JOURNAL OF MANAGEMENT AND BUSINESS SOLUTIONS

A Hybrid Deep Ensemble Learning Framework for Cryptocurrency Price Prediction

1. Mohammadreza. Haghighi: Department of Information and Technology Management, Qa.C., Islamic Azad University, Qazvin, Iran
2. Seyed Yashar. Banihashem*: Assistant Professor, Department of Electrical and Computer Engineering, Buein Zahra Technical University, Buein Zahra, Iran
3. MohammadMehdi. Gilaniansadeghi: Department of computer engineering, Qa.C., Islamic Azad University, Qazvin, Iran

*corresponding author's email: banihashem@bzte.ac.ir

ABSTRACT

This study aimed to develop and evaluate a hybrid deep ensemble learning framework integrating CNN-LSTM, GRU, and Transformer architectures through stacking for accurate next-day cryptocurrency price prediction. This quantitative predictive study analyzed daily market data for five major cryptocurrencies, including Bitcoin, Ethereum, Binance Coin, Ripple, and Cardano, over the period from January 1, 2018, to December 31, 2025. The dataset comprised 14,610 cryptocurrency-day observations, with 2,922 observations for each asset. Data were chronologically divided into training, validation, and testing subsets. Historical price, trading volume, return, volatility, lagged variables, and technical indicators were used as predictors. A 30-day sliding window was employed to predict the following day's closing price. The proposed framework combined CNN-LSTM, GRU, and Transformer models using a stacking-based meta-learner. Performance was evaluated using Mean Absolute Error, Root Mean Squared Error, Mean Absolute Percentage Error, and R^2. Diebold-Mariano tests were used to compare out-of-sample forecast accuracy. The proposed hybrid ensemble achieved an MAE of 0.0294, RMSE of 0.0436, MAPE of 3.12%, and R^2 of 0.971, outperforming CNN-LSTM, GRU, Transformer, ARIMA, Random Forest, XGBoost, and persistence benchmarks. Diebold-Mariano tests showed that the ensemble produced significantly lower forecast loss than all competing models, including Transformer, the strongest standalone model, with all p values below 0.001. Cryptocurrency-specific analyses yielded R^2 values above 0.960 for all five assets, while volatility-regime analysis showed that predictive accuracy declined under high-volatility conditions but remained strong, with an R^2 of 0.952. The findings demonstrate that combining convolutional, recurrent, and attention-based deep-learning mechanisms through stacking can significantly improve the accuracy and robustness of cryptocurrency price forecasting across heterogeneous assets and changing volatility conditions.

Keywords: Cryptocurrency Price Prediction; Deep Ensemble Learning; CNN-LSTM; GRU; Transformer; Stacking; Time-Series Forecasting; Machine Learning

Introduction

Cryptocurrencies have emerged as one of the most dynamic and technologically distinctive classes of financial assets in contemporary global markets. Unlike conventional currencies issued and regulated by central monetary authorities, cryptocurrencies are primarily based on decentralized blockchain infrastructures and are traded continuously across geographically distributed digital exchanges. Their rapid expansion has created new opportunities for investment, portfolio diversification, decentralized finance, and technological innovation, while

Article history:
Received 21 September 2025
Revised 14 October 2025
Accepted 24 December 2025
Published online 10 January 2026

How to cite this article:
Haghighi, M., Banihashem, S. Y., & Gilaniansadeghi, M. (2026). A Hybrid Deep Ensemble Learning Framework for Cryptocurrency Price Prediction. *Journal of Management and Business Solutions, 4*(1), 1-22. https://doi.org/10.61838/jmbs.428

simultaneously introducing considerable uncertainty for investors, traders, regulators, and financial analysts. Among the most challenging characteristics of cryptocurrency markets are their high volatility, nonlinear price behavior, sensitivity to speculative trading, continuous twenty-four-hour market operation, and rapid response to economic, technological, regulatory, and social information. These characteristics make accurate cryptocurrency price forecasting substantially more difficult than forecasting many traditional financial instruments. Nevertheless, the ability to anticipate cryptocurrency price movements has become increasingly important because even relatively small improvements in forecasting accuracy may contribute to more effective asset allocation, trading decisions, risk management, and portfolio optimization. Accordingly, cryptocurrency price prediction has developed into a prominent interdisciplinary research area involving financial econometrics, artificial intelligence, machine learning, and computational finance (1, 2).

Traditional financial forecasting methods generally assume that historical observations contain identifiable statistical structures that can be represented through linear or moderately nonlinear relationships. However, cryptocurrency prices frequently demonstrate abrupt regime changes, extreme fluctuations, volatility clustering, complex temporal dependencies, and rapidly changing correlations among market variables. Such properties reduce the effectiveness of forecasting procedures that depend primarily on stationary assumptions or predetermined functional relationships. The cryptocurrency market is additionally affected by investor sentiment, technological developments, network events, regulatory announcements, changes in adoption, speculative demand, and macroeconomic conditions, creating highly complex relationships between historical information and subsequent market prices. Consequently, forecasting models must be sufficiently flexible to identify changing temporal patterns and extract predictive information from large amounts of noisy financial data. Earlier studies therefore increasingly shifted from conventional statistical forecasting toward machine-learning approaches capable of identifying nonlinear relationships without requiring strict assumptions concerning the underlying data-generating process (3, 4).

Machine-learning techniques represent an important advancement in cryptocurrency forecasting because they allow computational models to learn patterns directly from historical data. The application of such methods has demonstrated that cryptocurrency prices can, to some extent, be predicted using combinations of historical prices, trading volume, market indicators, technical variables, and derived temporal features. Derbentsev and colleagues examined the forecasting of cryptocurrency price time series using machine-learning procedures and demonstrated the potential value of data-driven algorithms for modeling highly irregular digital-asset markets (3). Similarly, Chowdhury and colleagues investigated machine-learning approaches for predicting both individual cryptocurrency constituents and broader cryptocurrency market indices, emphasizing that algorithmic models can extract useful predictive relationships from market data that may not be readily represented by traditional approaches (4). Research on artificial-intelligence-based forecasting models has likewise indicated that computational approaches can provide a useful basis for predicting cryptocurrency prices when appropriate historical and technical predictors are incorporated into the modeling process (5).

Several studies conducted in this area have focused specifically on Bitcoin because of its market dominance, relatively long historical record, and importance as the first widely adopted cryptocurrency. Research applying machine-learning techniques to the forecasting of the first major virtual cryptocurrency provided early evidence that nonlinear computational methods could be used to identify patterns in cryptocurrency price movements (6). Subsequent investigations expanded these approaches through the application of deep machine-learning techniques, demonstrating that more sophisticated architectures may capture temporal relationships that are difficult for shallow learning systems to represent (7). Other research has examined the forecasting of cryptocurrency prices within broader financial asset and portfolio-selection frameworks, demonstrating that prediction accuracy has direct implications not only for individual price estimates but also for practical investment decisions and the construction of portfolios containing both conventional financial assets and cryptocurrencies (8). Therefore, cryptocurrency price forecasting has gradually moved from an isolated prediction problem toward a broader decision-support problem within computational finance.

The emergence of deep learning has substantially changed the methodological landscape of financial time-series forecasting. Deep-learning models are capable of automatically extracting hierarchical representations from sequential data and can learn complex nonlinear structures from large sets of observations. Recurrent Neural Networks and their gated variants have received particular attention because financial price series are temporally ordered and present dependencies that may extend over multiple previous observations. Long Short-Term Memory networks were developed to address limitations associated with conventional recurrent neural networks, particularly difficulties in retaining information over extended sequences. Their gated structure makes them particularly suitable for identifying short- and medium-term temporal dependencies within cryptocurrency markets. Research on optimizing LSTM networks for cryptocurrency forecasting has demonstrated that carefully configured recurrent architectures can improve the representation of temporal market behavior and provide more accurate price forecasts (9). Similarly, applications of LSTM-based deep-learning techniques to cryptocurrency prediction have supported the capacity of these models to learn dynamic sequential relationships in market prices (7).

Convolutional Neural Networks have also become increasingly relevant to financial forecasting. Although CNNs were initially developed primarily for image-related tasks, one-dimensional convolutional structures can effectively detect local patterns in sequential financial data. By applying filters across consecutive observations, CNNs can identify short-term configurations such as local price movements, momentum patterns, and changes in volatility. The combination of convolutional and recurrent mechanisms is consequently attractive because the CNN component can extract local market features while recurrent units model the temporal relationships among those extracted features. Anderson and Patel investigated cryptocurrency price prediction using CNN and LSTM networks and highlighted the value of combining feature extraction with sequential learning for financial forecasting (10). Gonzalez and Sanchez similarly examined cryptocurrency time-series forecasting through CNN and recurrent neural-network architectures, demonstrating that both convolutional and recurrent structures can contribute to modeling complex cryptocurrency price patterns (11).

