risk managers about the predictive accuracy of different modeling techniques, and support policymakers in understanding the underlying volatility structure of digital assets. This paper is structured as follows: (i) introduction, (ii) literature review, (iii) methodology, (iv) results & discussion, and (v) conclusion & recommendations.

2. Literature review

2.1 Background theories

Cryptocurrency markets, with their inherent volatility and rapid evolution, offer a unique challenge for researchers aiming to predict market dynamics. The interplay of various economic and behavioral factors, coupled with structural inefficiencies and chaotic patterns, creates an environment where traditional theories and modern techniques must converge to provide meaningful insights into volatility prediction.

The concept of market efficiency, as outlined in the Efficient Market Hypothesis, posits that asset prices reflect all available information (Fama 1970). While this hypothesis has guided much of the research in traditional financial markets, cryptocurrency markets frequently exhibit inefficiencies due to their speculative nature, lack of regulation, and susceptibility to sentiment-driven trading (Urquhart 2016). These inefficiencies create opportunities for models that can identify patterns amidst the noise. However, Efficient Market Hypothesis alone cannot fully explain cryptocurrency volatility; behavioral biases, as addressed in Behavioral Finance Theory, add another layer of complexity. Investors in cryptocurrency markets often exhibit herd mentality and overconfidence, leading to pronounced volatility clustering, periods of heightened volatility followed by similar patterns (Thaler 1980, Bouri, Azzi and Dyhrberg 2017). These behavioral tendencies, combined with the deviations from Efficient Market Hypothesis, justify the need for hybrid approaches that integrate statistical rigor with the adaptability of machine learning.

Moreover, Market Microstructure Theory and Modern Portfolio Theory provide complementary perspectives on the structural and risk-related aspects of cryptocurrency volatility (Markowitz 1991, O'hara 1998). The fragmented and decentralized nature of cryptocurrency markets, characterized by varying

https://doi.org/10.71058/jodac.v9i11005

liquidity levels and trading environments, exacerbates short-term price swings and increases systemic risk (Auer and Claessens 2021). For instance, high trading volumes and wide bid-ask spreads often serve as precursors to volatility spikes. Incorporating these microstructural features into predictive models enhances their ability to capture the nuances of cryptocurrency markets. Simultaneously, Modern Portfolio Theory underscores the importance of accurate volatility predictions for risk management and portfolio optimization, as high volatility complicates diversification and increases exposure to extreme losses (Liu and Tsyvinski 2021). These theories collectively suggest that volatility prediction is not merely a technical exercise but a critical tool for managing risk in speculative assets.

Adding to the complexity is the observation that cryptocurrency markets often exhibit chaotic behavior, aligning with principles from Chaos Theory (Peters 1996). This perspective suggests that while market movements may appear random, they are often driven by deterministic but highly sensitive patterns. For example, sudden regulatory announcements or technological developments can trigger disproportionately large market responses due to the interconnectedness of participants. Traditional models, such as GARCH, are well-suited to capturing clustering in volatility but may struggle to account for the chaotic and non-linear relationships inherent in these markets (Kristjanpoller and Minutolo 2015). Deep learning models, particularly LSTM and GRU, excel in identifying these complex dependencies, making them valuable tools for understanding volatility in a chaotic environment.

Furthermore, recent empirical studies reinforce the need for combining traditional theories with modern predictive approaches. For example, research by Urquhart (2016) confirms that Bitcoin markets exhibit inefficiencies, challenging the assumptions of Efficient Market Hypothesis. Meanwhile, Bouri, Azzi and Dyhrberg (2017) demonstrate that behavioral factors significantly influence cryptocurrency volatility, highlighting the role of fear and greed. Auer and Claessens (2021) emphasize the importance of market structure in shaping volatility dynamics, while Liu and Tsyvinski (2021) underscore the implications of volatility for portfolio construction. Collectively, these studies validate the integration of Efficient Market Hypothesis, Behavioral Finance, Market Microstructure, Modern Portfolio Theory, and Chaos Theory as a cohesive framework for understanding and predicting cryptocurrency volatility.

2.2 Empirical studies

A growing body of empirical research has sought to understand the dynamic behavior of cryptocurrency prices and volatility, particularly through the lens of predictive modeling. Early work by Auer and Claessens (2021) utilized regression models to reveal how market announcements and regulatory developments significantly influence cryptocurrency prices, trading volumes, and investor behavior. Similarly, Liu and Tsyvinski (2021) extended traditional factor models such as CAPM, Fama-French, and Carhart to the cryptocurrency domain, discovering that cryptocurrency returns are largely driven by momentum and investor sentiment, with minimal linkage to traditional financial assets such as equities, bonds, or commodities. Although regression models provide a straightforward interpretation of macro and micro drivers, their inability to capture nonlinear and regime-dependent structures has led researchers to explore more sophisticated econometric approaches. Consequently, autoregressive models such as ARIMA have been commonly employed to model and forecast cryptocurrency volatility. Studies by Chu, Chan et al. (2017) and Catania, Grassi and Ravazzolo (2019) applied ARIMA and ARFIMA models and found limited predictive performance due to the non-stationary and noisy nature of crypto time series. Nonetheless, these models proved useful for short-horizon volatility estimation under stable market conditions. Yet, their linearity and lack of memory effects render them less suitable for capturing the frequent regime shifts in digital asset markets.

To better model volatility clustering and conditional heteroskedasticity, GARCH family models have been widely adopted. Bouri, Azzi and Dyhrberg (2017), for instance, demonstrated asymmetric volatility effects in Bitcoin using GARCH and TGARCH models, where positive shocks prior to the 2013 crash had a larger

https://doi.org/10.71058/jodac.v9i11005

impact on volatility than negative shocks. Further refinements in the GARCH framework have been explored by Katsiampa (2017), who compared several GARCH variants (e.g., EGARCH, CGARCH) and confirmed their usefulness in modeling persistent volatility patterns. Similarly, Conrad, Custovic and Ghysels (2018) and Ejder and Özel (2024) emphasized that such models, while suitable for capturing short-memory dynamics, often fall short when structural breaks or nonlinearities are present. In contrast, stochastic volatility (SV) models offer a more flexible framework by incorporating latent variables, which allow for more adaptive modeling. Yen and Cheng (2021), for example, employed an SV model to evaluate the influence of Economic Policy Uncertainty (EPU) indices on cryptocurrency volatility, finding that China's EPU had a statistically significant impact on Bitcoin and Ethereum volatility, while U.S. and Japanese indices showed negligible effects.

