International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

An Investigation of The Contribution of Attention-Based Hybrid Deep Learning Models to Prediction Errors in Cryptocurrency Markets

Aynur İNCEKIRIK [1]
Mohammad Luqman YOUSUFİ [2]
Nihat ALTUNTEPE [3]

Received: 13.03.2026, Accepted: 15.05.2026
DOI Number: 10.5281/zenodo.20800529

Abstract

This study comparatively analyzes deep learning architectures for predicting the daily closing prices of Ethereum, Solana, and BNB cryptocurrencies. Given the high volatility, non-linear price movements, and noise-intensive data structures of cryptocurrency markets, accurate price modeling holds significant importance for both academic research and investment decisions. The models examined include LSTM, GRU, CNN, CNN-LSTM, CNN-GRU, Attention-LSTM, and Attention-GRU. The dataset covers the period from April 10, 2020, to February 12, 2026, at daily frequency, split into 70% training and 30% testing. Hyperparameter optimization was conducted across varying window lengths, epochs, batch sizes, neuron counts, and dropout rates. Model performance was evaluated using MAE, MSE, MAPE, and R^2 metrics. Results indicate that GRU-based models consistently yield lower error values and more stable performance on cryptocurrency time series. The hybrid CNN-GRU architecture produces competitive results on certain series; however, attention mechanism-integrated models generally increase error values. These findings suggest that model complexity does not guarantee superior prediction accuracy in financial time series, and that compact architectures may offer stronger generalization in highly volatile data environments. The study also addresses a methodological gap by benchmarking multiple deep learning models under identical datasets and experimental conditions.

Key words: Cryptocurrency, Deep Learning, LSTM, GRU, Attention Mechanism, Price Forecasting

JEL Codes: C45, C53, C58, G15, G17

---

[1] Assist Prof., PhD, Manisa Celal Bayar University, Türkiye, aynur.incekirik@cbu.edu.tr, https://orcid.org/0000-0002-5029-6036
[2] Graduate Student, Manisa Celal Bayar University, Türkiye, m.loqman75@gmail.com, https://orcid.org/0009-0000-8460-4415
[3] Assoc Prof. PhD, Isparta University of Applied Sciences, Türkiye, nihataltuntepe@isparta.edu.tr, https://orcid.org/0000-0002-2774-315X

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

1. Introduction

Cryptocurrency markets have rapidly grown into a market segment within the global financial system in recent years, attracting increasing investor interest. Their decentralized structures, continuously active market mechanisms, and high liquidity levels significantly distinguish the price dynamics of cryptocurrency assets from those of traditional financial assets. In particular, cryptocurrencies such as Ethereum (ETH), Solana (SOL), and Binance Coin (BNB) which are digital assets with high market capitalization are increasingly finding a place not only in individual investors' portfolios but also in those of institutional investors. However, since cryptocurrency markets are characterized by high volatility, sudden price spikes, and non-linear price movements, accurately predicting the prices of these assets presents a significant challenge for both academic research and investment strategies (Seabe et al., 2025). Factors inherent to the nature of financial markets such as uncertainty, structural breaks, and sensitivity to external shocks make price forecasting particularly challenging in cryptocurrency markets. Cryptocurrency markets are recognized as one of the most difficult market types to forecast in the financial time series literature (Baur et al., 2018). These markets exhibit volatile price movements due to factors such as their relatively short history compared to traditional financial assets, the intensity of speculative behavior, and limited market depth (Katsiampa, 2019). These characteristics frequently violate the fundamental assumptions of classical econometric models such as stationarity, linearity, and constant variance thereby increasing the need for more flexible and data-driven methods.

Traditional methods for forecasting financial time series are largely based on statistical models that rely on linear assumptions. However, the complex, nonlinear, and high-noise data structure of cryptocurrency markets can limit the forecasting performance of classical econometric models (Zhao et al., 2026). Cryptocurrency prices are shaped by the interaction of numerous factors, including market sentiment, speculative behavior, technological advancements, regulatory decisions, and global macroeconomic shocks. This multidimensional structure indicates that price movements arise within a complex and dynamic system rather than a deterministic process. Consequently, there is a need for methods that are more flexible and capable of capturing nonlinear relationships when modeling such data structures. In recent years, machine learning and, in particular, deep learning-based models have become a significant area of research in the analysis of financial time series (Gao, 2025). Deep learning architectures are widely used in financial forecasting studies due to their ability to automatically learn complex patterns and nonlinear relationships within the data. In particular, Recurrent Neural Networks (RNN) and their advanced variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, have secured a prominent place in the financial forecasting literature due to their success in modeling temporal dependencies in time series. These models have the potential to generate stronger predictions regarding future price movements by storing information learned from past observations through memory mechanisms. However, the complex nature of financial time series makes it difficult for a single model architecture to produce optimal results in every scenario. Consequently, hybrid models created by combining different deep learning architectures have been gaining increasing attention in recent years. In particular, hybrid architectures that combine Convolutional Neural Networks (CNNs) with recurrent neural networks offer significant advantages in both capturing local patterns within the data and modeling temporal dependencies (Rezaei et al., 2021).

In recent years, deep learning-based approaches have begun to be widely used in financial time series analysis due to their superior performance in capturing complex and nonlinear patterns (Goodfellow et al., 2016). In particular, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models are frequently preferred in cryptocurrency price forecasting studies due to their ability to simultaneously learn long- and short-term dependencies (McNally et al., 2018). However, the literature highlights that models based solely on recurrent neural networks have limitations in capturing short-term local patterns. To overcome this limitation, hybrid architectures combining Convolutional Neural Networks (CNNs) with recurrent neural networks have been proposed. While CNN-based structures can effectively capture short-term and local patterns in time series, recurrent structures such as LSTM and GRU based on RNNs enable the modeling of long-term dependencies (Borovykh et al., 2017). The combination of these two architectures allows for a more effective representation of the multi-layered structure in financial time series. However, many studies in the literature evaluate the performance of hybrid architectures solely in terms of prediction accuracy; they do not sufficiently discuss which information components these architectures utilize to what extent or under which market conditions they demonstrate superiority. A comparative analysis of the performance of different deep learning architectures in cryptocurrency markets is important both for understanding the impact of model complexity on prediction accuracy and for identifying model architectures better suited to the structure of financial time series. In particular, examining the impact of integrating the attention mechanism into hybrid deep learning models on prediction errors has the potential to contribute to methodological discussions in this field (Priadinata et al., 2025).

Another approach that has gained prominence in the field of deep learning in recent years is attention mechanisms. It stands out as a key component that enhances the information processing capacity of deep learning models. It aims to reduce information loss by enabling the model to place greater emphasis on time steps that are more critical for the prediction process, rather than treating all past information equally (Vaswani et al., 2017). In other words, attention mechanisms aim to make the learning process more efficient by enabling the model to focus on time steps within the historical data that are more important for prediction (Wei et al., 2026). Theoretically, this approach could help the model better distinguish meaningful signals from noisy data. However, the structure of financial time series which are highly volatile and often exhibit stochastic properties also demonstrates that model complexity does not always translate to better prediction performance.

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

Particularly in data environments with high noise, such as cryptocurrency markets, the generalization ability of overparameterized architectures may weaken, and the model may encounter overfitting issues. Therefore, the empirical evaluation of the actual contribution of attention mechanisms in financial time series has emerged as an important research topic (Ataei et al., 2025). While the use of attention mechanisms in the context of financial time series is increasing, studies systematically comparing the marginal contribution of this mechanism to LSTM and GRU architectures across cryptocurrency assets with different market structures remain limited.

The aim of this study is to compare the performance of different deep learning architectures in predicting the prices of Ethereum, Solana, and Binance Coin leading digital assets in the cryptocurrency market and to empirically examine the contribution of attention-based hybrid models to prediction errors. The primary reason for selecting Ethereum, Solana, and Binance Coin instead of Bitcoin is that these assets are large-scale altcoins representing different blockchain ecosystems and market dynamics. Thus, the study aims to comparatively evaluate the performance of deep learning models across different market structures. In this context, the study utilized basic architectures such as LSTM, GRU, and CNN, as well as hybrid models like CNN-LSTM and CNN-GRU, alongside Attention-LSTM and Attention-GRU architectures. The models' prediction performance was evaluated using error metrics such as MSE, MAE, MAPE, and R^2, and the effectiveness of different architectures in cryptocurrency price forecasting was analyzed comparatively. Thus, the study aims to contribute to the literature by presenting empirical findings regarding the relationship between model complexity and prediction accuracy in cryptocurrency markets. Additionally, the study comprehensively examines the contribution of attention-based hybrid deep learning models to prediction performance in cryptocurrency markets.

