JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

BITCOIN PRICE PREDICTION WITH TECHNICAL INDICATORS: A HYBRID TRANSFORMER-RIDGE REGRESSION APPROACH

Dio Richard Prastiyo[1]; Muhammad Zaky Darajat[1]; Christian Sri Kusuma Aditya[1*]

Department of Informatics[1]
Universitas Muhammadiyah Malang, Malang, Indonesia[1]
http://umm.ac.id[1]
diorichard@webmail.umm.ac.id, zakydarajat@webmail.umm.ac.id, christianskaditya@umm.ac.id[*]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Bitcoin has emerged as the dominant cryptocurrency, exhibiting rapid adoption alongside extreme price volatility that complicates investment strategies, risk management, and regulatory decision-making. While prior hybrid studies have predominantly combined multiple deep learning components such as CNN–LSTM or Transformer–GRU architectures, the integration of a deep neural architecture with a regularized linear model remains underexplored in Bitcoin price forecasting. To address this gap, this study proposes a hybrid framework combining a Transformer neural network with Ridge Regression, wherein the Transformer captures nonlinear temporal dependencies while Ridge Regression introduces L2 regularization to mitigate overfitting and enhance interpretability—an integration explicitly motivated by the bias–variance trade-off. The model is trained on technical indicators including MACD, Bollinger Bands, and RSI, and an ensemble weighting parameter α is systematically optimized via grid search. Empirical evaluation demonstrates that the hybrid model consistently outperforms standalone baselines, achieving an MAE of 1,251.572, RMSE of 1,623.004, R^2 of 0.991, and MAPE of 1.701%, with performance differences confirmed statistically via the Diebold–Mariano test. Economic validation reveals that the hybrid model is the only strategy to demonstrate statistically significant directional accuracy, although absolute trading returns remain below passive benchmarks under trending market conditions—a dissociation consistent with established findings in financial forecasting research. These results indicate that the model's primary contribution lies in forecast reliability and directional signal quality rather than return maximization under simple trading rules. Sensitivity to macroeconomic shocks and computational demands remain limitations for real-time deployment, suggesting directions for future research.*

Keywords: *Bitcoin, Ridge Regression, Technical Indicators, Time Series Forecasting, Transformer.*

Intisari— *Bitcoin telah berkembang menjadi aset kripto dominan dengan tingkat adopsi yang pesat disertai volatilitas harga yang sangat tinggi, sehingga menyulitkan strategi investasi, manajemen risiko, dan pengambilan keputusan regulasi. Meskipun berbagai penelitian hybrid sebelumnya umumnya mengombinasikan beberapa komponen deep learning seperti CNN–LSTM atau Transformer–GRU, integrasi antara arsitektur neural mendalam dengan model linear terregularisasi masih relatif jarang dieksplorasi dalam peramalan harga Bitcoin. Untuk mengatasi kesenjangan tersebut, penelitian ini mengusulkan kerangka hybrid yang mengombinasikan jaringan saraf Transformer dengan Ridge Regression, di mana Transformer menangkap dependensi temporal nonlinier sementara Ridge Regression menerapkan regularisasi L2 untuk mengurangi overfitting dan meningkatkan interpretabilitas—integrasi yang didasarkan pada konsep trade-off bias–variance. Model dilatih menggunakan indikator teknikal MACD, Bollinger Bands, dan RSI, dengan parameter bobot ensemble α yang dioptimalkan melalui grid search. Hasil empiris menunjukkan bahwa model hybrid secara konsisten mengungguli model baseline dengan capaian MAE sebesar 1.251,572, RMSE sebesar 1.623,004, R^2 sebesar 0,991, dan MAPE sebesar 1,701%, dengan perbedaan performa yang terverifikasi melalui uji Diebold–Mariano. Validasi ekonomi menunjukkan bahwa model hybrid menjadi satu-satunya model yang memiliki directional accuracy signifikan secara statistik, meskipun return trading absolut masih* *berada di bawah strategi pasif pada kondisi pasar yang sedang tren. Temuan ini menunjukkan bahwa kontribusi utama model terletak pada reliabilitas peramalan dan kualitas sinyal arah, bukan pada maksimalisasi return. Sensitivitas terhadap guncangan makroekonomi dan kebutuhan komputasi yang tinggi masih menjadi keterbatasan untuk implementasi real-time dan dapat menjadi arah penelitian selanjutnya.*

***Kata Kunci**: Bitcoin, Regresi Ridge, Indikator Teknis, Peramalan Deret Waktu, Transformer.*

INTRODUCTION

Bitcoin, as the first and most prominent digital currency, has experienced rapid growth accompanied by substantial price volatility over the past decade. Its decentralized structure and market dynamics contribute to frequent and unpredictable price fluctuations, posing significant challenges for accurate forecasting [1]. Reliable Bitcoin price prediction is therefore essential not only for investors and traders seeking to improve decision-making but also for policymakers concerned with maintaining financial market stability. Bitcoin's volatility is driven by multiple interacting factors, including market sentiment, investor behavior [2], and macroeconomic variables such as inflation, interest rates, and global uncertainty [3], [4].

Technological developments like blockchain innovation and algorithmic trading further amplify these dynamics [5]. To better anticipate price movements, technical indicators are commonly used to capture both short- and long-term volatility patterns [6]. However, despite their widespread application, the non-stationary and nonlinear nature of cryptocurrency markets continues to challenge traditional forecasting models. Classical statistical models such as Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) have been widely utilized in financial forecasting. Nevertheless, these methods struggle to represent nonlinear dependencies and regime shifts that characterize digital asset prices [7], [8].

The emergence of machine learning (ML) techniques has allowed for more flexible modeling of temporal dependencies [9], while deep learning (DL) architectures particularly Long Short-Term Memory (LSTM) networks have further improved predictive accuracy for complex sequential data [10]. Among deep learning architectures, Recurrent Neural Networks (RNNs) including LSTM and Gated Recurrent Units (GRU) have demonstrated substantial promise in financial time-series forecasting [11]. More recently, Transformer-based architectures leveraging self-attention mechanisms have gained traction due to their ability to model long-range dependencies efficiently, outperforming recurrent structures in several sequential domains [12], [13]. Despite these strengths, Transformer models remain susceptible to overfitting, high computational demand, and limited interpretability, which are particularly problematic in volatile cryptocurrency markets [14].

Hybrid frameworks that combine deep learning with conventional regression techniques offer a principled remedy to these limitations [8]. From a bias–variance perspective, nonlinear deep models exhibit low bias but high variance, whereas regularized linear models trade higher bias for substantially lower variance. Integrating these complementary characteristics within a unified framework therefore yields improved generalization and stability—an advantage empirically supported by prior studies showing that hybrid architectures consistently outperform single-model approaches in cryptocurrency markets [15], [16]. Ridge Regression is especially well suited as the linear component, given its strong L2 regularization and inherent interpretability, which effectively mitigate overfitting and multicollinearity common in purely deep learning pipelines [17]. Building on this rationale, this study proposes a hybrid forecasting framework that integrates a Transformer neural network with Ridge Regression for Bitcoin price prediction. The Transformer captures long-range temporal dependencies in price dynamics, while Ridge Regression contributes regularization and transparency. Both components are trained on a curated set of technical indicators—MACD, Bollinger Bands, and RSI—selected to represent momentum and volatility patterns in cryptocurrency markets.