Given the nonlinear, high-frequency, and sentiment-driven nature of cryptocurrency markets, recent empirical studies have increasingly turned to deep learning techniques, particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. Lahmiri and Bekiros (2019) found that LSTM outperformed traditional neural networks in predicting Bitcoin returns, due to its memory retention and sequence learning abilities. Likewise, Seabe, Moutsinga and Pindza (2023) and Murray, Rossi et al. (2023) confirmed the superiority of LSTM over shallow learners and rule-based models, especially in capturing chaotic price behavior and volatility bursts. GRU, while architecturally simpler, demonstrated competitive accuracy and faster convergence in studies by Verma, Tyagi and Aneja (2023) and Girsang (2023), particularly under time-constrained environments and on high-frequency data. These neural models, when combined with external features such as trading volume, macroeconomic signals, and social sentiment, yielded robust and generalizable forecasts. Furthermore, Goutte, Le et al. (2023)showed that integrating LSTM models with social media signals and technical indicators like RSI and MACD substantially improves prediction accuracy.

The performance of hybrid and ensemble models has also gained attention, combining the interpretability of econometric models with the adaptability of machine learning. Seo and Kim (2020) pioneered the ANN-GARCH framework for volatility forecasting, demonstrating its superiority over individual models in predicting commodity price dynamics, with later extensions into the cryptocurrency space. Similarly, Darley, Yussuff and Adenowo (2021) combined ARIMA with artificial neural networks, while Akyildirim, Goncu and Sensoy (2021) applied Random Forest and Gradient Boosting Machines to price direction classification. Sun, Liu and Sima (2020) and Basher and Sadorsky (2022) further enriched ensemble models with text mining and sentiment analysis, confirming that incorporating diverse data sources leads to better performance in both price prediction and volatility forecasting. More recent innovations leverage transformer architectures and attention-based networks (Patel, Tanwar et al. 2020, Vaddi, Neelisetty et al. 2020, Tanwar, Patel et al. 2021), which have demonstrated their potential in learning long-range dependencies in cryptocurrency data, outperforming even LSTM models under certain configurations.

Complementing predictive models, several studies have investigated the efficiency and predictability of cryptocurrency markets. Urquhart (2016) employed a range of statistical tests including Ljung-Box, Runs, BDS, and Hurst exponent, concluding that Bitcoin exhibited inefficiencies during its early years but gradually approached semi-strong form efficiency. Extending this analysis, Caporale et al. (2018) evaluated Litecoin, Ripple, and Dash using R/S and fractional integration methods, identifying that while these markets were still not fully efficient, they demonstrated signs of convergence. Ripple, in particular, was noted for its stability, while Litecoin showed improvement over time. Moreover, Bouri, Azzi and Dyhrberg (2017) and Kuznetsov, Kryvinska et al. (2023) used long memory and entropy-based techniques to show that cryptocurrency efficiency is time-varying, often linked to market cycles and sentiment-driven phenomena, this aligned with Anamika, Chakraborty and Subramaniam (2023). Similarly, Vidal-Tomás (2021) argued that market efficiency in cryptocurrencies follows an adaptive pattern, evolving in response to innovation, regulatory developments, and broader financial integration.

https://doi.org/10.71058/jodac.v9i11005

3. Methodology

3.1 Data

The data for this study is obtained from a secondary source, specifically the Yahoo Finance database via the yfinance Python API, covering the period from January 2018 to March 2025. The dataset consists of daily historical price data for Ethereum (ETH), one of the most actively traded and capitalized cryptocurrencies. The primary variables include the opening price, closing price, highest and lowest prices of the day, and daily trading volume. Based on this raw market data, several additional variables are computed to support the predictive modeling framework. These include return measures (simple and logarithmic), volatility metrics (calculated using a 5-day rolling standard deviation), and a suite of technical analysis indicators such as moving averages (SMA, EMA), momentum indicators (RSI, Stochastic Oscillator), trend-following measures (MACD), and volatility-based bands (Bollinger Bands, ATR).

Table 1. Input variables measurement

Factors	Variables	Description	Measure	References
Return & Volatility Metrics	Return	Log_return = $(P_t / P_{t-1}) - 1$	%	Bouri, Azzi and Dyhrberg (2017); Lahmiri and Bekiros (2019); Catania, Grassi and Ravazzolo (2019); Katsiampa (2017);
	Rolling Volatility	5-day rolling standard deviation of returns	%	
	Parkinson Volatility	High-low range-based volatility estimator	%	
	Garman Klass Volatility	Improved volatility estimator using OHLC prices	%	
	Rogers Satchell Volatility	Adjusted volatility estimator accounting for drift	%	
Price & Volume Variables	Open	Daily opening price	USD ($)	Kristjanpoller and Minutolo (2015); Lahmiri and Bekiros (2019); Bouri, Azzi and Dyhrberg (2017)
	High	Daily high price	USD ($)	
	Low	Daily low price	USD ($)	
	Close	Daily close price	USD ($)	
	Adj Close	Daily adjusted close price	USD ($)	
	Volume	Number of units traded daily	Units	
Trend & Moving Average Indicators	SMA_14	Simple Moving Average over 14 periods	USD ($)	Auer and Claessens (2021); Girsang (2023); Goutte, Le et al. (2023); Basher and Sadorsky (2022)
	EMA_14	Exponential Moving Average over 14 periods	USD ($)	
	EMA_12	Exponential Moving Average over 12 periods	USD ($)	
	EMA_26	Exponential Moving Average over 26 periods	USD ($)	
	MACD	Moving Average Convergence Divergence (EMA12 - EMA26)	USD ($)	
Momentum Indicators	Momentum Indicators	Relative Strength Index (14-day)	0 - 100	Lahmiri and Bekiros (2019); Patel, Tanwar et al. (2020)
	Stochastic_%K	Stochastic Oscillator (momentum indicator)	0 - 100	
	Lowest_Low	Lowest price over 5 days	USD ($)	
	Highest_High	Lowest price over 5 days	USD ($)	
Volatility & Volume Indicators	BB_Mid	Bollinger Bands Middle Line (SMA-based)	USD ($)	Seabe, Moutsinga and Pindza (2023); Murray, Rossi et al.
	BB_Upper	Upper Bollinger Band	USD ($)	
	BB_Lower	Lower Bollinger Band	USD ($)	

https://doi.org/10.71058/jodac.v9i11005

	TR	True Range (Volatility indicator)	USD ($)	(2023); Verma, Tyagi and Aneja (2023)
	ATR_14	A verage True Range over 14 periods	USD ($)	
	OBV	On-Balance Volume (trend confirmation)	Cumulative sum	