2. Literature Review

In recent years, with the digitalization of cryptocurrency markets, the application of machine learning and deep learning models has shown a significant rise in the academic literature. Cryptocurrency markets, as a digital marketplace where digital assets are bought and sold, have long been addressed in the academic literature. The studies cited in the literature are discussed below.

In their 2025 study, Aruwaji and Swanepoel present a new deep learning framework inspired by attention mechanisms that can simultaneously predict price trajectories across financial instruments. The study examines the five major cryptocurrencies Bitcoin, Ethereum, Binance Coin, Ripple, and Litecoin as well as the five leading stock indices: NASDAQ, S&P 500, Dow Jones, FTSE 100, and Nikkei 225. The model was constructed using multidimensional time-series data, historical prices, trading volumes, macroeconomic inputs, and social sentiment indicators. Experimental tests are based on daily data from 2020 to 2024. The dataset was split sequentially by date to reflect real-world forecasting conditions:

the first 80% of the data from 2020 through the end of 2023 was set aside for model training, while the remaining 20% covering all of 2024 was reserved for performance evaluation. In the cryptocurrency sector, it outperforms GRU and LSTM by achieving a MAPE of 2.0% for Ethereum and 3.0% for Bitcoin. The findings suggest that attention-based architectures hold significant potential for navigating market details in financial forecasting and provide value to asset managers, analysts, and strategic decision-makers in dynamic investment environments.

Gautam (2025) proposed a hybrid deep learning and machine learning model in his work that combines LSTM trees and XGBoost for cryptocurrency price forecasting. The LSTM component captures temporal dependencies in historical price data, while XGBoost improves the forecast by modeling nonlinear relationships using auxiliary features such as sentiment scores and macroeconomic indicators. The model was evaluated on historical datasets for Bitcoin, Ethereum, Dogecoin, and Litecoin, incorporating both local and global market data. A comparative analysis using the Mean Absolute Percentage Error (MAPE) and Min-Max Normalized Root Mean Square Error (MinMax RMSE) revealed that the LSTM+XGBoost hybrid consistently outperformed standalone models and traditional forecasting methods.

In their study, Mahdi et al. (2025) proposed a new deep learning hybrid model that integrates attention-based Transformer and Gated Recurrent Unit (GRU) architectures to improve the accuracy of cryptocurrency price forecasts. By combining the Transformer's strength in capturing long-term patterns with the GRU's ability to model short-term and sequential trends, the hybrid model offers a comprehensive approach to time series forecasting. They applied the model to predict the daily closing prices of Bitcoin and Ethereum based on historical data, including past prices, trading volumes, and the Fear and Greed Index. They evaluated the performance of the proposed model by comparing it with four different machine learning models: two non-sequential feedforward models the radial basis function network (RBFN) and the general regression neural network (GRNN) and two bidirectional sequence-based models the bidirectional long short-term memory (BiLSTM) and the bidirectional gated recurrent unit (BiGRU). The model's performance was evaluated using the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) metrics. The results demonstrate that the hybrid model consistently achieves superior accuracy and highlights its effectiveness for financial forecasting tasks. These findings provide valuable insights for improving real-time decision-making processes in cryptocurrency markets and support the increasing use of hybrid deep learning models in financial analysis.

Kumar et al. (2025) propose Attention_GRU+LSTM, a robust and adaptable deep learning model, to forecast ambient PM2.5 concentrations in their study. The hybrid model integrates gated recurrent units (GRU) and long short-term memory (LSTM) networks with attention mechanisms to effectively capture both temporal and spatial dependencies in air quality data. Experimental validation was conducted using a publicly available Beijing air pollution dataset. The model was evaluated using multiple performance metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE), and Coefficient of Determination (R^2), across short- and medium-term forecast horizons (2 days, 5 days, and 10 days). For 2-day forecasts, the proposed model achieved an RMSE of 10.735, an MAE of 6.401, and an R^2, outperforming state-of-the-art models such as LSTM, GRU, and hybrid deep learning models, whose RMSE values ranged from 12.568 to 14.107, MAE values from 8.305 to 11.003, and R^2 values from 0.335 to 0.472. The model continued to demonstrate robust performance for long-term forecasts; with RMSE values of 24.779 for 5-day and 27.649 for 10-day forecasts, it maintained its superiority over baseline and attention-based architectures. Furthermore, comparative results against models reported in previous studies confirm the consistent superiority and robustness of the proposed Attention_GRU+LSTM model.

In his study, Lee (2024) presents a comparative analysis of two advanced deep learning models with attention mechanisms (Attention-LSTM and Attention-GRU) for forecasting cryptocurrency price movements. The models combine historical OHLCV data with four technical indicators (SMA, EMA, TEMA, and MACD) to improve classification accuracy across three categories (uptrend, downtrend, and neutral). While both models incorporate attention mechanisms to focus on relevant time steps, they aim to capture market dynamics through sequential data. Experimental results show that incorporating technical indicators significantly improves model performance, with MACD providing the highest accuracy. While the Attention-GRU model demonstrates computational advantages, the Attention-LSTM model excels at capturing long-term dependencies.

Saqware and Beary (2024) applied deep learning methods to forecast and predict the adjusted closing prices of the Bitcoin (BTC-USD) and Ethereum (ETH-USD) cryptocurrency markets. Based on the root mean square error (RMSE), the hybrid CNN-LSTM model incorporating the Attention Mechanism outperformed both the CNN and LSTM models in forecasting the adjusted closing price of ETH-USD. Additionally, the traditional LSTM model also performed well in forecasting the adjusted closing price of BTC-USD. In forecasting, the hybrid CNN-LSTM model produced better results for both the adjusted closing prices of BTC-USD and ETH-USD compared to individual models. Furthermore, while the hybrid model performs well over shorter forecast horizons, it loses its predictive ability as the horizon extends. This finding plays a significant role in the analysis of future cryptocurrency markets.

In his study, Sun (2024) evaluates the performance of five advanced deep learning (DL) models using ten years of Amazon closing price data: Long Short-Term Memory (LSTM), Self-Attention, Convolutional Neural Network with Attention Mechanism-LSTM (CNN-LSTM-attention), Gated Recurrent Unit with

Attention Mechanism-LSTM (GRU-LSTM-attention), and Bidirectional Convolutional Neural Network with Attention Mechanism-BiLSTM-GRU (CNN-BiLSTM-GRU-attention). The results indicate that the CNN-BiLSTM- GRU-attention model demonstrated superior performance, achieving a root mean square error (RMSE) of 1.054589 and a coefficient of determination (R^2) of 0.970123; this demonstrates its proficiency in processing complex financial data. The significance of this study lies not only in validating the effectiveness of attention-based recurrent models in stock market forecasting but also in introducing the innovative application of the CNN-BiLSTM-GRU-Attention model in financial forecasting; this application holds broad potential for implementation.

Yousufi and İncekırık (2024) focused in their study on the fundamental components of the digital economy and the prediction of Ethereum's price movements in the cryptocurrency market using deep learning techniques. Within the scope of the study, RNN, LSTM, and GRU models were used for Ethereum price predictions, and their performance on time-series data was evaluated. The data underwent various preprocessing steps, such as scaling, sequence creation, and model training, while performance evaluation was conducted using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results demonstrated the potential and reliability of deep learning algorithms in financial market forecasting.

Albayrak and Saran (2023) compared the Autoregressive Integrated Moving Average (ARIMA) method with three different recurrent neural network models (LSTM, GRU, and attention-layered LSTM) in their study. The dataset was constructed using data from the Istanbul Stock Exchange for three different stocks (ASELS, AKBNK, and AKEN) over the 2001–2020 period. In the study, intraday forecasts were made using 28 financial indicators on Borsa Istanbul data, and the results of the four different models were compared. ARIMA, a statistical and linear model, was compared with nonlinear RNN models in time series forecasting, and it was determined that it had a higher average error rate compared to the three neural network models. Although the LSTM model produced results similar to those of the GRU, the GRU model was found to be slightly superior in terms of performance. However, the neural network model incorporating an attention mechanism did not yield better results compared to other basic neural network models. In a study conducted by Raşo and Demirci (2019), a 9-layer Support Vector Machine (SVM)-based machine learning model was used to forecast stock prices of the Borsa İstanbul 30 (BIST 30) index. The study utilized a dataset of stock prices from January 1, 2016, to April 11, 2018, and aimed to forecast five-day stock prices. When determining the model's input variables, indicator and oscillator values calculated using economic analysis methods were utilized. According to the study's results, the model's prediction performance was evaluated with mean squared error (MSE) values of 0.0322, 0.109, 0.09, 0.1069, and 0.2581, respectively, from the first to the fifth day.