The novelty of this research lies in three contributions that distinguish it from prior hybrids such as CNN–LSTM, Transformer–GRU, or stacked neural ensembles. First, the integration is explicitly motivated by the bias–variance trade-off rather than heuristic stacking of deep components. Second, the ensemble weighting parameter $\alpha \in [0, 1]$, which governs the relative contribution of each model, is systematically optimized via grid search with increments of 0.01, ensuring a controlled and reproducible fusion. Third, beyond conventional error metrics (MAE, RMSE, R^2, and MAPE), the evaluation framework incorporates statistical validation through the Diebold–Mariano (DM) test

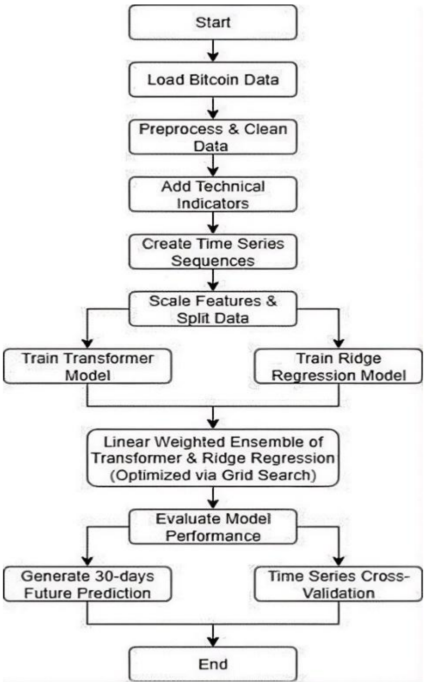
Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

and financial-performance validation—including directional accuracy, Sharpe ratio, maximum drawdown, and Calmar ratio under a simulated trading strategy with transaction costs—benchmarked against both a naive random-walk forecast and a buy-and-hold strategy [12], [13]. Together, these evaluations assess both the statistical reliability and the practical economic relevance of the proposed model in highly volatile cryptocurrency markets.

MATERIALS AND METHODS

Before detailing each methodological component, Figure 1 illustrates the overall workflow of the proposed hybrid forecasting framework, including data acquisition, preprocessing, feature engineering, sequence construction, model training, ensemble fusion, and forecasting. Historical Bitcoin price data were collected and transformed into supervised sequences, which were subsequently used to train two parallel models: a Transformer network designed to capture nonlinear temporal dependencies and a Ridge Regression model that provides linear regularization to enhance model stability and generalization. The predictions generated by both models were then integrated using a weighted ensemble strategy to produce the final price forecasts.



Source: (Research Results, 2025)
Figure 1. Methodology Framework for Bitcoin Price Prediction using Hybrid Transformer-Ridge Regression Approach

A. Data Collection and Preparation

Daily Bitcoin market data were retrieved from the Binance Exchange API covering the period from January 2018 to December 2025, yielding a total of 2,677 daily candlestick records. Each record corresponds to one calendar day and includes open, high, low, close, and volume (OHLCV) fields, along with supplementary trading activity variables such as quote asset volume, number of trades, and taker buy volumes, as detailed in Table 1. Data were collected programmatically using the daily interval setting, and the resulting records were parsed directly into a structured tabular format, eliminating manual data entry errors. Missing values were handled using linear interpolation to maintain sequence continuity. All features were subsequently normalized using RobustScaler, applied independently to each column, to mitigate the influence of extreme outliers commonly present in cryptocurrency price series. Features with inherently bounded ranges, specifically daily returns and log returns, were retained in their original scale as they are already approximately normalized.

Table 1. Summary of Variables used in This Study, Illustrating Comprehensive Coverage of Price, Volume, and Trading Activity Information Required for Robust Bitcoin Price Forecasting.

No	Variable	Data Type	Description
1	Open time	Nominal	Date and time of candle opening
2	Open	Continuous	Opening price (USD)
3	High	Continuous	Highest price (USD)
4	Low	Continuous	Lowest price (USD)
5	Close	Continuous	Closing price (USD)
6	Volume	Continuous	Daily trading volume
7	Close time	Nominal	Date and time of candle closing
8	Quote asset volume	Continuous	Asset volume in quote currency (USD)
9	Number of trades	Ordinal	Number of trades (integer count)
10	Taker buy base asset volume	Continuous	Buy volume (base asset)
11	Taker buy quote asset volume	Continuous	Buy volume (quote asset)

Source: (Research Results, 2025)

B. Feature Engineering and Data Preparation

From the cleaned dataset, a set of technical indicators was engineered to capture price behavior and market dynamics in the cryptocurrency market. These indicators were grouped according to their functional roles in time-series forecasting, including

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

trend indicators (moving averages and exponential moving averages), momentum indicators such as the Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) [18], volatility indicators including Bollinger Bands and Average True Range (ATR), and volume-based indicators such as On-Balance Volume (OBV) [19]. In addition, return-based features, price ratios, and lagged variables of closing prices and trading volumes were incorporated to capture temporal dependencies and multi-scale market dynamics. The parameters of these indicators followed commonly used configurations in financial analysis, such as RSI (14), MACD (12, 26, 9), Bollinger Bands (20-period moving average with 2 standard deviations), and ATR (14). A complete summary of the engineered features is presented in Table 2, organized by functional category to improve methodological transparency and reproducibility. Given the inherent multicollinearity commonly observed among technically derived indicators, a subsequent feature selection stage based on multicollinearity diagnostics was applied to determine the final set of input variables used for model training, as described in the following subsection.

Table 2. Engineered Feature Set Grouped by Functional Roles to Capture Multi-Scale Market Dynamics Prior to Multicollinearity Reduction

No	Feature Group	Variables	Description
1	Return-based Features	Returns, Log_Returns	Capture price changes in absolute and logarithmic scales
2	Moving Averages	MA5, MA10, MA20, MA50, MA100, MA200	Identify short- to long-term market trends
3	MA Ratios	MA5_ratio, MA10_ratio, MA20_ratio, MA50_ratio, MA100_ratio, MA200_ratio	Compare moving averages with current prices to confirm trend direction
4	Exponential Moving Averages	EMA12, EMA26	Capture trend dynamics using exponentially weighted averages
5	Volatility Indicators	Volatility (20), ATR (14), BB_Upper, BB_Lower, BB_Width, BB_Position	Measure price variability and detect potential breakout conditions
6	Momentum Indicators	RSI (14), RSI_MA (14), MACD, MACD_Signal, MACD_Hist	Evaluate momentum strength and potential trend reversals
7	Volume-based Indicators	OBV, Volume_MA (20), Volume_Ratio	Incorporate trading volume to identify accumulation and distribution patterns
8	Price Ratios	HL_Ratio, CO_Ratio	Capture relative intraday price relationships
9	Lag Features (Price)	Close_Lag_1, Close_Lag_2, Close_Lag_3, Close_Lag_5, Close_Lag_7	Encode historical price dependency
10	Lag Features (Volume)	Volume_Lag_1, Volume_Lag_2, Volume_Lag_3, Volume_Lag_5, Volume_Lag_7	Encode historical volume dependency

Source: (Research Results, 2025)

C. Final Feature Selection for Model Training

Following the feature engineering stage, a feature selection procedure was conducted to determine the final set of input variables for model training. Since many technical indicators are derived from the same underlying price series, multicollinearity is inherently present in financial time-series data, particularly in cryptocurrency markets. To evaluate the degree of multicollinearity among candidate variables, the Variance Inflation Factor (VIF) was calculated. VIF is widely used as a diagnostic measure to assess the inflation of coefficient variance caused by linear dependence among predictors, with VIF ≤ 10 commonly used as a reference threshold for potentially problematic multicollinearity [20]. However, in technical-indicator-based financial modeling, strict elimination based solely on VIF may remove variables that carry meaningful market signals. Indicators such as Moving Averages, MACD, and Bollinger Bands are mathematically derived from the same price series and therefore naturally exhibit strong linear dependencies — a structural characteristic of engineered financial features rather than a modeling deficiency.

