Research Article

Bitcoin Price Forecasting: An Integrated Approach Using Hybrid LSTM-ELM Models

Changqing Luo,1 Lurun Pan,1 Binwei Chen,2 and Huiru Xu1

1Finance School, Hunan University of Technology and Business, Changsha 410205, Hunan Province, China
2School of Economics, University of East Anglia, Norwich, UK

Correspondence should be addressed to Lurun Pan; 1539348626@qq.com

Received 19 July 2022; Revised 9 October 2022; Accepted 26 October 2022; Published 9 November 2022

Academic Editor: Junwei Ma

Copyright © 2022 Changqing Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, digital currencies have flourished on a considerable scale, and the markets of digital currencies have generated a nonnegligible impact on the whole financial system. Under this background, the accurate prediction of cryptocurrency prices could be a prerequisite for managing the risk of both cryptocurrency markets and financial systems. Considering the multiscale attributes of cryptocurrency price, we match the different machine learning algorithms to corresponding multiscale components and construct the ensemble prediction models based on machine learning and multiscale analysis. The Bitcoin price series, respectively, from 2017/11/24 to 2020/4/21 and 2020/4/22 to 2020/11/27, is selected as the training and prediction datasets. The empirical results show that the ensemble models can achieve a prediction accuracy of 95.12%, with better performance than the benchmark models, and the proposed models are robust in upward and downward market conditions. Meanwhile, the different algorithms are applicable for components with varying time scales.

1. Introduction

In recent years, cryptocurrency has developed quickly around the world. According to CoinMarketCap, the total number of digital currencies in the world reached 1.93 trillion dollars as of April 2nd, 2021, of which the Bitcoin market value is 1.10 trillion dollars, accounting for 57.12% of the total market value of digital currencies. Unlike fiat currency, the price of cryptocurrency fluctuates sharply. For example, the Bitcoin price exceeded $4,000 in early September 2017, rising as high as $19,800 at the end of 2017, increasing 400%. While in January 2018, its price fell to $9,800, with a drop of 50.5%. As Bitcoin has become one of the most important financial assets in the financial markets, Bitcoin futures were officially listed and traded on the Chicago Mercantile Exchange (CME) in 2017. Considering the importance of cryptocurrency, The People’s Bank of China focused on promoting the research and development and pilot work of Central Bank Cryptocurrency in 2020, and the nation-owned banks, such as Bank of China, Agricultural Bank of China, and China Construction Bank, have tested digital currency wallets. Powell, The Chairman of the Federal Reserve, acknowledged in 2020 that Central Bank Digital Currencies (CBDCs) might improve the payment system in the United States. In May 2020, The Swiss Financial Markets Supervisory Authority (FINMA) authorized InCore Bank to allow the economy, custody, transfer, and development of digital assets, making the bank the first commercial bank that offered digital assets services in the world. In this context, accurate prediction of cryptocurrency prices can provide investors with decision support in portfolio management and risk management and provide a basis for the government to formulate regulatory policies. Therefore, providing accurate price predictions of cryptocurrency has become an essential task of academia and the industry.

Artificial neural networks (ANN), support vector machines (SVM), deep learning (DL), and other artificial intelligence technologies have gradually been applied to quantitative investment, risk management, and other financial fields. Because artificial intelligence technology is a data-driven method that is adaptive to the nonlinear features
of financial data, it can effectively improve the performance of predictive models. Considering the nonlinearity of cryptocurrency price, we decompose price time series into multiple time-scale series: high-, medium-, and low-frequency components. Then, we match each component with different deep learning algorithms and construct the ensemble models by integrating each component predicted from heterogeneous algorithms on different time scales. The Bitcoin price series, respectively, from 2017/11/24 to 2020/4/21 and 2020/4/22 to 2020/11/27, is selected as the training and prediction datasets. The empirical results show that the ensemble models can achieve a prediction accuracy of 95.12%, with better performance than the benchmark models. The proposed models are robust in the upward and downward market conditions. Meanwhile, the different algorithms are applicable for components with varying time scales.