In their study (Saracık and İncekırık, 2023), the authors applied LSTM and GRU techniques in Google Colab to forecast stock prices for five companies listed on the XELKT index of Borsa İstanbul. The dataset covers the period from January 2, 2013, to December 30, 2022. In analyses conducted over four different days, an effort was made to determine the day yielding the most accurate predictions. Since the model performance metrics MSE below 1 and MAPE below 5% indicated that both techniques successfully predicted stock prices, it was concluded that both methods were effective. A comparison of the two techniques revealed that the LSTM technique performed slightly better than the GRU technique.

Zhang et al. (2023) proposed a new price prediction model based on transfer learning in their study. The model was pre-trained using the GRU algorithm on mature market data and then fine-tuned with target market samples. An integrated framework was presented that includes complexity decomposition for data preprocessing, sample entropy for feature selection, and support vector regression for result post-processing. In an empirical analysis on China's new market, the model's RMSE, MAE, MAPE, and R^2 values were found to be 0.529, 0.476, 0.717%, and 0.501, respectively.

Wen and Li (2023) note that existing time series forecasting methods suffer from issues such as low accuracy when dealing with certain non-stationary multivariate time series data. In their study, they proposed a new time series forecasting model called LSTM-Attention-LSTM, aimed at addressing the shortcomings of existing methods. The model uses two LSTM models as an encoder and a decoder and incorporates an attention mechanism between the encoder and decoder. The model has two distinctive features: First, by using the attention mechanism to calculate the relationship between sequential data, it eliminates the disadvantage of the decoder not being able to obtain sufficiently long input sequences; second, it is suitable for sequential forecasting with long time steps. In this study, the proposed model was validated using various real-world datasets. The results show that the LSTM-attention-LSTM model is more accurate in forecasting than some currently dominant models. In their work, they also evaluated the effect of the attention mechanism at different time steps by varying the time step.

Shi et al. (2022) propose a CNN-LSTM and XGBoost hybrid model based on the attention mechanism for stock price prediction, citing the strong nonlinear generalization capabilities of neural networks. The study's model enhances prediction accuracy by integrating a time series model, attention-based Convolutional Neural Networks, a Long Short-Term Memory network, and an XGBoost regressor into a nonlinear relationship. Stock data is first preprocessed using ARIMA. Subsequently, a deep learning architecture developed within a pre-training-fine-tuning framework is adopted. The pre-training model is an Attention-based CNN-LSTM model based on a sequence-to-sequence framework. The model first uses convolutions to extract deep features from the original stock data and then uses Long Short-Term Memory (LSTM) networks to extract long-term time series features. Finally, the XGBoost model is used for fine-tuning. The results indicate that the hybrid model is more effective and achieves relatively high prediction accuracy; this can assist investors or institutions in making decisions and achieving their goals of increasing returns and mitigating risk.

This literature review provides a comprehensive conceptual foundation covering cryptocurrency markets, blockchain infrastructures, price forecasting models, investor sentiment, and market dynamics. By comprehensively modeling large, multi-variable, long-term datasets using modern deep learning techniques, this study aims to contribute to the literature both methodologically and practically through its extensive dataset, multi-dimensional variable structure, and hybrid forecasting models.

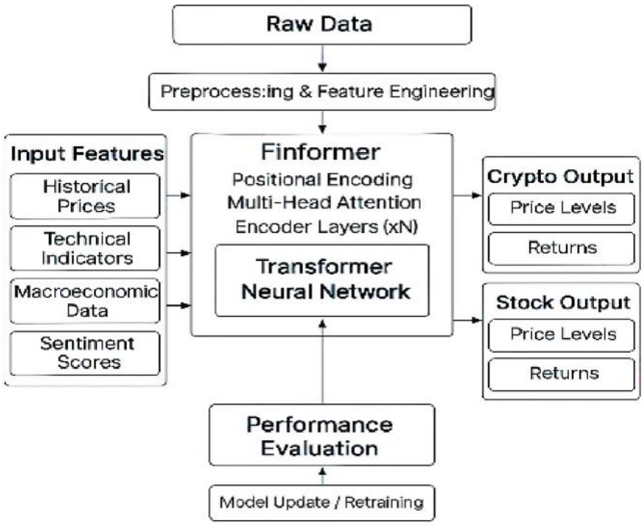
3. Deep Learning Techniques

With the rapid development of the digital economy, traditional analytical methods have proven insufficient for managing complex and volatile market conditions; this situation has heightened the strategic importance of artificial intelligence-based decision support systems. Deep learning is an advanced subfield of machine learning that can automatically learn complex and nonlinear relationships in data through multi-layered artificial neural networks (LeCun et al., 2015). Thanks to capabilities such as feature extraction, long-term dependency analysis, and complex behavior modeling, it demonstrates superior performance in image, language, and financial time series data, enabling more accurate analysis of cryptocurrency market price dynamics. These techniques are particularly used in areas such as sequential data analysis and time series forecasting.

In this context, a Recurrent Neural Network (RNN) is a type of deep learning architecture used to process sequential data. In an RNN, the connections between neurons form a directed graph. This architecture utilizes an internal state to process the input sequence. Therefore, the architecture can be successfully used for sequential tasks such as speech recognition. In an RNN, each output is determined by repeatedly applying the same operation to the examples in the sequence. The output is determined based on all previous computations. Long Short-Term Memory (LSTM) networks are another deep learning architecture based on recurrent neural networks. In the traditional RNN architecture, the problem of exploding or vanishing gradients can be observed. In RNNs, very long input sequences cannot be processed appropriately. In response, LSTM uses forget gates to overcome these issues. In the LSTM architecture, error backpropagation is allowed up to a limited number of time steps. For a typical LSTM unit, there are three types of gates in a cell: an input gate, an output gate, and a forget gate. The opening and closing operations on these gates are used to control which information should be retained and when to access that information (Zhang et al., 2018). The Gated Recurrent Unit (GRU) architecture is another deep learning architecture based on recurrent neural networks. In a typical GRU architecture, there are two gates: a reset gate and an update gate.

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

Figure 1. Attention-Based Deep Learning Model



Source: (Aruwaji & Swanepoel, 2025)

Figure 1 illustrates the conceptual framework of how deep learning performs predictive analyses by integrating an attention-based algorithm. The deep learning techniques used in this study are described below.

Gated Recurrent Unit (GRU) Model

Recurrent Neural Networks (RNNs), particularly Gated Recurrent Units (GRUs), are effective at modeling sequential data by preserving temporal dependencies; GRUs address fundamental challenges of standard RNNs, such as vanishing and exploding gradients (Zaremba et al., 2014).

In a GRU, the update gate and reset gate are the two gate units that constitute each GRU memory cell; Equations 1–4 define the operators for storing data under various conditions. These equations denote the input at time step t as$t$ and the hidden state from the previous time step as x_{t-1} . The variables z_t and r_t represent the update gate and reset gate, respectively, and control the flow of past information. The candidate hidden state is represented by $\widehat{h_t}$, and the updated hidden state at the current time step is denoted by h_t. The functions \`$t$\` and \`tanh\` denote the \`$sigmoid$\` and \`hyperbolic tangent\` activation functions, respectively. The parameters \`$W_z, W_r$\` and \`$W_h$\` are the weight matrices corresponding to the update gate, reset gate, and candidate hidden state; \`$B_z, B_r$\` and \`$B_h$\` are the associated bias vectors. The notation \`$[x_{t-1}, a_t]$\` denotes the concatenation of the previous hidden state and the current input.

$$z_t = \sigma(W_z \cdot [x_{t-1}, a_t] + B_z) \tag{1}$$

$$r_t = \sigma(W_r \cdot [x_{t-1}, a_t] + B_r) \quad (2)$$

$$\widehat{h_t} = tanh(W_h \cdot [x_{t-1}, a_t] + B_h) \quad (3)$$

$$h_t = (1 - z_t) \cdot x_{t-1} + z_t \cdot \widehat{h_t} \quad (4)$$

Long Short-Term Memory (LSTM) Model

The memory cell is the most important component of an LSTM because it possesses long-term memory. Each LSTM memory cell has three gate components: the forget gate, the input gate, and the output gate. While the input gate regulates the inflow of new information into the memory cell, the forget gate determines which data to discard. The output gate manages the flow of information from the cell. Memory cells play a vital role in enabling the LSTM to store information for extended periods and to more effectively identify sequential patterns in time-series data. All parameters used in the LSTM are defined using Equations 5 through 11:

$$i_t = \sigma(Z_i(x_{t-1} + c_t) + B_i) \quad (5)$$

$$f_t = \sigma\left(Z_f(x_{t-1} + c_t) + B_f\right) \quad (6)$$

$$o_t = \sigma(Z_o(x_{t-1} + c_t) + B_o) \quad (7)$$

$$\widehat{c_t} = tanh(Z_c(x_{t-1} + c_t) + B_c) \quad (8)$$

$$m_t = f_t * c_{t-1} + i_t * \widehat{c_t} \quad (9)$$

$$h_t = o_t * tanh(m_t) + i_t * \widehat{c_t} \quad (10)$$

$$\widehat{y_t} = h_t \quad (11)$$

The input gate is denoted by i_t, the forget gate by f_t, and the output gate by o_t. The variable t represents the current time step. The input at time$t$ is c_t, and the previous hidden state is x_{t-1}. The candidate cell state is denoted by $\widehat{c_t}$, and the updated memory cell state by m_t. The final hidden state output is h_t, and the model prediction at time t is $\widehat{y_t}$. The weight matrices associated with each gate and candidate cell state are Z_i, Z_f, Z_o and Z_c, respectively. Their corresponding error vectors are B_i, B_f, B_o and B_c. The functions $tanh$ represent the $sigmoid$ and hyperbolic tangent activation functions, respectively (Kumar et al., 2025).