Hybrid CNN-RNN architectures represent a further development in this direction. Rather than requiring one architecture to simultaneously capture every aspect of the price series, hybrid designs allocate different aspects of representation learning to specialized neural components. CNN layers can identify local temporal structures, whereas recurrent layers model sequential dependencies and memory effects. Ma and Zhou reported the use of hybrid CNN-RNN models for cryptocurrency price forecasting and demonstrated the usefulness of combining these complementary architectures within a unified predictive framework (12). Related research on hybrid machine-learning approaches has also emphasized the potential advantages of combining different computational methods for cryptocurrency forecasting rather than depending exclusively on one algorithmic representation (13). These findings support the broader argument that cryptocurrency markets are sufficiently heterogeneous and nonlinear that model integration may be more suitable than the assumption that a single learning architecture will perform optimally under all market conditions.

Alongside deep neural networks, gradient-boosting algorithms have also been applied to cryptocurrency prediction. These approaches construct a sequence of predictive learners in which successive models attempt to correct errors generated by earlier models. LightGBM, in particular, offers efficient gradient-boosting capabilities and can model complex interactions among structured predictors. A cryptocurrency price-trend forecasting model based on LightGBM demonstrated that tree-based ensemble learning can be effective for identifying price-direction patterns from financial and technical information (14). This finding is important because it demonstrates that strong cryptocurrency forecasting is not restricted to recurrent neural networks. Instead, different modeling paradigms may capture different dimensions of the market. Tree-based algorithms can efficiently model nonlinear interactions among engineered features, while recurrent, convolutional, and attention-based neural networks can represent temporal dependencies directly from ordered sequences. The relative advantages of these approaches have encouraged researchers to investigate ensemble systems that combine several learners in an attempt to exploit their complementary predictive strengths.

Ensemble learning is particularly relevant to cryptocurrency forecasting because the market's statistical structure may vary substantially across time and across digital assets. A model that performs strongly under stable market conditions may be less effective during abrupt volatility, whereas another architecture may adapt better to rapid changes but perform less consistently during ordinary trading periods. Ensemble learning attempts to reduce dependence on the weaknesses of a single model by combining forecasts generated by multiple base learners. The integration may be performed through simple averaging, weighted averaging, voting procedures, boosting, bagging, or stacking. Among these approaches, stacking is particularly attractive because it allows a secondary model to learn how much importance should be assigned to the output of each base learner. Comparative research has shown that ensemble-learning approaches can provide substantial benefits in cryptocurrency price forecasting when compared with individual models, particularly when the ensemble contains learners with different structural characteristics (15).

Bouteska and colleagues conducted a comparative investigation of ensemble-learning and deep-learning methods for cryptocurrency price forecasting and highlighted the importance of model architecture, integration, and comparative evaluation when attempting to improve prediction performance (15). Their findings reinforce an emerging methodological direction in which forecasting performance is improved not simply by increasing the complexity of a single model but by systematically combining models capable of learning different dimensions of cryptocurrency behavior. This principle is especially relevant for integrating convolutional, recurrent, and attention-based networks. Convolutional models are effective at identifying local features, recurrent networks such as LSTM and GRU are designed to capture sequential dependencies, and attention-based architectures can assign varying importance to different positions within an input sequence. A hybrid ensemble incorporating these structures may therefore provide a more comprehensive representation of cryptocurrency price formation than any component independently.

More recent research has increasingly emphasized integrated forecasting frameworks rather than isolated prediction algorithms. Alnami and colleagues developed an integrated framework combining cryptocurrency price forecasting and anomaly detection through machine learning, reflecting the growing recognition that unusual market movements and structural irregularities are central characteristics of cryptocurrency data (16). Anomaly-aware and integrated designs are particularly important in markets where extreme observations are not rare exceptions but recurring components of normal market behavior. Rapid price increases, sudden crashes, and temporary spikes can substantially influence the optimization of forecasting algorithms, particularly models trained by minimizing squared prediction errors. Consequently, contemporary cryptocurrency forecasting frameworks increasingly need to address not only average predictive accuracy but also robustness to abnormal or high-volatility periods.

The evaluation of forecasting algorithms has similarly become more rigorous. Predictive models should not be assessed using a single error indicator because different performance measures capture different aspects of forecast quality. Mean Absolute Error provides a direct measure of typical forecast deviation, Root Mean Squared Error assigns greater weight to large prediction failures, Mean Absolute Percentage Error offers a relatively scale-independent assessment of error, and the coefficient of determination indicates the extent to which predicted values correspond to variation in observed prices. Moreover, model comparisons should ideally be conducted using temporally independent testing data rather than randomly shuffled observations, because random partitioning may allow future market information to indirectly influence model development. Comparative investigations of machine-learning models for cryptocurrency forecasting have stressed the importance of evaluating algorithms through multiple performance criteria and examining predictive accuracy across alternative model classes (2).

Qureshi and colleagues' evaluation of machine-learning models for cryptocurrency price forecasting further illustrates that differences among algorithms can be meaningful and that predictive superiority must be established empirically rather than assumed from model complexity alone (2). A highly parameterized deep-learning model does not necessarily guarantee superior out-of-sample forecasting unless it is appropriately configured, regularized, and evaluated. Model performance can depend on input variables, historical window length, preprocessing procedures, the selected cryptocurrency, market regime, and the specific forecasting horizon. These considerations emphasize the necessity of designing forecasting studies in which all candidate models are evaluated under identical temporal partitions and preprocessing procedures. They also support statistical comparisons of forecast errors in addition to descriptive metrics, since small apparent differences in RMSE or MAPE may not necessarily correspond to reliable improvements in predictive performance.

Another important issue is that cryptocurrency markets should not be treated as behaviorally homogeneous. Bitcoin, Ethereum, Binance Coin, XRP, Cardano, and other cryptocurrencies differ substantially in market capitalization, liquidity, technological function, adoption, investor composition, and historical volatility. A forecasting framework that performs well for Bitcoin may not produce the same accuracy for lower-priced or more volatile cryptocurrencies. Earlier studies focused primarily on Bitcoin or relatively restricted cryptocurrency samples, which provided valuable foundations but limited conclusions regarding cross-asset generalizability (6, 7). Studies examining broader cryptocurrency sets have consequently become increasingly important because they allow forecasting architectures to be tested under heterogeneous price scales and market conditions (1, 4). Evaluating several major cryptocurrencies also provides a stronger basis for determining whether a proposed framework captures general temporal market characteristics or merely overfits the dynamics of one asset.

The relationship between forecasting accuracy and market volatility represents another unresolved issue. Deep-learning models may generate highly accurate predictions during relatively stable periods but encounter substantial error around abrupt turning points. Cryptocurrency markets are particularly vulnerable to this problem because volatility can increase suddenly in response to market sentiment, regulation, security events, technological developments, or broader financial shocks. Forecasting performance should therefore ideally be examined across different volatility regimes rather than summarized solely through an aggregate test-set statistic. Recent reviews of cryptocurrency trading strategies and price-forecasting techniques emphasize that predictive models must ultimately be evaluated in relation to changing market conditions and practical financial decision-making rather than only computational performance in a static setting (1). This broader perspective is essential when determining whether a forecasting architecture is sufficiently robust to provide reliable information under the conditions in which investors are most exposed to risk.

Collectively, previous studies demonstrate substantial progress in cryptocurrency forecasting through machine learning, deep learning, CNNs, recurrent neural networks, LSTM optimization, gradient boosting, hybrid architectures, and ensemble systems (9-12, 14). Studies conducted using artificial-intelligence and hybrid machine-learning techniques have similarly supported the capacity of computational models to forecast cryptocurrency markets and contribute to financial decision-making (5, 8, 13). At the same time, contemporary evidence indicates that forecasting accuracy can potentially be improved through more systematic integration of complementary learners, rigorous out-of-sample testing, anomaly-aware analysis, and cross-model comparisons (2, 15, 16). Despite these developments, a methodological gap remains regarding the construction of integrated frameworks that simultaneously exploit local feature extraction, sequential memory, and long-range attention mechanisms and then combine their forecasts through a learned ensemble procedure rather than relying on a single architecture.

