# Cryptocurrency Price Prediction Based on CNN-BiLSTM-AM Model

Meijun Gao
*School of Electronics Engineering and Computer Science, Peking University, Beijing, China*

Keywords: Cryptocurrency Price Prediction, CNN, BiLSTM, Attention Mechanism.

Abstract: Contemporarily, as the cryptocurrency market has experienced unprecedented rapid growth, its high price volatility and complex nonlinear dynamic characteristics have made cryptocurrency price prediction a focus of concern. Accurate price prediction is crucial for investors for it helps investors effectively manage investment risks and significantly increase investment returns. This study innovatively proposes a hybrid prediction model which integrates convolutional neural network (CNN), bidirectional long short-term memory network (BiLSTM) and attention mechanism (AM), named as the CNN-BiLSTM-AM model, aiming to accurately predict the price of three typical cryptocurrencies, BTC, ETH, and LTC. CNN is used in feature extraction. BiLSTM is introduced to process time series data. AM tracks how feature states affect cryptocurrency closing prices over history. With the aim of verifying the effectiveness of the hybrid model, the next day's closing price of BTC, ETH, and LTC is selected as the test dataset for this model and three other mainstream prediction models. Experimental results indicate that this hybrid model ranks first in terms of prediction accuracy, specifically manifested in the smallest Mean Absolute Error (MAE) and Root Mean Squard Error (RMSE), along with the highest R-Square ($R^2$). These results indicate that the hybrid CNN-BiLSTM-AM model can be adopted as a powerful tool for investors to formulate investment strategies and make actual investment decisions.

## 1 INTRODUCTION

Since the launch of Bitcoin in 2009, the cryptocurrency market has profoundly reshaped the financial ecology and created a new paradigm of decentralized and highly secure value transfer and storage. Unlike typical financial assets, which are mainly supported by the government or central banks, cryptocurrencies rely on advanced peer-to-peer network architecture and blockchain technology to operate independently. The rapid development of representative cryptocurrencies such as BTC, ETH, and LTC has not only established their solid position in the trading field, but also stimulated widespread investment and speculation enthusiasm. With the continuous expansion and increasing influence of the cryptocurrency market, accurate prediction of its price fluctuations has become a cutting-edge research topic across multiple disciplines such as economics, finance and computer science, attracting attention from many scholars and professionals. Financial market forecasting, a traditional problem in economics and finance, is even more difficult in the rapidly evolving cryptocurrency market environment. Among traditional methods, econometric models such as ARIMA are widely used in financial markets, but their forecasting effectiveness is severely challenged in the emerging field of cryptocurrency, because this market is highly dynamic and complex, far beyond the traditional scope (Zhang, 2003). In the face of the unique decentralized characteristics of cryptocurrency and its extreme sensitivity to external factors such as regulatory policy adjustments, market sentiment fluctuations and technological innovation, its price trends show significant non-linear and highly volatile characteristics. These features limit the availability of traditional linear models in accurately capturing the dynamic price patterns of cryptocurrencies, prompting the research community to continuously explore forecasting methods that are more suitable for this emerging market.

Faced with high complexity and dynamic changes of the cryptocurrency market, traditional prediction methods seem to be unable to cope with it. However, the latest achievements in machine learning field, especially models such as SVM, random forests, and various neural networks are playing an increasingly important role in cryptocurrency price prediction with



Gao, M.
Cryptocurrency Price Prediction Based on CNN-BiLSTM-AM Model.
DOI: 10.5220/0013269800004568
In *Proceedings of the 1st International Conference on E-commerce and Artificial Intelligence (ECAI 2024)*, pages 516-523
ISBN: 978-989-758-726-9

their outstanding ability to analyse complex nonlinear relationships. The application of these new methods has opened up new ways to improve forecast accuracy. In particular, deep learning technology, with its advantage in processing massive sequence data, has shown extraordinary potential in cryptocurrency price prediction. Scholars have actively tried plenty of deep learning architectures, such as LSTM, CNNs, and their fusion models, aiming to more accurately grasp market dynamics.