Attention Mechanism

Humans typically focus on salient information. The attention mechanism is a technique used in deep learning based on the human cognitive system. $X = X_1, X_2, \cdots, X_N$ For the input $X = x_1, x_2, \cdots, x_N$, a query vector q is provided, the indices of the selected information are denoted by $z = 1, 2, \cdots, N$, and the attention distribution is then obtained.

This concept is inspired by the human visual system and has found applications in various fields, including neural machine translation (Bahdanau et al., 2014) and image processing (Xu et al., 2015). The primary goal of the training phase is to highlight specific features by applying a weighted sum method based on a context vector. The context vector "A" is defined as follows at each step "t":

$$A_t = \sum \propto_t \cdot H_t \tag{12}$$

Typically, we use a recurrent network, and \`$H_t$\` represents the model that generates hidden states and feeds them into the attention mechanism. We compute \`$\propto_t$\` from the normalized attention mechanism.

$$\propto_t = softmax(e_t) \tag{15}$$

The weights of the attention model, denoted by \`$e_t$\` and computed using a feedforward neural network, depend on the previously hidden state \`$H_{t-1}$\`.

$$e_t = \sigma(Z_a H_{t-1} + B_a) \tag{14}$$

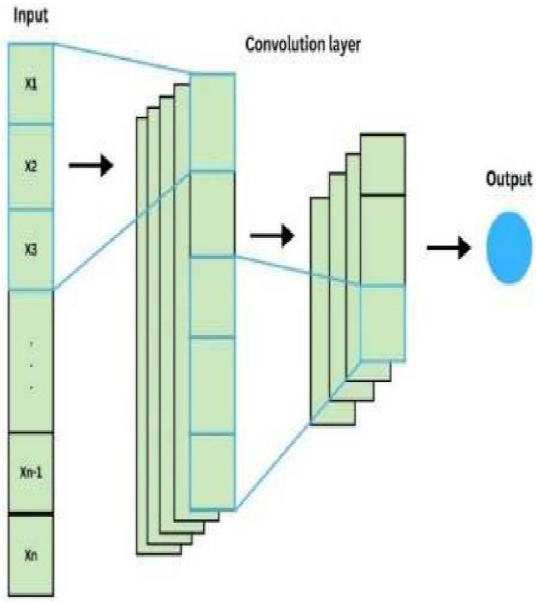
During training, we calculated the weight matrix (Z_a) and the bias vector (B_a) for the attention model. The attention model's context vector "A" dynamically represents the relevant portion of the time-series input at time t. It computes a normalized, probabilistic attention model weight that indicates the importance of a specific data point in the sequence via a weighted sum. The attention mechanism allows the model to focus on specific segments within the data sequence during the learning process by referencing its memory during the prediction phase. This allows the model to highlight critical components in the feature space, improve its ability to recognize important patterns, and make accurate predictions (Luong et al., 2015). The neural network layer uses only one layer for aggregate attention. A feedforward network with a nonlinear function specifically the hyperbolic tangent $(tanh)$ as the default activation function forms this layer. The additive attention mechanism enhances the network's ability to focus on relevant information and recognize important patterns; this allows the network to dynamically assign importance or weights to various sections of the input sequence. In contrast, in multiplicative attention, we use matrix multiplication to reduce hidden states and calculate attention scores. We can then apply various activation functions to the dot product using adjustable parameters.

Convolutional Neural Network (CNN) Model

Although Convolutional Neural Networks (CNNs) are primarily used for image processing tasks, they have also proven effective in the analysis and prediction of time series data. CNNs typically consist of two main layers:

convolutional layers and pooling layers. In the context of time series forecasting, one-dimensional CNNs are used to process sequential data; here, the convolutional layers slide across the time series to identify recurring patterns and hidden trends (Lahuddin et al., 2025). Figure 2 below illustrates a CNN architecture.

Figure 2. CNN Architecture



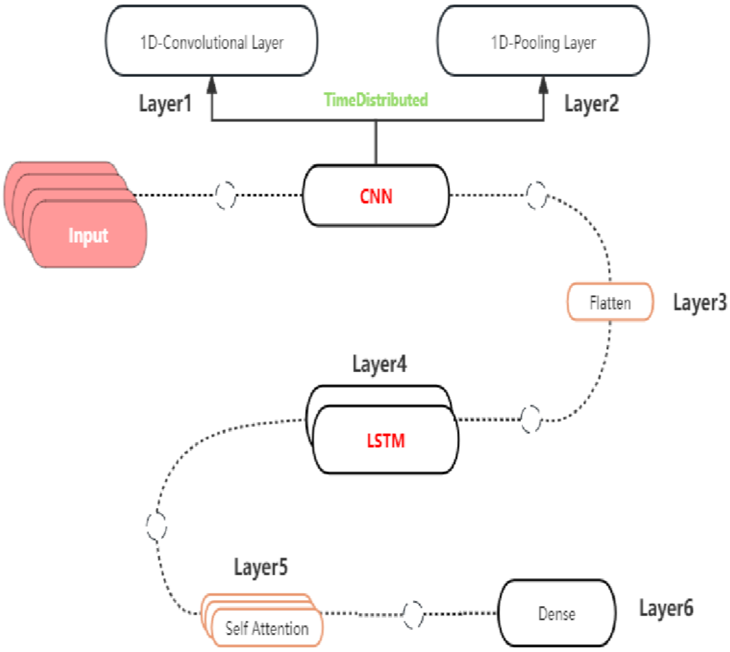
Source: (Lahuddin et al., 2025).

Attention-Mechanism-Based Hybrid CNN-LSTM Model

The attention mechanism-based hybrid CNN-LSTM model was developed by combining CNN and LSTM to improve the prediction accuracy of cryptocurrency data. This study incorporates the Attention Mechanism to obtain the most suitable models for time series predictions. Unlike an independent LSTM model, the CNN-LSTM-attention model begins by reshaping the input data into a three-dimensional array to facilitate convolutional operations. This preprocessing allows each data sub-sequence to be analyzed and transformed separately; this is crucial for effectively capturing local and temporal features. Data preprocessing adjusts the window size in the input and predicts a single output value. The preprocessed data will serve as the input for the CNN model layers. The CNN layer will be followed by maximum pooling, flattening, and recurrent vector layers. Subsequently, the recurrent vector, which is the same size as the window, will be connected to the Attention Mechanism. Finally, the output will be sent to the LSTM layers after being processed by the Attention Mechanism. The LSTM will consist of several layers followed by a fully connected layer. Finally, the final output of the hybrid model will be processed. The complete workflow of this model is detailed in Figure 3 (Sun, 2024).

IJCEAS
International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

Figure 3. CNN-LSTM-attention framework



Source: (Sun, 2024)

Performance Metrics

Three fundamental metrics, which are standard in financial time series forecasting, were used to measure the performance of deep learning models: Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). These metrics provide a standard basis for comprehensively evaluating model accuracy. MSE measures error magnitude on an absolute scale, while MAE and MAPE measure relative error, and R^2 identifies observed variations in model performance. For financial instruments exhibiting particularly high volatility, these metrics play a critical role in assessing the model's generalization ability.

4. Dataset and Application

In this study, deep learning-based time series models were used to forecast the daily price movements of leading digital assets in the cryptocurrency market. Since Bitcoin serves as the dominant and benchmark asset in the cryptocurrency market, altcoins representing different market behaviors were selected for this study. For the analysis, the daily closing prices of the digital assets Ethereum (ETH), Solana (SOL), and Binance Coin (BNB) were considered. The primary reason for selecting these assets was to model heterogeneous price behaviors by jointly evaluating digital assets that exhibit volatility within the crypto ecosystem in terms of market capitalization, trading volume, and global investor interest.