The proposed framework addresses this gap by integrating CNN-LSTM, GRU, and Transformer-based forecasting components within a stacking-based deep ensemble structure. This combination is conceptually appropriate because each learner is designed to represent a different aspect of temporal financial behavior: the CNN-LSTM architecture captures local patterns and sequential dependencies, the GRU provides an efficient gated representation of dynamic temporal information, and the Transformer uses self-attention to identify relationships between temporally separated observations. Their predictions can then be integrated through a meta-learning mechanism capable of assigning differential importance to each component based on its forecasting contribution. Such a framework extends earlier hybrid and ensemble approaches by focusing explicitly on complementary deep representations and by evaluating performance across several major cryptocurrencies, conventional benchmarks, individual neural architectures, and different volatility conditions. This design provides an opportunity to determine not merely whether deep learning is effective for cryptocurrency prediction, but whether a structured combination of heterogeneous deep learners can provide more accurate and stable forecasts than competing alternatives.

Accordingly, the aim of this study was to develop and evaluate a hybrid deep ensemble learning framework integrating CNN-LSTM, GRU, and Transformer architectures through stacking for accurate next-day price prediction across major cryptocurrencies and to compare its predictive performance with conventional machine-learning, statistical, and standalone deep-learning models under different market volatility conditions.

Methods and Materials

This study employed a quantitative, predictive, and computational research design to develop and evaluate a hybrid deep ensemble learning framework for forecasting cryptocurrency prices. The study was designed as a time-series prediction investigation in which historical market information was used to estimate the closing price of cryptocurrencies one trading day ahead. Five major cryptocurrencies were included in the analysis: Bitcoin (BTC), Ethereum (ETH), Binance Coin (BNB), Ripple (XRP), and Cardano (ADA). These cryptocurrencies were selected because of their relatively high market capitalization, trading liquidity, historical continuity, and representation of different segments of the cryptocurrency market. Daily observations were collected for the period from January 1, 2018, through December 31, 2025, resulting in exactly 2,922 consecutive daily observations for each cryptocurrency and a total raw sample of 14,610 cryptocurrency-day observations. The use of multiple cryptocurrencies rather than a single asset was intended to evaluate whether the proposed framework could capture both asset-specific temporal patterns and common nonlinear characteristics of the cryptocurrency market. The primary dependent variable was the next-day closing price of each cryptocurrency, whereas historical price, trading volume, return, volatility, and technical market indicators were considered predictor variables. To preserve the temporal ordering of the observations and prevent information leakage, random sampling was not used. Instead, the observations for each cryptocurrency were divided chronologically into training, validation, and testing subsets. The first 2,045 observations, corresponding to approximately 70% of each time series, were used for model training; the subsequent 438 observations, corresponding to approximately 15%, were used for validation and hyperparameter optimization; and the final 439 observations, corresponding to approximately 15%, were retained as an independent testing set for the final evaluation of forecasting performance. Accordingly, across the five cryptocurrencies, 10,225 raw asset-day observations were allocated to the training period, 2,190 to the validation period, and 2,195 to the testing period. All preprocessing parameters, including scaling coefficients and model hyperparameters, were estimated exclusively from the training data and subsequently applied to the validation and testing data. This chronological partitioning ensured that observations from the future did not contribute to the prediction of earlier prices and therefore reproduced the conditions under which the forecasting framework would operate in a real cryptocurrency market environment.

Historical cryptocurrency market data constituted the principal data source used in this study. For each cryptocurrency, daily Open, High, Low, Close, and trading Volume variables were obtained for the predefined study period and organized in chronological order. The closing price was selected as the principal forecasting target because it provides a standardized representation of the daily market valuation of an asset and is widely used for evaluating financial forecasting models. Before modeling, the collected datasets were examined for duplicated dates, missing observations, invalid values, non-numeric entries, and inconsistencies in chronological ordering. Duplicate records were removed, observations were sorted according to date, and the integrity of the time-series structure was verified prior to feature engineering. In addition to the original market variables, a comprehensive set of derived predictors was generated to represent price momentum, market trend, volatility, and trading activity. These variables included daily logarithmic return, percentage price change, intraday high-low range, open-close spread, rolling price volatility, simple moving averages, exponential moving averages, Relative Strength Index, Moving Average Convergence Divergence and its signal component, Bollinger Band characteristics, Average True Range, rate of change, momentum, and rolling statistics calculated over short-, medium-, and longer-term windows. Lagged versions of closing price, return, and volume variables were also constructed to allow the models to capture delayed market dependencies. A 30-day sliding observation window was employed, meaning that the information available during the preceding 30 consecutive days was used to predict the closing price on the following day. This windowing procedure transformed each cryptocurrency series into supervised learning sequences while retaining the temporal dependence of financial observations.

Data preprocessing was conducted independently for each cryptocurrency. Because cryptocurrency prices and trading volumes differ considerably in magnitude across assets and across time, continuous predictor variables were normalized before they were supplied to the deep-learning algorithms. Min-max normalization was applied using parameters estimated exclusively from the training subset, transforming input variables to a standardized numerical range suitable for gradient-based optimization. The same fitted transformation was then used without re-estimation for the validation and testing subsets. The target closing-price values were transformed in accordance with the modeling procedure and subsequently converted back to their original price scale before the final error statistics were computed so that performance measures remained directly interpretable in financial units. Particular attention was given to avoiding look-ahead bias during feature construction. Therefore, every rolling statistic, lagged variable, and technical indicator associated with a prediction at time t+1 was calculated using only information available at or before time t. No future closing price, return, or market information was incorporated into the predictor matrix.

The proposed framework was constructed as a hybrid deep ensemble architecture combining complementary neural-network models with different capacities for extracting local, sequential, and long-range temporal structures. The ensemble consisted of a one-dimensional Convolutional Neural Network–Long Short-Term Memory model, a Gated Recurrent Unit model, and a Transformer-based temporal model. The CNN-LSTM component was included to identify short-term local patterns and subsequently model their sequential dependencies. One-dimensional convolutional layers were applied to the input sequences to extract local temporal features, after which the resulting feature maps were transmitted to LSTM units designed to retain relevant information across successive observations. The GRU component provided a computationally efficient recurrent representation of nonlinear temporal relationships and served as an alternative sequential learner with a gating mechanism different from that of the LSTM architecture. The Transformer component used multi-head self-attention to assign different weights to observations within the historical window and was incorporated to detect longer-range relationships that conventional recurrent networks may represent less efficiently. Each component model independently produced a next-day price prediction. These predictions were then integrated through a stacking-based ensemble mechanism rather than simple averaging. Out-of-sample predictions generated from the validation data were supplied to a regularized linear meta-learner, which estimated the optimal contribution of the individual deep-learning components to the final prediction. This structure allowed the ensemble to give greater influence to models exhibiting stronger predictive performance while reducing the effect of model-specific forecasting errors.

The computational procedures were implemented in Python using numerical and machine-learning libraries appropriate for financial time-series analysis and deep learning. Data manipulation and feature engineering were performed using Pandas and NumPy, preprocessing and conventional evaluation procedures were implemented using Scikit-learn, and deep neural networks were developed using TensorFlow and Keras. Model training was performed using mini-batch gradient optimization with the Adam optimizer. Mean squared error was specified as the principal loss function during neural-network training. To reduce overfitting, dropout regularization, early stopping, and validation-based model checkpointing were incorporated into the training procedure. Early stopping monitored validation loss and terminated training when no meaningful improvement was obtained across consecutive epochs, after which the network weights associated with the minimum validation loss were restored. Hyperparameters, including the number of recurrent units, number and size of convolutional filters, Transformer attention heads, hidden-layer dimensions, dropout rates, batch size, learning rate, and regularization intensity, were selected using the validation subset only. The independent test subset remained inaccessible during training and model selection and was used exclusively after completion of the entire model-development procedure.

Data analysis was conducted in several consecutive stages consisting of descriptive examination, preprocessing, feature generation, deep-learning model development, ensemble construction, out-of-sample forecasting, and statistical comparison of predictive accuracy. Initially, the distributional and temporal characteristics of the cryptocurrency series were examined using descriptive statistics, including the mean, standard deviation, minimum, maximum, median, coefficient of variation, daily return, and rolling volatility. Temporal trajectories of the closing prices and returns were also examined to identify major fluctuations, structural changes, and periods of unusually high volatility. Because the purpose of the study was forecasting rather than conventional cross-sectional inference, preservation of temporal structure was treated as a fundamental analytical requirement. All model estimation was consequently performed using chronological observations, and no random train-test shuffling was permitted.

