Scientific Journal of Informatics
Vol. 12, No. 4, November 2025

p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Comparative Performance Analysis of Deep Learning Models for Cryptocurrency Price Forecasting

Ryo Pambudi [1]*, Dinar Mutiara Kusumo Nugraheni[2], Aris Puji Widodo[3]

[1]Master of Information System, Universitas Diponegoro, Indonesia
[2,3]Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Diponegoro, Indonesia

Abstract.
Purpose: This research aims to find an accurate cryptocurrency price prediction model to mitigate financial risks caused by high price volatility. This research compares the predictive capabilities of five Deep Learning model, namely LSTM, GRU, BiLSTM, Transformer, and Performer, for predicting cryptocurrency prices with the highest accuracy in the digital financial market.
Methods: The methods applied in this research are dataset, preprocessing data, model training, and model evaluation. The dataset used in this study, namely the price per minute data for BTC, ETH, BNB, and XRP, was obtained from Kaggle. Data processing includes normalization using MinMaxScaler and sequence generation through the Sliding Window technique. To validate each deep learning model, and four metrics consisting of MAE, MSE, RMSE, and MAPE are used for evaluation.
Result: The Transformer model created the best results for the lowest MAPE value across all datasets, the smallest being BTC and ETH at 0.20%, BNB at 0.29%, and XRP at 0.36% demonstrating high accuracy and generalization. The BiLSTM was ranking second since it captured effectively the bidirectional temporal dependencies; the GRU was moderate but stable in its performance. The data showed that the accuracy of LSTM and Performer varied.
Novelty: This study offers a comprehensive comparison of various Deep Learning models in detail, enabling it to find the best model for predicting cryptocurrency prices with high accuracy. This study provides valuable insights for the development of advanced deep learning-based price forecasting systems in the field of digital financial analysis.

Keywords: Cryptocurrency price prediction, LSTM, BiLSTM, GRU, Transformer, Performer
Received May 2020 / **Revised** November 2020 / **Accepted** March 2021

INTRODUCTION

Technological advances in the economic sector have led to the emergence of digital innovations such as cryptocurrency, which has become a viable alternative to conventional currencies. Cryptocurrency is a decentralized digital financial system built on blockchain technology, which is an encrypted, transparent, secure, and immutable data recording system that guarantees security [1]. Compared to traditional currencies, crypto assets offer several advantages, namely ease of use across countries, fast transaction speeds, higher transparency, and cost efficiency in the management process.

It is one of the different crypto assets that have been in existence and endorsed by mankind, The two top digital currency only commodities accepted for the use in global financial networks are BTC and ETH. Both are critically important in developing blockchain-based currencies, particularly smart contracts and Decentralized Finance (DeFi) [2], DeFi is essentially an ecosystem of financial applications built on blockchain technology that aims to disintermediate traditional finance by replacing centralized intermediaries with smart contracts. Its dominance can be observed from its market capitalization of USD 90 billion making Bitcoin as one of the main cryptocurrencies studied in the cryptocurrency ecosystem [3]. Furthermore, the wide usage of cryptocurrencies has inspired many countries to make their own a Central Bank Digital Currency (CBDC), this is a digital form of a country's fiat currency that is centrally issued, regulated, and backed by the nation's central bank.it really shows how now these blockchain financial systems are being recognized publicly.

*Corresponding author.
Email addresses: ryopambudi@students.undip.ac.id (Pambudi)
DOI: [10.15294/sji.v12i1.35653](https://doi.org/10.15294/sji.v12i1.35653)

Even though it has a huge potential, one of the major problems for cryptocurrency is still its high price volatility. Such an unstable situation is a significant financial risk for investors, users and the whole market, which can cause losses [4]. Addressing this problem requires a system or algorithm with the precision of forecasting price changes. This solution would also reduce the amount of money lost financially and it would give a hand to the market players in making the right trading decisions [5].

Various previous studies on cryptocurrency price forecasting have addressed different kinds of prediction models. To qualify for predictive technology standards, a model should be able to conduct analysis on historical cryptocurrency price data and recognize extremely complicated price movement patterns. Deep learning methods, for instance, have been found effective and suitable for the estimation of cryptocurrency prices by past researchers [6]. There was a finding that the two models LSTM and GRU as predictive models could detect the long-term patterns in the cryptocurrency price data through their experiments [7]. This is in line with other studies, where LSTM has been proven effective in predicting Ethereum prices based on the model's ability to recognize the characteristics and transaction volume of this cryptocurrency [8].

Other research that were carried using the GRU model concluded that this model has a potential to achieve a more consistent performance in down or bear market [9]. However, previous studies have also indicated that LSTM provides superior prediction performance over other models when predicting cryptocurrency prices, for BTC, XRP and LTC especially [10], [11], [12], [13], which demonstrates the strong capability of the model to manage non-linear and highly volatile data. The BiLSTM model also had better performance in predicting BTC price than the other approaches [14]. The Performer and BiLSTM models also generated positive forecasts outcomes for BTC, ETH, and LTC [15]. In a different study, it was observed that the GRU model was the most effective and the most accurate in predicting the prices of Bitcoin, ETH, and ZEC [16]. In contrast, some studies stated that LSTM and BiLSTM were the most accurate deep learning models for predicting the price of Bitcoin [17].

Before conducting experiments comparing the most accurate deep learning models, we reviewed various previous research literature to understand the areas of research that had been studied by previous researchers on cryptocurrency rate prediction. This literature review aimed to identify the most accurate types of prediction models, and researchers also evaluated the approaches used in previous studies as reference material. Researchers found that there is still a research gap in the comparative analysis of these deep learning models due to the different types of cryptocurrencies. It was concluded that even though previous literature has been published providing information about the predictive capabilities of a deep learning model, a more in-depth comparative analysis is still needed, as evidenced by conducting direct experiments and evaluating the results with more mathematical methods.

Based on the above description, researchers are interested in further exploring the predictive capabilities of deep learning models. This study aims to analyze and compare the performance of five deep learning models in crypto prediction, namely LSTM, GRU, BiLSTM, Transformer, and Performer, on four types of cryptocurrencies, namely Bitcoin BTC, ETH, XRP, and BNB. This study is expected to find the most accurate deep learning model that can provide the most effective results for cryptocurrency price forecasting. Through experimental evaluation and literature review, this study provides practical recommendations for investors, analysts, developers, and policymakers in determining the most effective deep learning prediction model to anticipate the high volatility of the cryptocurrency trading market so as to minimize the risk of user losses.