The study encompasses the stages of data collection, preprocessing, model development, and performance evaluation. The dataset used in the study covers the period from April 10, 2020, to February 12, 2026, and consists of a total of 2,135 observations comprising daily closing prices. The data, obtained from the Yahoo Finance database, allows for the incorporation of periods of high volatility into the modeling process. The contribution of co-movements among assets and potential cross-interactions to predictive performance was evaluated comparatively.

Preparing the Dataset for Modeling and the Training Infrastructure

During the modeling phase, the dataset was divided chronologically into two subsets while preserving its time-series structure. The dataset, consisting of a total of 2,135 observations, included 1,495 observations from April 10, 2020, to May 12, 2024, which formed the training dataset (70%), while the 640 observations from May 13, 2024, to February 12, 2026, formed the test dataset (30%). In this division, random splitting was not used; instead, chronological splitting was applied to preserve the time-dependent structure. The training dataset was used during the model's parameter learning process; validation was performed using an early stopping mechanism within the training process. The test dataset was reserved solely for final performance evaluation. To prevent scale differences from negatively affecting the learning process during model training, the data set was normalized to the [0,1] range using the MinMax method. The scaling process was applied only to the training data set, and the obtained parameters were applied to the test data set. This approach eliminated the risk of data leakage. Considering the time series structure, the dataset was restructured using a sliding window approach. In this method, past observations were defined as model inputs using a specific lookback window, and the price for the next day was predicted. Thus, the models were trained to perform forward-looking predictions by learning past price dynamics.

In this study, LSTM, GRU, CNN, CNN-LSTM, CNN-GRU, Attention-LSTM, and Attention-GRU architectures were trained on the normalized dataset, and their performances were evaluated comparatively. The prediction results were converted back to the original price level by applying inverse scaling at the model output stage and compared with actual observations. Model performance was analyzed using MSE, MAE, MAPE and R^2 error metrics. The Python programming language was used for the implementation process. The Pandas and NumPy libraries were used for data processing and analysis, while the Matplotlib library was used for visualization. Deep learning models were created and trained using the TensorFlow and Keras frameworks. The training process was conducted in a GPU-enabled Google Colab environment to reduce computational costs and increase processing speed.

Model Training and Hyperparameter Optimization

Before proceeding to predict ETH, SOL, and BNB prices, hyperparameter optimization was performed on the training dataset. The lookback window size, number of epochs, and batch size were identified as the key hyperparameters affecting model performance. For the lookback window, historical data series spanning 15, 30, 60, and 90 days were evaluated. The number of epochs was tested at 20, 50, and 100, while the batch size was tested at 16, 32, and 64. These hyperparameter combinations were systematically tested to analyze their effects on model performance. These parameter ranges were selected by considering values commonly used in deep learning-based financial time series forecasting studies.

To reduce overfitting, a 25% dropout rate was applied to all architectures. The Adam optimization algorithm was used to optimize model parameters, and the mean squared error (MSE) was chosen as the loss function. An early stopping mechanism was employed to terminate training when no improvement in validation loss was observed during the training process. In model setup, a time series approach was applied by including variables in the models one by one. Each cryptocurrency ETH, SOL, and BNB was modeled and forecasted separately using its own historical price data. This approach allows the dynamics of each asset to be learned independently, thereby enhancing the interpretability of the models and eliminating the risk of potential data leakage between different assets.

The LSTM and GRU models consist of two recurrent layers. Each layer uses 50 hidden neurons. In the first recurrent layer, a “return_sequences = True” structure was used to preserve temporal information, while a “return_sequences = False” was employed in the second layer to obtain the final temporal representation. A 25% dropout rate was applied between layers to mitigate the risk of overfitting. Subsequently, a 25-neuron dense layer and a single-neuron output layer were used to generate the prediction output.

The CNN model includes two 1D convolutional layers to extract local features from time-series data. In the first layer, features were extracted using 64 filters and a 3-dimensional kernel; "same padding" was applied, followed by dimension reduction via a 2-dimensional max-pooling layer. The second convolutional layer uses 32 filters. The resulting feature maps are flattened and fed into a fully connected layer with 50 neurons and ReLU activation. A 25% dropout is applied before the output layer.

In models with an attention mechanism, self-attention was applied to the hidden state sequences obtained from the recurrent layers. This allows the model to learn by weighting the most important information across time steps. The attention output was flattened and fed into a dense layer with 25 neurons and ReLU activation. In the final stage, dropout was applied, the model was flattened, and predictions were generated using a single-neuron output layer.

Obtaining the Models' Prediction Results

The primary variables evaluated for prediction performance in this study are the prices of ETH, SOL, and BNB. Training and testing processes were conducted using LSTM, GRU, CNN, CNN-LSTM, CNN-GRU, Attention-LSTM, and Attention-GRU architectures. After training was completed for each model, forward-looking predictions were generated on the test dataset. The obtained forecast values were converted to their original price levels via an inverse scaling process and compared with actual observations.

Model performance was calculated using the MSE, MAE, MAPE, and R^2 error metrics. Table 1 presents the performance results of each model on the test dataset. The models with the lowest error values were identified, and the contribution of architectural components to prediction success was analyzed comparatively. Thus, the effectiveness of different deep learning architectures for price prediction in the cryptocurrency market was systematically evaluated.

Table 1. Model Performance Comparison

Models		Variables					
		ETH		SOL		BNB	
Performance Metrics		Day-Epoch-Batch		Day-Epoch-Batch		Day-Epoch-Batch	
LSTM	MAE	15-100-64	86.4644	60-100-64	5.5975	30-100-64	16.0482
	MSE		14,731.5787		57.0276		617.1380
	MAPE		2.8730%		3.3971%		2.1955%
	R^2		0.972889		0.955668		0.977023
GRU	MAE	30-100-64	81.2627	30-20-16	5.423	90-20-16	**14.8750**
	MSE		**13267.3432**		55.0327		**513.6444**
	MAPE		2.7126%		3.2777%		**2.0370%**
	R^2		0.975584		0.957219		**0.980876**
CNN	MAE	30-50-32	93.3356	60-50-32	5.6482	30-20-16	19.1812
	MSE		15,982.8361		56.8363		816.7147
	MAPE		3.1828%		3.4033%		2.5778%
	R^2		0.970587		0.955817		0.969592
CNN-LSTM	MAE	90-50-32	81.6222	90-50-32	5.6026	30-50-32	16.6647
	MSE		13,604.6936		57.9639		644.6252
	MAPE		2.6998%		3.3624%		2.2656%
	R^2		0.974963		0.954940		0.975999
	MAE	9 0	**81.1447**	3 0	**5.4002**	6 0	15.4467

IJCEAS
International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

CNN-GRU	MSE		13,533.3413		**54.4558**		557.3781
	MAPE		**2.6913%**		**3.2490%**		2.1010%
	R^2		**0.975095**		**0.957667**		0.979248
Attention-LSTM	MAE	15-100-64	139.7729	30-100-64	8.0006	30-100-64	36.5382
	MSE		36,309.2104		115.6892		2115.9042
	MAPE		4.6916%		4.8779%		4.9837%
	R^2		0.933180		0.910066		0.921220
Attention-GRU	MAE	60-100-64	162.4729	15-100-64	9.0999	15-20-16	40.2419
	MSE		45,641.2702		149.944		3,156.6248
	MAPE		5.2228%		5.3124%		5.3561%
	R^2		0.916006		0.883437		0.882472

Source: Authors' calculations

The findings presented in Table 1 demonstrate that the performance of deep learning architectures used to forecast daily closing prices in the cryptocurrency market varies significantly both across assets and across model architectures. It is clear that no single architecture outperforms others across all series. Furthermore, the fact that window length, epoch, and batch size combinations produce different optimal values for each asset highlights that cryptocurrency assets possess heterogeneous and asset-specific temporal dynamics.

For the Ethereum (ETH) series, the lowest error values were produced by the GRU and CNN-GRU architectures. In particular, the GRU model, with a 30-day window length, 100 epochs, and a 64-batch configuration, delivered more successful results compared to LSTM and CNN-based models, achieving values of 81.2627 MAE, 13,267.3432 MSE, 2.7126% MAPE, and 0.975095 R^2. While the LSTM model also produced relatively low errors, the classic CNN architecture demonstrated weaker performance on the ETH series. It is observed that error values in attention-based models increased dramatically; for example, in Attention-GRU, MSE: 45,641.2702. This suggests that the ETH price series, while complex, responds better to more compact recurrent structures that do not require overparameterized attention mechanisms.