During the past decade, the advancement of deep learning technologies has significantly impacted time series forecasting. In this context, the LSTM network has performed well in capturing complex long-term correlations in sequence data with its outstanding capability. Therefore, it has gradually become a popular tool in cryptocurrency price forecasting field. The research of Shah and Zhang strongly proved that the LSTM network can precisely grasp the time series traits of Bitcoin price data. Meanwhile, its prediction precision is significantly improved compared with traditional models (Shah & Zhang, 2014). In a 2018 study, McNally pioneered a new hybrid model by combining the advantages of LSTM and Gated Recurrent Unit (GRU). Experimental data clearly showed that this hybrid model surpassed the single LSTM model in prediction accuracy (McNally, 2016).

Another significant breakthrough is the application of CNN to time series data to extract characteristic patterns of price fluctuations. Researchers pointed out that CNN is good at capturing local characteristics in cryptocurrency price trends. These characteristics are then input into LSTM to deeply mine the long-term dependence of the price sequence. Recent studies have shown that mixed models integrating CNN and LSTM have better performance than single models when dealing with the high volatility and uncertainty of cryptocurrency prices (Zhang et al, 2021). In addition, the introduction of attention mechanisms (AMs) has added new vitality to these models. AM gives different importance weights to different parts of the input data, allowing the model to concentrate on the historical information that is most crucial to the prediction results (Chen et al., 2020). This not only makes the prediction result more accurate, but also enhances the interpretability of the model. At present, the application of CNNs, BiLSTMs and AMs has become effective frameworks for parsing the complex market dynamics of cryptocurrencies (Seabe et al, 2023).

Cryptocurrency price prediction remains a challenging field because of its high volatility and reliance on external factors like news, regulations, and technology. This research introduces a novel model that incorporates CNN, BiLSTM, and AM to address these challenges. By integrating the three different models, a more precise and robust prediction can be made for financial investors and analysts.

# 2 DATA AND METHOD

This research selected three typical cryptocurrencies, BTC, ETH and LTC. The detailed historical price data for BTC and LTC is from January 2014 to January 2024. Since ETH was first issued in July 2014, the historical price data for ETH is from January 2019 to January 2024. They are all downloaded from Yahoo Finance. These data cover key information such as daily opening and closing prices, as well as price percentage changes and trading volume. During the experiment, the whole dataset was separated into two sections: the first 80% were used for model training, and the remaining 20% were utilized as a test set to check the model's prediction accuracy.

Following part details the specific principles of CNN, LSTM, AM, as well as the hybrid CNN-BiLSTM-AM model. Lecun first introduced CNN to the public in 1998 (Lecun et al, 1998). CNNs are used to capture spatial information from time series data and identify short-term trends in cryptocurrency price changes. The mathematical representation of the convolutional layer is as follows:

$$z_t = tanh(W_c \cdot x_t + b_c) \tag{1}$$

Here, $z_t$ is the output of the convolution layer, $W_c$ represents the convolution filter weights, $x_t$ refers to the input vector (such as price features), and $b_c$ is the bias term. A pooling layer is adopted after the CNN layer to lower dimensionality of the output, ensuring computational efficiency while retaining the most critical features.

Schmidhuber first introduced the concept of LSTM model in 1997 (Hochreiter & Schmidhuber, 1997). It was invented to address the persistent problems with gradient disappearance and gradient explosion in RNNs (Ta et al, 2020). LSTM networks are ideal for cryptocurrency price prediction beacause of their powerful capability of extracting long-term dependencies in sequential data. Each LSTM memory cell consists of Forget Gate, Input Gate, and Output Gate. The function of Forget Gate is to determine what information should be removed from the memory cell. It takes in the current input and output value from the previous moment and returns a value between 0 and 1, indicating how much information should be forgotten. The calculation formula is as follows:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t]\right) + b_f \quad (2)$$

Here, $W_f$ is a weight matrix which connects the prior hidden state and the current input. $b_f$ is the bias term.