METHODS

The framework for this research comprises multiple stages which are dataset, preprocessing data, model training, and model evaluation as shown in Figure 1. For this research, a dataset from Kaggle was used and later checked for validation. The MinMaxScaler method and the Sliding Window technique for sequence generation was used. After processing the dataset, training and testing subsets were created. The training data was used to fit the proposed models which include GRU, LSTM, BiLSTM, Transformer, and Performer. The predicted outputs were visualized for each selected cryptocurrency, and for the final step, all models were evaluated comprehensively using the MAE, MSE, RMSE, and MAPE metrics to assess the performance and results. Of all these features, the Close variable was chosen as the primary target variable because it is most indicative of the price trends and is a key variable in numerous scholarly works related to financial predictions. The frequency per minute was preserved to analyze the short-term volatility inherent in high-frequency trading in digital assets, enabling the model to scrutinize price fluctuation patterns more intensively.

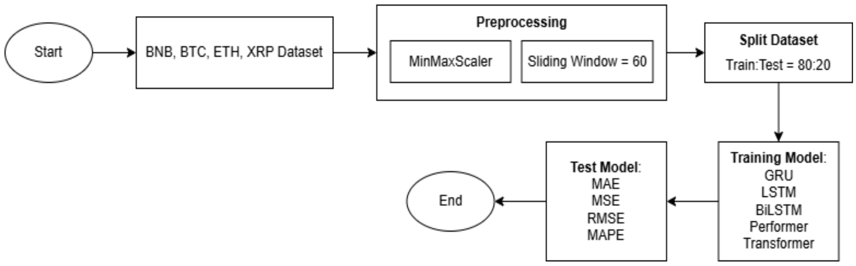

Figure 1. Research Method Scheme

Dataset

This research dataset was downloaded in CSV format from an open-source resource on Kaggle [18], [19], [20]. The dataset captures the historical trading figures of the primary four cryptocurrencies, namely BTC, ETH, BNB, and XRP. Because this is a minute-level recording of an entire dataset for every active trading day and hour, the developed dataset captures the minute dynamics of the cryptocurrency market in real-time functionality with unprecedented temporal granularity. The observation period varies for every single asset, with BTC being from 2015 to 2025, ETH from 2016 to 2024, BNB from 2018 to 2025 and XRP from 2017 to 2025. Each dataset holds millions of records due to the high-frequency trading and long-term trading, with roughly 4,911,806 records for BTC, 4,235,499 for ETH, 3,667,636 for BNB, and 4,140,180 for XRP. The dataset in this scale offers historical price volatility and shifting trends across the entire period in the crypto market. The specifications of the datasets in this research are detailed in Table 1. The analyses were performed using the “Closing” parameter of the dataset, denoting the closing price of the cryptocurrency every minute.

Table 1. Dataset Specifications

Parameter	Description	Data Type
Unix Timestamp	Unix timestamp (seconds since 1-Jan-1970), time mark for this value	Integer (int64)
Date	Time and date of the transaction, datetime in UTC time zone format	DateTime (datetime64[ns, UTC])
Symbol	Name of the cryptocurrency pair (e.g., BTC/USD) Test Symbol	Text (object)
Open	Opening cryptocurrency price at start of date range chosen	Number (float64)
High	Highest prices the crypto achieved in the selected period	Number (float64)
Low	Lowest price reached by the cryptocurrency during the specified period	Number (float64)
Close	Closing price the crypto achieved in the selected period	Number (float64)
Volume	Transaction volume during the specified period, usually in the base currency	Number (float64)

Of all these features, the Close variable was selected as the main target variable because it is the most representative of price trends and is a commonly used indicator in financial forecasting research. The frequency per minute was maintained to capture the short-term volatility that is characteristic of high-frequency digital asset trading, allowing the model to study price fluctuation patterns in greater depth.

Preprocessing Data

The data preprocessing tasks in this research are designed to be systematically organized in terms of deep learning models to analyze time series data. The research begins with the application of pandas, numpy, matplotlib, and tensorflow scientific libraries. These libraries are critical in terms of data processing, numerical modeling, data visualization, and modeling RNN-based models. The data to be processed in this task entails cryptocurrency data with 1-minute data intervals, which are essential in terms of implementing the price prediction process. The data preprocessing task applies an entire workflow process related to the preprocessing step of cryptocurrency time series data. The process begins with data extraction from the CSV file that contains 1-minute data intervals essential in terms of cryptocurrency data. The data preprocessing task applies Exploratory Data Analysis (EDA) to assess its structure and determine data values that are missing, while also examining the distribution and characteristics of the dataset to identify potential outliers and anomalies. The time data preprocessing stage begins with the transformation process that entails making the ‘Date’ column datetime format with the data set as the index essential in terms of implementing efficient time data analysis. The data preprocessing task on cryptocurrency data was applied in this research to ensure that the data was qualified to be processed by deep learning models. The parameters that defined the computer setup installed with the process to execute Python 3 within the Visual Studio Code environment are illustrated in Table 2.

Table 2. Computer Specifications Used for Model Testing

No	Components	Description
1	Operating System Name (OS)	Windows 10 Pro Education 64-bit
2	Processor	AMD Ryzen 9 7900X 12-core dan 24-thread
3	Motherboard	MSI B650 Gaming Plus WiFi
4	Installed Physical Memory (RAM)	32 GB 6400 MHz DDR5
5	VGA Card	NVIDIA GeForce RTX 4070 12GB VRAM
6	Storage	SSD 500 GB
7	Power Supply	750 Watt

The normalization process is carried out with the aid of the ‘Min-Max Scaling Method’ with a scaling range from 0 to 1. The method is adopted to standardize the scale among the observations and to increase the efficiency of the learning process within the LSTM models. The use of the ‘Min-Max Scaling Method’ ensures that the proportion and distribution are maintained among the values without affecting the extreme values and to increase the rate of convergence during the training process. The dataset is split into 80% for training and 20% for testing to maintain temporal properties, and the validation loss shown in the results actually comes from using the test set as the validation set during training rather than having a separate dedicated validation split from the training data. The dataset is structured using a sliding window approach where sequences of 60 time steps are used to predict the next single time step, and the final 60 time steps of the training set serve as the initial input context to generate predictions for the test set, maintaining temporal continuity and providing the model with sufficient historical context for forecasting unseen data in this time series framework. This process has a higher potential for identifying relationships between data instances in the aforementioned models, as it can detect both short-term and long-term relationships.

The obtained sequence data is then normalized to the format required by the input structure to facilitate the detection of temporal patterns by models that use LSTM. The same holds true for the testing data to ensure that there is no scaling difference between the two sets of data. Only after completing the modeling process are the prediction results transformed back to the original scale to enable direct comparison with the true prices since the initial prediction results are still in the transformed scale.

