Implementations of Deep Learning in Cryptocurrency Based on Comparisons of Various Scenarios

Zhanyu Li[1*], Siwei Mao[2]

[1]CATs, Shanghai, China
[2]School of electrical engineering, Xi'an university of technology, Xi'an, China
*lizhanyu8888@gmail.com

Abstract. As a matter of fact, the cryptocurrency market attracts plenty of investors on account of its ultra-high volatility. To realize accurate prediction of the price for cryptocurrency, lots of state-of-art machine learning scenarios are applied. With this in mind, this study explores deep learning's role in Bitcoin price prediction, focusing on LSTM, GRU, CNN and transformer models. These models excel at identifying long-term patterns and nonlinear features in time series data. In reality, bitcoin's volatility makes price prediction challenging, as traditional methods fall short. According to the analysis, LSTM and GRU address gradient issues in RNNs, improving accuracy through specialized gating mechanisms. Hybrid models like DL-GuesS, LSTM with Self-Attention, IFA-BiLSTM, and MFB are also examined. Transformer Frameworks are applied for prediction and achieve good results. These integrate multi-source data (e.g., historical prices, social media sentiment, news) to enhance prediction accuracy. The MFB model, combining BiLSTM and BiGRU with BorutaShap for feature selection, represents a cutting-edge approach. Overall, these results shed light on guiding further exploration of cryptocurrency price prediction.

Keywords: Cryptocurrency prediction, CNN, LSTM, GRU, self-attention, Transformer.

1 Introduction

Deep learning has revolutionized artificial intelligence, achieving remarkable success across domains including image recognition, natural language processing, and speech processing. Its potential in financial applications, particularly for cryptocurrency markets, has attracted significant research interest. As the pioneering cryptocurrency, Bitcoin presents unique challenges for price prediction due to its extreme volatility and complex market dynamics. Traditional forecasting models often fail to capture these nonlinear patterns effectively, creating opportunities for deep learning approaches to provide more accurate predictions [1, 2].

This paper systematically examines deep learning models for Bitcoin price forecasting, beginning with fundamental architectures like LSTM and GRU [3]. We explore how neural network variants, including convolutional and recurrent networks, can be adapted for financial time series analysis. Special attention is given to hybrid

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_77

approaches and advanced feature selection techniques that enhance prediction accuracy.

The cryptocurrency market's exponential growth has intensified demand for reliable price forecasting tools. Bitcoin's blockchain technology has not only transformed digital currency concepts but also advanced distributed ledger systems [4]. However, its technical limitations, such as low transaction throughput (approximately 7 transactions per second), create unique market behaviors that challenge conventional analysis methods. Deep learning models, particularly LSTM networks, have demonstrated superior capability in identifying long-term patterns in Bitcoin's volatile price movements by incorporating diverse inputs: historical price data, trading volume metrics, sentiment analysis from social media, and key macroeconomic indicators.

Our investigation extends to cutting-edge architectures including Transformer-based models and innovative hybrid systems like DL-GuesS, MFB, and IFA-BiLSTM. These approaches leverage multi-source data integration and advanced optimization techniques to achieve unprecedented prediction robustness, addressing critical gaps in cryptocurrency market analysis.

2 Models

2.1 Convolutional Neural Networks

CNNs are mainly used to process data with multilayered structure [5]. The traditional convolutional neural network consists of four parts: convolution layer, pooling layer, fully connected layer and output layer (seen from Fig. 1) [7]. Convolution is the core of CNN. The convolution layer consists of multiple feature planes, and extracts array features through a set of weights called filters on the feature plane. Features use the same weights on the same feature plane. CNN has the characteristics of local connection, weights sharing, pooling and multilayered structure [6], which reduces the calculation parameters and the calculation difficulty of the network through reducing the complexity of the network, extracting part of the image features, and sharing weight parameters. The pooling operation makes the input change have a certain invariance and enhances the robustness of the network.

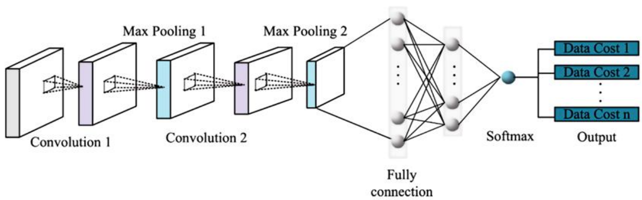

Fig. 1. Architecture of CNN [7].