Instead of applying VIF as a strict exclusion criterion, this study adopts a hybrid feature selection strategy that combines statistical diagnostics with financial domain relevance. Key financial indicators — including EMA26, MACD-based signals, Bollinger Band measures, RSI, ATR, and volatility indicators — were retained despite elevated VIF values because their market representations are theoretically grounded in

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

established technical analysis principles. The potential impact of multicollinearity is explicitly mitigated by the use of Ridge Regression within the proposed hybrid forecasting framework. Unlike ordinary least squares estimation, Ridge Regression applies L2 regularization that constrains coefficient magnitudes and stabilizes estimation under high inter-predictor correlation, making it particularly well suited as the linear component in settings where technically derived features exhibit strong linear dependencies. Based on this approach, sixteen features were selected as the final input variables. The VIF values and selection rationale for the retained features are presented in Table 3.

Table 3. VIF Values of the Final Selected Features, Illustrating That Features With Extreme Multicollinearity (VIF>100) Were Retained Based On Domain Relevance and Mitigated Through Ridge Regression's L2 Regularization

No	Feature	VIF	Market Representation
1	EMA26	>100	Long-term trend indicator
2	MACD_Signal	>100	Momentum confirmation signal
3	MACD	>100	Momentum indicator
4	BB_Width	>100	Volatility regime indicator
5	MA20_ratio	33.16	Relative trend strength
6	ATR	12.28	Market volatility measurement
7	RSI	7.65	Momentum oscillator
8	BB_Position	6.04	Price position within volatility band
9	Number of trades	5.60	Market activity level
10	Volume_Lag_1	4.81	Short-term trading volume signal
11	Volume_Lag_2	4.58	Short-term volume persistence
12	OBV	4.12	Volume-based momentum indicator
13	Volume_Lag_7	4.03	Weekly volume pattern
14	Volume_Lag_3	3.99	Short-term volume dynamics
15	Volume_Lag_5	3.72	Medium-term volume behavior
16	Volatility	3.46	Market price fluctuation

Source: (Research Results, 2025)

D. Sequence Formation and Data Splitting

Time-series sequences were constructed using a sliding-window approach. Since window size selection plays a critical role in capturing temporal dependencies in cryptocurrency markets [21], [22], multiple candidate sizes were evaluated and a 10-day window was adopted as the final configuration based on sensitivity analysis results. For each position in the scaled dataset, the input sequence consists of 10 consecutive daily observations with the closing price of the following day as the target, yielding input tensors of shape (number of sequences × 10 days × 16 features).

The dataset was partitioned chronologically into training (80%) and test (20%) sets to prevent data leakage and preserve temporal integrity. Within the Transformer training phase, an additional 20% of the training set was held out as an internal validation split to monitor convergence and trigger early stopping. The ensemble weight α was optimized on a separate temporally held-out validation segment drawn from within the training period, ensuring the test set remained entirely unseen during all stages of model development. Random seeds were fixed throughout to ensure full reproducibility.

E. Transformer Model

The Transformer encoder [7] was implemented to model long-term dependencies in sequential price data. Unlike recurrent architectures, the Transformer processes input sequences in parallel and relies on a self-attention mechanism to capture relationships across all time steps. The core operation of the Transformer is the scaled dot-product attention, which computes attention weights based on query (Q), key (K), and value (V) representations. The attention mechanism is defined as:

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Where d_k denotes the dimensionality of the key vectors used for scaling. This mechanism allows the model to dynamically assign importance to different time steps when generating contextual representations. Since Transformers lack inherent temporal awareness, sinusoidal positional encoding was incorporated to inject information about the order of observations. The positional encoding functions are defined as:

$$PE_{(pos,2i)}=sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (2)$$

$$PE_{(pos,2i+1)}=cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \quad (3)$$

Where pos represents the position in the sequence and d_{model} denotes the embedding dimension. Each encoder block consisted of multihead self-attention, a two-layer feed-forward network, residual connections, and layer normalization. Dropout regularization was applied within attention and feed-forward layers to reduce overfitting.

In this study, the Transformer architecture comprised two encoder blocks with four attention heads and a head dimension of 112, resulting in an embedding dimension of 448. The feed-forward network employed a hidden dimension of 448 with GELU activation. After the encoder layers, global average pooling and global max pooling were applied and concatenated to aggregate temporal representations, followed by a multilayer perceptron with hidden sizes of 256 and 128 units. The model was trained using the AdamW optimizer with a custom learning-rate schedule and Huber loss. Training was performed for up to 200 epochs with a batch size of 16, and a 20% validation split was used during training. Early stopping with a patience of 12 epochs was applied based on validation loss, and random seeds were fixed to ensure experimental reproducibility.

F. Ridge Regression Model

Ridge Regression model was employed as a linear baseline and an interpretability anchor within the proposed hybrid forecasting framework. Ridge Regression estimates model parameters by minimizing the sum of squared residuals augmented with an L2 regularization penalty applied to the regression coefficients, as formulated in Equation (4). This regularization mechanism effectively mitigates multicollinearity among correlated predictors by constraining the magnitude of model coefficients.

$$L=\|y\text{-}X\beta\|_2^2+\alpha\|\beta\|_2^2 \qquad (4)$$

The L2 regularization term reduces coefficient variance and improves model stability when predictors exhibit strong correlations. This property is particularly relevant in financial forecasting tasks where technical indicators often contain overlapping or highly correlated information. Ridge Regression has been widely applied in financial prediction problems due to its robustness and regularization capability. Prior empirical studies on Bitcoin price prediction have also demonstrated that Ridge Regression can achieve strong predictive performance compared with other machine learning models, including Random Forest, Support Vector Regression, and Lasso regression [23]. These findings highlight its effectiveness in handling financial datasets with correlated features.

Within the proposed hybrid architecture, the Ridge model serves as a stabilizing linear component that complements the nonlinear sequence-learning capability of the Transformer. By combining the structured linear regularization of Ridge Regression with the deep temporal representation learning of the Transformer, the hybrid framework aims to produce more robust and reliable forecasts under highly volatile cryptocurrency market conditions.

G. Hyperparameter Optimization

Transformer**.** Hyperparameters were determined through random search over 15 iterations, covering embedding dimension (64–256), number of attention heads (2–8), feed-forward dimension (128–512), dropout rate (0.1–0.3), and learning rate (1×10^{-4} to 1×10^{-3}). The final configuration comprises two encoder blocks, four attention heads with head dimension 112 (embedding dimension 448), feed-forward dimension 448 with GELU activation, and an MLP head of 256 and 128 units with batch normalization and dropout of 0.2. Training used the AdamW optimizer with a warmup learning rate schedule and Huber loss, for up to 200 epochs with batch size 16 and early stopping at patience 12.

Ridge Regression**.** The regularization parameter was selected via 5-fold TimeSeriesSplit cross-validation with R^2 as the criterion, evaluated over six candidate values ranging from 0.001 to 100.0. The optimal value of α = 0.01 yielded the highest mean cross-validation R^2 across all temporal folds, and the final model was retrained on the full training set using this value.

H. Ensemble Strategy and Weight Optimization

To merge the complementary strengths of the two models, a weighted ensemble was formulated:

$$\acute{y}_{hybrid}=\alpha\acute{y}_{Transformer}+(1\text{-}\alpha)\acute{y}_{Ridge} \qquad (5)$$

The where $\alpha \in [0, 1]$ controls the relative contribution of each component. The optimal α was determined through grid search over 101 discrete values from 0 to 1 in increments of 0.01, evaluated on a temporally held-out validation segment using RMSE as the selection criterion. This optimization was conducted exclusively on the validation segment, which lies chronologically after the training period and before the test period, ensuring no look-ahead bias [13]. The selected optimal weight was α = 0.40, corresponding to a 40% Transformer and 60% Ridge composition, which

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

was subsequently applied to generate all test set predictions reported in the Results section.