Model Training

The training dataset was used to train each deep learning model during the model implementation stage. Training was used to optimize the weights of the model for better accuracy in predicting cryptocurrency prices. Each model took sequential cryptocurrency price data as input, where each element of the sequence represented a single time point in the time-series data. We had GRU, LSTM, BiLSTM, Transformer, and Performer as deep learning models. Despite the differences between models, all deep learning models had an equivalent target prediction objective which was to predict exactly what will be the next price.

LSTM

The Long Short-Term Memory network is a Recurrent Neural Network that was created to solve the problem of long-term dependencies in time-series data. It brings in three main gates: input, forget, and output gates that help the model selectively save or get rid of information and control its flow through memory cells. This kind of mechanism helps LSTM learn effectively about long-range dependencies within sequential data [21], [22], [23]. LSTM has been very much used for modeling time-series because it can deal with complicated and various temporal patterns. With its cell state and gating structure, LSTM can keep important information for a long time while throwing away unimportant details; hence, it works very well on tasks related to predicting sequential data like cryptocurrency price forecasting.

At this stage of research, the LSTM model was created with two stacked layers (stacked LSTM) composed of 64 neuron units and ReLU activation functions for every layer. The first layer was set up with return_sequences=True; thus, it would create output at each time step to be passed on to the next layer, while the second one produced a single vector representation summarizing all temporal information from previous 60 time steps. A Dropout mechanism of 0.2 was applied after each LSTM layer in order to reduce the risk of overfitting by randomly deactivating some neurons during training. The last layer was a single Dense layer that generated closing price predictions on a continuous scale. Its ability to capture long-term contextual relationships gives a very big advantage in enhancing predictive performance in very volatile and nonlinear market conditions.

GRU

The Gated Recurrent Unit (GRU) is another Recurrent Neural Network (RNN) architecture that uses two key gates: reset and update gates to control the flow of data and prevent the vanishing gradient problem that occurs in standard RNNs. The use of GRU can be more computationally efficient and practical in scenarios involving computing and resource-limited environments [24], [25], [26]. In other applications such as battery prediction, fault detection, and stock price prediction, the use of GRU has proven more accurate compared to LSTM [27], [28].

At this stage of research works, the initial layer of the GRU network processes data in sequence form, with each step involving the activation of the dual gate function in the GRU network—the reset gate defines how much past data needs to be forgotten (e.g., noise within the price movement), and the update gate regulates the amount of past memories to be retained versus how much new data needs to be added. The result is the production of the hidden state that recognizes complex observations within financial data to identify trend, volatility, and momentum within financial transactions. The output produced from the initial layer with 64 temporal features is then further processed in the other GRU layer to gather data from the whole sequence prior to the transformation of data by the dense layer to produce one price prediction within cryptocurrency transactions during training. The model then trains to optimize its gate parameters to weigh critical data observations that affect cryptocurrency price transactions through backpropagation during training processes, with the reset gate aiming to disregard randomly fluctuating data and the other gate aimed to retain trend data within financial transactions.

BiLSTM

BiLSTM The Bidirectional Long Short-Term Memory network is an extension to the standard LSTM architecture that processes input data in two different ways forward and backward passes thus permitting the model to efficiently exploit contextual data from past and future instances concurrently. The bidirectional property of BiLSTM helps improve the model's capability to extract complex temporal relationships present within data. Experiments have proved that BiLSTM has been useful in improving the prediction accuracy on highly volatile financial data such as cryptocurrency price series over unidirectional models such as regular LSTM and traditional statistical models such as ARIMA [29], [30], [31]. Recent research works have proved that BiLSTM leads to significant improvements in accuracy during cryptocurrency prediction tasks such as Bitcoin, Ethereum, and Litecoin prediction due to its proven capacity to process complex data with highly volatile and nonlinear patterns [32], [33], [34]. BiLSTM achieves this by using bidirectional information flow which, in effect, gives it a more detailed temporal representation. Thus, it is a tool that can be used with great success in tasks related to the prediction of financial time series that are usually volatile and uncertain.

The operations in this research, the first layer of an LSTM unit are divided into two parallel streams that handle data in opposite directions. The forward LSTM examines the time-dependent patterns from t0 to t59 and does so; by consecutively processing the Open, Close, and Volume features. On the other hand, the backward LSTM takes the same series, but in an opposite order from t59 to t0, to not only detect reversal patterns but also to figure out support-resistance levels from a different angle. Next, these two outputs are connected via the feature fusion concatenation method at each timestep which results in a new temporal representation combining both past and future patterns. Next, the top-layer network extracts higher-level patterns from this decorated temporal sequence, the final hidden states from both directions being concatenated to form the final output representation. This is especially important when we model the prices of cryptocurrencies in the machine learning model. The model can therefore be used to see the price changes that were caused by the past patterns when performing a forward pass. It can also find the reversal points and the status of the higher levels by reversing. As a result, it gives you full picture of what your products and services are going to be and can take on for reference when anticipating future prices with much greater accuracy than before with old one-way solvers.

Transformer
The design of the Transformer, namely the Temporal Fusion Transformer (TFT) and Multi-Head Attention renditions, has been very successful in working with multivariate data and remembering long-term dependencies. Various works in the literature have demonstrated the superiority of the Transformer-based model over traditional architectures such as LSTM, especially when combined with other techniques or enriched by auxiliary features [35], [36], [37]. The ability of transformers to model global contextual relationships and identify intricate price patterns across long time intervals makes them a useful tool for managing complex time-series data, such as historical cryptocurrency prices [38]. However, other studies also reported that LSTM may still be superior in certain scenarios, especially on Bitcoin and Ethereum price predictions [10], [39]. It suggests that while attention mechanisms bring strong representational power, the performance of Transformers may vary depending on the characteristics of the data and model configurations.

The Transformer model of this research starts the process of Multi-Head Self-Attention, where every timestep in the sequence can interact with the rest of the timesteps, thus being able to calculate the level of relevance and dependency. For example, in the context of cryptocurrency price prediction, if the model is to forecast the price at time t+1, then each data point from the last 60 timesteps through the attention mechanism "communicates" with each other in order to find the historical points that have the most influence on the prediction. The attention results are subsequently combined with the original information via a residual connection and are normalized by Layer Normalization to keep the system stable during training. After that, the output goes to a Feed Forward Network which operates as a separate non-linear change for each position in the sequence, thus deepening the feature representation with complex patterns that the attention mechanism might have not recognized. The work comes to an end with another residual connection and normalization before the sequence is processed further.

