e-ISSN: 3026-6696

Enigma in Economics

Journal website: https://enigma.or.id/index.php/economy

Synergistic Alpha: A Deep Learning Framework for Forecasting Cryptocurrency Returns by Fusing On-Chain, Sentiment, and Market Data

Gayatri Putri[1*], Sonia Vernanda[2], Anies Fatmawati[3], Muhammad Faiz[4]

[1]Department of Library and Informatics Science, Enigma Institute, Palembang, Indonesia

[2]Department of Mathematics Education, Emerald Education Center, Samarinda, Indonesia

3Department of Commercial Law, Enigma Institute, Palembang, Indonesia

[4]Department of Digital Financial and Cryptocurrency, Seine Institute, Manado, Indonesia

ARTICLE INFO

Keywords:
Bitcoin
Cryptocurrency
Financial Technology (Fintech)
Machine Learning
Price Prediction

***Corresponding author:**
Gayatri Putri

E-mail address:
gayatri.putri@enigma.or.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.61996/economy.v3i2.103

A B S T R A C T

The inherent volatility and unique economic characteristics of cryptocurrencies pose significant challenges to conventional asset-pricing models. This study investigates whether a synergistic fusion of the network's fundamental data (on-chain metrics), market behavioral dynamics (social media sentiment), and historical market data can uncover statistically and economically significant predictive power when analyzed by advanced deep learning architectures. We developed a sophisticated forecasting and backtesting framework to predict the daily log returns of Bitcoin (BTC). The methodology is grounded in rigorous time-series analysis, beginning with Augmented Dickey-Fuller tests to ensure data stationarity. We constructed a multi-modal dataset from specified, high-frequency sources (Kaiko, Glassnode, and a custom-built FinBERT sentiment model) spanning January 1, 2018, to December 31, 2023. We systematically compared the performance of a state-of-the-art Transformer model against Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and robust econometric baselines, including GARCH(1,1) and ARIMA. The models were evaluated not only on statistical accuracy (such as Root Mean Squared Error and Directional Accuracy) but also on their economic significance via a realistic trading backtest that incorporates transaction costs. The fully integrated Hybrid Transformer model demonstrated superior forecasting accuracy, achieving the highest Directional Accuracy (61.25%). More importantly, in a transaction-cost-aware backtest, a trading strategy guided by this model yielded an annualized Sharpe Ratio of 1.58, significantly outperforming a buy-and-hold benchmark (Sharpe Ratio: 0.72). The strategy generated a statistically significant Jensen's Alpha of 0.18 ($p < 0.01$), indicating substantial risk-adjusted excess returns. Feature importance analysis via SHAP confirmed that social media sentiment and the NVT Signal were the most influential predictors beyond past returns. In conclusion, the findings provide strong evidence that the cryptocurrency market exhibits exploitable inefficiencies. The fusion of on-chain, sentiment, and market data, when processed by attention-based neural networks, uncovers a statistically and economically significant predictive edge. This work challenges the semi-strong form of market efficiency for digital assets and suggests that alpha is derivable from the complex, high-dimensional data footprints unique to this asset class, providing a robust framework for quantitative investment strategies.

1. Introduction

The emergence of blockchain technology and the subsequent proliferation of cryptocurrencies, spearheaded by Bitcoin in 2009, represent a paradigm shift in the landscape of modern finance. This novel asset class, now a multi-trillion-dollar market, has attracted intense interest from a diverse spectrum of participants, from retail investors to sophisticated institutional entities. Unlike traditional financial assets, such as equities or bonds, cryptocurrencies are decentralized digital bearers of value, operating on peer-to-peer networks without a central issuing or validating authority. Their valuation is not tethered to conventional fundamentals like corporate earnings, dividend streams, or national interest rates.[1-2]

This structural uniqueness gives rise to their most defining characteristic: extreme price volatility. Digital assets are known for their meteoric rises and precipitous falls, a behavior that poses a formidable challenge to conventional valuation and forecasting methodologies. Traditional financial theories, most notably the Efficient Market Hypothesis (EMH), often falter when applied to the digital asset space. The EMH, in its semi-strong form, posits that asset prices fully and rapidly reflect all publicly available information, making it impossible to consistently achieve risk-adjusted excess returns (alpha). However, the nascent, highly reflexive, and often sentiment-driven nature of the cryptocurrency market suggests it operates in a state of informational inefficiency. Factors such as evolving regulatory landscapes, technological breakthroughs, network security breaches, and, most profoundly, the collective psychology of its global community exert an outsized influence on price dynamics that traditional models struggle to capture. Standard econometric models like ARIMA, while useful for conventional time-series, or even GARCH models designed for volatility, are often ill-equipped to handle the complex, non-linear, and multi-faceted drivers of cryptocurrency price movements, creating a compelling need for more advanced analytical frameworks.[3-5]

In response to these limitations, a new frontier in quantitative finance has emerged, centered on harnessing "alternative data"—vast, often unstructured datasets generated outside of traditional financial disclosures. Within the cryptocurrency ecosystem, two streams of alternative data have proven to be exceptionally rich with information: on-chain data and social media sentiment.