Source: by author

Table 1 presents the measurement details of the input variables used in this study, categorized into five key analytical groups: return and volatility metrics, price and volume variables, trend and moving average indicators, momentum indicators, and volume-based technical indicators. The return variable is measured using logarithmic returns, a standard approach in financial time series analysis, while volatility is captured through multiple estimators including a 5-day rolling standard deviation, Parkinson's, Garman-Klass, and Rogers-Satchell methods, each offering varying sensitivity to price fluctuations and noise. Price-based variables such as open, high, low, close, and volume represent the foundational market data collected directly from Yahoo Finance. Technical indicators are then derived to capture different dimensions of market behavior: moving averages (SMA, EMA) and MACD reflect price trends; RSI and Stochastic Oscillator capture momentum; and Bollinger Bands, ATR, and True Range provide insights into volatility dynamics. On-Balance Volume (OBV) is also included as a volume-based trend confirmation tool.

3.2 Models

This study employs a combination of traditional econometric models and advanced machine learning algorithms to predict both returns and volatility of Ethereum. The selection of these models is grounded in extensive empirical literature on financial forecasting, particularly in the domain of cryptocurrency markets, where nonlinearity, volatility clustering, and structural breaks are prevalent. To model the time-series nature of returns, the Autoregressive Integrated Moving Average (ARIMA) model is employed as a benchmark traditional forecasting tool. ARIMA is widely used in financial econometrics for capturing linear dependencies in time series data (Chu, Chan et al. 2017). Despite its limitations in handling non-stationarity and structural breaks common in crypto markets, ARIMA offers a transparent baseline model. In this study, the ARIMA model was configured with an order of (1, 0, 1), selected based on the lowest Akaike Information Criterion (AIC) during training, and used to forecast returns.

For volatility forecasting, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is used, given its capability to capture volatility clustering and heteroskedasticity, which are key features in cryptocurrency price behavior. GARCH(1,1) was selected as it has consistently shown strong performance in modeling crypto volatility across multiple studies (Bouri, Azzi and Dyhrberg 2017, Katsiampa 2017). The model parameters were estimated using maximum likelihood, and volatility was forecasted over the test period using the conditional variance output. To incorporate multivariate dependencies between different volatility measures (e.g., rolling volatility, Parkinson's, Garman-Klass), the Vector Autoregression (VAR) model is utilized. VAR allows modeling of multiple interdependent time series and has been shown effective in forecasting volatility spillovers and co-movements (Conrad, Custovic and Ghysels 2018). In this study, a VAR(1) model was fitted using the lag length selected by the Bayesian Information Criterion (BIC), and forecasts were made using the fitted values and test-period simulations.

As a non-parametric machine learning method, the Random Forest (RF) regression algorithm is used for its ability to model nonlinear relationships without assuming any prior data distribution. RF has proven robust in financial time series prediction due to its ensemble structure and resistance to overfitting. In this study, the RF model was trained with 100 estimators (trees), using the Gini criterion for splitting and a maximum depth tuned via cross-validation. To capture temporal dynamics and nonlinear interactions more effectively, two recurrent neural network models are employed: Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). These models are particularly suitable for sequential data, as they are designed to

https://doi.org/10.71058/jodac.v9i11005

retain information over long time lags, which is crucial for capturing dependencies in financial time series. The LSTM model used in this study consisted of a single hidden layer with 50 units, a dropout rate of 0.2, and was trained over 50 epochs with a batch size of 32, using the Adam optimizer and Mean Squared Error (MSE) as the loss function. Similarly, the GRU model shared the same architecture and training configuration. These deep learning architectures have been shown to outperform traditional models in various cryptocurrency forecasting tasks (Basher and Sadorsky 2022, Girsang 2023, Seabe, Moutsinga and Pindza 2023, Verma, Tyagi and Aneja 2023). The performance of each model was evaluated using multiple metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).

4. Results & Discussion

4.1 Descriptive Analysis

Figure 1 illustrates the historical daily adjusted closing price of Ethereum from January 2018 to March 2025. The price series reflects the characteristic volatility of the cryptocurrency market, with pronounced peaks and troughs corresponding to major speculative cycles and market corrections. Notably, Ethereum experienced significant surges in price during late 2020 and early 2021, reaching an all-time high above $4,000, followed by sharp declines, indicating strong market overreactions and corrections. The price subsequently stabilized before another bullish trend emerged in 2024. These fluctuations underscore Ethereum’s sensitivity to macroeconomic conditions, market sentiment, and technological developments within the crypto ecosystem. The dynamic behavior captured in the figure justifies the use of advanced modeling approaches to analyze and forecast Ethereum’s return and volatility.