Performer
The Performer is a modified Transformer architecture that aims to address the problem of efficiency when using a typical self-attention mechanism. Fast Attention Via Positive Orthogonal Random Features (FAVOR+) is a linear-time approximation algorithm that uses positive orthogonal random features to efficiently calculate the Transformer's self-attention, enabling scalable processing of long sequences. To this end, it uses the FAVOR+ algorithm, which allows the calculation of attention to have linear complexity with respect to the input length, while the complexity in the case of a standard Transformer model is quadratic [40], [41]. The Performer's applicability extends to real-time systems and large-scale financial time-series data, e.g. crypto-currency price prediction [15]. Being a development of the Transformer, the Performer provides faster and more efficient computation of complicated sequential data, a crucial feature due to the high volatility of crypto currency markets [42].

The Performer in this research processes cryptocurrency time series data through several stages of efficient transformation. The model first projects the input using linear projection into 64-dimensional embeddings to prepare the data for a more complex attention process. The essential part of Fast Attention is the most important part of Performer. It uses the ReLU kernel function as a feature map to approximate traditional softmax calculations, thus the computational complexity is changed from quadratic to linear with respect to sequence length. Another important element in the process of protecting against causality is time dependent, to summarize the information could not be accessed beyond the previous instant enabling the transactions to happen temporally, keeping the standards of finance prediction.

Model Evaluation
Evaluation of each model was conducted after the end of the training process using the testing dataset that had not been used during training. This evaluation aimed to test how well each model could predict cryptocurrency prices on new data and therefore how well it generalizes. The following evaluation metrics were used, along with explanations of their significance in the context of cryptocurrency forecasting:

Mean Absolute Error (MAE) is an average size measure for all absolute differences between predicted and actual values. MAE describes intuitively what average prediction error means in original price units without considering its direction. In very volatile markets for cryptocurrencies, MAE would help quantify a typical deviation between predicted and actual prices and thus be able to show clearly how accurate a model is overall in predicting. The MAE is mathematically defined as follows Eq.(1).

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

Mean Squared Error (MSE) is the average of the squared differences between actual and predicted values. Squaring the errors, MSE punishes massive deviations respectively to small ones and particularly sensitive for outliers and extreme prediction errors both of what can have a huge influence on cryptocurrency price prediction. The squaring highlights large errors by giving higher weight to predictions which are far from the actual range. The MSE is given mathematically as Eq.(2).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

Root Mean Squared Error (RMSE) is the square root of the MSE, so it still has the same units as the data. This makes it easier to understand the size of the mistakes made by predictions in terms of the actual price. RMSE gives an estimate of how large typical prediction errors can be while keeping the price scale the same. This is more straightforward and useful for financial and cryptocurrency forecasting. RMSE mathematically in Eq.(3).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Mean Absolute Percentage Error (MAPE) is defined as the average absolute error in percentage terms based on the actual value. It is especially suitable for assessing predictive performance across different cryptocurrencies that have diverse price scales. MAPE gives a relative measure of error, which helps in easily understanding how much deviation exists between predicted and actual prices in proportional terms. Mathematically, MAPE can be defined as Eq.(4).

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4}$$

Notes:

y_i = actual value
$\hat{y}_i$ = predicted value
n = total number of data points
$|y_i - \hat{y}_i|$ = absolute difference between the actual and predicted values
$(y_i - \hat{y}_i)^2$ = squared difference between the actual and predicted values
$\left|\frac{y_i - \hat{y}_i}{y_i}\right|$ = relative error per data point

Besides the metrics listed above for quantitative evaluation, comparing the actual prices with predicted values were also presented through image visualization to demonstrate the performance of model qualitatively. These visualizations can be used to discover error patterns and long or short term trends between predicted and real cryptocurrency prices. The integration of both quantitative metrics and qualitative visual analysis helps establish an all around methodology to identify the optimal deep learning model for cryptocurrency price prediction.

RESULT AND DISCUSSION

This study compared several state-of-the-art deep learning architectures for the task of cryptocurrency closing price estimation. Experiments were conducted on four different datasets, BNB, BTC, ETH, and XRP, in order to get a broader perspective on model effectiveness across different crypto-asset classes. Each dataset was minute-level resolution, representing real-time movement within the cryptocurrency market. These datasets were relatively large, with 3,667,636 entries for BNB, 4,911,806 for BTC, 4,235,499 for ETH, and 4,140,180 entries for XRP. The datasets had some historical features like Open, Close, and Volume, with Close being the value used for prediction.e), and transaction volume (Volume); the predictive focus was on the value of Close.

Five different models were considered for the experiments: GRU, LSTM, BiLSTM, Transformer, and Performer. The GRU, LSTM, and BiLSTM network structures each had two primary layers with 64 units, the ReLU activation function, and a dropout rate of 0.2 to reduce overfitting. The BiLSTM used bidirectional processing to get the temporal context not only from the past but also from the future time steps. On the other hand, the Transformer model used a self-attention mechanism with two attention heads and a 64-dimensional feed-forward layer. The Performer, being a resourceful version of the Transformer, substituted the standard self-attention mechanism with kernel-based linear attention to lower the computational complexity. The effectiveness of each model was judged based on the following four metrics: MSE, MAE, RMSE, and MAPE. Training results and evaluation scores for models on four cryptocurrency datasets are depicted in Table 3.

Table 3. Model Training and Evaluation Results

Crypto currency	Training Model	Train Loss	Val Loss	MSE	MAE	RMSE	MAPE
BNB	GRU	7.0410	4.1041	358637	54449	59886	1.36%
	LSTM	6.0239	0.8111	31147	13119	17649	0.37%
	BiLSTM	3.1477	0.7982	122076	28423	34939	1.02%
	Performer	2.9700	0.8720	415713	62679	64476	1.68%
	Transformer	1.0652	0.1294	29250	11776	17103	0.29%
BTC	GRU	4.897	4.127	208650.410	345.424	456.783	0.99%
	LSTM	5.077	1.056	256868.024	390.207	506.822	1.54%
	BiLSTM	2.740	1.583	295335.289	440.167	543.448	1.30%
	Performer	2.550	0.296	141146.857	367.701	375.695	1.31%
	Transformer	0.898	0.061	29134.732	88.583	170.689	0.20%
ETH	GRU	5.0817	0.2987	234.1676	13.1109	15.3025	0.59%
	LSTM	4.9966	9.0574	8109.1732	83.4584	90.0509	4.50%
	BiLSTM	0.1856	2.4618	505.9427	19.1464	22.4932	0.96%
	Performer	2.4300	0.4730	1120.3892	33.3754	33.4722	1.65%
	Transformer	0.9025	0.1864	155.9380	9.7845	12.4875	0.45%
XRP	GRU	0.0001	0.0001	0.0001	0.0090	0.0103	1.67%
	LSTM	0.0001	0.0031	0.0002	0.0114	0.0131	1.98%
	BiLSTM	0.0001	0.0000	0.0000	0.0039	0.0048	0.72%
	Performer	0.0001	0.0000	0.0001	0.0069	0.0072	1.26%
	Transformer	0.0000	0.0000	0.0000	0.0020	0.0034	0.36%

The best results, from the BNB dataset, are provided in Table 3, on the Transformer model. MSE of 29250, MAE of 11776, RMSE of 17103, and MAPE of 0.29%, are all the lowest values in compared to other models. These low error values indicate that the transformer is able to predict accurately and can generalize well. This achievement shows that the Transformer is highly accurate for interpreting both short- and long-term temporal patterns of crypto price data. Figure 2 shows the graph of price prediction data from the BNB dataset.