The function of Input Gate is to decide which new information should be appended to the memory cell. The Input Gate receives two input values: the current input value and the previous output value. It works in two steps: Generating the Candidate Memory Value,

$$\widetilde{C_t} = tanh\ (W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

where $W_c$ and $b_c$ refer to the weight and bias terms. They are used to compute the candidate value and calculating Output:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i\ ) \quad (4)$$

The Input Gate output $i_t$ decides which parts of the candidate memory value will be added to the current memory state. LSTM refreshes the cell state $C_t$ after the two operations above. The update process is as follows:

$$C_t = f_t C_{t-1} + i_t \widetilde{C_t} \quad (5)$$

The former term represents the amount of previous memory to retain based on the Forget Gate's output. The latter term represents the new information being appended to the memory state, controlled by the Input Gate. The function of Output Gate is to receive inputs including the current input value and the previous output value. The calculation formula is:

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o\ ) \quad (6)$$

Here, $O_t$is an output that determines how much of the current memory to retain. The final output hidden state ht is computed as:

$$h_t = o_t tanh\ (C_t) \quad (7)$$

In 1980, Treisman et al. first proposed AM (Treisman & Gelade, 1980). The important information is chosen from a vast amount of information by computing the probability distribution of attention. AM is employed to focus on the most relevant portions of the time series by learning which features from the past are most important for predicting the future. Then it assigns different weights to different time steps due to their importance. The attention score calculation is shown by this formula:

$$\alpha_t = exp(e_t)\ /\Sigma_k exp\ (e_k) \quad (8)$$

Here, $e_t$ represents the relevance between the current and past hidden states. $\alpha_t$ is the weight assigned to each input time step. The hybrid model integrates CNN, BiLSTM, and the attention mechanism to discover both short-term trends and long-term dependencies in the training data. The CNN layer extracts spatial patterns, while the BiLSTM captures both forward and backward dependencies from the time series. The attention mechanism then refines the predictions by emphasizing the most critical historical data points. Figure 1 visually outlines this framework. The model framework consists of:

- Input Layer. Receives the historical price data of BTC, ETH, and LTC.
- CNN Layer. Extracts short-term features.
- BiLSTM Layer. Captures long-term dependencies bi-directionally.
- Attention Layer. Focuses on the most relevant time steps.
- Output Layer. Generates the predicted cryptocurrency prices.

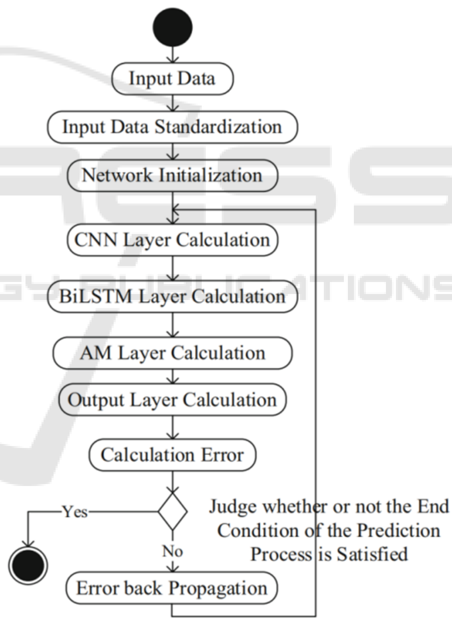


Figure 1: Model framework of CNN-BiLSTM-AM (Photo/Picture credit: Original).

The parameters' settings of the CNN-BiLSTM-AM model are as follows. The convolution layer uses 64 filters and 3 kernels. The activation function for both convolution layer and pooling layer is ReLU. The pool size is 1. The number of BiLSTM hidden units is 64. The training process consists of the following parameters: batch size of 64, time step of 5, learning rate of 0.001, epoch of 100, loss function of