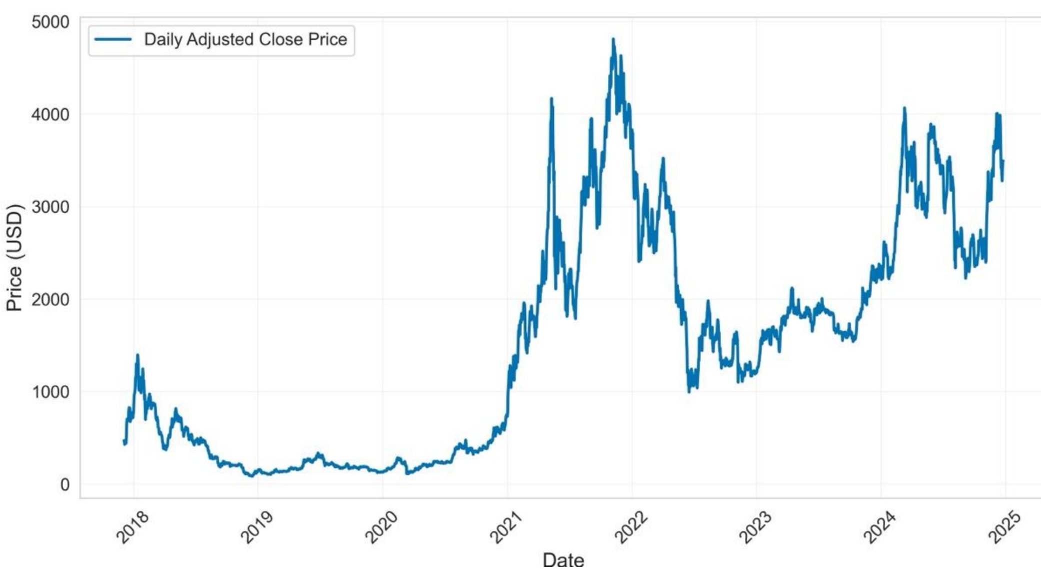

Figure 1. Etherium historical price from 2018 to 2025

Srouce: by author

Figure 2 presents the daily return of Ethereum from 2018 to 2025, capturing the high-frequency fluctuations in percentage changes relative to the previous day’s price. The series clearly exhibits substantial volatility, with frequent sharp spikes and drops, particularly evident during periods of market stress such as the sharp drawdowns in early 2020 and 2022. The return distribution is characterized by heavy tails and volatility

https://doi.org/10.71058/jodac.v9i11005

clustering—features commonly observed in financial time series, especially in cryptocurrency markets. These patterns highlight the limitations of linear models and underscore the need for more advanced approaches such as GARCH and deep learning algorithms to model and predict return dynamics effectively. Overall, the figure provides strong visual evidence of the erratic and high-risk nature of Ethereum's price behavior, justifying the study's focus on volatility forecasting.

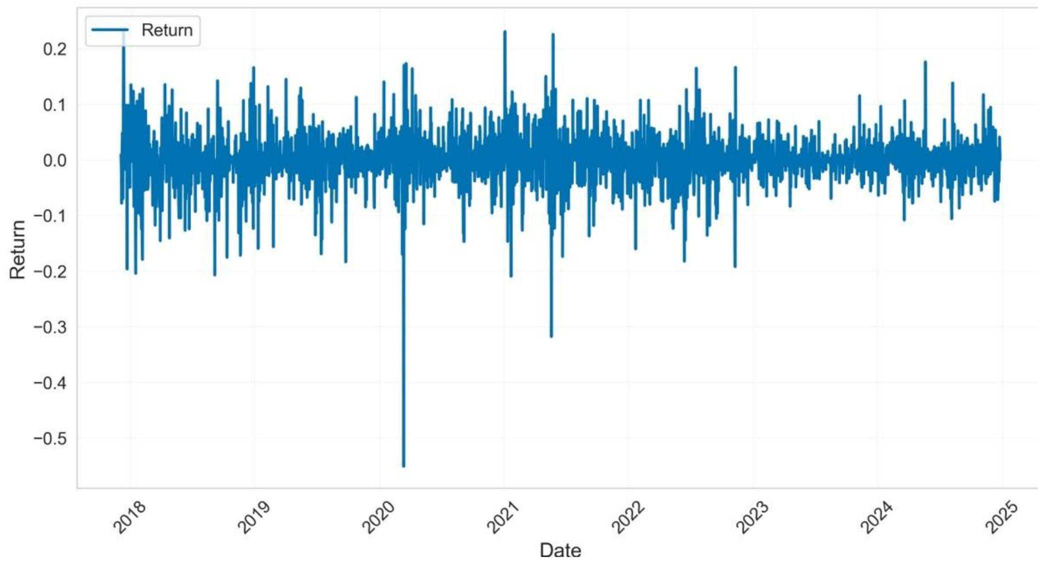

Figure 2. Etherium daily return from 2018 to 2025

Srouce: by author

Figure 3 displays Ethereum's return volatility over the period from 2018 to 2025, as measured by four distinct approaches: Rolling Volatility, Parkinson Volatility, Garman-Klass Volatility, and Rogers-Satchell Volatility. These estimators, derived from daily price data, each offer unique insights into the magnitude and frequency of market fluctuations. Rolling Volatility, calculated as the 5-day standard deviation of log returns, provides a basic time-series view of volatility clustering, which is a key characteristic captured effectively by GARCH-type models. Meanwhile, range-based estimators like Parkinson, Garman-Klass, and Rogers-Satchell incorporate daily high-low and open-close prices, producing more sensitive measures of intraday variation. The frequent and sharp spikes across all four methods—particularly during early 2020, mid-2021, and early 2022—highlight periods of extreme market stress, underscoring the necessity of using models capable of capturing such abrupt shifts in volatility.