Performer model, on the other hand, exhibited the worst prediction performance for this dataset: MSE 415713 and MAE 62679. This underperformance may stem from the archetypal-attention structure of Performer, which focuses on improving computational efficiency via using kernelized attention mechanism but are less suitable to capture complex dynamics in BNB price volatility.

The prediction results of the LSTM, BiLSTM, and GRU models are at an intermediate level. Of these, LSTM is a bit more effective with a MAPE value of 0.37% and an RMSE of 17649, thus, it is the second-best model after the Transformer. This outcome confirms the research already done where it has been demonstrated that LSTM is more powerful than GRU in learning long-term temporal dependencies.

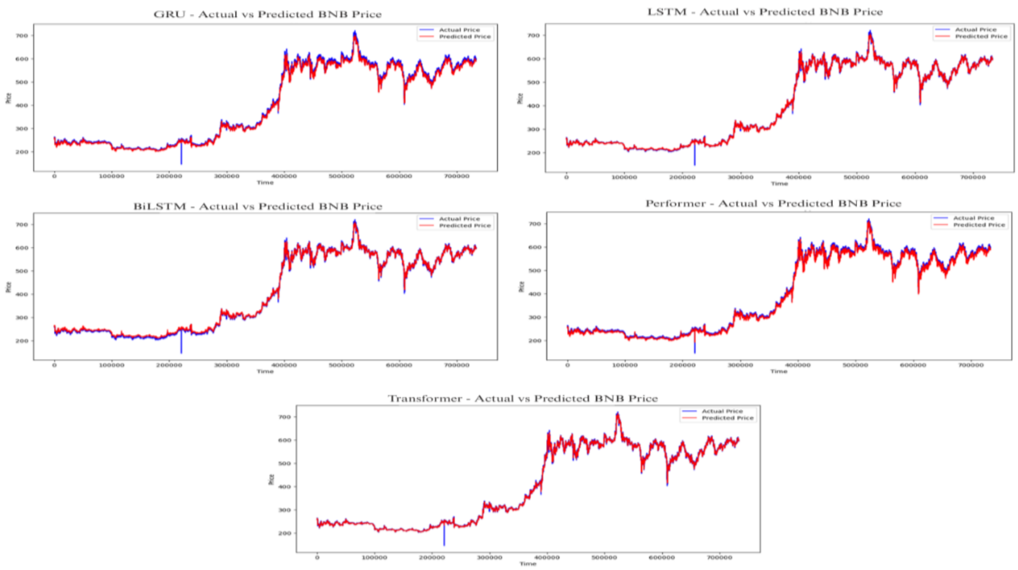

Figure 2. Predicted Price Graph for the BNB Dataset

Similar to the pattern shown in Table 3 for the BTC dataset, it is evident that the Transformer model also ranked the highest among all models. In addition to recording an MSE of 29134.732, an MAE of 88.583 and an RMSE of 170.689; this model also had the lowest MAPE (0.20%). Therefore, these results have further strengthened the evidence for how well the Transformer model can handle modeling the volatile time-series data of the Bitcoin price using its adaptive attention mechanism to dynamically allocate contextual information weight to each input time point for its predictions. The graph of price forecasting results for the BTC dataset can be found in Figure 3.

Among these, the BiLSTM and LSTM models gave less stable performances on this dataset. Specifically, the LSTM recorded the highest MAPE of 1.54% with an RMSE of 506.822. This suboptimal performance was probably contributed by overfitting during training, as observed by the large difference between the train loss value of 5.077 and the validation loss value of 1.056. The Performer and GRU models had moderate prediction performances; GRU performed slightly better in terms of MAPE, which was 0.99%, while its RMSE value was still comparably high at 456.783.

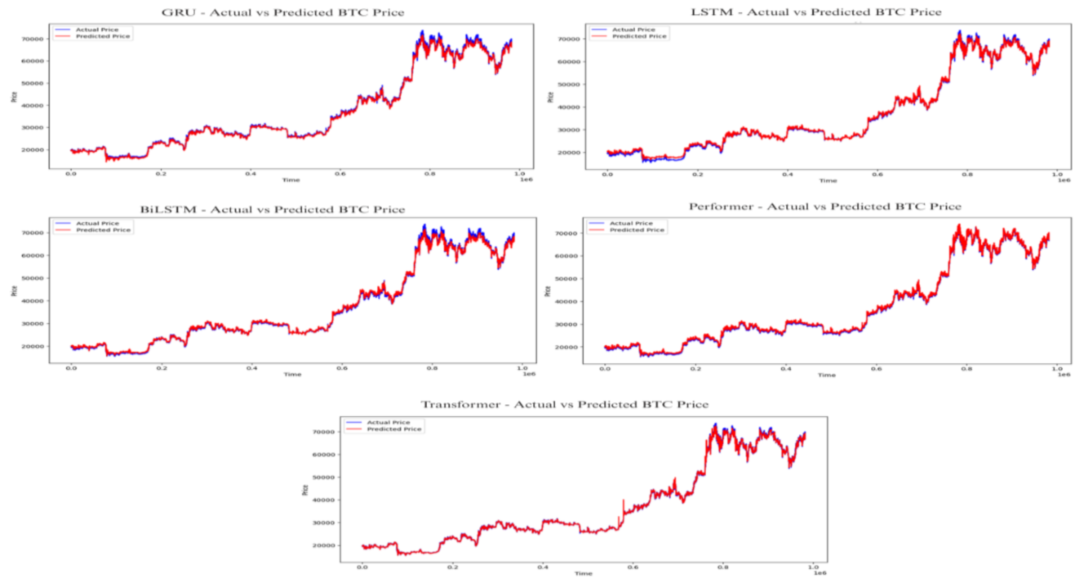

Figure 3. Predicted Price Graph for the BTC Dataset

In the ETH dataset, the model forecast, that can be found in Table 1, illustrates the persistent performance of the Transformer architecture over again. The model that we are talking about here gives us an MSE value of 155.938, an MAE value of 9.7845, an RMSE value of 12.4875, and an MAPE of 0.45%. These figures show that the Transformer was consistent in understanding patterns related to the price of Ethereum very well. The result, in other words, guarantees to one that the transformer architecture that is based on attention enables both the time series of finance to be modeled accurately and also prediction to be made correctly. The Predicted Price Graph for the ETH Dataset can be found Figure 4.