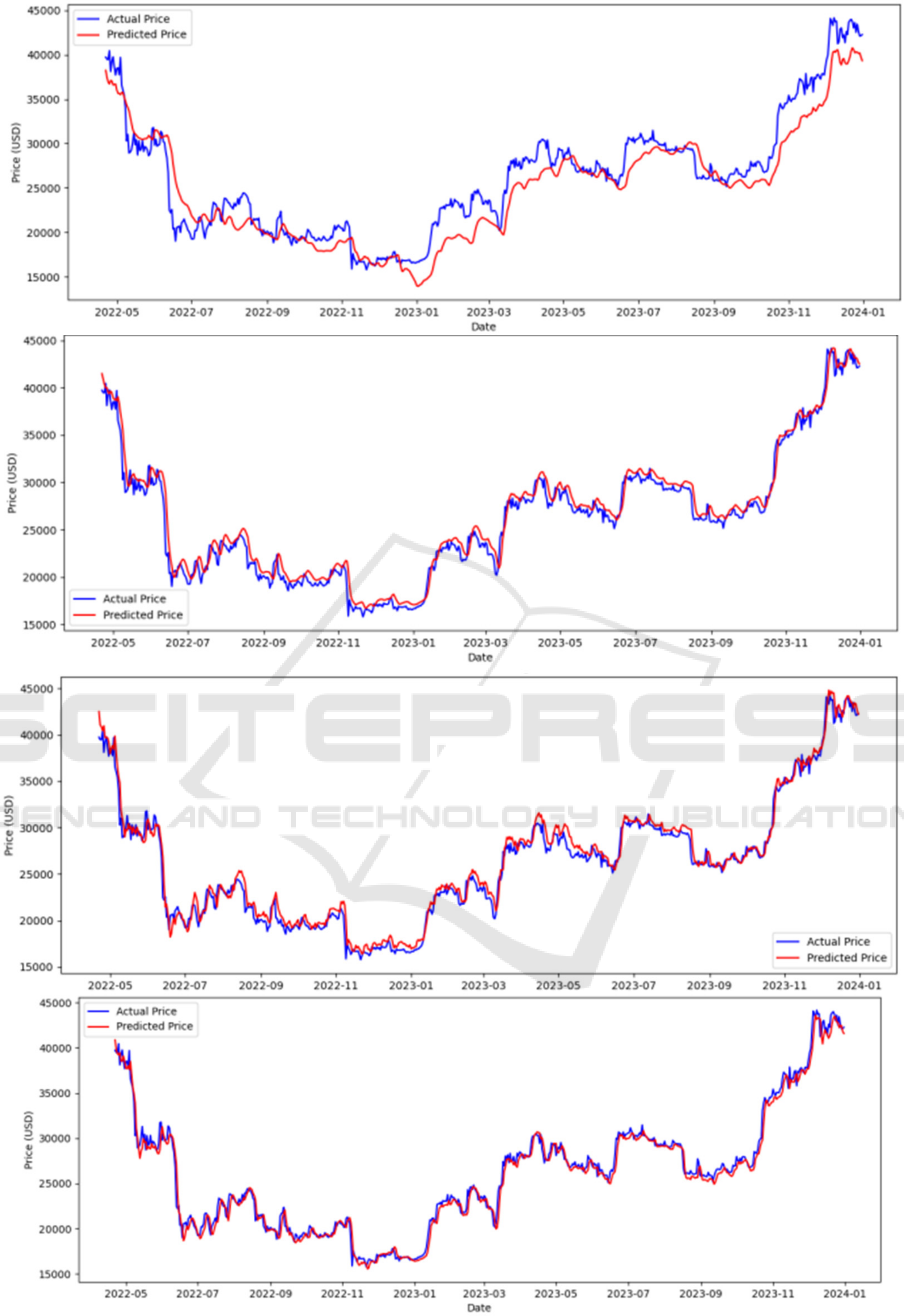




Figure 2: The CNN, LSTM, CNN-BiLSTM and CNN-BiLSTM-AM for BTC (left to right and then upper to lower) forecasted and actual BTC prices (Photo/Picture credit: Original).