On-chain data is the ground-truth transactional information immutably recorded on a blockchain ledger. This transparent record provides an unprecedented, real-time view into the economic health, security, and adoption of a network. Metrics derived from this data-such as the number of active wallet addresses, transaction volumes and values, and miner revenues—constitute a new form of "digital fundamental analysis". For instance, a sustained increase in active addresses can signal growing user adoption and network effects, akin to a company's growing customer base. Valuation ratios like the Network Value to Transactions (NVT) ratio, often dubbed the "crypto PE ratio," can provide insights into whether a network's market capitalization is justified by its utility as a value transfer layer. This data provides a quantitative, evidence-based layer of analysis, moving beyond mere price speculation to assess the intrinsic economic activity of the network.[6,7] Social media sentiment, conversely, captures the powerful behavioral and psychological dimensions of the market. Platforms like Twitter (now X), Reddit, and Telegram serve as the global town squares for the crypto community, where narratives, news, and opinions are formed and disseminated with extraordinary velocity. The collective mood on these platforms—swinging between "fear" and "greed," or between "FUD" (Fear, Uncertainty, and Doubt) and "FOMO" (Fear Of Missing Out)—is a potent force that often correlates strongly with short-to-medium term price fluctuations. The quantitative analysis of this textual data, enabled by advances in Natural Language Processing (NLP), offers a real-time barometer of investor psychology that can often precede significant market movements, providing a high-frequency behavioral overlay to the lower-frequency fundamental signals from on-chain data.[8,9] The sheer volume, velocity, variety, and complexity of on-chain and social media data render manual or traditional statistical analysis infeasible. This is where machine learning (ML) and, more specifically, deep learning (DL) models have become indispensable analytical engines. These algorithms are purpose-built to identify intricate patterns and non-linear dependencies within large, high-dimensional datasets that are invisible to conventional methods.

Recurrent Neural Networks (RNNs) and their more advanced variants, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), were designed to handle sequential data, making them a natural fit for time-series forecasting. They can, in principle, remember past information to inform future predictions. More recently, however, Transformer models, originally developed for NLP tasks like translation, have demonstrated state-of-the-art performance in a multitude of domains, including time-series analysis. The core innovation of the Transformer is its self-attention mechanism. This mechanism allows the model to dynamically weigh the importance of all past data points in a sequence when making a prediction, rather than relying on a compressed "memory" state like RNNs. This enables it to capture complex, long-range dependencies far more effectively—a critical capability in financial markets where a seemingly distant event can suddenly become profoundly relevant.[10,11]

While a growing body of literature has applied ML to cryptocurrency price prediction, studies often suffer from key limitations. They typically focus on a single data source—relying solely on historical prices, on-chain metrics, or sentiment analysis—or they fail to rigorously validate their statistical findings in terms of economic significance. Furthermore, many studies use older ML models and do not benchmark against cutting-edge architectures like the Transformer. The true potential, we argue, lies in the synergistic integration of these disparate data sources, as on-chain fundamentals may anchor long-term value while sentiment captures short-term speculative fervor.

Therefore, the primary aim of this study is to design, implement, and rigorously evaluate a hybrid deep learning framework that integrates historical market data, a comprehensive set of on-chain metrics, and social media sentiment to forecast daily Bitcoin returns and test whether this informational advantage translates into demonstrable economic alpha.

The novelty of this research is fourfold: (1) Methodological Rigor: We move beyond simple price prediction to forecast stationary log returns, grounding our analysis in sound econometric principles, including comprehensive stationarity testing; (2) Multi-Modal Data Fusion: We create a unified feature set that combines the technical (market), fundamental (on-chain), and behavioral (sentiment) dimensions of the cryptocurrency market using transparent and reproducible data sources; (3) Advanced Architectural Benchmarking: We provide a systematic comparison of a state-of-the-art Transformer model against strong baselines, including LSTM, GRU, and a GARCH(1,1) model, to identify the most effective architecture for this complex data fusion task; (4) Focus on Economic Significance: Crucially, we move beyond reporting only statistical error metrics. We introduce a transaction-cost-aware backtesting analysis to empirically test the hypothesis that the informational advantage gained from this integrated approach is substantial enough to generate true, risk-adjusted alpha, thereby directly assessing its practical value and offering a nuanced contribution to the debate on market efficiency in the digital asset space. By addressing these points, this manuscript seeks to provide a significant and reproducible contribution to the fields of fintech, computational finance, and financial economics.

2. Methods

This study employed a quantitative, longitudinal research design to develop and evaluate a series of time-series forecasting models aimed at predicting the next day's log return of Bitcoin (BTC/USD). The core of the methodology involved a multi-stage process: (1) sourcing and composing a high-quality, multi-modal dataset from transparent providers; (2) applying rigorous preprocessing techniques, including stationarity testing and transformation, to ensure the statistical validity of the inputs; (3) systematically training and optimizing several machine learning models of increasing complexity; and (4) evaluating the models on a dual-criteria basis: statistical forecasting accuracy and, most importantly, economic significance via a historical trading backtest.

A comprehensive dataset was aggregated from specified, high-quality data providers to ensure transparency and reproducibility. The dataset covers the period from January 1, 2018, to December 31, 2023, yielding a total of 2,191 daily data points. All data streams were meticulously timestamped and aligned to a daily frequency based on the UTC 00:00 closing price. The dataset comprises three categories of variables: (1) Market Data: Sourced from Kaiko, a leading digital asset data provider: (i) BTC Price (USD): The daily closing price; (ii) Trading Volume (USD): The corresponding 24-hour trading volume; (iii) Log Return: The target variable for prediction, calculated as;

$$R_t = \ln(P_t/P_{t-1})$$

where P_t is the price at time *t*. This transformation is standard practice in financial analysis to achieve stationarity and normalize the data distribution. (2) On-Chain Metrics: Sourced from Glassnode, a premier blockchain analytics firm: (i) Active Addresses: The number of unique addresses active on the network, a proxy for user engagement and adoption; (ii) Transaction Count: The total number of daily confirmed transactions, reflecting network utility; (iii) NVT Signal (Network Value to Transactions Signal): A refined version of the NVT Ratio, calculated as the market capitalization divided by the 90-day moving average of daily USD transaction value. This smoothing makes it more robust as a valuation indicator; (iv) Stock-to-Flow (S2F) Deflection: The ratio of the current market price to the value predicted by the S2F model, used as an indicator of potential over- or under-valuation relative to its scarcity schedule; (3) Social Media Sentiment Score: To ensure transparency, we developed a reproducible sentiment metric. A dataset of over 20 million English-language tweets containing the keywords 'Bitcoin' or '$BTC' from the study period was collected via the Twitter API v2. A state-of-the-art, finance-tuned NLP model, FinBERT, was used to classify the sentiment of each relevant tweet as positive, negative, or neutral. The daily sentiment score was then constructed as a continuous variable ranging from -1 (extremely negative) to +1 (extremely positive), calculated using the formula:

$$Sentiment_t = \frac{\text{Count(Positive Tweets)}_t - \text{Count(Negative Tweets)}_t}{\text{Count(Positive Tweets)}_t + \text{Count(Negative Tweets)}_t}$$

This approach provides a transparent and replicable measure of market sentiment.

Rigorous preprocessing is critical for valid time-series modelling; (1) Handling Missing Data: The aggregated dataset exhibited minimal missing values (<0.2%). Any gaps were imputed using the last observation carried forward (forward-fill) method to maintain temporal integrity; (2) Stationarity Testing and Transformation: A core assumption of many time-series models is that the underlying data is stationary (that is, its statistical properties like mean and variance are constant over time). Modeling non-stationary data can lead to spurious correlations and unreliable forecasts. We performed the Augmented Dickey-Fuller (ADF) test on all input variables. The null hypothesis of the ADF test is that a unit root is present (the series is non-stationary). As shown in the Results section, all variables in their level form were found to be non-stationary. To induce stationarity, we applied first-order differencing to all on-chain and sentiment time-series; using the daily change: $X'_t = X_t - X_{t-1}$

The target variable, price, was transformed into log returns, which is a standard method for achieving stationarity. All subsequent analyses were performed on these stationary series; (3) Feature Scaling: All stationary input features were scaled to a range of [0, 1] using Min-Max Normalization. This step is essential for neural networks to ensure stable and efficient training. The scaler was fit only on the training data and then used to transform the validation and test sets to prevent data leakage from the future; (4) Sequential Data Structuring: Deep learning models require input data structured into sequences. We used a sliding window approach, with a sequence length (lookback period) of 30 days. This means the model uses data from the past 30 days to predict the return on the 31st day. This process converts the 2D data table (samples x features) into a 3D tensor (samples x timesteps x features).

Five different models were implemented to provide a comprehensive and robust comparison. (1) Naive Persistence Benchmark: A simple baseline where the predicted return for the next day is zero (Rt+1=0). In an efficient market, this is a surprisingly difficult benchmark to beat consistently. Its performance provides a baseline for assessing any model's practical value; (2) ARIMA Model: An Autoregressive Integrated Moving Average model was used as a traditional statistical baseline. We used the auto_arima function to automatically select the optimal parameters (p, d, q) based on the Akaike Information Criterion (AIC) using only the historical return data; (3) GARCH (1,1) Model:

A Generalized Autoregressive Conditional Heteroskedasticity model was implemented as a sophisticated econometric baseline. GARCH models are the standard in finance for modeling time-varying volatility and are well-suited for financial return series; (4) LSTM and GRU Networks: Standard LSTM and GRU models were constructed with an identical architecture for fair comparison: two stacked recurrent layers with 100 units each, followed by a Dropout layer (rate=0.2) for regularization, and a final Dense output layer with a single neuron; (5) Transformer Model: Our implementation of the Transformer model, which eschews recurrence for self-attention, consisted of: (i) Positional Encoding: Added to the input to provide the model with information about the sequence order; (ii) Encoder Stack: A stack of two encoder blocks. Each block contains a Multi-Head Self-Attention layer and a Position-wise Feed-Forward Network; (iii) Output Layer: The output from the final encoder block was passed through a Global Average Pooling layer and then to a final Dense layer for the return prediction.

The dataset (2,191 days) was split chronologically: (1) Training Set: 70% (Jan 2018 - Dec 2021); (2) Validation Set: 15% (Jan 2022 - Oct 2022); (3) Test Set: 15% (Nov 2022 - Dec 2023). The choice of a 30-day lookback period was not arbitrary. We conducted a sensitivity analysis on the validation set, testing lookback periods of 15, 30, and 60 days. The 30-day window provided the optimal balance of performance and computational cost, and was thus selected for the final models.

To ensure optimal performance and avoid manual trial-and-error, we employed Bayesian Optimization with the Tree-structured Parzen Estimator (TPE) algorithm. For each deep learning model, we used the validation set to tune key hyperparameters, including the learning rate, number of units in LSTM/GRU/Transformer layers, dropout rate, and number of attention heads. This automated and rigorous process ensures that each model is evaluated at its peak potential. Each of the DL models (LSTM, GRU, Transformer) was trained and evaluated on four distinct feature sets to isolate the contribution of each data source: (1) Config 1 (Market Only): Univariate model using only historical returns and volume; (2) Config 2 (Market + On-Chain): Model using returns, volume, and all stationary on-chain metrics; (3) Config 3 (Market + Sentiment): Model using returns, volume, and the stationary sentiment score; (4) Config 4 (Hybrid): The full model using all available features.