In the Solana (SOL) series, the lowest error was achieved by the CNN-GRU model in the 30-100-64 combination (MAE: 5.4002, MSE: 54.4558, MAPE: 3.2490%, R^2: 0.957667). The GRU model also demonstrated very similar performance. While LSTM and CNN-LSTM architectures produced relatively higher errors, models with attention mechanisms also performed notably poorly on the SOL series. This result indicates that GRU-based architectures are more effective at capturing the short- and medium-term temporal dependencies of the

SOL series. The addition of the attention mechanism increased model complexity but did not improve generalization performance.

In the Binance Coin (BNB) series, the lowest error was also achieved with the GRU model in the 90-20-16 configuration (MAE: 14.8750, MSE: 513.6444, MAPE: 2.0370%, R^2: 0.980876). The CNN-GRU model is also relatively successful, but the pure CNN architecture produced one of the weakest results in the BNB series (MAE: 19.1812). Error values have increased dramatically in the Attention-LSTM and Attention-GRU models (; for example, Attention-GRU MSE: 3156.6248). This finding indicates that BNB price dynamics are better represented by simpler recurrent structures rather than deep attention layers.

Overall, GRU-based architectures are observed to produce more stable and lower error rates for all three crypto assets. While LSTM and hybrid CNN-GRU/CNN-LSTM models offer competitive performance in some series, the addition of attention has not provided a systematic improvement. This clearly demonstrates that model complexity does not automatically increase prediction accuracy. It is understood that overparameterized architectures can increase generalization error in time series with high volatility, such as cryptocurrency. Therefore, it is concluded that model selection must be made by considering each asset's unique volatility structure and degree of temporal dependence.

The error metrics presented in Table 1 reveal significant differences in the forecasting performance of various deep learning architectures for cryptocurrency closing prices. However, an evaluation based solely on numerical error metrics may not fully reflect the models' learning behavior during the training process. Therefore, to examine the learning dynamics exhibited by the models during their training and validation processes in greater detail, the loss curves obtained for each model were analyzed graphically. These graphs enable a visual assessment of the models' convergence rate, stability, and potential overfitting during the training process.

Figure 4 presents the loss curve of the CNN-GRU Model for ETH closing prices.

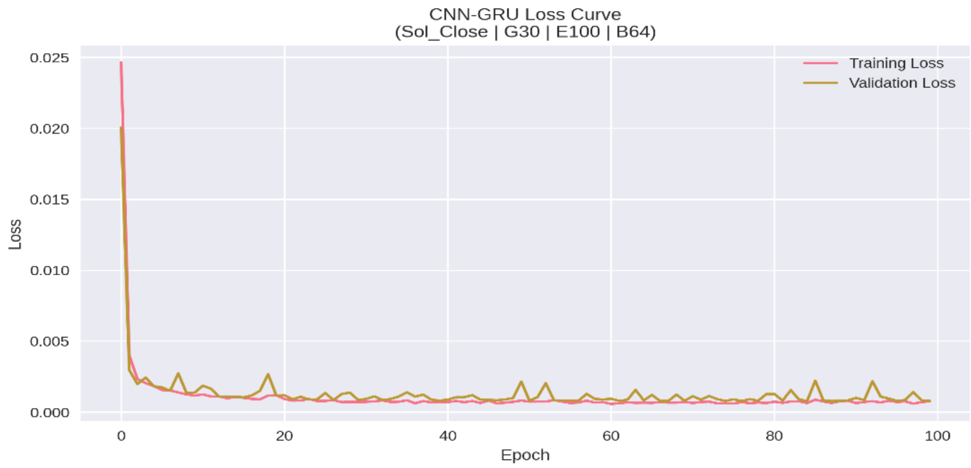
Figure 4. ETH – CNN-GRU Model Loss Curve

Source: Authors' calculations

Upon examining the ETH loss graph of the CNN-GRU model in Figure 4, it is evident that the loss values decrease very rapidly during the first epochs of the training process. This sharp drop in both training and validation loss indicates that the model was able to learn the fundamental patterns in the dataset in a short period of time. In subsequent epochs, both loss values are observed to remain at very low and stable levels. The fact that the training and validation curves are very close to each other suggests that the model demonstrates good generalization performance without overfitting.

Figure 5 presents the loss graph of the CNN-GRU model for SOL closing prices.

Figure 5. SOL – CNN-GRU Model Loss Curve

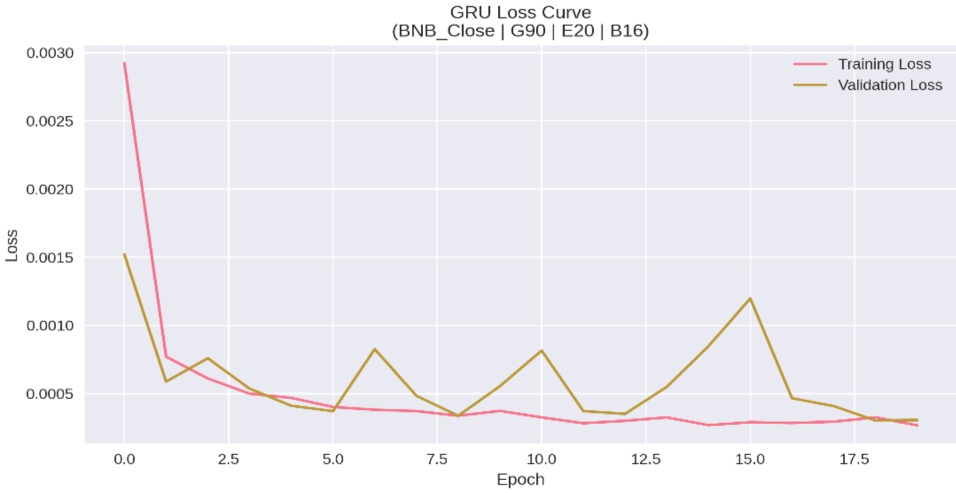


Source: Authors' calculations

Figure 5 The loss curve for the CNN-GRU model shows that loss values decrease very rapidly during the first epochs of training. This rapid decline indicates that the model was able to learn the fundamental patterns in the dataset at an early stage. In subsequent epochs, while the training loss remains at a very low and stable level, small fluctuations are observed in the validation loss. However, the fact that the two curves generally progress closely together indicates that the model does not exhibit a significant overfitting problem and possesses an acceptable generalization performance.

Figure 6 presents the loss graph of the CNN-GRU model for BNB closing prices.

Figure 6. BNB – GRU Model Loss Curve



Source: Authors’ calculations

Figure 6: The loss curve for the BNB GRU model shows that loss values decrease rapidly as epochs progress during the training process. A significant drop in both training and validation loss during the first few epochs indicates that the model quickly learned the fundamental patterns in the dataset. While the training loss follows a low and stable trend throughout the process, small fluctuations are observed in the validation loss during certain epochs. Nevertheless, the fact that the two curves progress closely together indicates that there is no significant overfitting issue in the model and that the model demonstrates reasonable generalization performance.

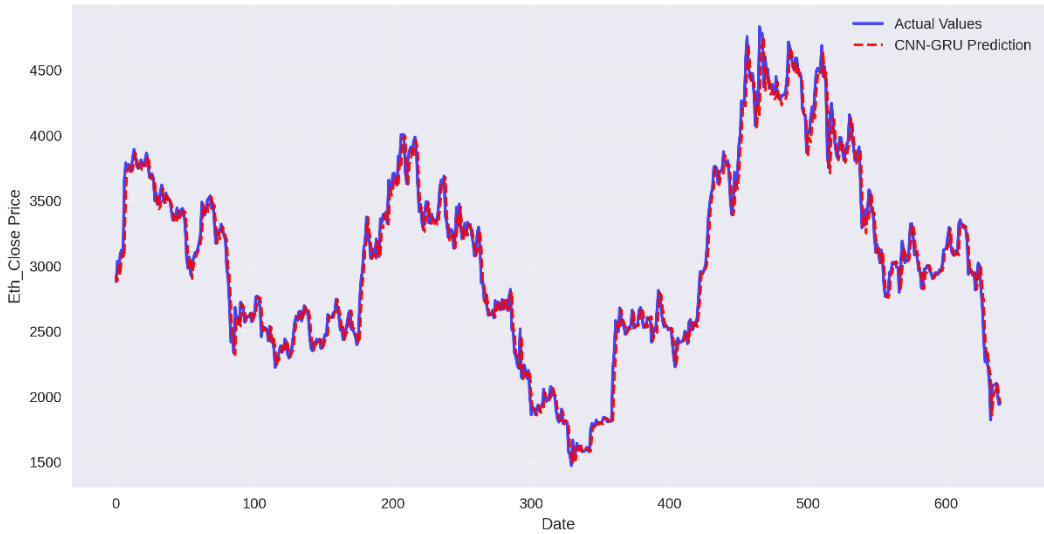
An analysis of the loss curves reveals that the models generally followed a stable learning process during training and showed a convergence trend with error values decreasing significantly after certain epoch levels. However, evaluating model performance based solely on error values during the training process is insufficient. To evaluate prediction success more comprehensively, the predicted values generated by the models must be compared with actual market prices. Accordingly, graphs showing the actual closing prices alongside model predictions for the best-performing models are presented below and analyzed visually.