https://doi.org/10.71058/jodac.v9i11005

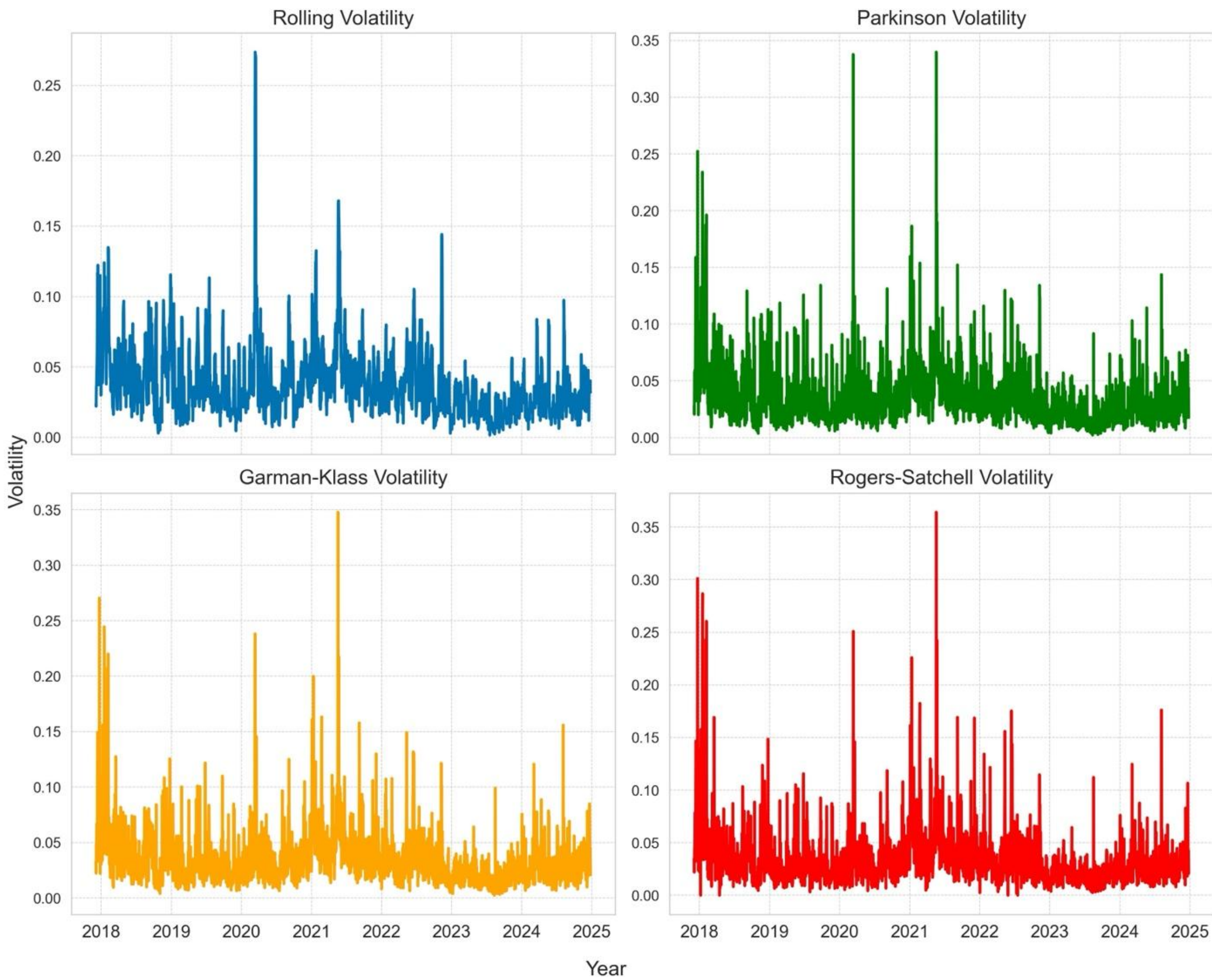

Figure 3. Etherium return volatility by different measures from 2018 to 2025

Srouce: by author

The diverse behaviors observed in these volatility measures motivate the application of multiple predictive modeling frameworks in this study. GARCH and ARIMA are employed to capture the linear and autoregressive dynamics of volatility and returns, respectively, while the Vector Autoregression (VAR) model is used to account for multivariate interactions among the different volatility measures. However, given the evident nonlinearity and temporal dependencies in the data, machine learning algorithms such as Random Forest, LSTM, and GRU are also implemented. These models are designed to handle non-stationary patterns and complex relationships, with LSTM and GRU particularly well-suited for sequential financial data with memory effects. By integrating both traditional econometric and advanced deep learning approaches, the study aims to assess the relative effectiveness of these models in forecasting volatility under varying market conditions.

4.2 Results

Table 2 presents the evaluation results for six predictive algorithms—ARIMA, GARCH, VAR, Random Forest, LSTM, and GRU—across five target variables: log return and four different measures of volatility

https://doi.org/10.71058/jodac.v9i11005

(Rolling, Parkinson, Garman-Klass, and Rogers-Satchell). Among the traditional econometric models, both ARIMA and VAR exhibit similar performance in return prediction, with MAE values of approximately 0.0307 and RMSE values around 0.0441. However, their coefficient of determination (R^2) values are negative, suggesting a poor fit and limited explanatory power. GARCH, while theoretically tailored for modeling volatility, performed poorly across all volatility measures, especially in terms of R^2, which reached values as low as -2.83 in the case of Rolling Volatility. These results suggest that traditional linear models, although useful as benchmarks, struggle to capture the complex and nonlinear patterns inherent in cryptocurrency price behavior.

Table 2. Evaluation metrics for models with return and different volatility measures prediction

Models	**Metrics**	**Log Return**	**Rolling volatility**	**Parkinson volatility**	**Garman Klass volatility**	**Rogers Satchell volatility**
ARIMA	MAE	0.030684	0.017465	0.017805	0.016818	0.017043
	RMSE	0.044123	0.023133	0.026272	0.026207	0.027934
	R^2	-0.001481	-0.001336	-0.000320	-0.000063	-0.000008
GARCH	MAE	0.030691	0.038882	0.033935	0.033717	0.033114
	RMSE	0.044177	0.045236	0.042913	0.042703	0.043322
	R^2	-0.003957	-2.829060	-1.668915	-1.655286	-1.405221
VAR	MAE	0.030694	0.017506	0.017750	0.016777	0.017017
	RMSE	0.044112	0.023150	0.026262	0.026197	0.027924
	R^2	-0.000973	-0.002868	0.000459	0.000688	0.000738
Random Forest	MAE	0.023604	0.010036	0.002589	0.004791	0.007509
	RMSE	0.033908	0.013448	0.006050	0.009285	0.013288
	R^2	0.408563	0.661606	0.946961	0.874469	0.773713
LSTM	MAE	0.026653	0.013201	0.009434	0.009282	0.010707
	RMSE	0.037373	0.017494	0.013552	0.013312	0.015268
	R^2	0.281483	0.427324	0.733830	0.741974	0.701261
GRU	MAE	0.025353	0.011386	0.004024	0.006642	0.008858
	RMSE	0.036632	0.015728	0.005683	0.009625	0.012659
	R^2	0.309707	0.537134	0.953198	0.865100	0.794629

Source: by author

In contrast, machine learning models demonstrate markedly superior performance. Random Forest achieved the best results across all volatility measures, with notably high R^2 values: 0.95 for Parkinson volatility, 0.87 for Garman-Klass, and 0.77 for Rogers-Satchell. Its ability to handle nonlinear relationships and variable interactions likely contributes to this performance advantage, particularly in multi-dimensional feature spaces. Deep learning models, including LSTM and GRU, also show competitive results. GRU slightly outperforms LSTM in most metrics, with lower MAE and RMSE and higher R^2 across all targets, particularly for Parkinson volatility where it achieved an R^2 of 0.95. These findings validate the use of sequence-based neural networks in modeling the temporal dependencies and high-frequency fluctuations characteristic of cryptocurrency markets.