I. Evaluation Metrics and Validation Strategy

Performance was comprehensively assessed using multiple metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2):

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}|y_i\text{-}\acute{y}_i| \tag{6}$$

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i\text{-}\acute{y}_i)^2} \tag{7}$$

$$\text{MAPE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i\text{-}\acute{y}_i}{y_i}\right| \times 100\% \tag{8}$$

$$R^2=1\text{-}\frac{\sum_{i=1}^{n}(y_i\text{-}\acute{y}_i)^2}{\sum_{i=1}^{n}(y_i\text{-}\bar{y})^2} \tag{9}$$

These metrics capture absolute errors, squared deviations, relative percentage errors, and explained variance, with R^2offering holistic insight into regression accuracy [24]. TimeSeriesSplit cross-validation [25] was used to evaluate model stability across temporal segments.

RESULTS AND DISCUSSION

Following the methodology outlined previously, this section presents an empirical evaluation of the proposed hybrid Transformer–Ridge Regression model. The hybrid approach is systematically compared with its individual components using standard forecasting metrics (MAE, RMSE, MAPE, and R^2), with statistical significance assessed through the Diebold–Mariano test. Model robustness is further examined via ensemble weight optimization and window-size sensitivity analysis, while interpretability is addressed using feature importance analysis. Together, these results demonstrate the hybrid model's predictive accuracy, statistical reliability, and practical relevance for cryptocurrency forecasting applications.

A. Hybrid Transformer–Ridge Regression Model Outperforms Individual Models

This study evaluates whether integrating a Transformer architecture with Ridge Regression improves Bitcoin price forecasting performance compared to using each model independently. Model accuracy is assessed using complementary metrics—MAE, RMSE, R^2, and MAPE—to capture absolute error, variance sensitivity, explanatory power, and relative forecasting accuracy under volatile market conditions.

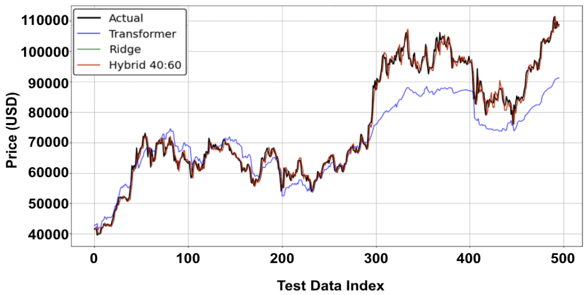
Table 4. Model comparison

Model	MAE	RMSE	R^2	MAPE (%)
Transformer	6481.971	8484.303	0.768	7.711
Ridge	1580.643	2163.425	0.984	2.125
Hybrid	1251.572	1623.004	0.991	1.701

Source: (Research Results, 2025)

Table 4 summarizes the comparative forecasting performance of the Transformer, Ridge Regression, and the proposed hybrid model. The hybrid Transformer–Ridge framework achieves the lowest error values (MAE = 1,251.6; RMSE = 1,623.0) and the highest explanatory power (R^2 = 0.991), along with the smallest proportional error (MAPE = 1.70%). These results indicate that the hybrid approach provides more accurate and stable predictions than either standalone model.

The performance differences can be attributed to the complementary modeling characteristics of the two components. The Transformer architecture is capable of capturing complex nonlinear temporal dependencies but exhibits higher sensitivity to noise and scale variability when applied independently in highly volatile cryptocurrency markets. In contrast, Ridge Regression benefits from linear regularization that promotes robustness and stability, although its linear structure limits its ability to fully capture nonlinear price dynamics. By combining nonlinear sequence learning with linear regularization, the hybrid framework achieves a more balanced bias–variance trade-off, leading to improved generalization performance, consistent with prior findings in hybrid financial forecasting studies [9].



Source: (Research Results, 2025)
Figure 2. Comparison of Actual and Predicted Bitcoin Prices

Figure 2. Comparison of actual Bitcoin prices and model predictions, illustrating that the

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

hybrid Transformer–Ridge model most closely tracks price movements during both stable and highly volatile periods. The qualitative comparison shown in Figure 2 further supports the quantitative results. The hybrid model closely follows the actual Bitcoin price trajectory across both stable and volatile periods, maintaining alignment during abrupt price movements. Compared to the standalone Transformer, which shows larger deviations under heightened volatility, and Ridge Regression, which tends to lag during rapid trend reversals, the hybrid model demonstrates a balance between responsiveness and stability.

From a practical forecasting perspective, this balance is particularly relevant for financial applications, where excessive sensitivity to short-term fluctuations or overly rigid trend-following behavior may reduce predictive reliability. The results indicate that the proposed hybrid Transformer–Ridge architecture provides a more dependable representation of Bitcoin price dynamics than either model in isolation.

B. Statistical Significance Analysis

While descriptive error metrics provide an initial indication of forecasting accuracy, they do not establish whether observed performance differences between models are statistically meaningful. To address this limitation, the Diebold–Mariano (DM) test was employed to formally evaluate whether the forecast errors produced by competing models differ significantly over time. The DM test compares the predictive accuracy of two forecasting models by examining the average difference between their respective loss values across the evaluation period. In this study, two loss specifications were considered: absolute error corresponding to Mean Absolute Error (MAE) and squared error corresponding to Mean Squared Error (MSE).

The test was conducted using a one-step-ahead forecasting horizon ($h = 1$), corresponding to a single daily prediction interval in the dataset. Because the forecasting horizon is limited to one step ahead, the Newey–West truncation lag was set to zero, indicating that no serial-correlation correction was required in the variance estimation of the loss differential.

Table 5. Summary of Diebold–Mariano Test Results

Model Comparison	Loss Function	DM Statistic	p-value	Significance
Hybrid vs Transformer	MAE	-53.33	<0.01	Highly Significant
Hybrid vs Transformer	MSE	-36.88	<0.01	Highly Significant
Hybrid vs Ridge	MAE	-16.54	<0.01	Highly Significant
Hybrid vs Ridge	MSE	-13.52	<0.01	Highly Significant
Transformer vs Ridge	MAE	52.82	<0.01	Highly Significant
Transformer vs Ridge	MSE	36.84	<0.01	Highly Significant

Source: (Research Results, 2025)

Table 5 presents the DM test results for all pairwise comparisons among the evaluated models under both MAE-based and MSE-based loss functions. The results show that the Hybrid Transformer–Ridge model significantly outperforms the standalone Transformer model under both loss specifications, as indicated by the negative DM statistics and highly significant p-values ($p < 0.01$). The negative values of the DM statistics indicate that the Hybrid model consistently produces lower forecast errors compared to the Transformer baseline.

Similarly, the comparison between the Hybrid model and Ridge Regression also reveals statistically significant differences, suggesting that the integration of nonlinear temporal learning and linear regularization contributes to improved predictive performance. These findings indicate that the hybrid architecture provides a more effective balance between model flexibility and stability when forecasting highly volatile cryptocurrency prices.

Furthermore, the statistically significant difference observed between the Transformer and Ridge Regression models suggests that higher model complexity does not necessarily guarantee improved predictive performance in volatile financial markets. Although the Transformer model is capable of capturing complex temporal dependencies, its forecasting errors remain significantly larger than those produced by the Ridge Regression model in this dataset. This result highlights the importance of model robustness and regularization when dealing with noisy financial time-series data.