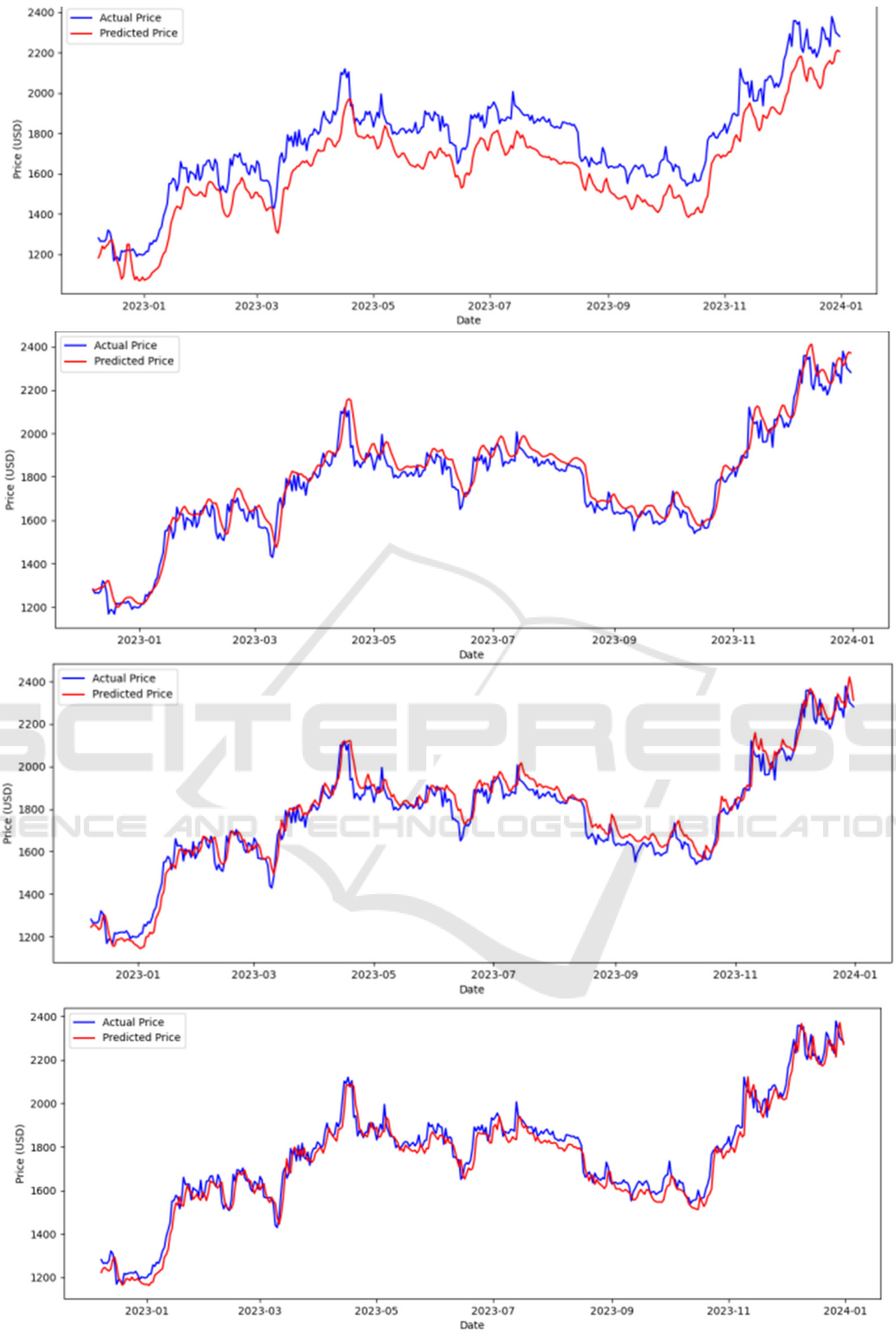


Figure 3: The CNN, LSTM, CNN-BiLSTM and CNN-BiLSTM-AM for ETH (left to right and then upper to lower) forecasted and actual BTC prices (Photo/Picture credit: Original).