The comparison underscores the limitations of traditional time series models in handling volatile, nonlinear, and high-dimensional financial data. While ARIMA and VAR can provide basic trend forecasting, they fail to account for the rapid structural changes and complex patterns evident in Ethereum's returns and

https://doi.org/10.71058/jodac.v9i11005

volatility. GARCH, despite its popularity in financial econometrics, performs poorly in this context, possibly due to its reliance on specific distributional assumptions. On the other hand, Random Forest offers strong baseline performance with interpretability, while LSTM and GRU provide advanced capabilities to capture long-range dependencies and nonlinear behaviors. These results collectively suggest that hybrid and deep learning approaches are more suitable for forecasting in the cryptocurrency domain, particularly when the goal is to predict volatile assets with high noise-to-signal ratios.

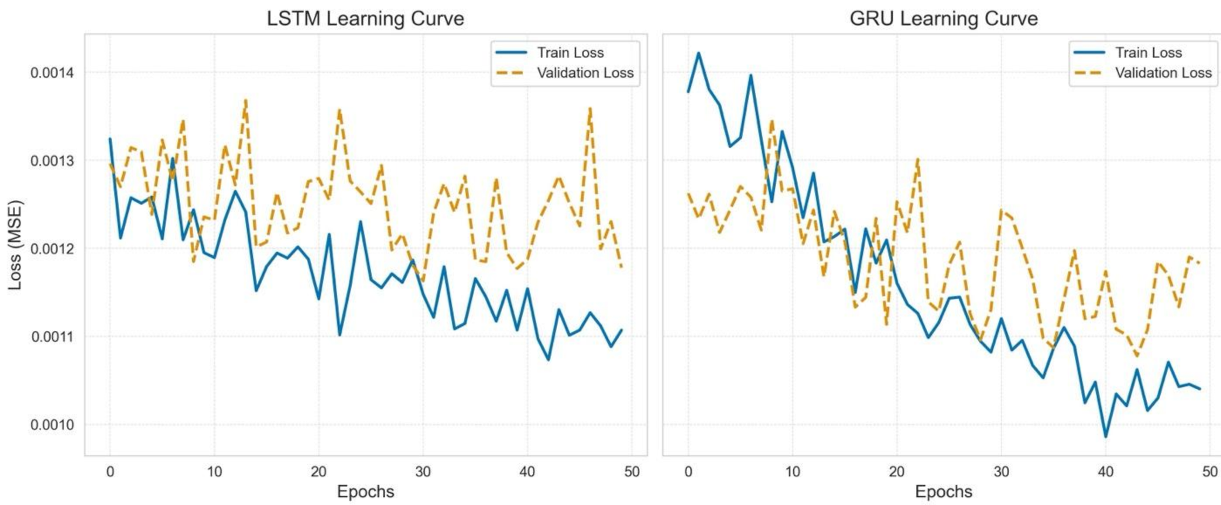

Figure 4. Learning curve of LSTM and GRU models

Srouce: by author

Figure 4 presents the learning curves for the LSTM and GRU models over 50 training epochs, showing both training and validation loss in terms of mean squared error (MSE). The gradual decline in loss for both models indicates successful convergence and learning stability, with GRU demonstrating a slightly smoother and more consistent decrease in both training and validation loss compared to LSTM. This aligns with the performance metrics in Table 2, where GRU consistently outperformed LSTM across all target variables, achieving lower MAE and RMSE values along with higher R^2 scores. Moreover, GRU's more stable learning behavior suggests better generalization capability, as evidenced by its relatively smaller gap between training and validation loss in the final epochs. These results support the robustness of GRU in capturing the complex temporal patterns of cryptocurrency volatility, making it a suitable model for high-frequency and highly volatile financial time series data.

Table 3. Cross train-test evaluation metrics

	Train MAE	Test MAE	Train R^2	Test R^2	Cross-Validation MAE
Random Forest	0,008866	0,023533	0,914064	0,395487	0,024264
ARIMA	0,031379	0,030559	0,000181	-0,00175	0,031406
GARCH	0,039832	0,030562	-0,28757	-6,4E-05	0,031384
VAR	0.030142	0.031853	-0.000913	-0.000272	0.030228
LSTM	0,0226	0,023708	0,53461	0,379081	0,023511
GRU	0,021412	0,023585	0,583726	0,376397	0,022406

Source: by author

https://doi.org/10.71058/jodac.v9i11005

Table 3 provides a comprehensive comparison of model performance in terms of training and testing accuracy, including Mean Absolute Error (MAE), R-squared (R^2), and cross-validation MAE. Among all models, Random Forest exhibits the lowest training MAE (0.008866) and the highest R^2 (0.914064), suggesting strong in-sample fit. However, the substantial drop in R^2 during testing (0.395487) and a significant increase in test MAE (0.023533) indicate potential overfitting. Traditional models such as ARIMA, GARCH, and VAR consistently underperform, showing low or even negative R^2 scores in both training and testing phases. Particularly, the GARCH model displays a negative training R^2 (-0.28757) and a test R^2 close to zero, confirming its limited ability to generalize across volatile and nonlinear cryptocurrency data. Although VAR slightly improves upon ARIMA and GARCH in terms of cross-validation MAE (0.030228), it still lacks predictive strength, as seen by the near-zero R^2 values.