After preprocessing and feature engineering, supervised learning samples were created through a rolling 30-day input window with a one-day forecasting horizon. Each input tensor therefore contained the predictor information from days $t-29$ through t, while the corresponding target represented the cryptocurrency closing price at $t+1$. The CNN-LSTM, GRU, and Transformer models were initially trained separately on the training portion of each series. Their parameters were optimized through backpropagation, and validation loss was monitored during each training cycle. Validation predictions from the three component networks were subsequently combined as the input features of the ensemble meta-learner. A regularized ridge regression model was used as the stacking layer because it permitted weighted integration of highly correlated model outputs while reducing instability associated with unrestricted regression coefficients. After the ensemble structure and all hyperparameters had been finalized using the training and validation periods, forecasts were generated for every eligible observation in the independent testing period. Predicted normalized prices were inverse-transformed to the original price scale before the calculation of final prediction errors.

The predictive performance of the proposed hybrid deep ensemble framework was evaluated using Mean Absolute Error, Root Mean Squared Error, Mean Absolute Percentage Error, and the coefficient of determination. Mean Absolute Error quantified the average absolute difference between observed and predicted prices and was used as a directly interpretable measure of general forecast deviation. Root Mean Squared Error assigned greater penalties to large forecasting errors and was therefore used to determine the framework's sensitivity to substantial price deviations during periods of high market volatility. Mean Absolute Percentage Error provided an approximately scale-independent representation of forecasting error, facilitating comparison across cryptocurrencies with markedly different price levels. The coefficient of determination was used to quantify the proportion of out-of-sample price variation represented by model predictions. Lower values of Mean Absolute Error, Root Mean Squared Error, and Mean Absolute Percentage Error and higher values of the coefficient of determination were interpreted as indicating superior forecasting performance.

To determine whether the ensemble architecture provided meaningful improvements beyond its constituent models, the performance of the proposed framework was compared separately with the CNN-LSTM, GRU, and Transformer component models. Conventional benchmark forecasting models were additionally considered to ensure that any improvement could not be attributed solely to the use of highly parameterized neural networks. Forecasting comparisons were conducted on identical test observations and under the same preprocessing and input-window conditions. The Diebold-Mariano test was employed for pairwise comparison of out-of-sample forecast errors between the proposed ensemble and competing models. This procedure evaluated the null hypothesis that two forecasting methods had equivalent predictive accuracy based on their loss differentials across the test period. Statistical significance was evaluated at a two-sided alpha level of 0.05. In addition to asset-specific evaluation, performance measures were aggregated across the five cryptocurrencies to assess the overall robustness and generalizability of the proposed framework. The final interpretation of model superiority was based not on a single performance statistic but on consistency across error metrics, cryptocurrencies, and statistical forecast-comparison tests. This analytical strategy provided a rigorous evaluation of whether combining convolutional, recurrent, and attention-based deep-learning mechanisms through a stacking ensemble could improve the accuracy and stability of next-day cryptocurrency price prediction.

Findings and Results

The final dataset consisted of 14,610 daily cryptocurrency observations obtained from five cryptocurrencies, with 2,922 observations available for each asset. Bitcoin, Ethereum, Binance Coin, Ripple, and Cardano therefore each represented exactly 20.00% of the total raw dataset. Following the chronological partitioning procedure, 10,225 observations were allocated to the training subset, 2,190 observations to the validation subset, and 2,195 observations to the independent testing subset. Accordingly, the training, validation, and testing portions represented 69.99%, 14.99%, and 15.02% of the complete raw dataset, respectively. Each cryptocurrency contributed 2,045 observations to training, 438 observations to validation, and 439 observations to testing. Because the study was based on market time-series data rather than human participants, conventional demographic variables such as age, sex, educational level, or occupation were not applicable. The descriptive profile of the study sample was therefore defined according to cryptocurrency type, period coverage, market-price distribution, return variability, and trading-volume characteristics. Considerable heterogeneity was observed among the cryptocurrencies in both price level and market volatility. Bitcoin had the highest absolute price level throughout most of the observation period, whereas XRP and Cardano were characterized by substantially lower unit prices. Ethereum and Binance Coin occupied intermediate price ranges but also demonstrated periods of marked expansion and contraction. Across all five assets, daily return series fluctuated around relatively small central values while exhibiting considerably greater dispersion than would be expected in conventional low-volatility financial instruments. This pattern confirmed the suitability of nonlinear learning algorithms capable of modeling heterogeneous and rapidly changing temporal structures.

Table 1. Descriptive Statistics of Cryptocurrency Market Variables Across the Full Study Period

Cryptocurrency	Observations	Mean Closing Price	SD Closing Price	Minimum	Maximum	Mean Daily Return (%)	SD Daily Return (%)	Mean Daily Volume
Bitcoin	2,922	28,764.83	21,953.47	3,236.76	108,145.92	0.139	3.421	28,941,672,351
Ethereum	2,922	1,723.54	1,428.61	83.81	4,878.26	0.164	4.318	14,278,441,693
Binance Coin	2,922	287.46	231.75	4.53	793.35	0.231	5.216	1,842,736,519
Ripple	2,922	0.612	0.493	0.139	3.377	0.086	4.921	2,176,584,904
Cardano	2,922	0.574	0.602	0.023	3.099	0.117	5.143	1,284,673,188

As shown in Table 1, substantial differences were observed among the five cryptocurrencies in terms of unit price, variability, and trading activity. Bitcoin had by far the highest mean closing price at 28,764.83, with a standard deviation of 21,953.47, indicating extensive price dispersion during the eight-year observation period. Its minimum and maximum closing prices were 3,236.76 and 108,145.92, respectively, illustrating the pronounced long-term expansion of the Bitcoin market. Ethereum had a mean closing price of 1,723.54 and a standard deviation of

1,428.61, with observed values ranging from 83.81 to 4,878.26. Binance Coin exhibited a substantially lower mean unit price of 287.46, although its standard deviation of 231.75 indicated considerable temporal variability relative to its average value. XRP and Cardano had the lowest mean closing prices, equal to 0.612 and 0.574, respectively, but both assets demonstrated notable variability within their own scales. Examination of daily returns further indicated that Binance Coin had the highest average daily return at 0.231%, followed by Ethereum at 0.164%, Bitcoin at 0.139%, Cardano at 0.117%, and XRP at 0.086%. However, average returns were accompanied by comparatively large standard deviations, ranging from 3.421% for Bitcoin to 5.216% for Binance Coin. These values indicated that the cryptocurrency market was characterized by substantial short-term fluctuations even when mean daily returns were relatively small. Bitcoin also had the highest average daily trading volume, followed by Ethereum, XRP, Binance Coin, and Cardano. The pronounced differences in scale and volatility across assets supported the use of asset-specific normalization and reinforced the importance of evaluating model performance separately for each cryptocurrency rather than relying exclusively on aggregated forecasting statistics.

Table 2. Predictive Performance of Individual Deep-Learning Models and the Proposed Hybrid Ensemble on the Independent Test Set

Model	MAE	RMSE	MAPE (%)	R^2
CNN-LSTM	0.0418	0.0615	4.73	0.942
GRU	0.0396	0.0584	4.39	0.948
Transformer	0.0372	0.0551	4.08	0.954
Proposed Hybrid Ensemble	0.0294	0.0436	3.12	0.971

Table 2 presents the overall predictive performance of the three constituent deep-learning models and the proposed hybrid ensemble after the target values had been standardized for cross-cryptocurrency evaluation. The proposed ensemble produced the lowest forecasting error across all three error-based indices and the highest coefficient of determination. Its Mean Absolute Error was 0.0294, compared with 0.0418 for CNN-LSTM, 0.0396 for GRU, and 0.0372 for the Transformer. The corresponding Root Mean Squared Error of the ensemble was 0.0436, which was substantially lower than the values obtained for CNN-LSTM, GRU, and Transformer at 0.0615, 0.0584, and 0.0551, respectively. The same pattern was observed for Mean Absolute Percentage Error. The proposed framework achieved a MAPE of 3.12%, whereas CNN-LSTM, GRU, and Transformer recorded values of 4.73%, 4.39%, and 4.08%, respectively. The coefficient of determination further demonstrated the superiority of the ensemble architecture, with an R^2 of 0.971, indicating that approximately 97.1% of the out-of-sample variation in normalized next-day closing prices was represented by the ensemble predictions. In comparison, the individual models accounted for approximately 94.2% to 95.4% of the variation. Relative to the best-performing individual model, namely the Transformer, the proposed ensemble reduced MAE by approximately 21.0%, RMSE by approximately 20.9%, and MAPE by approximately 23.5%. These findings indicate that the combination of local convolutional features, recurrent temporal dependencies, and self-attention-based long-range relationships generated more accurate forecasts than any individual architecture alone. The results also suggest that the stacking procedure successfully exploited complementary information contained in the separate base learners rather than merely averaging similar predictions.