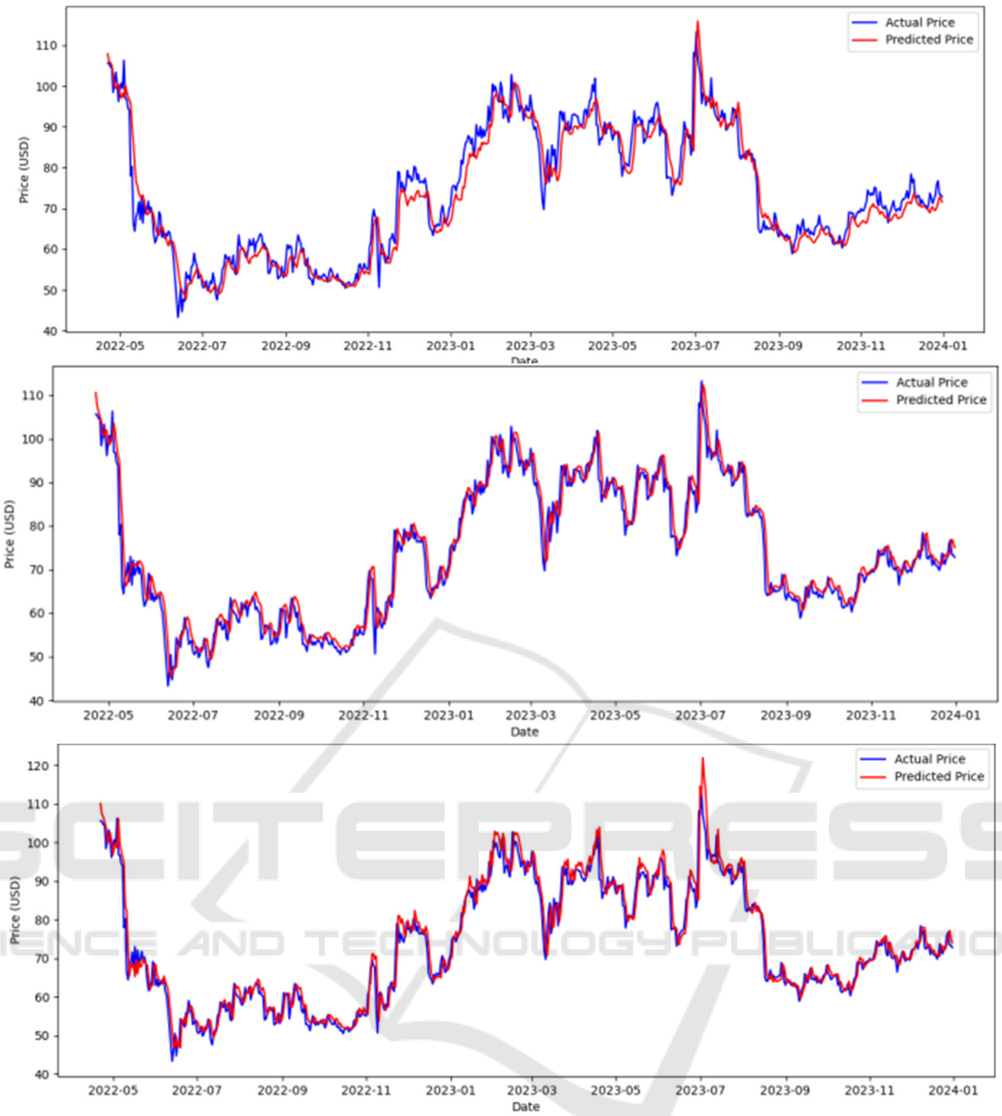




Figure 4: The CNN, LSTM, CNN-BiLSTM and CNN-BiLSTM-AM for LTC (left to right and then upper to lower) forecasted and actual BTC prices (Photo/Picture credit: Original).

MAE, and optimizer selection of Adam. All of the four models' training parameters are the same.

The research uses MAE, RMSE and $R^2$ as the evaluation indexes to assess the performance of different models. They are calculated as follows:

$$MAE = \Sigma_i |y_i - \hat{y}_i|/n \tag{9}$$

$$RMSE = \sqrt{\Sigma_i (y_i - \hat{y}_i)^2/n} \tag{10}$$

$$R^2 = 1 - (\Sigma_i (y_i - \hat{y}_i)^2)/(\Sigma_i (y_i - \bar{y})^2) \tag{11}$$

In these formulae, $y_i$ is the real price and $\hat{y}_i$ is the forecasted price. Lower values of MAE and RMSE, along with an $R^2$ closer to 1, means the model has a better performance.