The research on predictions of cryptocurrency prices is still in the nascent stage (Sun Yin et al. [1] and Guo et al. [2]). Based on the multiscale analysis and deep learning methods, we propose a prediction model that boosts the prediction accuracy for Bitcoin price. Our study extends the current literature from two aspects. First, under the new perspective of frequent analysis, we decompose the original Bitcoin price series and reconstruct the subseries into high-, medium-, and low-frequency components based on their similarity and complexity, thus, providing an understanding of a microscopic scale for Bitcoin price fluctuations. Second, the deep learning method is incorporated into the prediction system, and the advantages of selected deep learning methods and the features of the high-, medium-, and low-frequency components are matched to enhance the model performance.

The rest of the study is organized as follows. Section 2 reviews the relevant literature. Section 3 is the construction of ensemble predictive models for cryptocurrency prices based on multiscale analysis and deep learning. Section 4 conducts the empirical study and analyzes the results. Section 5 is the conclusions and implications.

2. Review of Relevant Literature

Two strands of literature have investigated the prediction method of cryptocurrency prices. The first category is econometric models, and the second is AI-based methods. More studies have shown the AI-based method’s competitive advantages in recent years. Meanwhile, some literature also demonstrates that the multiscale analysis could improve the prediction models’ performance for time series.

The econometric models include Autoregressive Integrated Moving Average Model (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and other time-series models. Ibrahim et al. [3] use ARIMA to predict the price trend of Bitcoin for the next five minutes to provide advice for high-frequency traders. Tan et al. [4] construct the Markov Switching GARCH model (MSGARCH) to predict the dynamic fluctuations of Bitcoin’s return series and show that the MSGARCH model has the highest goodness-of-fit. Aras [5] and other studies also show that the econometric models are suitable for predicting the price of digital currencies. However, most econometric models require the data to follow a specific distribution, and price data usually exhibit nonlinear characteristics, constraining the further improvement of the econometric models.

To overcome these limitations, some scholars have developed the second strand of models, which are artificial intelligence models, including ANN, support vector regression (SVR), Long Short-Term Memory networks (LSTM), and other emerging methods to predict Bitcoin price. McNally et al. [6] find LSTM is better than ARIMA. Indera et al. [7] accurately predict the price of Bitcoin by building a nonlinear autoregressive model with exogenous inputs based on a multilayer perceptron (MLP). Nakano et al. [8] explore Bitcoin intraday earnings forecasting methods based on ANN. Silva de Souza et al. [9] and Cavalli and Amoretti [10] prove the convolutional neural network (CNN) model, and other machine learning methods can improve the investment performance of cryptocurrency. Basher and Sadorsky [11] use random forests (RF) to predict the Bitcoin price direction. Rathore et al. [12] applied machine learning methods to forecast closing prices using various features. The AI-based techniques used in the above literature allow hidden patterns to be extracted from large datasets without requiring prior knowledge of the data. Therefore, they can effectively handle nonlinear data and obtain a desirable prediction performance.

Due to the high accuracy of the AI-based model, other real-life applications besides the analysis of cryptocurrencies also apply AI-based methods for prediction. In the energy price or consumption area, Urolagin et al. [13] develop the multivariate LSTM model to predict the oil price. Dong et al. [14] use K-nearest neighbors (KNN) to forecast the electrical loading. Li et al. [15] and Iwabuchi et al. [16] use BP neural network, improved particle swarm optimization algorithm, wavelet transform, and LSTM to predict the electricity consumption or price. Chaturvedi et al. [17] forecast India’s total and peak monthly energy demand by using recurrent neural network (RNN) and other soft computing techniques. For predicting geohazards, Ma et al. [18] applied the decision tree C 5.0 and cluster algorithms to predict landslide. Later, Zhang et al. [19], Zhang et al. [20], Ma et al. [21], and Ma et al. [22] used different soft computing algorithms, such as ACO-SVR, CEEMD-LCSS, and ABC-SVR metaheuristic-based SVR to predict landslide displacements. As for the applications in fraud detection, Rtyali and Ennaya [23] used machine learning methods to catch fraudulent transactions. Liu et al. [24] constructed the credit risk prediction model based on XGBoost and a graph-based neural network. Achakzai and Juan [25] developed a classifier based on machine learning to detect financial frauds. AI-based methods are also applied in the field of disease detection and control; Ozt¨urk and Ozkaya [26] designed a classifier by using LSTM and CNN methods to diagnose gastrointestinal tract diseases. Kanipriya et al. [27] and Yudistira et al. [28] constructed different LSTM models to detect the malignant lung nodule or the growth of COVID-19 cases. Lee et al. [29] predicted Parkinson’s disease using gradient boosting
decision tree models (GBDT) and showed a better performance of GBDT model. Besides these areas, AI-based models are widely applied in nearly all fields. Most existing works have proven the advantages mentioned above of AI-based models. Therefore, due to the robustness, high accuracy, and adaptiveness of the AI-based models, the prediction performance for Bitcoin price could be enhanced by adopting these methods.