Overall, the DM test results provide statistical validation for the proposed hybrid forecasting framework. By integrating nonlinear temporal representation learning with linear regularization, the Hybrid Transformer–Ridge model achieves consistently lower forecasting errors with statistically significant improvements over the baseline models. These findings reinforce the robustness and reliability of the empirical results presented in the previous subsection.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

C. Economic and Financial Validation

To complement the statistical evaluation, an economic and financial validation was conducted to examine the practical implications of the forecasting models within a simulated trading environment. While conventional forecasting metrics such as MAE, RMSE, and R^2 evaluate predictive accuracy, they do not necessarily indicate whether the resulting forecasts can generate economically meaningful investment decisions. Therefore, several financial performance indicators were evaluated, including directional accuracy (DA), total return, Sharpe ratio, maximum drawdown, and Calmar ratio. A binomial test was applied to assess whether each model's directional accuracy statistically exceeds random prediction (H_0: DA = 50%). In addition, a random-walk forecast and a buy-and-hold investment strategy were included as benchmark references to contextualize the results under realistic market conditions. The detailed results of the economic validation are presented in Table 6.

Table 6. Economic and Financial Validation of Forecasting Model

Model	Directional Accuracy (%)	DA p-value	Total Return (%)	Sharpe Ratio	Max Drawdown (%)	Calmar Ratio
Transformer	52.76	0.1198	134.54	1.8010	−33.62	2.6459
Ridge Regression	50.10	0.5000	36.36	0.8271	−39.35	0.6619
Hybrid Transformer–Ridge	63.19	<0.01	87.02	1.4348	−34.28	1.7375
Random Walk	47.13	0.9054	51.37	1.0148	−38.78	0.9375
Buy-and-Hold (Benchmark)	—	—	172.08	1.6937	−28.38	3.9144

Source: (Research Results, 2025)

Three principal insights emerge from Table 6. First, the Hybrid model is the only strategy that demonstrates statistically significant directional predictive ability, suggesting that the ensemble integration of nonlinear temporal learning and linear regularization produces directional signals that are meaningfully superior to random prediction. The remaining models, including the Transformer and Ridge Regression, fail to reject the null hypothesis of random directional accuracy, indicating that strong point-forecast performance does not automatically translate into reliable price-direction signals.

Second, a clear dissociation exists between statistical forecasting accuracy and trading profitability. Despite the Hybrid model's superiority in conventional error metrics and directional accuracy, it does not achieve the highest absolute return among model-based strategies. The Transformer, which performs substantially worse in point-forecast metrics, generates stronger returns under the simulated trading framework. This pattern suggests that a model's sensitivity to short-term price momentum—a characteristic that increases noise susceptibility in regression tasks—may prove advantageous in trend-following trading contexts. This dissociation between statistical and economic performance is well documented in financial forecasting research and underscores the importance of evaluating models across both dimensions.

Third, the buy-and-hold benchmark outperforms all active model-based strategies on both return and risk-adjusted metrics, which largely reflects the broadly upward Bitcoin price trend during the evaluation period. Under sustained trending conditions, passive strategies naturally benefit from full market exposure, whereas signal-based strategies periodically exit the market and incur transaction costs, limiting cumulative gains. This outcome does not diminish the value of the forecasting models but rather contextualizes their practical utility: in trending markets, active strategies face an inherently high bar relative to passive benchmarks, and the Hybrid model's statistically significant directional accuracy may offer greater relative advantage during periods of higher market uncertainty or reversals. Overall, these findings reinforce that the primary contribution of the proposed hybrid framework lies in improving the reliability and statistical quality of Bitcoin price forecasts rather than maximizing short-horizon trading returns. Future research may explore more sophisticated trading strategies, transaction-cost-aware optimization, and dynamic risk management mechanisms to better translate forecasting improvements into economically meaningful investment outcomes.

D. Optimal Ensemble Weight Determined by Experimental Validation

A critical step in developing the Hybrid Transformer–Ridge Regression model is the determination of the ensemble weight parameter (α), which governs the relative contribution of the Transformer and Ridge components. This parameter directly controls the trade-off between nonlinear adaptability and statistical stability. To identify an optimal balance, α was optimized through a systematic grid search across 100 discrete values ranging from 0 (pure Ridge Regression) to 1 (pure Transformer), consistent with established hyperparameter tuning practices in financial forecasting models [26].

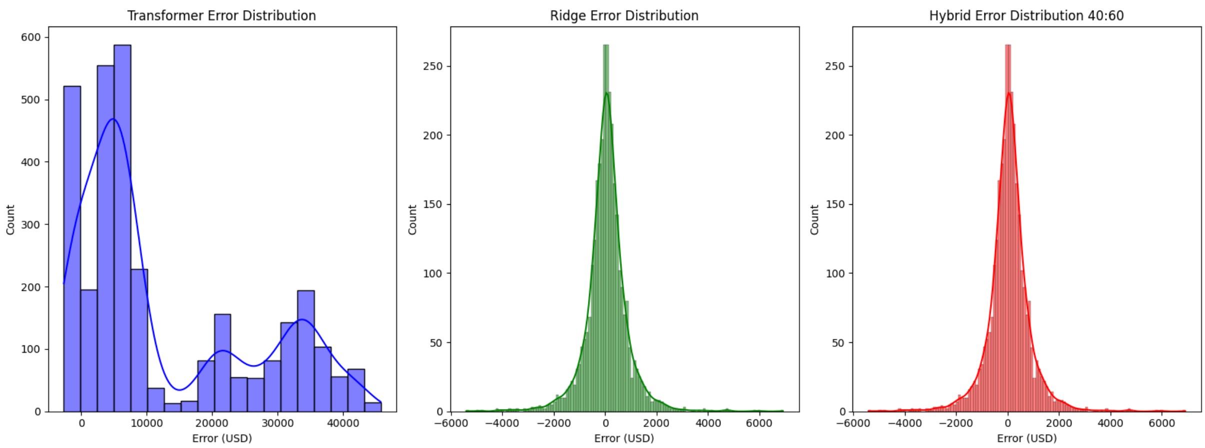
Table 7. Ensemble Weight Experiments Demonstrating That A 40:60 Transformer–Ridge Composition Achieves Optimal Forecasting Accuracy by Balancing Adaptability and Stability.

Transformer:Ridge (%)	MAE	RMSE	R^2	MAPE (%)
0:100	1580.643	2163.422	0.984	2.125
10:90	1345.741	1748.531	0.990	1.815
20:80	1288.062	1683.710	0.990	1.735
30:70	1262.784	1650.621	0.991	1.708
40:60	1251.572	1623.004	0.991	1.701
50:50	1274.024	1649.762	0.991	1.723

Source: (Research Results, 2025)

Table 7 reports the forecasting performance obtained under different Transformer–Ridge weight combinations. The results reveal that Ridge Regression, even as a standalone component (0:100), already provides a strong predictive baseline (R^2 = 0.984), underscoring its role as the primary driver of stability within the hybrid framework. Introducing a moderate Transformer contribution progressively improves forecasting accuracy, as reflected by declining error metrics, but this improvement plateaus and reverses beyond a certain threshold — indicating that the Transformer's contribution is most effective when constrained rather than dominant. The optimal configuration is empirically identified at α = 0.4, corresponding to a 40:60 Transformer–Ridge composition. This result indicates that Ridge Regression serves as the principal source of predictive stability, contributing the larger share of the ensemble weight, while the Transformer plays a complementary but secondary role by capturing nonlinear temporal dynamics that the linear model alone cannot represent. Increasing the Transformer weight beyond this threshold progressively degrades forecasting accuracy, confirming that the deep learning component's value lies in controlled supplementation rather than dominance. This behavior is consistent with the bias–variance rationale underlying the proposed framework: Ridge Regression provides the structural robustness necessary for generalization, while a constrained Transformer contribution refines predictions at the margin without introducing excessive sensitivity to noise.