IJCEAS

International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

Figure 7 presents the CNN-GRU Model's Actual vs. Forecast Comparison Chart for ETH closing prices.

Figure 7. ETH – CNN-GRU Model Actual vs. Forecast Comparison Chart



Source: Authors' calculations

In the graph in Figure 7, it can be seen that the values predicted by the CNN-GRU model closely align with the actual ETH closing prices. The model successfully tracked the general trend and price fluctuations, with only minor deviations occurring at certain points of sudden rises and falls.

Figure 8 presents the CNN-GRU Model's Actual vs. Predicted Comparison Chart for SOL closing prices.

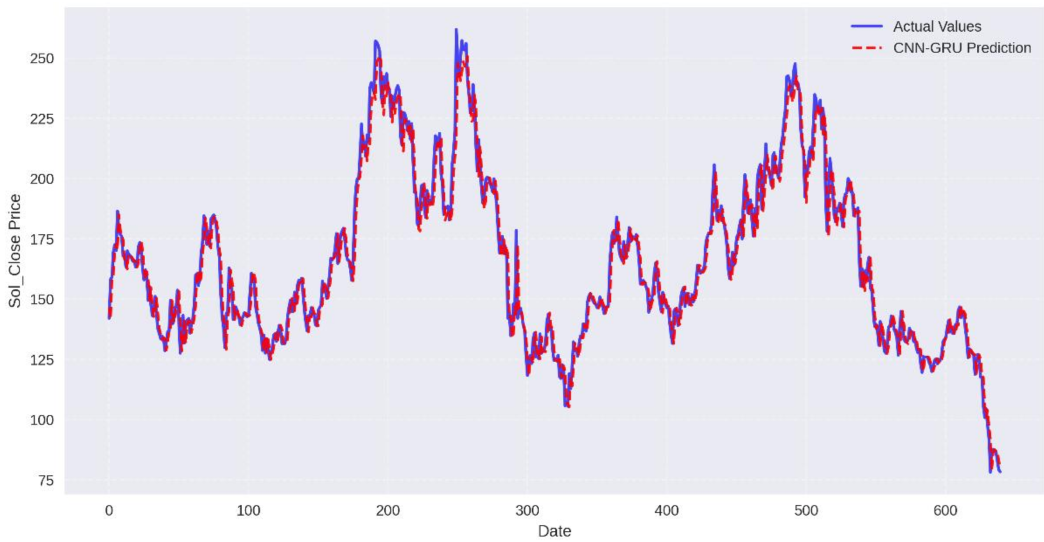
Figure 8. SOL – CNN-GRU Model Actual vs. Predicted Comparison Chart

Source: Authors' calculations

In the graph created for SOL closing prices in Figure 8, it is observed that the model's predictions show a high level of alignment with actual values. The CNN-GRU model demonstrated strong predictive performance by accurately capturing both short-term fluctuations and the overall price trend.

Figure 9 presents the CNN-GRU Model's Actual vs. Forecast Comparison Chart for BNB closing prices.

Figure 9. BNB – GRU Model Actual vs. Forecast Comparison Chart



Source: Authors' calculations

The graph in Figure 9 shows that the predictions generated by the GRU model for BNB closing prices largely track the actual values. However, it is notable that during some sharp price changes, the prediction curve reacts with a slight delay compared to the actual values.

The graphical analyses presented provide a visual means of evaluating the performance of deep learning models in predicting cryptocurrency closing prices. The fact that the actual values and the model's predictions largely follow a parallel trajectory indicates that GRU- and CNN-GRU-based architectures are particularly effective at capturing both the general trend and short-term fluctuations in price series. However, it was observed that limited deviations appeared in the prediction curves during some sharp price changes. This situation highlights that modeling sudden price movements in highly volatile cryptocurrency markets is relatively more challenging. When evaluated alongside the findings in the previous Table 1 and the graphs presenting the loss analysis results, these findings support the notion

IJCEAS
International Journal of Contemporary Economics and Administrative Sciences
ISSN: 1925 – 4423
Volume: XVI, Issue: 1, Year: 2026, pp. 866-894

that models containing simpler and more repetitive structures exhibit more stable forecasting performance in cryptocurrency time series.

5. Conclusion and Discussion

In this study, the performance of different deep learning architectures was comparatively analyzed in predicting the daily closing prices of altcoins Ethereum, Solana, and Binance Coin which are leading digital assets in the cryptocurrency market. To this end, recurrent neural network-based models (LSTM and GRU), convolutional neural network (CNN) architectures, hybrid CNN-based models (CNN-LSTM and CNN-GRU), and models incorporating attention mechanisms (Attention-LSTM and Attention-GRU) were tested under the same dataset and similar modeling conditions. The models' performance was evaluated using MAE, MSE, MAPE, and R^2 error metrics, and the optimal model configurations were determined by performing hyperparameter optimization for each model. The findings indicate that the predictive performance of deep learning architectures used to model cryptocurrency price series varies significantly depending on the model structure, dataset characteristics, and selected hyperparameter combinations.

When the analysis results are evaluated overall, it is observed that GRU-based models produce more stable results and lower prediction errors for all three cryptocurrency assets. In particular, the fact that the GRU model produces lower MAE, MSE, and MAPE values compared to other architectures, along with a high R^2 value, for the Ethereum and BNB price series demonstrates that this architecture can effectively learn the temporal dependencies in time series data. The fact that the GRU architecture has a simpler structure and contains fewer parameters compared to the LSTM model may enable the model to converge more quickly during the training process and capture the underlying patterns in the data more effectively. Considering that the data structure in financial time series often contains high noise and has limited deterministic patterns, it is believed that more compact architectures may exhibit stronger generalization performance compared to overparameterized models. In this context, based on the findings, it can be stated that increasing model complexity does not always improve prediction accuracy in high-volatility data environments such as the cryptocurrency market.

When examining the performance of hybrid architectures, it is observed that the CNN-GRU model, in particular, produces competitive results on some time series. This suggests that the strong ability of convolutional neural networks to capture local patterns and short-term dependencies within the data, when combined with the temporal dependency modeling capability of recurrent networks, can contribute to a more effective representation of financial time series. While CNN layers extract short-term fluctuations and local features from the price series, GRU or LSTM layers process these features along the time dimension to model longer-term dependencies. According to the findings of the study, it can be stated that hybrid architectures do not systematically produce the best performance for all assets. This indicates that cryptocurrency markets have a heterogeneous structure and that different assets may respond to different model architectures to varying degrees.

The most striking aspect of the study is that models incorporating an attention mechanism produce significantly higher prediction errors compared to other architectures. It was observed that error values increased significantly for all three cryptocurrency assets in both the Attention-LSTM and Attention-GRU models. The primary purpose of the attention mechanism is to enable the model to focus on time steps that are more important for prediction among past observations, thereby enhancing the model's ability to select relevant information. However, in data structures with high volatility and noise, such as cryptocurrency markets, distinct and recurring patterns are often limited. Consequently, the model's excessive focus on specific time steps via the attention mechanism can lead to the model learning random movements within the data. This situation can cause the model to overfit the training dataset and result in reduced generalization performance on the test dataset. Consequently, despite the theoretical advantages of the attention mechanism, it is evident that it does not always lead to performance improvements, particularly in financial time series with high-noise data environments.

The findings also indicate that the temporal dynamics of crypto assets exhibit distinct characteristics. The fact that the hyperparameter combinations yielding the best performance differ across all three assets suggests that a single model architecture should not be expected to produce optimal results for all assets in the cryptocurrency market. For example, while shorter moving windows yield more successful results in the Ethereum series, longer window lengths are seen to produce better results in the Binance Coin series. This situation indicates that the cryptocurrency assets under consideration possess different dynamics in terms of volatility structure, market depth, and investor behavior. Therefore, in modeling studies aimed at cryptocurrency price forecasting, model selection must be conducted by considering the structure of the dataset and the temporal characteristics of the asset. When the findings are evaluated overall, it is concluded that model complexity must be carefully addressed in the model development process for price forecasting in cryptocurrency markets. Although more complex architectures theoretically possess higher representational capacity, the high-noise structure of financial time series can limit the generalization performance of such models. Therefore, when selecting a model for cryptocurrency price forecasting, not only architectural complexity but also the number of model parameters, the size of the dataset, and the characteristics of the data structure must be considered. The results obtained in this study indicate that GRU-based models, in particular, could serve as an effective forecasting tool in cryptocurrency markets due to their ability to produce low prediction errors while containing fewer parameters.