# 3 RESULTS AND DISCUSSION

## 3.1 Model Performances

CNN, LSTM, CNN-BiLSTM and CNN-BiLSTM-AM were applied in this experiment. The performance of different models is evaluated by using each of them to predict cryptocurrency prices for BTC, ETH, and LTC. The prediction time range for BTC and LTC is from May 2022 to January 2024. Since ETH was first issued in July 2014, according to the ratio of 8:2 for training and testing sets, the prediction time range for ETH is from January 2023 to January 2024. The results for BTC, ETH and LTC are shown in Figure 2, Figure 3 and Figure 4.

As can be seen from the results, CNN-BiLSTM-AM, CNN-BiLSTM, LSTM, and CNN rank highest to lowest in terms of the degree of fitting for broken lines of the forecasted prices to the true prices for all three cryptocurrencies. CNN-BiLSTM-AM always has the highest and nearly perfectly coincident degree of fitting for broken lines of its predicted prices to real prices, whereas CNN always has the lowest. Based on each model's predicted prices and actual cryptocurrency prices, the evaluation metrics of each model can be calculated easily. Tables 1, 2, and 3 display the comparison results of the twelve studies for BTC, ETH and LTC, respectively. For all three cryptocurrencies, the CNN-BiLSTM-AM model continuously beats CNN, LSTM, and CNN-BiLSTM in terms of three evaluation metrics. These results demonstrate that combining CNN with BiLSTM networks and the Attention Mechanism improves prediction accuracy, with lower RMSE , MAE values and $R^2$ closer to 1.

Table 1: Comparison of evaluation error indices of different models for BTC.

| Model | RMSE | MAE | R2 |
|---|---|---|---|
| CNN | 2375.992 | 1884.418 | 0.873 |
| LSTM | 1076.398 | 822.402 | 0.974 |
| CNN-BiLSTM | 949.195 | 701.529 | 0.980 |
| CNN-BiLSTM-AM | 844.350 | 593.538 | 0.984 |

Table 2: Comparison of evaluation error indices of different models for ETH

| Model | RMSE | MAE | R2 |
|---|---|---|---|
| CNN | 151.266 | 140.014 | 0.641 |
| LSTM | 60.618 | 47.942 | 0.942 |
| CNN-BiLSTM | 55.770 | 44.222 | 0.951 |
| CNN-BiLSTM-AM | 52.909 | 40.977 | 0.956 |

Table 3: Comparison of evaluation error indices of different models for LTC.

| Model | RMSE | MAE | R2 |
|---|---|---|---|
| CNN | 4.095 | 2.904 | 0.928 |
| LSTM | 3.488 | 2.418 | 0.948 |
| CNN-BiLSTM | 3.805 | 2.684 | 0.937 |
| CNN-BiLSTM-AM | 3.426 | 2.323 | 0.950 |

## 3.2 Comparison

By comparing the prediction accuracy of all the four models on cryptocurrency price data, significant model performance difference can be observed. Specifically, the model's capability of capturing temporal dependencies in time series data is much improved with the addition of BiLSTM. In comparison, the basic CNN model performed the worst in the prediction of BTC, ETH and LTC, with highest RMSE and MAE values, while the $R^2$ is relatively low. This phenomenon shows that although CNN performs well in extracting short-term spatial features, it is insufficient for processing time series data when long-term trends need to be captured. The LSTM performs much better than CNN, especially in the prediction of ETH and LTC. Furthermore, by introducing BiLSTM, the model performance has been further improved. The capacity of BiLSTM to concurrently analyze data in both forward and backward directions sets it apart from other models. This trait significantly improves the model's comprehension of long-term interdependence. The CNN-BiLSTM-AM model performed best on all evaluation indicators and cryptocurrencies. The addition of AM enables the model to concentrate on the most influential time nodes in the data, thereby more accurately capturing key points and long-term trends in price data. It is worth mentioning that this model has achieved particularly outstanding results in predicting BTC, with the RMSE value reduced to 844.350 and the $R^2$ value as high as 0.984, fully verifying its excellent performance and applicability.