2.2 Recurrent Neural Networks

RNNs are mainly used to solve continuous and related data problems. Its characteristics are that it can pass part of the features of the previous layer of data to the current layer, and can also pass part of the features of the current layer of features to the next layer, so as to realize the connection between the past, present and future. The recurrent neural network includes input layers, hidden layers and output layers. Its architecture is shown in Fig. 2 [5]. However, due to the issues of gradients vanishing and exploding in traditional RNN, we need other RNN derivative models to deal with it, like Long Short-Term Memory Networks (LSTM) and Gated Recurrent Units (GRU). The following will introduce these two models.

Unlike general RNN, the hidden layer of the Long Short-Term Memory Networks is composed of a series of LSTM units. The core of the LSTM unit is the input gate, forget gate and output gate, as shown in Fig. 3 [8]. These series of "gate" operations can solve the vanishing and exploding gradient problems very well, especially in long-distance transmission, its performance is far better than RNN.

GRU is based on LSTM and simplifies the structure. The input gate, forget gate are simplified to update gate, and output gate are transformed to reset gate. Fig. 4 shows the architecture of GRU. The advantage of GRU is that it not only solves the problem of the vanishing and exploding gradient, but also has simpler structure than LSTM. From a computational point of view, the model is more efficient and can save computational costs [8].

2.3 Transformer Frameworks

Transformer is a different model based on Self-Attention mechanism, abandoning the structure of LSTM and CNN [9]. It is composed of encoder and decoder. The encoder has a series of layers. Every layer has two parts. One is multi-head self-attention. The other one is feed-forward networks. The decoder also has series of layers. It adds a third part, masked multi-head attention, based on the encoder layer. Fig. 5 shows the structure of Transformer. The transformer lacks the ability for sequence [9]. Positional encoding is an important element, which helps the model deal with the problem. Compared to CNN and LSTM, transformer has a faster training speed and better model performance. But it also needs more data and computing power.

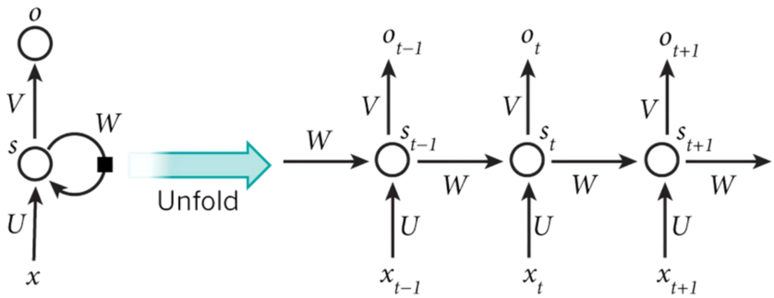

Fig. 2. Architecture of recurrent neural networks [5].

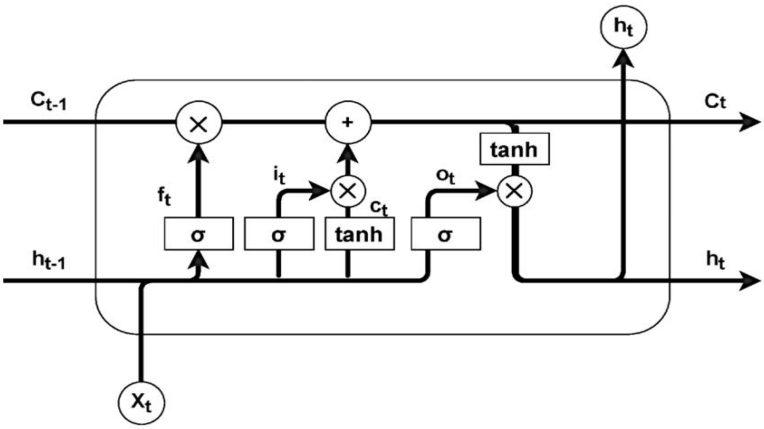

Fig. 3. Architecture of LSTM network [8].