For a better understanding of the price movement of cryptocurrency, some recent studies also investigate the multiscale properties of cryptocurrency prices. Since some cryptocurrencies have monetary, commodity, and speculative attributes, their prices are affected by various factors and show heterogeneous features on short-, medium-, and long-term horizons. Corbet et al. [30] verify the time-frequency relationship between Bitcoin and traditional financial markets. Haffar and Le Fur [31] argue that the Bitcoin market is significantly affected by financial markets in the short term. In the long run, this impact is weaker, and they also find the price of Bitcoin is affected by various factors. Maghyereh and Abdoh [32] show that comovement between Bitcoin and the stock market is most significant in a long-term period, and the degree of correlation under different time scales is dynamic. The previous research show that there are multiscale features for cryptocurrency, while the existing literature has not paid enough attention to these features.

By incorporating the multiscale features of time series with machine learning, the ensemble models are applied to predict the financial time series. They have been proven to enhance predictive performance to some extent. Yu et al. [33] combine empirical mode decomposition (EMD) and neural networks to predict crude oil prices. Premanode and Toumazou [34] proposed differential EMD to improve exchange rate prediction using the SVR method. Bisoï et al. [35] combined reduced kernel extreme learning machine (RKELM) with variational mode decomposition (VMD) and compare trend prediction results with Naive Bayes classifiers, ANN, and SVM. The empirical results show that the VMD-RKELM model is superior to other prediction methods. Zhang et al. [20] showed the ensemble EMD method could improve the performance of the SVR model. Luo et al. [36] apply the multiscale analysis to investigate the financial time series and find new characteristics of financial risk contagion. Rayi et al. [37] used VMD and deep learning to forecast wind power and achieve high prediction accuracy. Yu et al. [38] used the rolling decomposition-ensemble model to predict gasoline consumption and statistically prove the effectiveness and robustness of the prosed model. Liu et al. [39] combined multiscale analysis with LSTM and other methods to forecast the carbon price and prove that the proposed model effectively predicts interval-valued carbon prices. By synthesizing the multiscale analysis and machine learning or deep learning, the nonlinearity and nonstationarity of the time-series data could be well handled for prediction. Therefore, a more accurate forecast could be achieved.

Overall, the price prediction of cryptocurrency is essential for portfolio management and risk supervision, and more and more studies use AI-based models to predict the cryptocurrency price and prove the effectiveness of the ensemble predictive model. Nevertheless, three main questions related to cryptocurrency need to be further explored. The first is that multiscale features of cryptocurrency prices have not been fully investigated. For a deeper understanding of price movement, obtaining the high-, medium-, and long-term components of cryptocurrency prices is essential. The second is which machine learning methods are more appropriate for different components of time series. Based on comprehending the features of multiscale components and matching the corresponding deep learning algorithms, the third is whether the ensemble predictive methods are also effective for predicting cryptocurrency prices. Given this, we investigate the multiscale attributes of cryptocurrency and select different intelligent algorithms to predict cryptocurrency prices according to the different characteristics of high and low frequencies, thus attempting to achieve a higher accurate prediction of cryptocurrency price.