This behavior is consistent with the characteristics of the standalone models. Ridge Regression serves as the primary stabilizing component within the hybrid framework, providing strong baseline predictive performance through L2 regularization that effectively suppresses noise and controls overfitting. The Transformer, while capable of capturing nonlinear temporal dependencies, demonstrates increasing susceptibility to noise when its contribution is overemphasized, as evidenced by the progressive performance degradation observed beyond α = 0.4 in Table 7. The hybrid model therefore derives its advantage primarily from the structural robustness of Ridge Regression, with the Transformer contributing targeted nonlinear refinement in a secondary and constrained capacity — an integration that achieves improved generalization without sacrificing the stability that the linear component provides.



Source: (Research Results, 2025)
Figure 3. Error Distributions of the Models, Showing That the Hybrid Approach Produces More Symmetric and Concentrated Residuals, Indicating Improved Stability.

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

The qualitative implications of this weighting strategy are illustrated in Figure 3, which presents the error distributions of the models. The Transformer model exhibits wide and asymmetric residuals, indicating high variance and sensitivity to extreme price movements. Ridge Regression produces more compact and symmetric residuals, although with indications of bias associated with its limited capacity to model nonlinear dynamics. In contrast, the hybrid configuration produces residuals that are more symmetrically distributed and more tightly concentrated around zero, suggesting improved error consistency that reflects the stabilizing influence of the Ridge-dominant ensemble composition. To complement the primary accuracy metrics, Table 8 reports descriptive statistics of the residual distributions — mean error, standard deviation, skewness, and kurtosis — to provide diagnostic insight into the stability and distributional behavior of prediction errors across models.

Table 8. Descriptive Statistics Of Residual Distributions for Model Comparison

Model	Mean Error	Std. Deviation	Skewness	Kurtosis
Transformer	12,336.79	13,396.07	0.878	−0.705
Ridge	55.68	838.29	0.148	7.914
Hybrid	18.41	560.25	0.664	3.281

Source: (Research Results, 2025)

The residual distribution statistics in Table 8 provide diagnostic evidence complementing the primary error metrics. The Transformer exhibits large mean error and high dispersion, reflecting systematic bias and high variance consistent with overfitting tendencies in deep learning models applied to noisy financial data. Its positive skewness indicates systematic underestimation during upward price movements, while the slightly negative kurtosis confirms broadly dispersed errors diagnostically consistent with its high standalone RMSE.

Ridge Regression maintains a near-zero mean error and substantially lower dispersion, confirming that L2 regularization suppresses systematic bias effectively. Nevertheless, its markedly elevated kurtosis reveals heavy-tailed error behavior, suggesting that the model generates occasional large deviations during abrupt price transitions despite performing well under stable conditions — a limitation attributable to its linear structure. The hybrid model addresses both deficiencies simultaneously. Its near-zero mean error and lower standard deviation indicate tighter and more symmetric error concentration, while intermediate skewness and kurtosis values confirm that the ensemble moderates both the high-variance tendency of the Transformer and the heavy-tailed behavior of Ridge Regression. This balanced residual profile is consistent with the Ridge-dominant ensemble composition, wherein the stability of the linear component is preserved while the Transformer provides targeted nonlinear refinement. Non-Gaussian residual behavior observed across all models is a common characteristic of financial time-series and reflects inherent market properties rather than modeling deficiencies [26].

E. Sensitivity Analysis of Window Size

Window determines the temporal context used in time-series forecasting and should be empirically evaluated, particularly in volatile cryptocurrency markets [21], [22]. To assess robustness, the Hybrid Transformer–Ridge Regression model was evaluated across window sizes ranging from 10 to 60 days, with all other experimental settings held constant.

Table 9. Window Size Sensitivity Analysis of the Hybrid Model

Window Size (days)	MAE	RMSE	R^2	MAPE (%)
10	1251.572	1623.004	0.991	1.701
20	1289.844	1668.392	0.990	1.754
30	1334.216	1724.885	0.989	1.812
40	1378.905	1786.431	0.988	1.869
50	1426.317	1854.762	0.987	1.934
60	1479.684	1928.550	0.986	2.006

Source: (Research Results, 2025)

Table 9 demonstrates that the hybrid model remains stable across all tested window sizes, with R^2 consistently maintained between 0.986 and 0.991 — a narrow range that confirms robustness to temporal input length variation. Notably, the 10-day configuration reproduces the final results reported in Table 4 exactly, establishing internal consistency between the sensitivity analysis and the primary evaluation. Two patterns merit interpretation. First, error metrics deteriorate gradually and monotonically as the window size increases, rather than exhibiting abrupt drops, which suggests that the hybrid model does not catastrophically lose predictive structure when longer histories are supplied.

Second, the fact that shorter windows consistently outperform longer ones implies that recent price dynamics carry more concentrated predictive signal for Bitcoin forecasting than extended historical contexts, a finding consistent with the high-frequency regime shifts characteristic of cryptocurrency markets, where distant

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

observations may reflect structurally different market conditions and introduce rather than reduce forecasting noise. Based on these results, a 10-day window is adopted as the final configuration. This choice is empirically justified not only by its superior error metrics but also by the broader sensitivity analysis, which confirms that the selected window represents the most informative temporal context rather than an arbitrary design decision.

F. Model Interpretability and Feature Importance Analysis

Model interpretability is an important consideration in financial forecasting to promote transparency and economic plausibility of predictions. Feature importance analysis was therefore conducted using the Ridge Regression component, which provides inherent transparency through linear coefficients. Given that Ridge Regression contributes 60% of the ensemble weight, the hybrid framework retains a partially transparent decision structure while the Transformer captures complementary nonlinear temporal dependencies that remain less directly interpretable.

It should be acknowledged however, that the feature space consists of heavily engineered technical indicators derived from the same underlying price series, with several features — including EMA26, MACD_Signal, MACD, and BB_Width — exhibiting extreme multicollinearity as evidenced by VIF values exceeding 100 (Table 3). Under such conditions, individual Ridge coefficients cannot be interpreted as independent causal economic contributions, but rather reflect regularized weighting across a highly correlated predictor space. Consequently, the importance values reported in Table 10 are more appropriately understood as indicators of relative predictive relevance within the model rather than measures of direct economic influence.

Table 10. Key Feature Categories and Predictive Relevance Within the Model

Feature Category	Total Importance	Mean Importance	Interpretation
Close Price	0.4991	0.0083	Short-term price memory
EMA26	0.4934	0.0082	Trend persistence
MACD	0.3871	0.0065	Momentum reversal
ATR	0.3472	0.0058	Volatility sensitivity
BB Width	0.2385	0.0040	Market compression/expansion

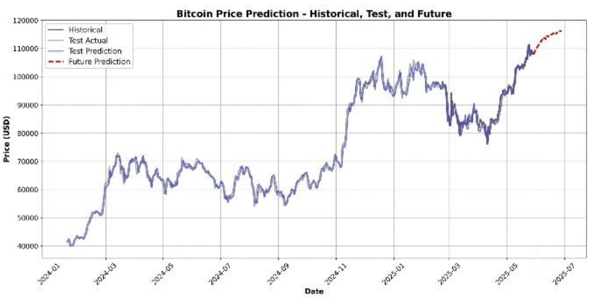
Source: (Research Results, 2025)

Table 10 summarizes the most influential feature categories based on aggregated absolute Ridge coefficients. Price-based and trend-related indicators dominate the model's predictive weighting, consistent with well-documented short-term price memory effects in cryptocurrency markets. Momentum and volatility indicators, including MACD, ATR, and Bollinger Band width, also show notable relative importance, suggesting that the model assigns meaningful weight to signals associated with trend reversals and volatility regime shifts — patterns broadly consistent with established principles of technical analysis. These patterns provide a degree of interpretive plausibility, indicating that the hybrid model's predictions are grounded in economically recognizable market signals rather than arbitrary feature representations.