## 3.3 Explanation and Implications

The study's findings indicate that the CNN-BiLSTM-AM model outperforms other models in applications that predict cryptocurrencies' prices. Its superiority is mainly attributed to the model's strong capacity to grab space and time dependence characteristics. Specifically, with the effective extraction of significant characteristics from the original input data, CNN creates a well-prepared data base for the ensuing analysis procedure. The introduction of the BiLSTM layer gives the model the capacity to simultaneously observe forward and backward event sequences, significantly enhancing its understanding of dynamic changes in time. The model may intelligently concentrate on the time nodes that have the greatest impact on future price fluctuations thanks to the integration of AM. By integrating all the advantages of the three methods, the accuracy of cryptocurrency price prediction is significantly improved.

### 3.4 Limitations and Prospects

Despite the CNN-BiLSTM-AM model shows significant advantages in prediction accuracy of cryptocurrency price, its limitations cannot be ignored. On one hand, the model is prone to be affected by the quantity and quality of input data. Given the nature of the cryptocurrency market being susceptible to multiple external factors, if the training data fails to fully capture these dynamic factors, the predictive robustness of the model will be greatly compromised. On the other hand, a significant limitation lies in the high computational complexity of the model. Integrating CNN, BiLSTM layers and AM results in a sharp increase in resource consumption. This not only prolongs the training cycle, especially when processing massive data, but also places higher demands on computing resources. Looking forward, future researchers can focus on the introduction of diversified training data. For example, the model's sensitivity and forecasting ability to sudden market changes can be significantly improved by taking into account the macroeconomic indicators and market sentiment analysis. Additionally, advanced technologies such as reinforcement learning can be combined to enhance the model's capability to adapt to market changes in real time. Meanwhile, expanding the application field of this hybrid model to other financial markets such as stocks and commodities, through cross-market verification, will not only further evaluate its universality and effectiveness, but also bring more innovation to other fields of financial market prediction.

## 4 CONCLUSIONS

To sum up, asset price forecasting is crucial in financial investment and decision activities. Given that cryptocurrency is a kind of financial assets with high volatility, accurate prediction is particularly challenging in this field. This paper innovatively raises a hybrid CNN-BiLSTM-AM model to forecast the cryptocurrency price of the next day. The research selects three widely known cryptocurrencies (BTC, ETH and LTC) to test different models' performance. Results indicate that the CNN-BiLSTM-AM model ranks first compared with CNN, LSTM, and CNN-BiLSTM according to prediction accuracy. This finding proves the superiority of this hybrid model in processing cryptocurrency price data and provides new ideas for subsequent research and practical applications. In conclusion, this study not only contributes new methods and insights to the prediction technology of the cryptocurrency market, but also provides valuable reference and inspiration for researchers in other related fields.

## REFERENCES

Chen, Q., Zhang, W., Lou, Y., 2020. *Forecasting stock prices using a hybrid deep learning model integrating attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network.* IEEE Access, 8, 117365-117376.

Hochreiter, S., Schmidhuber, J., 1997. *Long Short-Term Memory*. MIT Press, 9(8), 1735-1780

Lecun, Y., Botou, L., Bengio, Y., Haffner, P., 1998. *Gradient-based learning applied to document recognition.* Proc IEEE, 86(11), 2278–2324

McNally, S., 2016. Predicting the price of Bitcoin using Machine Learning. Doctoral dissertation, Dublin, National College of Ireland.

Seabe, P. L., Moutsinga, C. R. B., Pindza, E. (2023). *Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach.* Fractal and Fractional, 7(2), 203.

Shah, D., Zhang, K., 2014. *Bayesian regression and Bitcoin.* 52nd annual Allerton conference on communication, control, and computing (Allerton), 409-414.

Ta, V., Liu, C., Tadesse, D., 2020. *Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading.* Appl Sci, 10(2), 437–456

Treisman, A., Gelade, G., 1980. *A feature-integration theory of attention.* Cogn Psychol 12(1), 97–146

Zhang, G., 2003. *Time series forecasting using a hybrid ARIMA and neural network model*. Neurocomputing 50, 159–175.

Zhang, Z., Dai, H. N., Zhou, J., Mondal, S. K., García, M. M., Wang, H., 2021. Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels. Expert Systems with Applications, 183, 115378.