3. Construction of Ensemble Models Based on Multiscale Analysis and Deep Learning

The construction of ensemble models based on multiscale analysis and deep learning is described in the following sections.

3.1. Procedures for Constructing the Ensemble Models. Since the cryptocurrency price has nonlinearity, such as aperiodicity and stochastics, it is appropriate to predict the price using nonlinear models. Compared with the staple econometric models, the deep learning models, which are data driven, are more efficient in dealing with the nonlinearity of the cryptocurrency price. Therefore, we use the deep learning models as the benchmark models to predict the prices. Meanwhile, we introduce the multiscale analysis during the model construction. By doing this, the ensemble model has two functions. The first is that the model can decompose the original price series into subcomponents with different frequencies, and the second is that the matchup of subcomponents and different deep learning models can boost the model’s performance. The specific procedures are displayed in Figure 1.

According to Figure 1, the first step is to decompose the original cryptocurrency prices into different subcomponents. The second step is to combine the subcomponents according to their similarities and frequencies by using sample entropy methods; thus, the high-, medium-, and low-frequency components are, respectively, reconstructed. Then, thirdly, according to Yu et al. [40] and Yu et al. [41], the LSTM and extreme learning machine (ELM) are probably more suitable for high- and low-frequency components. Thus, we use different algorithms to predict each component, and the final results are integrated by combining each output of subcomponents.

3.2. Decomposition and Reconstruction of Cryptocurrency Price. The multiscale analysis of Bitcoin price requires signal decomposition analysis methods. Commonly used signal
decomposition methods include Wavelet Analysis, EMD, Ensemble EMD (EEMD), and VMD. Islam et al. [42] used the EMD method to decompose financial time series and believes that the decomposition effect of EMD is better by comparing it with wavelet decomposition. Huang et al. [43] believed that VMD overcomes the problems of endpoint effects and modal mixing better, and the empirical results showed that VMD can improve the prediction accuracy compared with the EMD method. Compared with other decomposition methods, the VMD method can overcome the problems that often occur in the decomposition process, such as modal aliasing, improper enveloping, and boundary instability, so it still has good stability when dealing with nonlinear and unstable data. Although existing studies have shown that the VMD method has certain advantages, whether the method is suitable for Bitcoin price prediction has not been systematically tested in the current research, so in the Bitcoin price prediction, it is necessary to compare the EMD and VMD decomposition and prediction effects. Therefore, we use two decomposition methods of EMD and VMD to decompose Bitcoin prices and reconstruct the price components.

EMD is a signal decomposition method proposed by Huang et al. [44], which can decompose signals into characteristic modes. Its advantage is that it does not need to choose the basis function but adaptively generates the natural modal function according to the analysed signal. It can analyze nonlinear and nonstationary signal sequences and has a high signal-to-noise ratio. EMD can extract intrinsic patterns from the original time series of Bitcoin and express each intrinsic pattern as an intrinsic mode functions (IMF), which should meet the following two conditions:

1. The extreme value of the function and the number of zero crossings must be equal or differ by not more than 1
2. The average value of the upper and lower envelopes formed by the IMF’s local maximum and minimum equals 0

By EMD decomposition, the original Bitcoin sequence is decomposed into \( n \) IMFs and a residual, which is expressed as follows:

\[
X_t = \sum_{j=1}^{N} c_{j,t} + r_{N,t},
\]

where \( n \) is the number of IMFs, \( r_{N,t} \) is the residual, and \( c_{j,t} \) \((j = 1, 2, ..., n)\) is the \( j \)th IMF at time \( t \).

VMD is an effective method to deal with nonlinear and nonstationary signals. Applying the VMD method, the cryptocurrency price data are decomposed into bidimensional IMF (BIMF). VMD defines the IMF of the EMD method as narrow-band BIMF, that is,

\[
x_k(t) = A_k(t) \cos[\phi_k(t)],
\]

where \( \phi_k(t) \) is nondecreasing. We envelope \( A_k(t) \geq 0 \), and the transformation of \( A_k(t) \geq 0 \) and instantaneous frequency \( \omega_k(t) = \phi'_k(t) \) with respect to \( k(t) \) is a phase moderate variation.