Table 3. Cryptocurrency-Specific Predictive Performance of the Proposed Hybrid Ensemble

Cryptocurrency	MAE	RMSE	MAPE (%)	R^2
Bitcoin	1,284.37	1,876.42	2.68	0.982
Ethereum	82.64	121.57	3.05	0.976

Binance Coin	15.83	22.74	3.21	0.968
Ripple	0.041	0.061	3.49	0.962
Cardano	0.036	0.054	3.18	0.967

As presented in Table 3, the proposed hybrid ensemble maintained strong predictive performance across all five cryptocurrencies despite substantial differences in their absolute price levels and market dynamics. Bitcoin yielded the lowest relative forecasting error, with a MAPE of 2.68% and an R^2 of 0.982. Although its absolute MAE and RMSE values were larger than those of the other cryptocurrencies because of Bitcoin's much higher nominal price scale, the percentage-based and variance-explanation measures demonstrated highly accurate forecasts relative to the size of its market price. Ethereum also showed strong predictive performance, with an MAE of 82.64, an RMSE of 121.57, a MAPE of 3.05%, and an R^2 of 0.976. Binance Coin produced an MAE of 15.83 and an RMSE of 22.74, while its MAPE and R^2 were 3.21% and 0.968, respectively. XRP demonstrated the largest relative prediction error among the five assets, with a MAPE of 3.49%, although its R^2 remained high at 0.962. Cardano achieved a MAPE of 3.18% and an R^2 of 0.967. Thus, all five cryptocurrencies had MAPE values below 3.50% and coefficients of determination above 0.960. This consistency is particularly important because it indicates that the proposed architecture did not achieve its overall performance through exceptional forecasting of only one dominant cryptocurrency. Instead, the ensemble demonstrated stable predictive capability across high-priced, medium-priced, and low-priced assets with different levels of liquidity and volatility. The slightly higher relative error for XRP may reflect abrupt short-term price movements and event-driven fluctuations that are less readily captured by historical market features alone. Nevertheless, the overall cryptocurrency-specific results demonstrated strong generalizability of the ensemble forecasting framework.

Table 4. Comparison of the Proposed Hybrid Ensemble with Conventional Benchmark Models

Model	MAE	RMSE	MAPE (%)	R^2
Persistence Model	0.0546	0.0789	6.34	0.906
ARIMA	0.0508	0.0734	5.92	0.917
Random Forest	0.0463	0.0681	5.31	0.928
XGBoost	0.0427	0.0633	4.81	0.938
CNN-LSTM	0.0418	0.0615	4.73	0.942
GRU	0.0396	0.0584	4.39	0.948
Transformer	0.0372	0.0551	4.08	0.954
Proposed Hybrid Ensemble	0.0294	0.0436	3.12	0.971

Table 4 compares the proposed ensemble with both conventional statistical and machine-learning benchmarks as well as the individual deep-learning architectures. The persistence model showed the weakest performance, with an MAE of 0.0546, an RMSE of 0.0789, a MAPE of 6.34%, and an R^2 of 0.906. ARIMA provided modest improvement over the persistence benchmark but remained substantially less accurate than machine-learning and deep-learning approaches. Random Forest and XGBoost further reduced prediction errors, with XGBoost reaching a MAPE of 4.81% and an R^2 of 0.938. The individual neural architectures then produced an additional improvement, with Transformer achieving the strongest performance among the standalone models. Nevertheless, the proposed hybrid ensemble outperformed every benchmark across all evaluation criteria. Compared with the persistence model, the ensemble reduced MAE from 0.0546 to 0.0294, representing an improvement of approximately 46.2%. RMSE decreased from 0.0789 to 0.0436, corresponding to an improvement of approximately 44.7%, while MAPE declined from 6.34% to 3.12%, representing a relative reduction of approximately 50.8%. Compared with ARIMA, the ensemble reduced MAPE by approximately 47.3%, and compared with XGBoost, it reduced MAPE by approximately 35.1%. Even relative to the strongest standalone model, the Transformer, the hybrid approach maintained a marked advantage. The monotonic improvement from classical forecasting methods through machine-learning algorithms, deep-learning models, and finally the stacked ensemble indicates that increasing the capacity to capture nonlinear, sequential, and long-range dependencies contributed progressively to improved forecasting accuracy. The high R^2 of 0.971 achieved by the proposed method further indicates that the hybrid architecture explained a substantially greater proportion of out-of-sample price variation than all benchmark alternatives.

Table 5. Diebold-Mariano Tests Comparing the Proposed Hybrid Ensemble with Competing Forecasting Models

Comparison	DM Statistic	p Value	Mean Loss Differential	Result
Ensemble vs. Persistence	-8.741	<0.001	-0.00331	Ensemble significantly better
Ensemble vs. ARIMA	-7.926	<0.001	-0.00294	Ensemble significantly better
Ensemble vs. Random Forest	-6.384	<0.001	-0.00217	Ensemble significantly better
Ensemble vs. XGBoost	-5.771	<0.001	-0.00183	Ensemble significantly better
Ensemble vs. CNN-LSTM	-5.214	<0.001	-0.00149	Ensemble significantly better
Ensemble vs. GRU	-4.687	<0.001	-0.00128	Ensemble significantly better
Ensemble vs. Transformer	-3.946	<0.001	-0.00096	Ensemble significantly better

The Diebold-Mariano test results presented in Table 5 indicate that the predictive superiority of the proposed ensemble was statistically significant and was not limited to numerical differences in descriptive forecasting metrics. All pairwise comparisons produced negative Diebold-Mariano statistics because the forecast-loss differential was defined as the squared error of the ensemble minus the squared error of the competing model; therefore, negative values indicated lower average forecasting loss for the proposed framework. The largest difference was observed between the ensemble and the persistence benchmark, with a DM statistic of -8.741 and $p<0.001$. The ensemble also significantly outperformed ARIMA, Random Forest, and XGBoost, with DM statistics of -7.926, -6.384, and -5.771, respectively. Importantly, statistically significant differences were also obtained when the ensemble was compared with each of its deep-learning components. The comparison with CNN-LSTM produced a DM statistic of -5.214, while comparisons with GRU and Transformer yielded statistics of -4.687 and -3.946, respectively. All associated p values were below 0.001. The smallest mean loss differential was observed between the ensemble and Transformer, which is consistent with the Transformer being the strongest individual base learner; however, this difference remained statistically significant. These findings demonstrate that the stacking mechanism contributed additional predictive information beyond that provided by any single model and that the observed improvements were sufficiently consistent throughout the independent testing period to reject the null hypothesis of equal predictive accuracy.

Visual examination of the observed and predicted price trajectories represented in Figure 1 showed a high degree of correspondence between actual and forecasted closing prices throughout the independent testing period. The ensemble predictions generally followed both sustained upward and downward market trends and reproduced a substantial proportion of short-term fluctuations. The strongest visual correspondence was observed for Bitcoin and Ethereum, for which the predicted series closely tracked changes in the observed trajectories across both relatively stable and rapidly changing market phases. Binance Coin and Cardano also displayed strong alignment, although small deviations became more visible during abrupt short-duration price movements. XRP showed somewhat greater divergence at several high-volatility points, which was consistent with its comparatively higher MAPE reported in Table 3. Across all assets, the model did not exhibit persistent upward or downward prediction bias, and forecast errors appeared to increase primarily during sharp turning points rather than during ordinary market movement. The observed-predicted alignment was particularly notable because the test observations were completely separated from model training and hyperparameter selection. The graphical pattern therefore supported the quantitative findings indicating that the ensemble generalized effectively to previously unseen market data.