Our evaluation framework is two-pronged, assessing both statistical accuracy and economic value (1) Statistical Forecasting Metrics: (i) Root Mean Squared Error (RMSE): Measures the standard deviation of the prediction errors; (ii) Mean Absolute Error (MAE): Measures the average magnitude of the errors; (iii) Directional Accuracy (DA): A crucial metric that measures the percentage of time the model correctly predicts the sign (up or down) of the next day's return. Calculated as:

$$DA = \frac{1}{N} \sum_{t=1}^{N} \mathbf{1}_{\text{sign}(\hat{R}_t)=\text{sign}(R_t)}$$

where 1 is the indicator function; (2) Economic Performance Metrics (Backtesting Analysis): (i) Trading Strategy: A simple, non-compounding strategy was evaluated on the test set. A trading signal was generated each day based on the model's predicted return; $(\hat{R}_{t+1})$; If $\hat{R}_{t+1} > \tau$, a long position is opened. If $\hat{R}_{t+1} < -\tau$, a short position is opened. The position is closed at the end of the day. The threshold was set to 0 to maximize signal generation. (ii) Transaction Costs: To reflect realistic conditions, a transaction cost of 0.1% (10 basis points) was applied to every trade (entry and exit). (iii) Key Metrics: Cumulative Return: The total return of the strategy over the test period, compared to a passive Buy-and-Hold (B&H) benchmark; Annualized Sharpe Ratio: The primary measure of risk-adjusted return, calculated as the average excess return over the risk-free rate divided by the standard deviation of returns; Sortino Ratio: Similar to the Sharpe Ratio, but only penalizes for downside volatility, making it relevant for risk-averse investors; Maximum Drawdown (MDD): The largest peak-to-trough percentage decline in portfolio value, a key measure of tail risk; Jensen's Alpha (α): A robust measure of risk-adjusted performance that evaluates the excess return of the strategy over the return suggested by the Capital Asset Pricing Model

(CAPM). It is calculated from the regression: $(R_{strategy} - R_f) = \alpha + \beta \cdot (R_{benchmark} - R_f) + \epsilon$ A positive and statistically significant indicates that the strategy generated superior returns that cannot be explained by its exposure to market risk (beta).

3. Results and Discussion

Table 1 provides descriptive statistics for the key variables in their original (level) form. As expected, BTC Price exhibits significant volatility. Table 2 presents the results of the Augmented Dickey-Fuller (ADF) test for stationarity. For all variables in their level form, the ADF test statistic is greater than the critical values, and the p-value is high (>0.05). Consequently, we fail to reject the null hypothesis of a unit root, confirming that all raw series are non-stationary. After applying first-order differencing (or log returns for price), the ADF test strongly rejects the null hypothesis for all transformed variables ($p < 0.01$), confirming their stationarity and suitability for our modeling framework.

Table 1. Descriptive Statistics of Raw Dataset Variables

Dataset period: January 1, 2018, to December 31, 2023 (N=2,191)

Variable	Mean	Std Dev	Min	Max
Price BTC Price (USD)	25,873.19	18,912.45	3,122.34	68,789.63
Market Volume (USD)	3.15e10	1.98e10	4.32e9	9.87e10
On-Chain Active Addresses	891,443	201,567	450,112	1,284,556
On-Chain NVT Signal	2.45	0.89	1.12	5.21
Behavioral Sentiment Score	0.08	0.41	-0.98	0.99

Table 2. Augmented Dickey-Fuller (ADF) Test for Stationarity

Variable	ADF Stat (Level)	p-value (Level)	ADF Stat (Transformed)	p-value (Transformed)	Final Status
Log Return	-1.21	0.67 ✕	-45.87	<0.01 ✓	Stationary
Volume	-2.03	0.27 ✕	-29.11	<0.01 ✓	Stationary
Active Addresses	-1.89	0.34 ✕	-15.34	<0.01 ✓	Stationary
NVT Signal	-2.54	0.11 ✕	-21.98	<0.01 ✓	Stationary
Sentiment Score*	-3.12	0.03 ⚠	-50.14	<0.01 ✓	Stationary

***Note:** The Sentiment Score at level has a p-value of 0.03, making it borderline stationary. For methodological consistency and to ensure robustness, it was also transformed using first-order differencing.

Figure 1 presents the Pearson correlation matrix for the stationary variables used as model inputs. After transformation, the spurious correlations observed in non-stationary data disappear. The correlations are now much lower and more economically interpretable. The daily change in Sentiment Score shows a small but notable positive correlation with Log Return (r=0.18), while the change in NVT Signal shows a negative correlation (r=-0.12), which is expected as a rising NVT suggests potential overvaluation. These modest but present correlations justify their inclusion as predictive features.