VMD method puts the signal into the constrained variational model to find the optimal solution and determines the center frequency (Dragomiretskiy and Zosso [45]). The objective function is expressed as follows:
\[
\min_{\{x_i, \omega_i\}} \left\{ \sum_k \| \delta^k \left( \left( \delta(t) + \frac{j}{m} \right) \ast x_k(t) \right) e^{-j\omega_i t} \|^2 \right\},
\]

\[\text{s.t.} \sum K x_k(t) = x(t),\]

where \(x_k(t)\) and \(\omega_k\), respectively, represent the kth modal component of the signal and its corresponding center frequency, \(\{x_1, x_2, \ldots, x_k\}\) and \(\{\omega_1, \omega_2, \ldots, \omega_k\}\) represent all modes and their related center frequency, \(\delta(t)\) is the Dirac function, and \(*\) denotes the convolution. VMD is more stable than the EMD method and can deal with mode mixing and boundary points in modal processing. Thus, it helps improve the performance of processing the nonlinear and nonstationary Bitcoin price data.

After the decomposition of EMD and VMD, we can obtain different components of the original price series. Predicting each component can improve the prediction accuracy. However, the prediction error of each component may accumulate and lead to a relatively large aggregate error. Thus, reducing the number of components and reconstructing the IMFs is necessary. We select the sample entropy method to combine the decomposed components in this study. Sample entropy, proposed by Richman and Moorman [46], is a method to measure the complexity of time series. If the sequence autocorrelation is higher, the sample entropy will be smaller. The more complex the sequence, the larger the entropy value of the sample. For the original time series of Bitcoin \(x(n) = \{x(1), x(2), \ldots, x(N)\}\), the specific algorithm of sample entropy is as follows:

1. Sequence \(x_i\) is sequentially composed into m-dimensional vector; that is, \(X(i) = [x(i), x(i+1), \ldots, x(i+m-1)]\), where \(i = 1, 2, \ldots, N-M+1\).

2. The distance between \(X(i)\) and \(X(j)\) is as follows:
\[
d_m(X(i), X(j)) = \max_{0 \leq m \leq 1} |x(i+k) - x(j+k)|.
\]

3. Given the tolerance \(r(r > 0)\), we count the number of \(d_m(X(i), X(j))\) which is less than \(r\) for each \(i\) value, where \(1 \leq j \leq N-m\) and \(j \neq i\), and the number is denoted as \(B_m\), and then, we calculate the ratio of \(B_m\) to \(N-m\), represented as \(B_m^{m}(r)\):
\[
B_m^{m}(r) = \frac{1}{N-m-1} B_m.
\]

4. The mean value of \(B_m^{m}(r)\) is denoted as \(B_m^{m}(r)\):
\[
B_m^{m}(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} B_m^{m}(r).
\]

5. We increase the dimension to \(m+1\) and repeat steps (1)–(3); then, the mean of \(B_m^{m+1}(r)\) is as follows:
\[
B_m^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_m^{m+1}(r).
\]

The theoretical sample entropy of this time series is defined as follows:
\[
\text{SampEn}(m, r) = \lim_{N \rightarrow \infty} \left\{ \ln \left( \frac{B_m^{m+1}(r)}{B_m^{m}(r)} \right) \right\}.
\]

When \(n\) is a finite value, the estimated value of sample entropy is as follows:
\[
\text{SampEn}(N, m, r) = -\ln \left( \frac{B_m^{m+1}(r)}{B_m^{m}(r)} \right).
\]

According to formula (9), it can be seen that the values of \(m\) and \(r\) have a specific correlation with the calculated results of sample entropy, but the trend of increasing or decreasing the entropy value is not affected by \(m\) and \(r\). Generally, \(m = 2\) and \(r = 0.1 \sim 0.25\)SD (SD is the standard deviation of the original time series) is suitable. This study sets the parameters \(m = 2\) and \(r = 0.2\)SD. By judging the sample entropy value of components with different frequencies, the components with the nearest entropy or similar complexity could be integrated as a new time series representing a type of decomposed components.