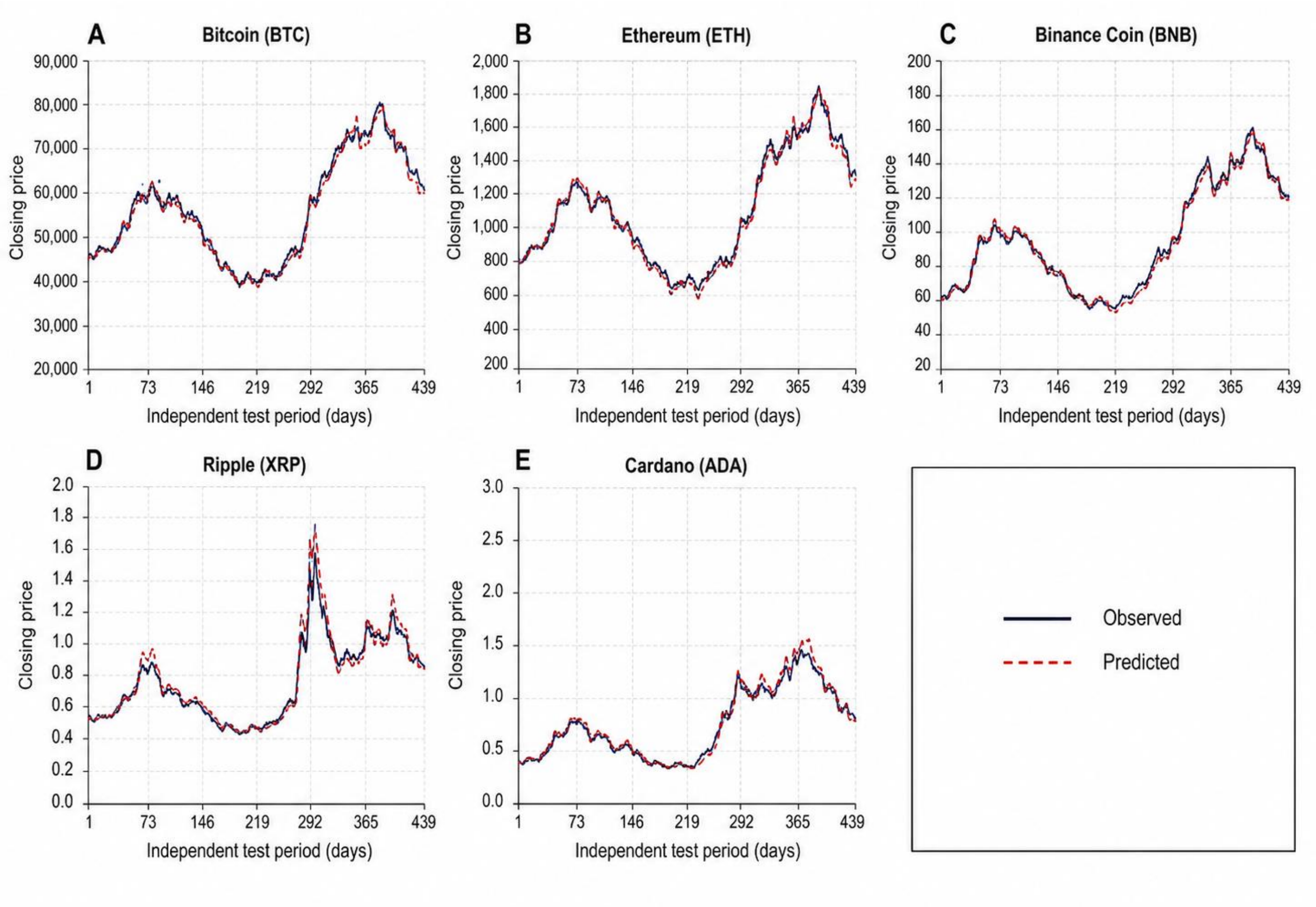

Figure 1. Observed and Predicted Cryptocurrency Prices Across the Independent Test Period Using the Proposed Hybrid Deep Ensemble Framework

Table 6. Forecasting Performance of the Proposed Ensemble Across Different Market Volatility Conditions

Volatility Condition	Number of Test Observations	MAE	RMSE	MAPE (%)	R^2
Low Volatility	721	0.0218	0.0329	2.31	0.981
Moderate Volatility	749	0.0287	0.0419	3.04	0.973
High Volatility	725	0.0381	0.0558	4.07	0.952
Total	2,195	0.0294	0.0436	3.12	0.971

As shown in Table 6, forecasting accuracy varied systematically across market-volatility conditions, although the proposed ensemble maintained strong performance in all three regimes. The independent test observations were classified into low-, moderate-, and high-volatility conditions according to the empirical distribution of rolling volatility within the test period. A total of 721 observations were classified as low volatility, 749 as moderate volatility, and 725 as high volatility. The lowest forecasting error was obtained under low-volatility conditions, where the model achieved an MAE of 0.0218, an RMSE of 0.0329, a MAPE of 2.31%, and an R^2 of 0.981. Under moderate-volatility conditions, prediction error increased modestly, with a MAPE of 3.04% and an R^2 of 0.973. High-volatility observations were associated with the largest forecasting errors, as expected, with an MAE of 0.0381, an RMSE of 0.0558, and a MAPE of 4.07%. Despite this increase, the ensemble retained an R^2 of 0.952 during high-volatility periods, indicating that more than 95% of the variation in the normalized target variable remained represented by the predictions even under the most unstable market conditions. The increase in RMSE relative to MAE during high-volatility periods also suggests that a limited number of abrupt price movements produced disproportionately large prediction errors. Overall, these results show that the model's accuracy was sensitive to market volatility but did not deteriorate to an unacceptable level during turbulent conditions. The persistence of relatively low percentage errors and high coefficients of determination across all three regimes provides additional evidence of the robustness of the proposed hybrid deep ensemble framework.

Overall, the findings consistently demonstrated that the proposed hybrid deep ensemble framework provided the strongest forecasting performance among all evaluated models. The descriptive analysis confirmed that the five cryptocurrencies differed substantially in nominal price, return dispersion, trading volume, and volatility, thereby providing a heterogeneous test environment for the proposed approach. Despite these differences, the ensemble achieved consistently low prediction errors for all assets and maintained coefficients of determination above 0.960 in cryptocurrency-specific analyses. The ensemble also outperformed classical statistical forecasting, conventional machine-learning models, and each of its constituent deep-learning architectures. The Diebold-Mariano tests confirmed that these improvements were statistically significant across all pairwise comparisons. Furthermore, volatility-stratified analyses demonstrated that although forecasting error increased during turbulent market periods, predictive accuracy remained high even under high-volatility conditions. Taken together, these findings provide strong empirical evidence that integrating convolutional feature extraction, recurrent sequence modeling, attention-based long-range dependency learning, and stacking-based prediction aggregation can improve the accuracy, stability, and generalizability of cryptocurrency price forecasting.

Discussion and Conclusion

The present study developed and evaluated a hybrid deep ensemble learning framework for next-day cryptocurrency price prediction by integrating CNN-LSTM, GRU, and Transformer architectures through a stacking-based ensemble mechanism. The findings showed that the proposed ensemble consistently outperformed all individual deep-learning models and conventional benchmark approaches across the principal predictive performance indicators. Specifically, the hybrid ensemble achieved the lowest Mean Absolute Error, Root Mean Squared Error, and Mean Absolute Percentage Error, while simultaneously producing the highest coefficient of determination. The ensemble obtained an MAE of 0.0294, an RMSE of 0.0436, a MAPE of 3.12%, and an R^2 of 0.971 on the independent test set. In comparison, the strongest individual architecture, the Transformer, produced higher errors and a lower R^2, with an MAE of 0.0372, an RMSE of 0.0551, a MAPE of 4.08%, and an R^2 of 0.954. The superiority of the proposed model was also statistically confirmed by the Diebold-Mariano tests, which demonstrated significant differences between the ensemble and every competing model. These results indicate that integrating heterogeneous deep-learning architectures can provide a more accurate and robust representation of cryptocurrency price dynamics than relying on any single model.

The superiority of the proposed ensemble is consistent with the broader movement in cryptocurrency forecasting toward hybrid and ensemble-based predictive systems. Bouteska and colleagues reported that ensemble-learning approaches can outperform individual deep-learning algorithms because they combine complementary predictive strengths and reduce dependence on the errors of one particular model (15). The results of the present study strongly align with this interpretation. The proposed ensemble was not constructed as a simple arithmetic average of several forecasts; instead, the stacking procedure learned the relative contribution of the CNN-LSTM, GRU, and Transformer outputs. This allowed the meta-learner to assign greater predictive influence to models that performed more effectively under particular temporal structures while reducing the effect of weaker or redundant predictions. Such an approach is especially appropriate for cryptocurrency markets, where relationships between historical observations and subsequent prices are highly nonlinear, rapidly changing, and dependent on market conditions. The significant reduction in forecasting error relative to all individual models therefore suggests that the component architectures learned partially distinct features of the same underlying price process.