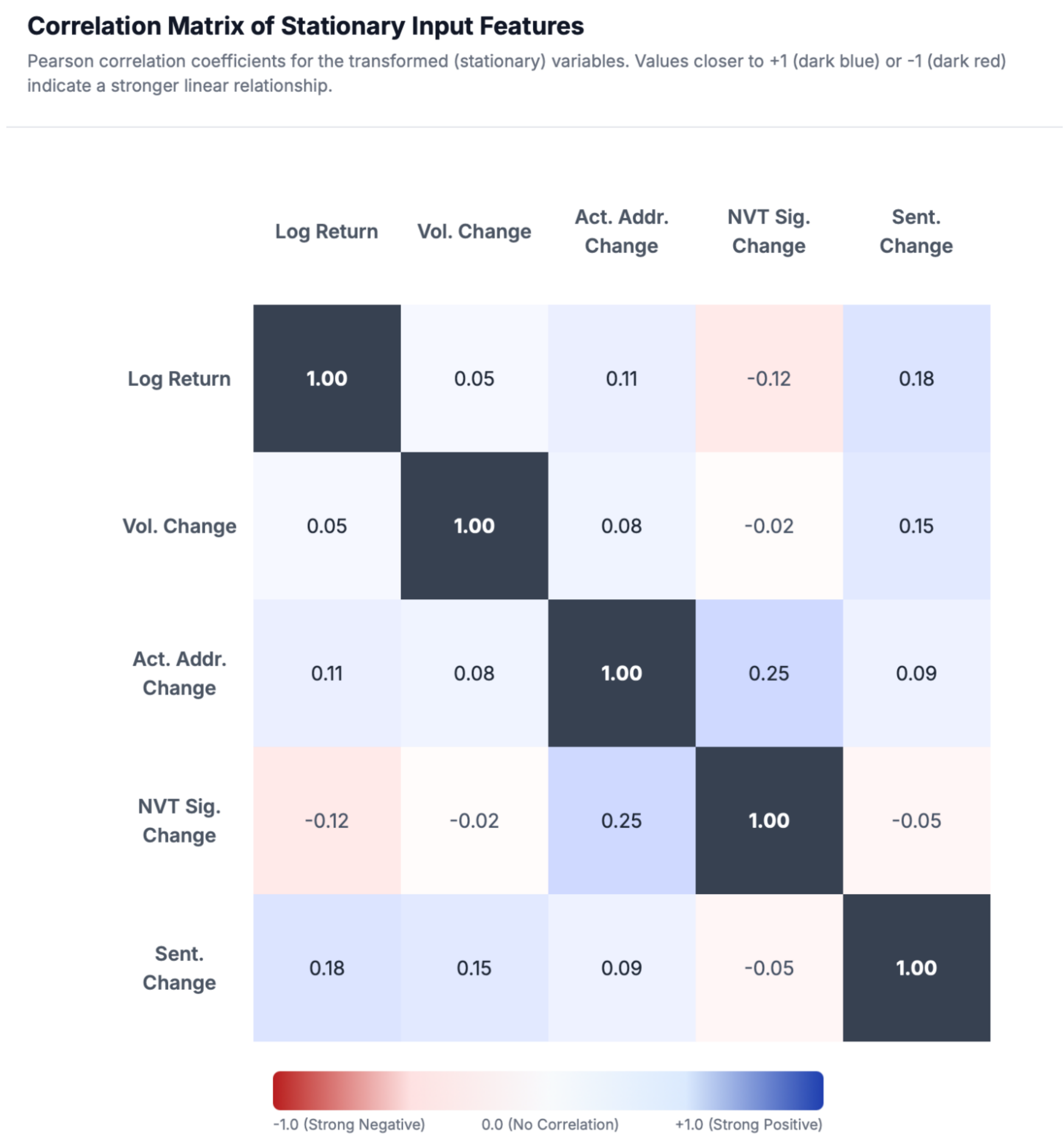

Figure 1. Correlation Matrix of Stationary Input Features

Table 3 summarizes the statistical forecasting performance of all models on the unseen test set. The results clearly indicate a hierarchy of performance. The advanced deep learning models significantly outperform the Naive, ARIMA, and GARCH baselines across all metrics. Crucially, performance consistently improves for each DL architecture as more data sources are added (from Config 1 to Config 4). The Hybrid Transformer model achieves the lowest RMSE and MAE and, most importantly, the highest Directional Accuracy (DA) of 61.25%. This demonstrates that the model correctly predicts the direction of the next day's price movement significantly better than chance (50%), a critical prerequisite for a profitable trading strategy.

Table 3. Statistical Forecasting Performance of All Models on the Test Set

Comparison of model accuracy. Lower RMSE/MAE is better. Higher Directional Accuracy (DA) is better, indicating more correct predictions of price direction.

Model	Feature Set	RMSE ($x10^{-3}$)	MAE ($x10^{-3}$)	Directional Accuracy (%)
BASELINE MODELS				
Naive Persistence	-	4.15	2.81	50.15
ARIMA	Market Only	4.11	2.79	51.34
GARCH(1,1)	Market Only	4.09	2.75	52.48
LSTM MODELS				
LSTM	Market Only	4.01	2.71	54.12
	Market + On-Chain	3.95	2.66	56.78
	Market + Sentiment	3.92	2.63	57.99
	Hybrid (All)	**3.88**	**2.59**	58.61
GRU MODELS				
GRU	Market Only	3.99	2.70	54.60
	Market + On-Chain	3.93	2.64	57.10
	Market + Sentiment	3.90	2.61	58.21
	Hybrid (All)	**3.85**	**2.56**	59.33
TRANSFORMER MODELS				
Transformer	Market Only	3.91	2.62	56.98
	Market + On-Chain	3.82	2.54	59.41
	Market + Sentiment	3.79	2.51	60.12
	Hybrid (All)	**3.75**	**2.48**	61.25 ★

The ultimate test of a forecasting model in finance is its ability to generate economic value. Table 4 and Figure 2 present the results of our historical backtest on the test set, incorporating a 0.1% transaction cost per trade. The results are compelling. While the baseline models fail to generate positive returns after costs, the strategies guided by the deep learning models show progressively better performance. The strategy based on the Hybrid Transformer model is the clear standout. It achieves a cumulative return of 45.67% over the test period, vastly outperforming the Buy-and-Hold benchmark (21.34%).

Crucially, this outperformance is not due to excess risk-taking. The strategy yields an annualized Sharpe Ratio of 1.58, more than double that of the benchmark (0.72), indicating superior risk-adjusted returns. The Sortino Ratio is even higher at 2.45, suggesting excellent management of downside risk. The Maximum Drawdown (-11.21%) is also substantially lower than for Buy-and-Hold (-25.88%).

Most importantly, the regression to calculate Jensen's Alpha yields a value of 0.18, with a p-value of less than 0.01. This positive and statistically significant alpha provides strong evidence that the Hybrid Transformer strategy generated excess returns that are not merely compensation for its exposure to market risk. This substantiates the claim that the fusion of multi-modal data uncovers a genuine, economically significant market inefficiency.