In contrast, deep learning models—LSTM and GRU—demonstrate a more balanced and generalizable performance across all metrics. Both models maintain relatively consistent train-test MAE values and similar R^2 scores around 0.38 in the test set, reflecting better stability in forecasting unseen data compared to Random Forest and traditional methods. Notably, GRU achieves the lowest cross-validation MAE (0.022406), supporting the earlier findings in Table 2 where it yielded the best predictive accuracy across multiple volatility measures. These findings are further supported by the learning curves in Figure 4, which show GRU's smooth and steadily declining loss trends, suggesting superior convergence and less variance between training and validation sets. Overall, the consistency of GRU across different evaluation settings highlights its reliability in capturing the sequential dependencies and nonlinear behaviors typical of cryptocurrency markets.

5. Conclusion & Recommendation

5.1 Conclusion

This study sets out to evaluate and compare the predictive performance of traditional econometric models and advanced machine learning techniques in forecasting the return and volatility of Ethereum, one of the most widely traded cryptocurrencies. Using historical data from January 2018 to March 2025, we applied six predictive algorithms—ARIMA, GARCH, VAR, Random Forest, LSTM, and GRU—across five target variables: log return, and four volatility measures including Rolling, Parkinson, Garman-Klass, and Rogers-Satchell volatility. The goal was to assess the relative forecasting capability of these models and determine which approaches are best suited for capturing the complex, nonlinear, and high-frequency dynamics characteristic of cryptocurrency markets.

The findings reveal that machine learning and deep learning models substantially outperform traditional econometric approaches in both return and volatility prediction. Among the traditional models, ARIMA and VAR showed relatively weak performance, with low R^2 values and higher MAE and RMSE scores. GARCH, while specifically designed to model volatility clustering, demonstrated poor generalization and underperformed across all volatility estimators. In contrast, Random Forest consistently produced the best performance in terms of both accuracy and R^2, particularly for Parkinson and Garman-Klass volatility measures. LSTM and GRU also performed well, with GRU slightly outperforming LSTM in both return and volatility forecasts. These outcomes align with the observed learning curves, where GRU exhibited more stable and consistent training dynamics, further affirming its robustness in modeling sequential financial data.

Compared to prior studies, these results reinforce and extend earlier findings. Like Seabe, Moutsinga and Pindza (2023) and Goutte, Le et al. (2023), our study confirms that deep learning models, especially recurrent architectures like LSTM and GRU, are highly effective in capturing the temporal dependencies

https://doi.org/10.71058/jodac.v9i11005

in crypto price movements. Furthermore, the superior performance of Random Forest echoes the conclusions of Basher and Sadorsky (2022) and Akyildirim, Goncu and Sensoy (2021), who emphasized the utility of ensemble methods in financial prediction tasks. However, our research uniquely contributes by offering a direct head-to-head comparison of these models within a unified framework, applying them not only to return forecasting but also to a diverse set of volatility estimation techniques, an aspect largely overlooked in earlier studies.

The superior performance of Random Forest can be attributed to its ability to handle nonlinearities and complex interactions without strong assumptions about data distribution. Its ensemble nature reduces variance and enhances generalization. GRU and LSTM, on the other hand, excel due to their internal gating mechanisms, which enable the retention and selective forgetting of information over time. GRU's simpler architecture and fewer parameters may explain its slightly better generalization and training efficiency in this study. In contrast, the poor results from ARIMA and GARCH models highlight their limitations in adapting to the erratic behavior and regime shifts typical of cryptocurrency markets.

This research contributes to the growing literature on cryptocurrency analytics in several ways. First, it provides a systematic and comparative evaluation of both traditional and advanced models for Ethereum forecasting. Second, it incorporates a diverse set of volatility estimators, offering a more holistic understanding of volatility dynamics in digital assets. Third, by using a rich dataset spanning over seven years, the study delivers empirically grounded insights into model performance under varying market conditions. The findings not only guide researchers in selecting appropriate models for crypto time series but also offer practical implications for investors, portfolio managers, and policymakers seeking to manage risk and understand market behavior in the digital asset space.

5.2 Recommendation

Based on the empirical findings and comparative analysis presented in this study, several practical and research-oriented recommendations can be proposed. First, given the consistent superiority of machine learning and deep learning models over traditional econometric approaches, practitioners such as quantitative analysts, crypto investors, and financial data scientists are encouraged to prioritize models like Random Forest, GRU, and LSTM when developing forecasting systems for cryptocurrency return and volatility. These models not only offer improved accuracy but also adapt better to nonlinear relationships and regime shifts, features that are highly characteristic of crypto markets like Ethereum. Moreover, Random Forest can be especially useful in operational settings where interpretability and faster implementation are required, while GRU is recommended in cases where sequential dependencies and training efficiency are critical.

Second, model selection should consider the nature of the volatility measure being used. As demonstrated in this study, models performed differently across volatility estimators such as Parkinson, Garman-Klass, and Rogers-Satchell. Therefore, financial modelers and risk managers should consider testing multiple estimators rather than relying on a single volatility definition, especially when applying models in high-risk environments or during market turbulence. Furthermore, the poor performance of GARCH across all settings suggests that its continued use in cryptocurrency research should be re-evaluated, particularly when more robust and adaptive alternatives are available.

Lastly, future research should explore hybrid modeling approaches that integrate traditional volatility theory with deep learning architecture. For example, incorporating GARCH-based features into neural network input structures or using ensemble learning strategies that combine LSTM with tree-based models could potentially capture both long-term volatility structure and short-term nonlinear patterns. In addition, expanding the feature set to include sentiment indicators, macroeconomic variables, and blockchain-specific data (e.g., transaction volume, gas fees) may enhance the explanatory power of these models. These

https://doi.org/10.71058/jodac.v9i11005

advancements would not only improve forecasting precision but also deepen our understanding of the multifaceted drivers of volatility in the digital asset ecosystem.