Overall, while the Ridge Regression component offers a useful window into the model's decision structure, the high degree of multicollinearity among engineered indicators limits the extent to which individual feature contributions can be reliably attributed. Future research may therefore extend this analysis using SHAP-based or Shapley–Lorenz decomposition methods, which are better suited to handle correlated feature spaces and provide more reliable predictive attribution in engineered financial settings [27], [28].

G. Future Forecasting Performance

Beyond retrospective evaluation on the test set, the hybrid Transformer–Ridge Regression model was applied to generate out-of-sample forecasts in order to examine its short-horizon forecasting behavior. Figure 4 presents the historical Bitcoin price series together with test set predictions and future forecasts produced by the hybrid model.

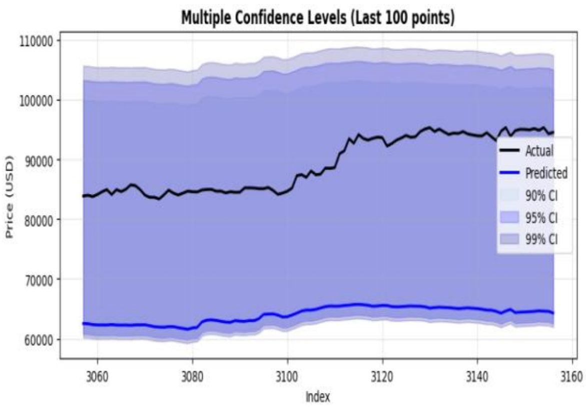


Source: (Research Results, 2025)
Figure 4. Hybrid Transformer–Ridge Regression Predictions of Bitcoin Prices

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

The results show that the predicted trajectory remains closely aligned with the recent historical trend, without exhibiting abrupt oscillations or excessive amplification of short-term fluctuations. This behavior indicates that the ensemble weighting strategy provides a stabilizing effect on the forecasting output when extending predictions beyond the observed data.



Source: (Research Results, 2025)
Figure 5. Forecasts Generated by the Hybrid Transformer–Ridge Regression Model With 90%, 95%, and 99% Confidence Intervals Over the Last 100 Observations.

Figure 5 further illustrates the forecast uncertainty by displaying prediction intervals at the 90%, 95%, and 99% confidence levels over the final observations. The intervals widen gradually as the forecast horizon increases, reflecting increasing uncertainty while maintaining a coherent central trend. Such behavior is consistent with expectations in financial time-series forecasting and suggests that the model does not produce unrealistically confident long-term projections.

Overall, the out-of-sample forecasting results demonstrate that the hybrid model produces stable and coherent future price trajectories under the examined conditions. These findings complement the retrospective performance evaluation and provide additional evidence of model robustness when generating forward-looking forecasts, without implying guaranteed predictive accuracy under unforeseen market conditions.

H. Limitations and Recommendations for Future Research

Despite achieving strong predictive performance and improved interpretability, this study acknowledges several limitations that warrant further investigation. The primary limitation arises from the reliance on historical price and technical indicator data, which may constrain the model's ability to respond effectively to unexpected market shocks, regulatory interventions, or abrupt macroeconomic shifts. Previous studies have highlighted that forecasting models which exclude exogenous variables—such as public sentiment or policy-related information may struggle to fully capture the real-time dynamics of cryptocurrency markets [29].

Another limitation concerns computational efficiency. Although the hybrid Transformer Ridge architecture balances accuracy and stability, the Transformer component still introduces notable computational overhead. This may restrict applicability in latency-sensitive or real-time trading environments, where rapid inference is critical. Similar concerns have been reported in prior deep learning–based financial forecasting studies employing complex architectures such as CNN–LSTM models, which require extensive hyperparameter tuning and computational resources [30]. Several directions for future research are therefore recommended. First, incorporating external information sources—such as social media sentiment, regulatory announcements, and macroeconomic indicators—may further enhance predictive robustness. Recent empirical evidence suggests that sentiment-aware deep learning models significantly improve Bitcoin price forecasting accuracy, underscoring the value of integrating non-price-based signals [29].

Second, while this study introduces interpretability through the Ridge Regression component and permutation-based feature importance, future work may extend explainable artificial intelligence (XAI) techniques to the deep learning component itself. Methods such as SHAP-based or Shapley–Lorenz value explanations have recently been shown to effectively interpret AI-driven financial time-series models, including Bitcoin price prediction [27], and could further enhance transparency and trust in hybrid architectures [28]. Finally, real-time deployment and validation under diverse and evolving market conditions remain an important avenue for future investigation. Continuous online evaluation would provide deeper insights into the adaptability and operational reliability of the proposed hybrid model, thereby strengthening its applicability in real-world financial forecasting systems.

CONCLUSION

This study proposes a hybrid forecasting framework that integrates a Transformer neural network with Ridge Regression for Bitcoin price prediction. The empirical results demonstrate that the hybrid approach achieves superior predictive accuracy and more stable error behavior compared to standalone models, indicating improved generalization performance in highly volatile Bitcoin markets. From a theoretical perspective, this work contributes to the literature by providing a bias–variance–motivated explanation of hybrid model advantages. By combining sequence-aware representation learning with linear regularization, the proposed framework balances model expressiveness and robustness, addressing overfitting tendencies commonly observed in deep learning–based financial forecasting models. This conceptual integration extends existing Transformer-based hybrid approaches by explicitly emphasizing regularization-driven stability rather than heuristic stacking of deep components. Methodologically, the study strengthens forecasting reliability through systematic feature engineering, multicollinearity diagnostics, and window-size sensitivity analysis, ensuring that model performance is not driven by arbitrary design choices. The inclusion of residual distribution analysis and statistical testing via the Diebold–Mariano test further enhances the rigor and interpretability of the evaluation process.

From a practical standpoint, the economic validation reveals that the hybrid model is the only evaluated strategy to demonstrate statistically significant directional accuracy, suggesting that its forecasting signals carry meaningful directional information. However, absolute trading returns remain below the passive buy-and-hold benchmark under the trending market conditions of the evaluation period, indicating that statistical forecasting improvements do not automatically translate into superior investment outcomes under simple trading rules. These findings suggest that the model's primary contribution lies in forecast reliability and directional signal quality rather than return maximization, and that translating predictive improvements into economically meaningful gains would require more sophisticated trading strategies, transaction-cost-aware optimization, and dynamic risk management mechanisms. Despite these contributions, several limitations remain. The model's sensitivity to macroeconomic shocks and its relatively high computational requirements may restrict real-time deployment. Future research may explore adaptive windowing strategies, uncertainty-aware forecasting, integration of external signals such as market sentiment, and more computationally efficient architectures to further enhance practical applicability.

REFERENCE

[1] O. Omole and D. Enke, "Deep learning for Bitcoin price direction prediction: models and trading strategies empirically compared," *Financ. Innov.*, vol. 10, no. 1, pp. 1–26, 2024, doi: 10.1186/s40854-024-00643-1.

[2] Y. Wang, G. Andreeva, and B. Martin-Barragan, "Machine learning approaches to forecasting cryptocurrency volatility: Considering internal and external determinants," *Int. Rev. Financ. Anal.*, vol. 90, no. August, p. 102914, 2023, doi: 10.1016/j.irfa.2023.102914.

[3] Z. C. Huang, I. Sangiorgi, and A. Urquhart, "Forecasting Bitcoin volatility using machine learning techniques," *J. Int. Financ. Mark. Institutions Money*, vol. 97, 2024, doi: 10.1016/j.intfin.2024.102064.

[4] M. Lin, Y. Liu, and V. N. K. Sheng, "Analysis of the impact of macroeconomic factors on cryptocurrency returns - Based on quantile regression study," *Int. Rev. Econ. Financ.*, vol. 97, no. October 2024, p. 103757, 2025, doi: 10.1016/j.iref.2024.103757.