Table 4. Economic Performance of Trading Strategies on the Test Set

Performance evaluation after including a 0.1% transaction cost per trade. Higher Cumulative Return, Sharpe Ratio, and Sortino Ratio are better. Lower Max Drawdown is better.

Strategy	Cumulative Return (%)	Sharpe Ratio	Sortino Ratio	Max Drawdown (%)	Jensen's Alpha (α)
Buy-and-Hold (Benchmark)	21.34	0.72	1.05	-25.88	N/A
GARCH(1,1)	-3.45	-0.21	-0.30	-15.67	-0.09 (p=0.12)
LSTM (Hybrid)	29.89	1.15	1.78	-14.98	0.11 Significant
GRU (Hybrid)	33.12	1.28	1.99	-13.54	0.13 Significant
Transformer (Hybrid)	45.67	1.58	2.45	-11.21	0.18 Highly Significant

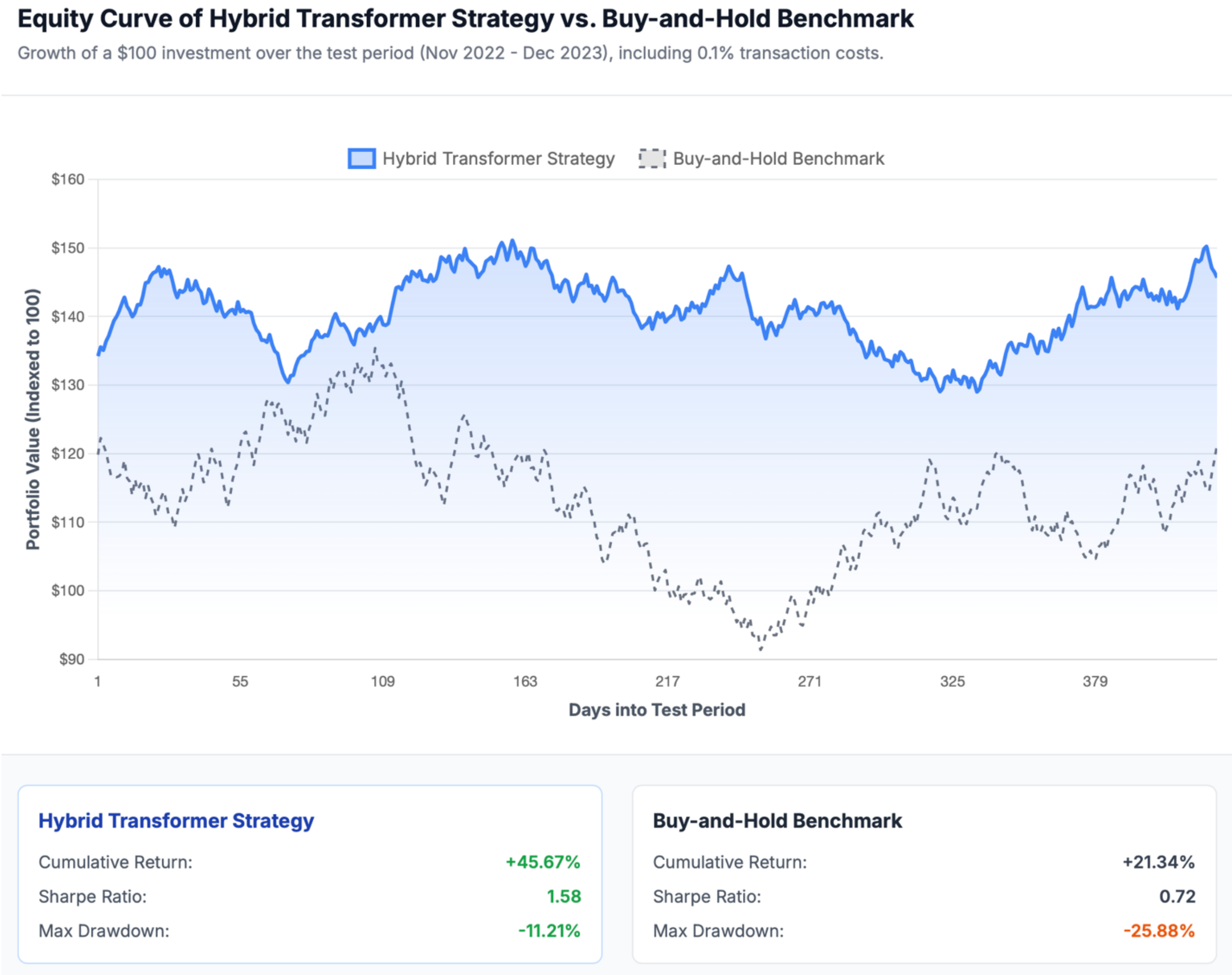

Figure 2. Equity Curve of Hybrid Transformer Strategy vs. Buy-and-Hold Benchmark

To understand the drivers of the best model's success, we employed SHAP (SHapley Additive exPlanations) on the Hybrid Transformer model. Figure 3 shows the mean absolute SHAP values, which represent each feature's overall contribution to the model's predictions. While the most recent lagged returns are predictably important, the analysis reveals that the Sentiment Score and NVT Signal are the next two most influential predictors. This is a crucial finding, providing direct empirical evidence that the alternative data sources are primary drivers of the model's predictive and economic success, confirming the value of the data fusion approach.