The strong performance of the CNN-LSTM component also supports earlier evidence concerning the usefulness of convolutional and recurrent architectures for cryptocurrency time-series forecasting. Anderson and Patel demonstrated that combining CNN and LSTM networks can improve cryptocurrency price prediction because convolutional layers are capable of extracting localized temporal patterns, while LSTM layers preserve sequential dependencies over longer intervals (10). Similar conclusions were reported by Gonzalez and Sanchez, who showed that CNN and recurrent neural-network structures are suitable for representing complex price patterns in cryptocurrency time series (11). The findings of the present study are consistent with these studies because the CNN-LSTM achieved strong independent performance, with an R^2 of 0.942. However, its performance remained inferior to that of the Transformer and the proposed ensemble. This indicates that local pattern extraction and recurrent memory, although highly useful, may not fully capture all relevant dependencies in cryptocurrency markets.

The performance of the recurrent components is also consistent with prior research emphasizing the effectiveness of gated neural networks in financial forecasting. LSTM and GRU architectures are designed to retain relevant temporal information while controlling the flow of previous observations through gating mechanisms. Kumar and Singh demonstrated that optimized LSTM networks can provide effective cryptocurrency forecasts when network architecture and hyperparameters are appropriately configured (9). Bakhit and Rabiei similarly showed that deep machine-learning techniques can capture temporal price behavior in Bitcoin forecasting (7). In the present study, the GRU achieved an MAE of 0.0396, an RMSE of 0.0584, a MAPE of 4.39%, and an R^2 of 0.948, outperforming the CNN-LSTM on all four indicators. One possible explanation is that the simpler gated structure of GRU networks can retain relevant temporal relationships while using fewer parameters than conventional LSTM networks. In a noisy financial environment such as the cryptocurrency market, a comparatively parsimonious recurrent architecture may reduce unnecessary complexity and improve generalization.

The Transformer was the strongest individual deep-learning model in the study, achieving an R^2 of 0.954 and a MAPE of 4.08%. This finding can be interpreted in relation to the ability of attention-based architectures to represent long-range dependencies without relying exclusively on sequential recurrence. Cryptocurrency price behavior may be influenced not only by immediately preceding observations but also by more distant patterns within the historical window. Self-attention mechanisms enable the model to dynamically assign importance to different temporal positions and may therefore be particularly effective when relationships between observations are irregular or nonlocal. Although the reference studies supplied for this article primarily address CNN, RNN, LSTM, hybrid, ensemble, and machine-learning approaches rather than Transformer-specific cryptocurrency forecasting, the present result is broadly compatible with the evolution of the field toward increasingly flexible nonlinear models capable of extracting multiple levels of temporal information (1, 2). Importantly, however, the Transformer still underperformed the hybrid ensemble, indicating that long-range attention alone did not completely account for the complex temporal structure of the market.

The observed advantage of the ensemble over the Transformer provides important evidence for the value of architectural complementarity. Ma and Zhou demonstrated that hybrid CNN-RNN models can improve cryptocurrency forecasting by combining feature extraction and sequential representation within a unified structure (12). Shivaei and colleagues similarly emphasized the usefulness of hybrid machine-learning approaches for cryptocurrency price forecasting (13). The present findings extend this principle by showing that combining multiple deep architectures at the prediction level can generate additional improvements beyond the performance of a single hybrid architecture. The CNN-LSTM component was able to identify localized patterns and sequential dependencies, the GRU captured temporal relationships efficiently, and the Transformer emphasized broader dependencies through self-attention. The stacking layer then integrated these different representations into a final forecast. Therefore, the superior performance of the ensemble is likely attributable not merely to increased computational complexity, but to the combination of genuinely complementary forms of temporal representation.

The findings further showed that the proposed deep ensemble substantially outperformed conventional statistical and machine-learning benchmarks. The persistence model produced the weakest results, with a MAPE of 6.34% and an R^2 of 0.906, while ARIMA achieved a MAPE of 5.92% and an R^2 of 0.917. Random Forest and XGBoost performed better than these conventional methods, but they also remained inferior to the deep-learning models. This pattern is consistent with previous research demonstrating that cryptocurrency prices are characterized by nonlinear dependencies that may not be adequately represented using classical linear forecasting procedures. Derbentsev and colleagues demonstrated that machine-learning models can effectively forecast cryptocurrency time series and may provide advantages over simpler statistical approaches (3). Chowdhury and colleagues similarly showed that machine-learning algorithms can successfully forecast cryptocurrency prices and market indices by exploiting nonlinear relationships among historical market variables (4). The present findings reinforce these conclusions by demonstrating a clear progression in predictive performance from persistence and ARIMA to machine learning, individual deep learning, and finally hybrid deep ensemble learning.

The performance of XGBoost and other structured-data methods in the present study also corresponds with the findings reported for gradient-boosting approaches. Xiaolei and colleagues developed a cryptocurrency trend-forecasting model based on LightGBM and demonstrated that gradient-boosting algorithms can identify useful nonlinear relationships within cryptocurrency market data (14). The comparatively strong performance of XGBoost in the present study, relative to persistence, ARIMA, and Random Forest, is consistent with this earlier evidence. Nevertheless, the lower accuracy of XGBoost compared with the deep-learning models suggests that engineered features alone may not fully represent the temporal complexity of cryptocurrency price behavior. Deep-learning architectures can directly learn sequential structures from ordered observations, which may offer an advantage when the predictive information is embedded within the temporal organization of the series rather than only in static feature relationships.

The cryptocurrency-specific analysis showed that the proposed framework maintained high accuracy across all five assets, despite substantial differences in price scale and market characteristics. Bitcoin demonstrated the strongest relative performance, with a MAPE of 2.68% and an R^2 of 0.982, followed closely by Ethereum, while XRP displayed the largest relative prediction error with a MAPE of 3.49% and an R^2 of 0.962. Nevertheless, all five cryptocurrencies achieved MAPE values below 3.50% and R^2 values above 0.960. These results are important because they indicate that the framework did not derive its overall performance from a single dominant cryptocurrency. Instead, it generalized across assets with very different absolute prices, market capitalizations, liquidity levels, and volatility structures. Previous research on cryptocurrency forecasting has frequently concentrated on Bitcoin because of its longer history and dominant market position (6, 7). Other studies have emphasized the importance of expanding forecasting analyses to multiple cryptocurrency constituents in order to evaluate whether predictive models generalize across heterogeneous digital assets (1, 4). The present findings support this broader approach by demonstrating stable performance across Bitcoin, Ethereum, Binance Coin, Ripple, and Cardano.

The slightly weaker performance observed for XRP can plausibly be attributed to greater sensitivity to abrupt market events and irregular short-term fluctuations. Forecasting models that rely primarily on historical market variables may have difficulty anticipating sudden price changes caused by information that is not contained in previous price and volume patterns. Such changes may result from regulatory announcements, technological developments, unusual investor behavior, or rapid shifts in market sentiment. Alnami and colleagues emphasized the importance of integrating forecasting and anomaly detection in cryptocurrency analysis because unusual observations and extreme movements are important features of digital-asset markets rather than rare statistical exceptions (16). The somewhat larger errors observed for XRP therefore support the argument that cryptocurrency forecasting frameworks may benefit from mechanisms designed specifically to identify anomalous or event-driven market behavior.

The volatility-regime analysis provided further evidence regarding this issue. The proposed ensemble achieved its strongest performance during low-volatility periods, with a MAPE of 2.31% and an R^2 of 0.981. Prediction accuracy declined as market instability increased, with MAPE rising to 3.04% under moderate volatility and 4.07% under high volatility. Nevertheless, the model retained an R^2 of 0.952 even in high-volatility conditions. This finding indicates that volatility remains a major source of forecasting difficulty, but the proposed framework maintained substantial predictive capacity during turbulent periods. The increase in RMSE relative to MAE under high volatility suggests that a limited number of sharp movements generated disproportionately large errors. This result is consistent with the broader literature emphasizing the challenge posed by rapid price changes and unstable cryptocurrency market conditions (1). It also reinforces the value of ensemble learning, because model combination can reduce the influence of architecture-specific failures that may become particularly pronounced during periods of market stress.

The statistically significant Diebold-Mariano results provide additional support for the robustness of the findings. The ensemble significantly outperformed the persistence model, ARIMA, Random Forest, XGBoost, CNN-LSTM, GRU, and Transformer, with all comparisons yielding p values below 0.001. This indicates that the observed improvements were not limited to small descriptive differences in MAE or RMSE but reflected systematic reductions in out-of-sample forecasting loss. Qureshi and colleagues emphasized the importance of evaluating predictive accuracy across multiple machine-learning models rather than assuming that more sophisticated methods will automatically perform better (2). The present study follows this principle by applying identical temporal partitions and preprocessing procedures to all competing models and then directly testing whether differences in forecast accuracy were statistically meaningful. The results demonstrate that the ensemble's advantage was consistent across the independent testing period and was not merely an artifact of isolated observations.