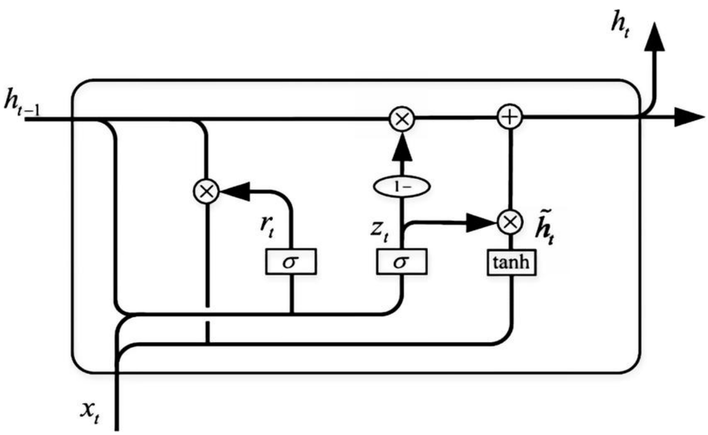

Fig. 4. Architecture of the GRU network [8].

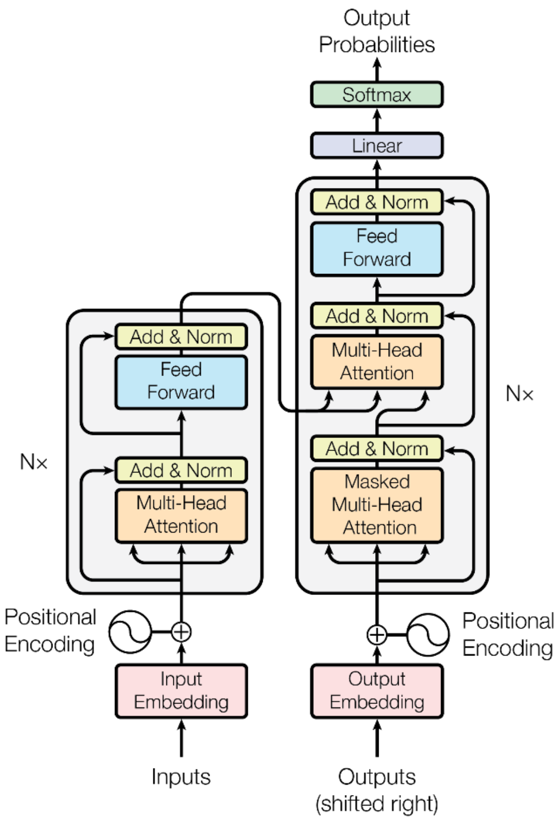

Fig. 5. Architecture of the Transformer [9]

3 Applications in Cryptocurrency

3.1 DL-GuesS

Previous study proposed a hybrid cryptocurrency price prediction model based on the Deep Learning and Sentiment (DL-GuesS). This model uses both historical price and Twitter text data for price prediction [10]. In the performance evaluation phase, predictions are made for Dash and Bitcoin respectively, and evaluates them through MSE, MAE, MAP. The authors also compared it with other prediction models. In Dash price prediction, DL-GuesS performs the best in the three indicators. Table 1 shows the differences in loss of each model. In Bitcoin-Cash Price Prediction, DL-GuesS also has the least loss among the three models. The data proves the advantage of the hybrid model (DL-GuesS). It has a higher accuracy than general models without text sentiments.

Table 1. Models and Loss

Model	Loss		
	MSE	MAE	MAPE
Dash	0.0208	0.0905	4.8102
Dash+Litecoin+Bitcoin	0.0205	0.0877	5.4018
DL-GuesS	0.0185	0.0805	4.7928
Bitcoin-Cash	0.0015	0.0224	4.6256
Bitcoin-Cash+Litecoin+Bitcoin	0.0017	0.0218	4.4449
DL-GuesS	0.0011	0.0196	4.4089

3.2 LSTM and Self Attention

Some Researchers added a self-attention mechanism to the general LSTM architectures [11]. The self-attention mechanism allows important statistics to be extracted from historical data, thereby making more accurate and adaptable predictions. The authors compared LSTM and LSTM with self-attention in training and testing, and the latter outperformed the former in both aspects. Table 2 shows the Self LSTM is obviously better than the regular LSTM in both training and testing. This reflects its reliability and effectiveness in prediction.