Unlike the previous case, the LSTM model showed a considerable drop in performance on this dataset, resulting in an MSE of 8109.1732 alongside a 4.50% MAPE. This is likely due to LSTM's difficulty in accommodating the short-term price volatility dynamics of ETH in the observed data. In contrast, the performance of the BiLSTM and GRU models is relatively stable; based on both MAE and RMSE values, BiLSTM performs slightly better than GRU. Meanwhile, the Performer model performed less competitively, with an MSE value of 1120.3892. Although the Performer architecture is designed to achieve great computational efficiency at large data scales, this linear attention approach seems less than optimal in capturing highly dynamic and externally sensitive characteristics of crypto time series.

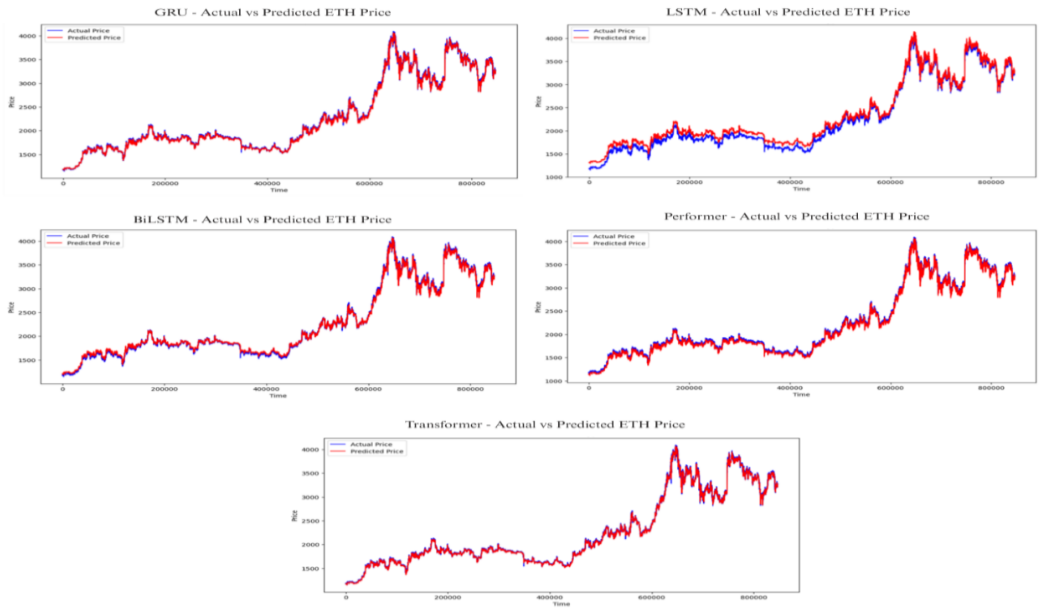

Figure 4. Predicted Price Graph for the ETH Dataset

The prediction outcomes on the XRP dataset, shown in Table 3, suggest that the relatively smaller price scale of XRP -- when compared to other crypto assets -- is a determining factor in a lower error metric value. In this regard, Transformer remains the best prediction model with an MSE of 0.0000, MAE of 0.0020, and MAPE of only 0.36%. This condition also lends credence to the flexibility of Transformer in modifying prediction performance across data scales. The chart of price prediction results on the XRP dataset is shown in Figure 5.

The BiLSTM model achieves the second best performance, with an MSE of 0.0000 and a MAPE of 0.72%, indicating that bidirectional mechanism is able to capture local price patterns more fully. The GRU and LSTM models give very good performance, however the accuracy of the LSTM model in this case has a ceiling due to its maximum observed MAPE value at 1.98%.At the same time, however, Performer sits within the middle category, displaying an MSE of 0.0001 and 1.26% for the MAPE. These results imply that the Performer architecture, while efficient in computation, has yet to fully refine predictions on low-volatility assets such as XRP.

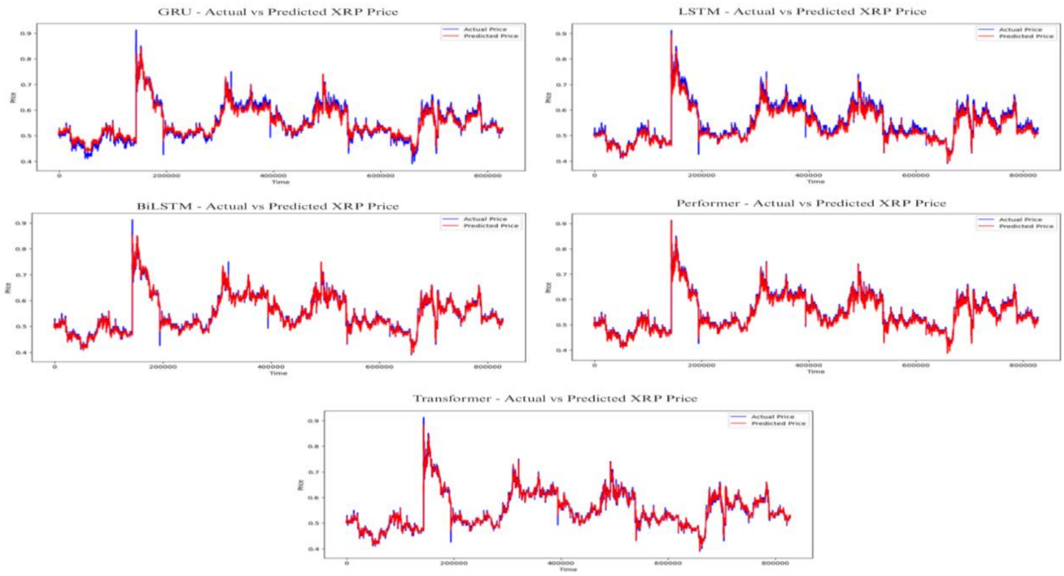

Figure 5. Predicted Price Graph for the XRP Dataset

Figure 6 presents the performance comparisons in a line plot. The Transformer consistently had the best predictive performance in the forecasting of cryptocurrency closing prices, as shown by the lowest MAPE values across all four evaluated datasets: BTC, ETH, BNB, and XRP. In detail, it reached a MAPE value of 0.20% for BTC, 0.45% for ETH, 0.29% for BNB, and 0.36% for XRP.