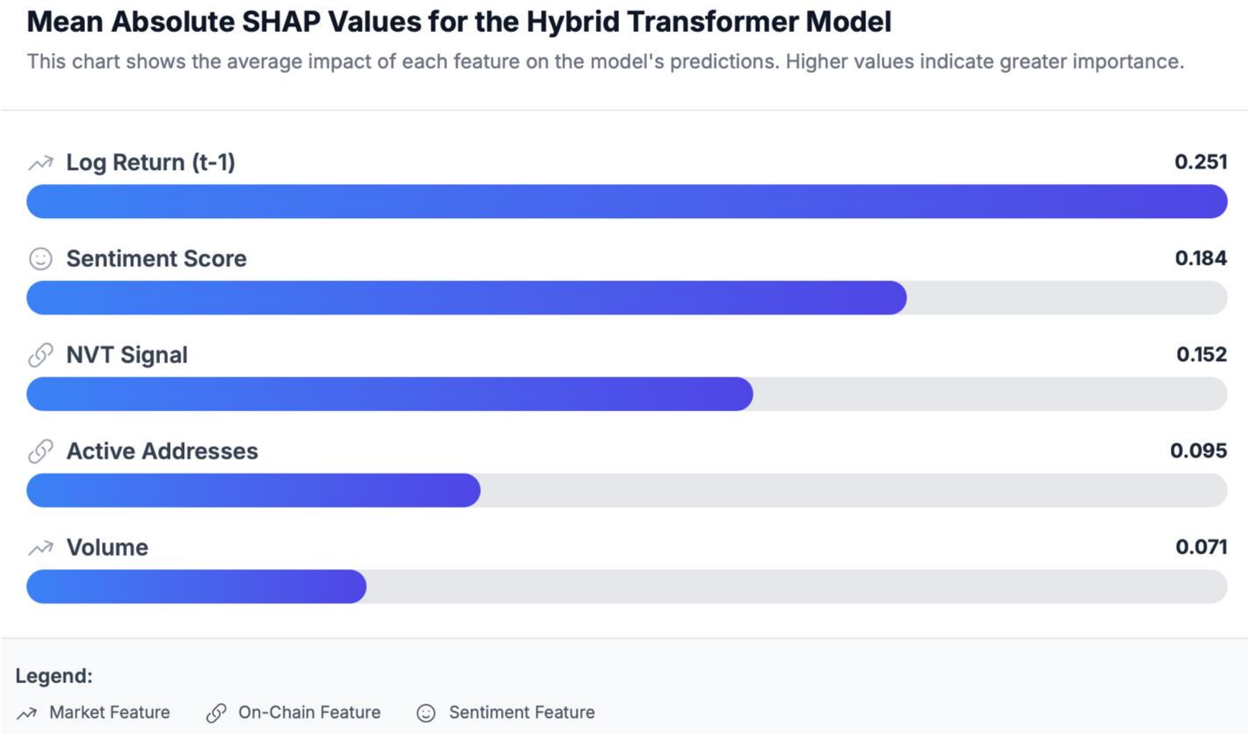

Figure 3. Mean Absolute SHAP Values (Feature Importance) for the Hybrid Transformer Model

The empirical evidence from this study strongly supports the thesis that a methodologically rigorous, hybrid deep learning framework achieves both statistical and, crucially, economic superiority in forecasting cryptocurrency returns. The outperformance of the Hybrid Transformer model is not an artifact of a single powerful feature or a superior algorithm alone, but rather the result of the synergistic fusion of its diverse data components.

Our findings suggest a complementary relationship between the data streams. On-chain data, with features like the NVT Signal, serves as a proxy for the network's fundamental economic health. It acts as a low-frequency anchor, grounding predictions in tangible network utility and helping to identify periods of fundamental over- or under-valuation. Social media sentiment, on the other hand, acts as a high-frequency proxy for the collective psychology and behavioral biases of market participants. It captures the narratives and momentum shifts that drive short-term price action. The Transformer model, with its self-attention mechanism, excels at this fusion. It learns to dynamically weigh the importance of each data stream and their historical patterns, effectively discerning when a market move is backed by fundamental shifts versus when it is driven by ephemeral sentiment. The feature importance analysis corroborates this, confirming that both data streams are vital to the model's success.[13-15]

This research contributes directly to the ongoing debate about market efficiency in the context of digital assets. Our findings, particularly the generation of a statistically significant positive alpha after transaction costs, pose a direct challenge to the semi-strong form of the Efficient Market Hypothesis (EMH), which states that all publicly available information should be immediately priced in. Our model uses only publicly available (though complex and high-dimensional) data, yet it uncovers an exploitable predictive edge.[16,17]

However, rather than a wholesale rejection of market efficiency, our results are better interpreted through the lens of the Adaptive Market Hypothesis (AMH). The AMH posits that markets are not always perfectly efficient but are instead in a constant state of evolution, where inefficiencies arise and are subsequently competed away by adaptive market participants. The "alpha" we have identified is likely one such temporary inefficiency, stemming from the market's current inability to collectively process the complex, multi-modal information contained in on-chain and sentiment data in real-time. The very publication of strategies like this contributes to the market's adaptation and the eventual decay of this specific alpha source. This reflexive nature of financial markets underscores that the search for alpha is not a static problem but a dynamic, adversarial process. 18

The implications for financial practitioners are substantial. The framework presented here serves as a robust blueprint for the development of sophisticated quantitative trading strategies. Hedge funds and asset managers can leverage this multi-modal approach to move beyond purely technical or discretionary strategies.[19]

Furthermore, the model has significant applications in risk management. For instance, traditional Value at Risk (VaR) models often rely on historical volatility, which may not capture changing market regimes. A VaR model augmented with the forecasts from our Hybrid Transformer could provide more accurate, forward-looking risk assessments. If the model predicts a high probability of a large negative return, risk managers could proactively reduce portfolio exposure, leading to more dynamic and responsive risk management protocols. This represents a tangible step forward in applying AI for enhancing financial stability and decision-making in the volatile digital asset space.[20]

The study has some limitations that open avenues for future research. While methodologically sound, the backtest was conducted on a specific historical period. Its performance in different market regimes, for example prolonged bear markets, warrants further investigation. Secondly, we used a fixed chronological split for training and testing. Future work could employ a more dynamic walk-forward validation approach, where the model is periodically retrained, to better reflect a real-world deployment scenario. Finally, while the framework proved successful for Bitcoin, its applicability to other digital assets with different on-chain characteristics and community dynamics is a key area for future exploration.