3.3 MFB

Others proposed the Multimodal Fusion Bitcoin (MFB), which combine BiLSTM and BiGRU [12]. The data used is composed of Bitcoin price data, tweets and news data. At the same time, the authors used the BorutaShap algorithm to further improve the feature selection process. They compared MFB with different kinds of models and got testing and validation loss images. Fig. 6 shows images of each model. MFB (e) has no obvious upward trend in validation loss, reflecting its ability to avoid overfitting. In terms of indicators, MFB almost outperform all the others, which underscore its accuracy (as given in Table 3). The authors also conducted experiments to verify the stability with a library called GroupTimeSeriesSplit. Table 4 shows the comparison of indicators for regular MFB and MFB with GTSS. The differences between both models are small in different aspects. It indicates the model possess excellent universality and stability.

3.4 IFA-BiLSTM

Some researchers used Improved Firefly Algorithm (IFA) and Bidirectional Long Short Term Memory (BiLSTM) to predict the cryptocurrency price, and made a comparison based on the criterion of RMSE and MAPE with other models [13]. Table 5 shows the comparison between LSTM,BiLSTM and BiLSTM with IFA. The IFA model outperforms among the models mentioned. This proves it has good accuracy and stability in terms of price prediction.

3.5 Transformer

Some researchers also used transformer models. They chose a few of different transformer like Standard Transformer 、 Informer 、 Autoformer, and made comparisons among them [14]. The models were trained by preprocessed relevant data. Each model is evaluated through a series of indicators, such as annualized return (AR), Sharpe ratio (SR). They get score by the weighted average of the standardized indicators and are ranked by the score. Table 6 shows the score and rank of each model. The Convolutional Transformer model ranks the first and the LogTrans model receives the lowest score. The Convolutional Transformer has advantages in cryptocurrency price prediction.

Table 2. Evaluation metrics for LSTM

Models	Training				Testing			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
LSTM with Self Predictions	0.02208	0.00107	0.03265	0.99496	0.02963	0.00180	0.04245	0.99284
LSTM	0.04405	0.00448	0.06690	0.97882	0.04476	0.00442	0.06652	0.98242

Table 3. Models and indicators.

Metrics	LSTM	GRU	Bi-LSTM	Bi-GRU	MFB
MAE	0.0070	0.0067	0.0070	0.0071	0.0065
MSE	9.6547e-05	9.2945e-05	0.0001	0.0001	8.2925e-05
R^2	0.6946	0.7060	0.5992	0.6787	0.7377
RMSE	0.0098	0.0096	0.0113	0.0101	0.0091
MAPE	2.5983	3.0445	2.7636	2.8028	2.7157

Table 4. Metrics comparison for MFB

Metrics	MAE	MSE	R^2	RMSE	MAPE
MFB with GTSS	0.0083	0.0001	0.7226	0.0093	2.6673
MFB without GTSS	0.0065	8.2925e-05	0.7377	0.0091	2.7157

Table 5. Prediction outcome and model comparison

Model	RMSE	MAPE	Accuracy
LSTM	2300.50	4.66%	95.34%
BiLSTM	2121.36	4.48%	95.52%
IFA-BiLSTM	2051.55	3.48%	96.52%

Table 6. General evaluation results

Frameworks	Score	Rank
Convolutional	7.50445	1
TFT	6.38027	2

Standard	4.69662	3
XL	4.37938	4
Informer	4.19039	5
Autoformer	3.48504	6
Sparse	2.61425	7
LogTrans	-2.46035	8