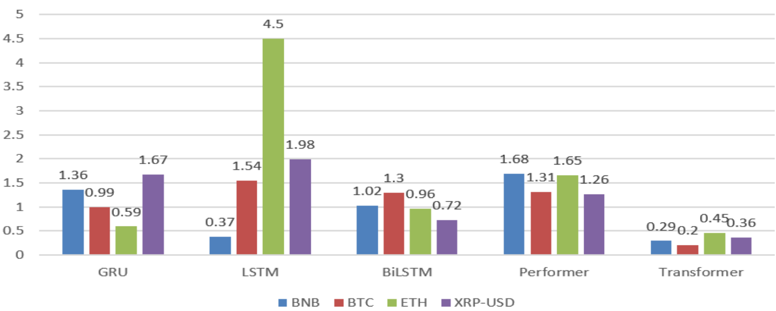

Figure 6. Comparative Prediction Performance Across All Cryptocurrency Datasets

Figure 6 presents the performance comparisons in a line plot. The Transformer consistently had the best predictive performance in the forecasting of cryptocurrency closing prices, as shown by the lowest MAPE values across all four evaluated datasets: BTC, ETH, BNB, and XRP. In detail, it reached a MAPE value of 0.20% for BTC, 0.45% for ETH, 0.29% for BNB, and 0.36% for XRP.
The transformer is stable with high prices and volatility, such as BTC, ETH, BNB, and relatively low prices and moderate fluctuations, such as XRP. This success is greatly dependent on the self-attention mechanism of the model that better captures long term dependencies and fully integrates time in the learning process. Another competitive performance of BiLSTM was XRP, which ran with a MAPE value of 0.72% in the XRP dataset. This achievement comes from BiLSTM's ability to process temporal information in two directions, forward and backward, which improves representation of temporal context compared to conventional LSTM.

The GRU model delivered relatively stable performance across all datasets, with MAPE ranging from 0.59% to 1.67%. However, GRU was never ranked as the model with the best prediction performance in all experiments. Performer has not yet delivered optimal results in crypto price prediction; its MAPE values tend to be higher, particularly on the BNB dataset at 1.68% and ETH at 1.65%. LSTM performed well only on the BNB dataset with a MAPE of 0.37%, but then it significantly dropped in performance on the ETH dataset with a MAPE of 4.50%. This indicates that this model is very sensitive to changes in data complexity being tested.

CONCLUSION

From this research, the Transformer model was the best for predicting the closing prices of cryptocurrencies across the four datasets I looked at. It had the lowest MAPE for each asset, whether those assets had high or low volatility. It was good at picking up historical clues and understanding long-term patterns due to its self attention feature, which makes me think it was the best overall. The BiLSTM was second, especially for small cap assets like XRP, while the GRU stayed consistent but never led in prediction accuracy. The Performer, even though it was designed to be quick, did not do well with the complex price movements of cryptocurrencies, and the LSTM showed a mixed bag depending on the dataset. These results show how important it is to pick the right model when dealing with financial time series data that can change so much, like crypto assets.

REFERENCES

[1] S. Mohapatra, N. Ahmed, and P. Alencar, "KryptoOracle: A Real-Time Cryptocurrency Price Prediction Platform Using Twitter Sentiments," *Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019*, pp. 5544–5551, 2019, doi: 10.1109/BigData47090.2019.9006554.

[2] Z. Zhou, "Research on the Relationship and Application of Cryptocurrency and Blockchain," *Highlights Business, Econ. Manag.*, vol. 21, pp. 382–388, Dec. 2023, doi: 10.54097/hbem.v21i.14492.

[3] Z. Y. Lin, "Investor attention and cryptocurrency performance," *Financ. Res. Lett.*, vol. 40, no. May 2020, p. 101702, 2020, doi: 10.1016/j.frl.2020.101702.

[4] M. Al Guindy, "Cryptocurrency price volatility and investor attention," *Int. Rev. Econ. Financ.*, vol. 76, no. October 2020, pp. 556–570, 2021, doi: 10.1016/j.iref.2021.06.007.

[5] F. Mehrdoust and M. Noorani, "Prediction of cryptocurrency prices by deep learning models: A case study for Bitcoin and Ethereum," *Int. J. Financ. Eng.*, vol. 10, no. 04, Dec. 2023, doi: 10.1142/S2424786323500329.

[6] E. Loginova, W. K. Tsang, G. van Heijningen, L. P. Kerkhove, and D. F. Benoit, "Forecasting directional bitcoin price returns using aspect-based sentiment analysis on online text data," *Mach. Learn.*, no. 0123456789, 2021, doi: 10.1007/s10994-021-06095-3.

[7] D. M. Gunarto, S. Sa'adah, and D. Q. Utama, "Predicting Cryptocurrency Price Using RNN and LSTM Method," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 1–8, 2023, doi: 10.32736/sisfokom.v12i1.1554.

[8] M. H. Abdul Hadi, N. A. Ramli, and Q. UI Islam, "Predicting Bitcoin and Ethereum prices using long short-term memory and gated recurrent unit," *Data Anal. Appl. Math.*, no. June, pp. 8–17, 2023, doi: 10.15282/daam.v4i2.10195.

[9] J. Kim, S. Kim, H. Wimmer, and H. Liu, "A Cryptocurrency Prediction Model Using LSTM and GRU Algorithms," *Proc. - 2021 IEEE/ACIS 6th Int. Conf. Big Data, Cloud Comput. Data Sci. BCD 2021*, pp. 37–44, 2021, doi: 10.1109/BCD51206.2021.9581397.

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

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

[12] J. Chen, "Analysis of Bitcoin Price Prediction Using Machine Learning," *J. Risk Financ. Manag.*, vol. 16, no. 1, 2023, doi: 10.3390/jrfm16010051.

[13] M. A. Ammer and T. H. H. Aldhyani, "Deep Learning Algorithm to Predict Cryptocurrency Fluctuation Prices: Increasing Investment Awareness," *Electron.*, vol. 11, no. 15, pp. 1–22, 2022, doi: 10.3390/electronics11152349.

[14] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," *Fractal Fract.*, vol. 7, no. 2, pp. 1–18, 2023, doi: 10.3390/fractalfract7020203.

[15] M. Ali, L. Khaniki, and M. Manthouri, "Enhancing Price Prediction in Cryptocurrency Using Transformer Neural Network and Technical Indicators," 2024.

[16] S. Hansun, A. Wicaksana, and A. Q. M. Khaliq, "Multivariate cryptocurrency prediction: comparative analysis of three recurrent neural networks approaches," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00601-7.