[5] A. Bouteska, M. Z. Abedin, P. Hajek, and K. Yuan, "Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods," *Int. Rev. Financ. Anal.*, vol. 92, no. November 2023, 2024, doi: 10.1016/j.irfa.2023.103055.

[6] L. Ø. Bergsli, A. F. Lind, P. Molnár, and M. Polasik, "Forecasting volatility of Bitcoin," *Res. Int. Bus. Financ.*, vol. 59, no. September 2021, 2022, doi: 10.1016/j.ribaf.2021.101540.

[7] C. Jin and Y. Li, "Cryptocurrency Price Prediction Using Frequency Decomposition and Deep Learning," *Fractal Fract.*, vol. 7, no. 10, pp. 1–29, 2023, doi: 10.3390/fractalfract7100708.

[8] E. Mahdi, C. Martin-Barreiro, and X. Cabezas, "A Novel Hybrid Approach Using an Attention-Based Transformer + GRU Model for Predicting Cryptocurrency Prices," *Mathematics*, vol. 13, no. 9, pp. 1–19, 2025, doi: 10.3390/math13091484.

[9] R. Bourday, I. Aatouchi, M. A. Kerroum, and A. Zaaouat, "Cryptocurrency Forecasting Using Deep Learning Models: A Comparative Analysis," *HighTech Innov. J.*, vol. 5, no. 4, pp. 1055–1067, 2024, doi: 10.28991/HIJ-2024-05-04-013.

[10] A. Jenefa, M. Mugilarasan, T. M. Thiyagu, R. Catherine Joy, P. Santhiya, and K. Vidhya,

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.7789

"DL-Crypto: Deep Learning Techniques for Cryptocurrency Price Prediction," *2025 8th Int. Conf. Circuit, Power Comput. Technol. ICCPCT 2025*, pp. 1763–1768, 2025, doi: 10.1109/ICCPCT65132.2025.11176771.

[11] A. Tanwar and V. Kumar, "Prediction of Cryptocurrency prices using Transformers and Long Short term Neural Networks," *2022 Int. Conf. Intell. Controll. Comput. Smart Power, ICICCSP 2022*, pp. 1–4, 2022, doi: 10.1109/ICICCSP53532.2022.9862436.

[12] P. K. Nagula and C. Alexakis, "A new hybrid machine learning model for predicting the bitcoin (BTC-USD) price," *J. Behav. Exp. Financ.*, vol. 36, p. 100741, 2022, doi: 10.1016/j.jbef.2022.100741.

[13] O. Omole and D. Enke, "Using machine and deep learning models, on-chain data, and technical analysis for predicting bitcoin price direction and magnitude," *Eng. Appl. Artif. Intell.*, vol. 154, no. November 2024, p. 111086, 2025, doi: 10.1016/j.engappai.2025.111086.

[14] A. Golnari, M. H. Komeili, and Z. Azizi, "Probabilistic deep learning and transfer learning for robust cryptocurrency price prediction," *Expert Syst. Appl.*, vol. 255, no. March, 2024, doi: 10.1016/j.eswa.2024.124404.

[15] D. L. John, S. Binnewies, and B. Stantic, "Cryptocurrency Price Prediction Algorithms: A Survey and Future Directions," *Forecasting*, vol. 6, no. 3, pp. 637–671, 2024, doi: 10.3390/forecast6030034.

[16] D. Lapitskaya, M. H. Eratalay, and R. Sharma, "Prediction of Cryptocurrency Prices with the Momentum Indicators and Machine Learning," *Comput. Econ.*, vol. 66, no. 3, pp. 2483–2501, 2025, doi: 10.1007/s10614-024-10784-1.

[17] M. Shahhosseini, G. Hu, and H. Pham, "Optimizing ensemble weights and hyperparameters of machine learning models for regression problems," *Mach. Learn. with Appl.*, vol. 7, no. January, p. 100251, 2022, doi: 10.1016/j.mlwa.2022.100251.

[18] M. Zatwarnicki and K. Zatwarnicki, "Timing Usage of Technical Analysis in the Cryptocurrency Market," *Appl. Sci.*, vol. 15, no. 23, p. 12802, Dec. 2025, doi: 10.3390/app152312802.

[19] A. Kumar, N. Sharma, R. Chauhan, and M. Sharma, "Predicting Cryptocurrency Prices: An Exploration of Blockchain's Impact on Traditional Currency," *2023 3rd Int. Conf. Smart Gener. Comput. Commun. Networking, SMART GENCON 2023*, pp. 1–6, 2023, doi: 10.1109/SMARTGENCON60755.2023.10442230.

[20] J. Jacob and R. Varadharajan, "Robust Variance Inflation Factor: A Promising Approach for Collinearity Diagnostics in the Presence of Outliers," *Sankhya B*, vol. 86, no. 2, pp. 845–871, Nov. 2024, doi: 10.1007/s13571-024-00342-y.

[21] I. Akouaouch and A. Bouayad, "A new deep learning approach for predicting high-frequency short-term cryptocurrency price," *Bull. Electr. Eng. Informatics*, vol. 14, no. 1, pp. 513–523, 2025, doi: 10.11591/eei.v14i1.7377.

[22] D. L. John, S. Binnewies, and B. Stantic, "Identifying Optimal Window Size Configurations for Big Data Time Series Forecasting," *Proc. - 2024 IEEE Int. Conf. Big Data, BigData 2024*, pp. 5138–5146, 2024, doi: 10.1109/BigData62323.2024.10825222.

[23] G. Dudek, P. Fiszeder, and W. Orzeszko, "Forecasting cryptocurrencies volatility using statistical and machine learning methods : A comparative study," vol. 151, no. March 2023, 2024, doi: 10.1016/j.asoc.2023.111132.

[24] K. Murray, A. Rossi, D. Carraro, and A. Visentin, "On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles," *Forecasting*, vol. 5, no. 1, pp. 196–209, 2023, doi: 10.3390/forecast5010010.

[25] A. P. Pratiwi, R. V. H. Ginardi, and A. Saikhu, "Enhancing Electricity Consumption Prediction with Deep Learning through Advanced Data Splitting Techniques," *Int. J. Artif. Intell. Res.*, vol. 8, no. 2, p. 180, 2024, doi: 10.29099/ijair.v8i2.1204.

[26] M. Yousaf, M. Tariq, A. Jabbar, and S. Q. Jalil, "A Comprehensive Survey of Cryptocurrency Forecasting: Methods, Trends, and Challenges," Dec. 03, 2024. doi: 10.20944/preprints202411.2330.v2.

[27] P. Giudici, A. Piergallini, M. C. Recchioni, and E. Raffinetti, "Explainable Artificial Intelligence methods for financial time series," *Phys. A Stat. Mech. its Appl.*, vol. 655, no. October, p. 130176, 2024, doi: 10.1016/j.physa.2024.130176.

[28] M. Sasikumar and E. D. Raj, "Investigating explainability of deep learning models for sequential data on stock price prediction,"

Procedia Comput. Sci., vol. 258, pp. 4190–4201, 2025, doi: 10.1016/j.procs.2025.04.669.

[29] H. S. Htay, M. Ghahremani, and S. Shiaeles, "Enhancing Bitcoin Price Prediction with Deep Learning: Integrating Social Media Sentiment and Historical Data," *Appl. Sci.*, vol. 15, no. 3, 2025, doi: 10.3390/app15031554.

[30] M. N. Tahmid Akhand, M. Ahsan Habib, and K. M. Rokibul Alam, "Analyzing Cryptocurrency Price Trends for Real-Time Price Predictions," *2023 26th Int. Conf. Comput. Inf. Technol. ICCIT 2023*, pp. 1–6, 2023, doi: 10.1109/ICCIT60459.2023.10441450.