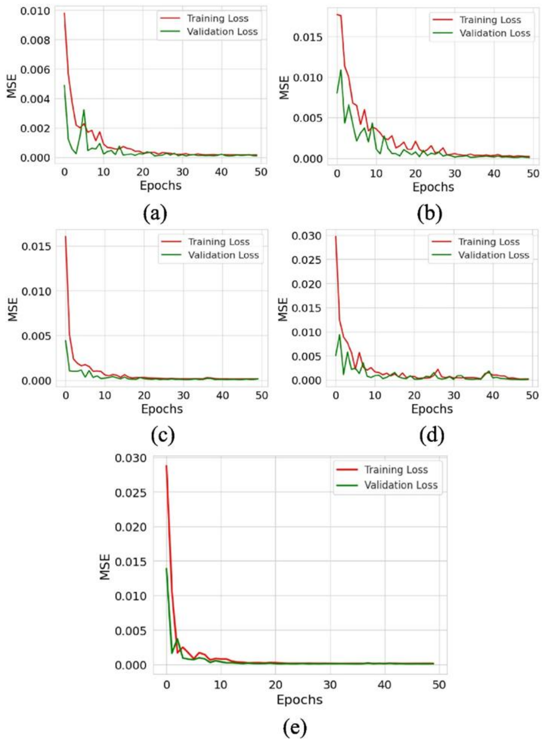

Fig. 6. Loss curves of different models [12].

4 Conclusion

Deep learning has demonstrated remarkable potential for cryptocurrency price prediction, overcoming limitations of traditional methods through advanced architectures. Our analysis reveals that LSTM networks, with their unique ability to capture long-term temporal dependencies, provide particularly effective frameworks for modeling volatile cryptocurrency markets. The evolution from basic RNN variants to sophisticated hybrid models (DL-GuesS, MFB) and Transformer-based approaches has significantly improved forecasting accuracy.

Future advancements should prioritize: (1) multi-source data integration combining market indicators with alternative data streams, (2) development of lightweight models for real-time trading applications, and (3) implementation of privacy-preserving techniques like federated learning. These directions promise to enhance both the precision and practicality of cryptocurrency prediction systems, providing market participants with more robust analytical tools.

Author Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Liu, H., Ma, B., Qin, L., Pang, J., Zhang, C., Huang, Q.: 'Set-label modeling and deep metric learning on person re-identification', Neurocomputing, 2015, 151, pp. 1283-1292.
2. Feng, L.: 'Data Analysis and Prediction Modeling Based on Deep Learning in E-Commerce', Scientific Programming, 2022, 1, p. 1041741.
3. He, Y., Chen, X., Miao, D., Zhang, H., Qin, X., Du, S., Lu, P.: 'Graph-enhanced anomaly detection framework in multivariate time series using Graph Attention and Enhanced Generative Adversarial Networks', Expert Systems with Applications, 2025, 271, p. 126667.
4. Li, C., Huang, Z.: 'The development process and risks for cryptocurrency', Social Science of China, 2022, 6, pp. 168-174.
5. LeCun, Y., Bengio, Y., Hinton, G.: 'Deep learning', Nature, 2015, 521, (7553), pp. 436-444.
6. Zhou, F., Jin, L., Dong, J.: 'Review of RNN', Acta of Computer Science, 2017, 40, (06), pp. 1229-1251.
7. Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., Parmar, M.: 'A review of convolutional neural networks in computer vision', Artificial Intelligence Review, 2024, 57, (4), p. 99.
8. Mienye, I. D., Swart, T. G., Obaido, G.: 'Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications', Information, 2024, 15, (9), p. 517.
9. 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.
10. Parekh, R., Patel, N. P., Thakkar, N., et al.: 'DL-GuesS: Deep learning and sentiment analysis-based cryptocurrency price prediction', IEEE Access, 2022, 10, pp. 35398-35409.
11. Pavani, G., Vyshnavi, K. S., Samhitha, M.: 'Cryptocurrency Price Prediction using LSTM with Self Attention', 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL) pp. 306-311.
12. Han, P., Chen, H., Rasool, A., Jiang, Q., Yang, M.: 'MFB: a generalized multimodal fusion approach for bitcoin price prediction using time-lagged

sentiment and indicator features', Expert Systems with Applications, 2025, 261, p. 125515.

13. Pratama, A. (2024, September). Bitcoin Cryptocurrency Price Prediction Using IFA-BiLSTM. In 2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA) (pp. 1-6). IEEE.
14. Yin, C., & Nisheva-Pavlova, M. (2024, June). From Features to Decisions: Analysis of Transformer Frameworks for Bitcoin Price Prediction. In 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA) (pp. 427-433). IEEE.