5.3 Limitations & Further research

One key limitation of this study lies in its reliance on historical price data and technical indicators derived solely from market variables, such as open, high, low, close, and volume. While these inputs are widely used and effective for time series modeling, they may not fully capture the influence of external factors such as regulatory announcements, macroeconomic shocks, investor sentiment, or blockchain-specific events (e.g., network upgrades or gas fee spikes), all of which can significantly affect cryptocurrency volatility. Additionally, although the deep learning models demonstrated strong performance, their interpretability remains limited compared to traditional models, posing challenges in explaining model decisions to stakeholders. Furthermore, the study focuses exclusively on Ethereum, which, while representative of major cryptocurrencies, may limit the generalizability of findings to other coins with different market behaviors or liquidity profiles.

For future research, it would be valuable to integrate alternative data sources such as social media sentiment, Google Trends, or blockchain network metrics to enrich the input space and potentially improve model performance. Researchers could also explore hybrid frameworks that combine traditional volatility estimators with neural architecture or apply attention mechanisms and transformer-based models to better capture temporal patterns and contextual dependencies. Moreover, extending the analysis to a broader set of cryptocurrencies, such as Bitcoin, Solana, or emerging altcoins, would provide comparative insights and test the robustness of the proposed models across different market conditions. Finally, applying explainable AI (XAI) techniques could help bridge the gap between predictive power and interpretability, enhancing both the transparency and utility of advanced forecasting models in financial decision-making.

Funding

The research is funded by the University of Economics and Law, Vietnam National University, Ho Chi Minh City, Vietnam.

References

Akyildirim, E., A. Goncu and A. Sensoy (2021). "Prediction of cryptocurrency returns using machine learning." Annals of Operations Research **297**: 3-36.

Anamika, M. Chakraborty and S. Subramaniam (2023). "Does sentiment impact cryptocurrency?" Journal of Behavioral Finance **24**(2): 202-218.

Auer, R. and S. Claessens (2021). Cryptocurrency market reactions to regulatory news 1. The Routledge handbook of FinTech, Routledge**:** 455-468.

Basher, S. A. and P. Sadorsky (2022). "Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?" Machine Learning with Applications **9**: 100355.

Bouri, E., G. Azzi and A. H. Dyhrberg (2017). "On the return-volatility relationship in the Bitcoin market around the price crash of 2013." Economics **11**(1): 2.

Catania, L., S. Grassi and F. Ravazzolo (2019). "Forecasting cryptocurrencies under model and parameter instability." International Journal of Forecasting **35**(2): 485-501.

https://doi.org/10.71058/jodac.v9i11005

Chu, J., et al. (2017). "GARCH modelling of cryptocurrencies." Journal of Risk and Financial Management **10**(4): 17.

Conrad, C., A. Custovic and E. Ghysels (2018). "Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis." Journal of Risk and Financial Management **11**(2): 23.

Darley, O. G., A. I. Yussuff and A. A. Adenowo (2021). "Price analysis and forecasting for bitcoin using auto regressive integrated moving average model." Annals of Science and Technology **6**(2): 47-56.

Ejder, U. and S. A. Özel (2024). "A novel distance-based moving average model for improvement in the predictive accuracy of financial time series." Borsa Istanbul Review **24**(2): 376-397.

Fama, E. F. (1970). "Efficient capital markets." Journal of finance **25**(2): 383-417.

Girsang, A. S. (2023). "Hybrid LSTM and GRU for cryptocurrency price forecasting based on social network sentiment analysis using FinBERT." IEEE Access **11**: 120530-120540.

Goutte, S., et al. (2023). "Deep learning and technical analysis in cryptocurrency market." Finance Research Letters **54**: 103809.

Katsiampa, P. (2017). "Volatility estimation for Bitcoin: A comparison of GARCH models." Economics Letters **158**: 3-6.

Kristjanpoller, W. and M. C. Minutolo (2015). "Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model." Expert systems with applications **42**(20): 7245-7251.

Kuznetsov, O., et al. (2023). Comparative Analysis of Cryptocurrency Trading Platforms Using the Analytic Hierarchy Process. ITTAP.

Lahmiri, S. and S. Bekiros (2019). "Cryptocurrency forecasting with deep learning chaotic neural networks." Chaos, Solitons & Fractals **118**: 35-40.

Liu, Y. and A. Tsyvinski (2021). "Risks and returns of cryptocurrency." The Review of Financial Studies **34**(6): 2689-2727.

Markowitz, H. M. (1991). "Foundations of portfolio theory." The journal of finance **46**(2): 469-477.

Murray, K., et al. (2023). "On forecasting cryptocurrency prices: A comparison of machine learning, deep learning, and ensembles." Forecasting **5**(1): 196-209.

O'hara, M. (1998). Market microstructure theory, John Wiley & Sons.

Patel, M. M., et al. (2020). "A deep learning-based cryptocurrency price prediction scheme for financial institutions." Journal of information security and applications **55**: 102583.

Peters, E. E. (1996). Chaos and order in the capital markets: a new view of cycles, prices, and market volatility, John Wiley & Sons.

Seabe, P. L., C. R. B. Moutsinga and E. Pindza (2023). "Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach." Fractal and Fractional **7**(2): 203.

https://doi.org/10.71058/jodac.v9i11005

Seo, M. and G. Kim (2020). "Hybrid forecasting models based on the neural networks for the volatility of bitcoin." Applied Sciences **10**(14): 4768.

Sun, X., M. Liu and Z. Sima (2020). "A novel cryptocurrency price trend forecasting model based on LightGBM." Finance Research Letters **32**: 101084.

Tanwar, S., et al. (2021). "Deep learning-based cryptocurrency price prediction scheme with inter-dependent relations." IEEE Access **9**: 138633-138646.

Thaler, R. (1980). "Toward a positive theory of consumer choice." Journal of economic behavior & organization **1**(1): 39-60.

Urquhart, A. (2016). "The inefficiency of Bitcoin." Economics Letters **148**: 80-82.

Vaddi, L., et al. (2020). "Predicting crypto currency prices using machine learning and deep learning techniques." Int. J **9**(4).

Verma, U., P. Tyagi and M. K. Aneja (2023). "Multi-branch cnn gru with attention mechanism for human action recognition." Engineering Research Express **5**(2): 025055.

Yen, K.-C. and H.-P. Cheng (2021). "Economic policy uncertainty and cryptocurrency volatility." Finance Research Letters **38**: 101428.