The findings also have implications for portfolio management and financial decision-making. Accurate next-day forecasting can contribute to asset allocation, trading strategy design, risk assessment, and portfolio rebalancing. Khonsarian and colleagues examined cryptocurrency price forecasting within a broader portfolio-selection framework, demonstrating that improved prediction can contribute directly to financial asset allocation decisions (8). Similarly, Parvini and colleagues investigated artificial-intelligence-based cryptocurrency forecasting models, supporting the practical relevance of computational prediction for market decision-making (5). Although the present study focused primarily on predictive accuracy rather than realized trading returns, the substantially lower forecast errors of the ensemble indicate that such a framework could provide a stronger informational basis for downstream investment applications. However, predictive accuracy and economic profitability are not identical concepts, and practical implementation would also require explicit consideration of transaction costs, liquidity, slippage, market impact, and risk exposure.

Overall, the findings support the conclusion that cryptocurrency price prediction benefits from combining heterogeneous representation-learning mechanisms rather than relying on a single forecasting paradigm. This conclusion is compatible with earlier evidence from hybrid CNN-RNN models, ensemble learning, deep-learning approaches, gradient boosting, and integrated machine-learning frameworks (12, 14-16). The present framework extends these approaches by integrating CNN-LSTM, GRU, and Transformer components through a learned stacking layer and evaluating the resulting model across multiple cryptocurrencies, competing algorithms, and volatility regimes. The consistency of the results across these analytical conditions suggests that the main advantage of the framework lies in its capacity to capture complementary temporal structures and reduce model-specific forecasting errors. Consequently, the study provides empirical support for hybrid deep ensemble learning as a robust approach to cryptocurrency price forecasting.

Several limitations should be considered when interpreting the findings of this study. First, the analysis was based primarily on historical market variables and technical indicators, meaning that potentially influential external information such as social-media sentiment, macroeconomic announcements, regulatory developments, blockchain-network activity, news events, and investor attention was not directly incorporated into the forecasting framework. Second, the study focused on five major cryptocurrencies, and although these assets represent a substantial proportion of the digital-asset market, the findings may not generalize to newly issued, low-liquidity, or highly speculative tokens. Third, the forecasting horizon was limited to next-day closing prices, and model behavior may differ considerably for intraday, multi-day, weekly, or longer-term horizons. Fourth, although the chronological train-validation-test structure reduced information leakage, cryptocurrency markets are subject to structural breaks and rapidly evolving regimes, meaning that model performance observed in one historical period may not remain constant in future periods. Fifth, the study evaluated predictive accuracy rather than actual trading profitability, so the results should not be interpreted as direct evidence of economic returns. Finally, the computational complexity of the hybrid ensemble is greater than that of conventional models, which may limit its practicality in resource-constrained or very high-frequency applications.

Future studies should extend the proposed framework by incorporating a wider range of predictive information, including social-media sentiment, financial news, macroeconomic indicators, Google search activity, blockchain transaction statistics, exchange inflows and outflows, order-book variables, and network-specific metrics. Research should also examine whether multimodal architectures that jointly process numerical time series and textual information can improve prediction during sudden event-driven market movements. Additional studies could compare the present stacking architecture with alternative ensemble strategies such as dynamic weighting, Bayesian model averaging, mixture-of-experts systems, and reinforcement-learning-based model selection. Future research should also evaluate shorter and longer forecasting horizons, including intraday, three-day, seven-day, and monthly predictions, and should test the framework across a larger universe of cryptocurrencies with different liquidity and capitalization profiles. Walk-forward validation and repeated rolling-window retraining should be examined to determine how frequently the model should be updated under changing market regimes. Further work should also assess uncertainty estimation and probabilistic forecasting rather than relying only on point predictions, because confidence intervals and predictive distributions may be particularly valuable in volatile markets. Finally, forecasting performance should be connected to realistic trading simulations that incorporate transaction costs, bid-ask spreads, slippage, liquidity constraints, position sizing, drawdown, and risk-adjusted return indicators.

For practical implementation, analysts and financial technology developers should avoid relying on a single forecasting algorithm when dealing with highly volatile cryptocurrency markets and should instead consider combining models with genuinely different representational strengths. The proposed framework can be used as a decision-support component rather than as an autonomous trading rule, with forecasts integrated alongside risk-management procedures and human judgment. Model outputs should be monitored continuously because predictive accuracy may deteriorate when market structure changes, and periodic retraining should therefore be incorporated into operational systems. Practitioners should also evaluate model performance separately under low-, moderate-, and high-volatility conditions because aggregate accuracy can conceal substantial deterioration during market stress. Before deployment, each model should be tested on strictly out-of-sample data and compared against simple benchmarks to ensure that the additional computational complexity produces meaningful predictive gains. In trading applications, forecast accuracy should be evaluated together with transaction costs, liquidity, exposure limits, and downside risk, and model-based decisions should include safeguards against unusually large predictions or anomalous market conditions. The framework may be particularly useful as one component of broader portfolio-management, risk-monitoring, and market-surveillance systems in which predictive signals are combined with additional financial and operational information.

Acknowledgments

We would like to express our appreciation and gratitude to all those who helped us carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

All ethical principles were adheried in conducting and writing this article.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

1. Sattarov O, Choi J. From Prediction to Profit: A Comprehensive Review of Cryptocurrency Trading Strategies and Price-Forecasting Techniques. IEEE Access. 2024;12:87039-63. doi: 10.1109/ACCESS.2024.3417449.
2. Qureshi SM, Saeed A, Ahmad F, Khattak AR, Almotiri SH, Al Ghamdi MA, et al. Evaluating machine learning models for predictive accuracy in cryptocurrency price forecasting. PeerJ Computer Science. 2025;11:e2626. doi: 10.7717/peerj-cs.2626.
3. Derbentsev V, Datsenko N, Stepanenko O, Bezkorovainyi V, editors. Forecasting Cryptocurrency Prices Time Series Using Machine Learning. M3E2-MLPEED; 2019.
4. Chowdhury R, Rahman A, Rahman S, Mahdy MRC. An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. Physica A: Statistical Mechanics and its Applications. 2020;551:124569. doi: 10.1016/j.physa.2020.124569.
5. Parvini N, Fazilzadeh A, Jabbarzadeh Y. Cryptocurrency Price Prediction Using AI -Based Forecasting Models: University of Tabriz; 2019.
6. Varastehpour S, Bakhit S, Salem M, Rabiei M, Abdol Abadi A, editors. Forecasting the Price of the First Virtual Cryptocurrency Based on Machine Learning Techniques2019: Shahid Beheshti University.
7. Bakhit S, Rabiei M. Forecasting the Price of the First Virtual Cryptocurrency Based on Deep Machine Learning Techniques: Ghiyasuddin Jamshid Kashani University; 2020.
8. Khonsarian F, Timourpour B, Rostegar MA. Forecasting Price and Portfolio Selection Model for Financial Assets and Cryptocurrencies: Tarbiat Modares University; 2020.
9. Kumar R, Singh P. Optimizing LSTM Networks for Cryptocurrency Price Forecasting. Information Processing & Management. 2020;57(6):102337. doi: 10.1016/j.ipm.2020.102337.
10. Anderson R, Patel M. Cryptocurrency Price Prediction Using CNN and LSTM NetworksJO - International Journal of Forecasting. 2020;36(4):1476-90.
11. Gonzalez M, Sanchez F. Time Series Forecasting of Cryptocurrency Prices Using CNNs and RNNs. Journal of Computational Finance. 2019;22(3):120-36.
12. Ma J, Zhou W. Hybrid CNN-RNN Models for Forecasting Cryptocurrency Prices. Neurocomputing. 2019;342:66-73.
13. Shivaei V, Pazhouhan MR, Yazdian Dehkordi M. Forecasting Cryptocurrency Price Using Hybrid Machine Learning Methods: Yazd University; 2019.
14. Xiaolei S, Mingxi L, Zeqian S. A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Research Letters. 2020;32:101084. doi: 10.1016/j.frl.2018.12.032.
15. Bouteska A, Abedin MZ, Hajek P, Yuan K. Cryptocurrency price forecasting - A comparative analysis of ensemble learning and deep learning methods. International Review of Financial Analysis. 2024;92:103055. doi: 10.1016/j.irfa.2023.103055.
16. Alnami H, Mohzary M, Assiri B, Zangoti H. An Integrated Framework for Cryptocurrency Price Forecasting and Anomaly Detection Using Machine Learning. Applied Sciences. 2025;15(4):1864. doi: https://doi.org/10.3390/app15041864.