Upon reviewing the literature, while there are numerous studies investigating the performance of deep learning-based models in cryptocurrency markets, it is observed that a significant portion of these studies use different data, different time intervals, and different features, making direct comparisons between models difficult. Additionally, findings regarding the impact of attention mechanisms on financial time series in the existing literature are inconsistent; while some studies report performance improvements, others indicate that they do not provide a meaningful contribution. Furthermore, while most studies focus on Bitcoin-centric analyses, systematic comparative studies on altcoins representing different market structures such as Ethereum, Solana, and Binance Coin are limited. In this context, the literature highlights a lack of comprehensive studies comparing different deep learning architectures (LSTM, GRU, CNN-based, and hybrid structures with attention mechanisms) under the same dataset and experimental conditions. This study aims to fill this gap by examining the relationship between model complexity and prediction accuracy in cryptocurrency markets within a more controlled experimental framework.

Consequently, this study presents significant findings regarding the relationship between model complexity and prediction accuracy by comparatively analyzing the performance of deep learning models for price forecasting in cryptocurrency markets. There is a research gap in the literature where different deep learning models are mostly evaluated under different datasets and experimental conditions, thus limiting direct comparisons. The results indicate that GRU-based architectures, in particular, may demonstrate more stable performance on financial time series with high volatility and provide methodological implications for model selection in cryptocurrency markets. Future studies examining broader cryptocurrency asset portfolios, utilizing high-frequency datasets, and evaluating newer deep learning approaches such as Transformer-based architectures could provide more comprehensive results regarding the predictability of cryptocurrency markets. Additionally, incorporating market sentiment data, social media indicators, and on-chain data sources into the model could contribute to a more comprehensive modeling of cryptocurrency price movements.

REFERENCES

Albayrak, E., and Saran, N. (2023). Stock Price Forecasting Using Statistical and Deep Learning Models. *Journal of Computer Science and Engineering,* 16(2), 161-169. https://doi.org/10.54525/tbbmd.1031017

Aruwaji, A. M., & Swanepoel, M. (July 2025). Finformer: An Attention-Based Deep Learning Model for Cryptocurrency and Stock Price Forecasting. In *2025 Conference on Information Communications Technology and Society (ICTAS)* (pp. 1-6). IEEE.

Ataei, S., Ataei, S. T., Omidmand, P., Karahroodi, H. H., & Nikzat, P. (2025). Applications of Deep Learning to Cryptocurrency Trading: A Systematic Analysis. *Soft Computing*: *Fusion with Applications*, *2*(4), 255-268.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative asset?. *Journal of International Financial Markets, Institutions and Money*, *54*, 177-189.

Borovykh, A., Bohte, S., & Oosterlee, C. W. (2017). Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv:1703.04691*.

Gao, M. (2025). Cryptocurrency Price Prediction Based on the CNN-BiLSTM-AM Model.

Gautam, M. (2025). Crypto Price Prediction Using LSTM+XGBoost. arXiv preprint arXiv:2506.22055.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep Learning* (Vol. 1, No. 2, pp. 1–800). Cambridge: MIT Press.

Katsiampa, P. (2019). An empirical investigation of volatility dynamics in the cryptocurrency market. *Research in International Business and Finance*, *50*, 322–335.

Kumar, S., Kour, V., Kumar, P*., et al.* GRU-LSTM with attention-based forecasting for enhanced air quality. *Nat Hazards* **121**, 15925–15947 (2025). https://doi.org/10.1007/s11069-025-07408-8

Lahuddin, H., Muliawan, M. R. A., Takemoto, K., Darwis, H., Jabir, S. R., & Adawiyah, R. (September 2025). Cryptocurrency Prices Forecasting Using LSTM, CNN, Transformer, TCN, and Hybrid Model: A Deep Learning Approach. In *2025 9th International Conference on Electrical, Electronics, and Information Engineering (ICEEIE)* (pp. 1–6). IEEE.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

Lee, M. C. (2024). Attention-Based Deep Learning Models for Cryptocurrency Price Prediction: A Comparative Analysis with Technical Indicators.

Luong, M. T., Pham, H., & Manning, C. D. (September 2015). Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1412–1421).

Mahdi, E., Martin-Barreiro, C., & Cabezas, X. (2025). A novel hybrid approach using an attention-based transformer+GRU model for predicting cryptocurrency prices. *Mathematics*, *13*(9), 1484.

McNally, S., Roche, J., & Caton, S. (March 2018). Predicting the price of Bitcoin using machine learning. In Proceedings of the *2018 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP)* (pp. 339–343). IEEE.

Priadinata, I. P. B., Sudipa, I. G. I., Meinarni, N. P. S., Radhitya, I. M. L., & Supartha, I. K. D. G. (2025). Comparative analysis of LSTM, GRU, and Bi-LSTM deep learning models for time series cryptocurrency price forecasting. *Sinkron: Journal of Computer Science Research*, *9*(3), 1024–1035.

Rezaei, H., Faaljou, H., and Mansourfar, G. (2021). Stock price prediction using deep learning and frequency decomposition. Expert Systems with Applications, 169, 114332. https://doi.org/10.1016/j.eswa.2020.114332

Saqware, G. J., & B, I. (2024, March). Hybrid deep learning model integrating attention mechanism for the accurate prediction and forecasting of the cryptocurrency market. In *Operations research forum* (Vol. 5, No. 1, p. 19). Cham: Springer International Publishing.

Saracık, Ö., & İncekırık, A. (2023). Stock Price Forecasting with Deep Learning Techniques. Alphanumeric Journal, 11(2), 137-156. https://doi.org/10.17093/alphanumeric.1357466

Seabe, P. L., Moutsinga, C. R. B., & Pindza, E. (2025). Sentiment-driven cryptocurrency forecasting: analyzing LSTM, GRU, Bi-LSTM, and temporal attention model (TAM). *Social Network Analysis and Mining*, *15*(1), 52.

Shi, Z., Hu, Y., Mo, G., & Wu, J. (2022). An attention-based CNN-LSTM and XGBoost hybrid model for stock price prediction. *arXiv preprint arXiv:2204.02623*.

Sun, X. (2024). Application of Attention-Based LSTM Hybrid Models for Stock Price Prediction. *Advances in Economics, Management and Political Sciences*, *104*, 77-91.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wei, Z., Wang, L., Liu, Z., & Zhou, Y. (2026). Prediction of extraction uranium concentration and optimization of injection acid concentration using a CNN–GRU–attention model in in-situ uranium leaching. Journal of Analytical Science and Technology, 17(1), 14.

Wen, X., & Li, W. (2023). Time series prediction based on an LSTM-attention-LSTM model. *IEEE Access*, *11*, 48322–48331.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. (June 2015). Show, attend, and tell: Neural image caption generation with visual attention. In *International Conference on Machine Learning* (pp. 2048–2057). PMLR.

Yahoo Finance. (2025). Historical daily closing prices for BNB-USD, ETH-USD, SOL-USD [Data set]. https://finance.yahoo.com/ Accessed on (02/14/2026)

Yousufi, M. L., & İncekırık, A. (2024). An analysis of Ethereum price prediction in the digital economy using deep learning techniques. In GENÇLERLE 360° 10th International Student Congress / The Circular Economy in a Globalizing World – Recycling / Circular Economy in the Globalizing World (Vol. 1, No. 1, pp. 58–66). Manisa Celal Bayar University. https://genclerle360.mcbu.edu.tr/wpcontent/uploads/sites/11/2025/01/01.08.-Onuncu-bildiri-kitabi.pdf

Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329*.

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.

Zhang, X., Dou, Y., Mao, J., Liu, W., & Han, H. (2023). Carbon Price Forecasting Approach Based on Multi-Scale Decomposition and Transfer Learning. Journal of Beijing Institute of Technology, 32(2), 242-255.

Zhao, X., Liu, J., Wang, Y., & Wang, J. (2026). CryptoMamba-SSM: Linear-Complexity State-Space Models for Cryptocurrency Volatility Prediction. *IEEE Open Journal of the Computer Society*, *7*, 226–243.