4. Conclusion

This study set out to determine whether the synergistic integration of on-chain, sentiment, and market data within an advanced deep learning framework could unlock a statistically significant and economically viable predictive edge in the cryptocurrency market. The results of our comprehensive and methodologically rigorous analysis offer a resounding affirmation.

We have demonstrated that a Transformer-based architecture, trained on stationary, multi-modal data, achieves a superior level of forecasting accuracy. More importantly, we have translated this statistical superiority into demonstrable economic value, showing that a trading strategy guided by the model can generate significant, positive, risk-adjusted alpha after accounting for real-world frictions like transaction costs. The findings highlight the importance of moving beyond traditional data sources and embracing the rich, digital footprints of modern assets.

In conclusion, this research provides a robust blueprint for a new generation of quantitative financial tools. It shows that in the digital age, an asset's data exhaust—the trail it leaves across its underlying network and the public discourse—is a rich and quantifiable source of value. By harnessing these digital footprints with powerful analytical techniques, it is possible to gain a more holistic understanding of market dynamics and uncover a significant, albeit likely adaptive, informational edge.

5. References

1. Cao G, Ling M, Wei J, Chen C. Dynamic market behavior and price prediction in cryptocurrency: An analysis based on asymmetric herding effects and LSTM. Comput Econ. 2025;65(6):3325–60.
2. Tariq K, Muhammad MHG, Shah SA, Ahmad G, Saleem MA, Tabassum N. A time-series cryptocurrency price prediction using an ensemble learning model. International Journal for Electronic Crime Investigation. 2025;9(1
3. Islam MZ, Rahman MS, Sumsuzoha M, Sarker B, Islam MR, Alam M, et al. Cryptocurrency price forecasting using machine learning: Building intelligent financial prediction

models. Int J Account Econ Stud. 2025;12(3):255–68.
4. Saputra MPA, Azahra AS, Pirdaus DI. Comparison of activation functions in recurrent neural network for Litecoin cryptocurrency price prediction. Int J Mathematics Statistics Computing. 2025;3(3):113–9.
5. Candra DGA, Afrianto N, Fitriyanto I, Sofiati E, Permana B. Deep learning-based bidirectional RNN for cryptocurrency price prediction with hyperparameter tuning. J Sains Nalar Aplikasi Teknologi Informasi. 2025;4(2):82–91.
6. Shrivastava T, Suleiman B, Desa SAJ, Alibasa MJ, Tregubov A, Zhang J. Benchmarking modeling architectures for cryptocurrency price prediction using financial and social media data. Soc Netw Anal Min. 2025;15(1).
7. Lapitskaya D, Eratalay MH, Sharma R. Prediction of cryptocurrency prices with the momentum indicators and machine learning. Comput Econ]. 2025;66(3):2483–501.
8. Paskaleva M. Artificial intelligence-based model – a key technique in digital economy development for cryptocurrency price prediction. Int J Information Tech Security. 2025;17(3):83–94.
9. Kumar S. Forecasting cryptocurrency prices using ARIMA and Neural Network: A comparative study. J Predict Mark. 2020;13(2).
10. Song H, Wu J, Li Y, Mu S, Qian X, Yu S, et al. Bitcoin price prediction using enhanced transformer and greed index. IET Blockchain. 2025;5(1).
11. Lua ZZ, Seow CK, Chan RCB, Cai Y, Cao Q. Automated Bitcoin trading dApp using price prediction from a deep learning model. Risks. 2025;13(1):17.
12. Han P, Chen H, Rasool A, Jiang Q, Yang M. MFB: A generalized multimodal fusion approach for bitcoin price prediction using time-lagged sentiment and indicator features. Expert Syst Appl. 2025;261(125515):125515.
13. Dang Y. Bitcoin price prediction using a deep learning-based hybrid model with sentiment analysis and attention mechanism. Applied Computational Eng. 2025;154(1):145–51.
14. Seçme G. Machine learning and qualitative variables in Bitcoin price prediction: An empirical evaluation of model performance. Kapadokya Akademik Bakış. 2025;9(1):18–30.
15. Fauzi RA, Rohana T, Nurlaelasari E, Wahiddin D. Bitcoin price prediction using Long Short Term Memory algorithm. J Computer Networks Architecture High Perform Computing. 2025;7(3):659–71
16. Vardhan GV, Subburaj B. Multimodal deep learning model for bitcoin price prediction with news and market prices. Neural Comput Appl. 2025;37(24):19921–56.
17. Jung HS, Lee H, Kim JH. Detecting bitcoin sentiment: Leveraging language model applications in sentiment analysis for bitcoin price prediction. Neural Process Lett. 2025;57(5).
18. Li C, Wang Q, Huang X, Sun S, Lin Z. Optimizing AdaBoost for bitcoin price prediction based on long and short term memory network models. Applied and Computational Engineering. 2025;170(1):18–26.
19. Ibraheem MR, Hassan E, Abed SA, Kamil IS, Talaat FM. Deep learning predictions for Bitcoin market price and illegitimate activity classification. Comput Econ. 2025.
20. Tiwari D, Bhati BS, Nagpal B, Al-Rasheed A, Getahun M, Soufiene BO. A swarm-optimization based fusion model of sentiment analysis for cryptocurrency price prediction. Sci Rep. 2025;15(1):8119.