[17] F. Tabe, A. Mansourabady, A. H. Rasekh, and B. Tanoori, "Comparison of Deep Learning Algorithms for 'Bitcoin Cash' Price Prediction," *2024 20th CSI Int. Symp. Artif. Intell. Signal*

Process. AISP 2024, pp. 1–5, 2024, doi: 10.1109/AISP61396.2024.10475222.
[18] I. Bukhari, "Binance Coin BNB, 1m Full Historical Data," 2025.
[19] I. Bukhari, "Ripple XRP, 7 Exchanges, 1m Full Historical Data," 2025.
[20] P. Gendotti, "BTC and ETH 1-min Price History," 2024.
[21] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, "A survey on long short-term memory networks for time series prediction," *Procedia CIRP*, vol. 99, pp. 650–655, 2021, doi: 10.1016/j.procir.2021.03.088.
[22] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, and A. Borodulin, "Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review," *Polymers (Basel).*, vol. 16, no. 18, p. 2607, Sep. 2024, doi: 10.3390/polym16182607.
[23] J. Xu and L. Cao, "Copula Variational LSTM for High-Dimensional Cross-Market Multivariate Dependence Modeling," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 35, no. 11, pp. 16233–16247, Nov. 2024, doi: 10.1109/TNNLS.2023.3293131.
[24] R. Rouhi Ardeshiri and C. Ma, "Multivariate gated recurrent unit for battery remaining useful life prediction: A deep learning approach," *Int. J. Energy Res.*, vol. 45, no. 11, pp. 16633–16648, Sep. 2021, doi: 10.1002/er.6910.
[25] Y. Li, Z. Zheng, and H. N. Dai, "Enhancing bitcoin price fluctuation prediction using attentive LSTM and embedding network," *Appl. Sci.*, vol. 10, no. 14, pp. 1–17, 2020, doi: 10.3390/app10144872.
[26] S. Nosouhian, F. Nosouhian, and A. Kazemi Khoshouei, "A Review of Recurrent Neural Network Architecture for Sequence Learning: Comparison between LSTM and GRU," Jul. 2021. doi: 10.20944/preprints202107.0252.v1.
[27] S. Mirzaei, J.-L. Kang, and K.-Y. Chu, "A comparative study on long short-term memory and gated recurrent unit neural networks in fault diagnosis for chemical processes using visualization," *J. Taiwan Inst. Chem. Eng.*, vol. 130, p. 104028, Jan. 2022, doi: 10.1016/j.jtice.2021.08.016.
[28] Q. Wang, "Stock Prediction Under GRU, LSTM and GNN," in *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*, IEEE, Jun. 2024, pp. 460–464. doi: 10.1109/ICIPCA61593.2024.10709241.
[29] M. Yang and J. Wang, "Adaptability of Financial Time Series Prediction Based on BiLSTM," *Procedia Comput. Sci.*, vol. 199, pp. 18–25, 2022, doi: 10.1016/j.procs.2022.01.003.
[30] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," in *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2019, pp. 3285–3292. doi: 10.1109/BigData47090.2019.9005997.
[31] P. K. Irlapale, "Elevating Cryptocurrency Predictions: Bidirectional LSTM Methodology," *INTERANTIONAL J. Sci. Res. Eng. Manag.*, vol. 08, no. 05, pp. 1–5, May 2024, doi: 10.55041/IJSREM34330.
[32] M. F. Bin Hossain, L. Z. Lamia, M. M. Rahman, and M. M. Khan, "FinBERT-BiLSTM: A Deep Learning Model for Predicting Volatile Cryptocurrency Market Prices Using Market Sentiment Dynamics," Nov. 2024.
[33] M. Nair, M. I. Marie, and L. A. Abd-Elmegid, "Prediction of Cryptocurrency Price using Time Series Data and Deep Learning Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, 2023, doi: 10.14569/IJACSA.2023.0140837.
[34] R. Patel, J. Chauhan, N. K. Tiwari, V. Upaddhyay, and A. Bajpai, "A Deep Learning Framework for Hourly Bitcoin Price Prediction Using Bi-LSTM and Sentiment Analysis of Twitter Data," *SN Comput. Sci.*, vol. 5, no. 6, 2024, doi: 10.1007/s42979-024-03112-9.
[35] E. Mahdi, C. Martin-Barreiro, and X. Cabezas, "A Novel Hybrid Approach Using an Attention-Based Transformer + GRU Model for Predicting Cryptocurrency Prices," *Mathematics*, vol. 13, no. 9, p. 1484, Apr. 2025, doi: 10.3390/math13091484.
[36] S. Gopali, S. Siami-Namini, F. Abri, and A. Siami Namin, "A Comparative Multivariate Analysis of VAR and Deep Learning-Based Models for Forecasting Volatile Time Series Data," *IEEE Access*, vol. 12, pp. 155423–155436, 2024, doi: 10.1109/ACCESS.2024.3447354.
[37] R. Bourday, I. Aatouchi, M. A. Kerroum, and A. Zaaouat, "Cryptocurrency Forecasting Using Deep Learning Models: A Comparative Analysis," *HighTech Innov. J.*, vol. 5, no. 4, pp. 1055–1067, Dec. 2024, doi: 10.28991/HIJ-2024-05-04-013.
[38] S. Penmetsa and M. Vemula, "Cryptocurrency Price Prediction with LSTM and Transformer Models Leveraging Momentum and Volatility Technical Indicators," in *2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA)*, IEEE, Oct. 2023,

pp. 411–416. doi: 10.1109/ICDSCA59871.2023.10393319.
[39] H. Zhao, M. Crane, and M. Bezbradica, “Attention! Transformer with Sentiment on Cryptocurrencies Price Prediction,” in *Proceedings of the 7th International Conference on Complexity, Future Information Systems and Risk*, SCITEPRESS - Science and Technology Publications, 2022, pp. 98–104. doi: 10.5220/0011103400003197.
[40] K. Choromanski *et al.*, “Rethinking Attention With Performers,” *ICLR 2021 - 9th Int. Conf. Learn. Represent.*, pp. 1–38, 2021.
[41] V. Likhosherstov, K. Choromanski, J. Davis, X. Song, and A. Weller, “Sub-Linear Memory: How to Make Performers SLiM,” *Adv. Neural Inf. Process. Syst.*, vol. 9, no. NeurIPS, pp. 6707–6719, 2021.
[42] Pappula Ashok, “Cryptocurrency Price Prediction using Deep Learning Algorithms: A Comparative Study,” *Commun. Appl. Nonlinear Anal.*, vol. 32, no. 4s, pp. 537–551, Dec. 2024, doi: 10.52783/cana.v32.2870.