3.3. Construction of the Forecasting Model Based on LSTM Algorithm. LSTM algorithm uses gate structure to process data, effectively utilizing the current and historical information in time-series data. The LSTM model has a unique network structure, so this method can better learn the past time-series data, find out the relationship between time series, and further dig out the inherent laws of time series by using the function of selective memory, thus having a good prediction effect. The primary mechanism is shown in Figure 2, and the pseudocode of the specific algorithm is illustrated in Figure 3.

As displayed in Figure 2, LSTM introduces input gate \((i_t)\), forget gate \((f_t)\), and output gate \((o_t)\) and cell activations \((C_t)\). \(C_t\) is the vector of new candidate values. All the three control gates and candidate states can be expressed as a function of the input eigenvalue \(x_t\) and the short-term memory function of the previous moment \(h_{t-1}\), which can be, respectively, expressed as follows:
\[
i_t = \sigma(W_i [h_{t-1}, x_t] + b_i),
\]
\[
f_t = \sigma(W_f [h_{t-1}, x_t] + b_f),
\]
\[
o_t = \sigma(W_o [h_{t-1}, x_t] + b_o),
\]
\[
C_t = f_t \ast C_{t-1} + i_t \ast \tilde{C}_{t-1},
\]
\[
\tilde{C}_t = \tanh(W_c [h_{t-1}, x_t] + b_c),
\]

where \(W_i, W_f, W_o\) represent the weights of the input data for the input gate, forget gate, and output gate, respectively, \(b_i, b_f, b_o\) are the bias weights for the input gate, forget gate, and output gate, respectively, and \(\sigma\) is the sigmoid activation function, which makes the
threshold range between 0 and 1. The three thresholds perform their duties, which determine the state in which feature information is stored, forgotten, and outputted, while the cell state and candidate state represents the short-term memory and the new knowledge to be stored in the cell state. Because of three control gates and one storage unit, LSTM can conveniently store, read, reset, and update long-term information. LSTM establishes a long-time delay between input and feedback, which updates the network weight of the front layer of the neural network in time and obtains the optimal parameters. That is, the gradient will neither explode nor disappear. In this sense, it is more suitable for dealing with complex time series and delay problems, so predicting high-frequency data components of cryptocurrency prices may have a better prediction effect.

3.4. Construction of the Forecasting Model Based on ELM Algorithm. ELM is a learning algorithm composed of a single-layer feedforward neural network, including input, output, and hidden layers. The corresponding output weights are calculated by randomly generating the input weights and offsets of the hidden layer. Figure 4 is the algorithm diagram of ELM, and Figure 5 shows the pseudocode of the ELM algorithm.
Given input samples \( ((x_j, t_j), j = 1, \ldots, M) \), where \( x_j \) and \( t_j \) respectively, represent the input vector of the \( j \)th sample and the corresponding expected output vector, and \( x_j = [x_{j1}, x_{j2}, \ldots, x_{jm}]^T \in R^m \) and \( t_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]^T \in R^m \). Let \( g(w, x, b) \) denote the activation function of the ELM. The structure of the ELM network consists of \( n \) input neurons, \( N \) hidden neurons, and \( m \) output neurons, which is expressed as follows:

\[
t_j = \sum_{i=1}^{N} \beta_i g_i(w_i \cdot x_j + b_i), \quad j = 1, \ldots, M, \tag{11}
\]

where \( \beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T \) and \( w_i = [w_{i1}, w_{i2}, \ldots, w_{im}]^T \) are weight vectors, which represent the weights between the neurons connecting the hidden layer, the input layer, and the output layer. \( b_i \) represents the \( i \)th hidden node offset, and \( w_i \cdot x_j \) represents the inner product operation of \( w_i \) and \( x_j \).

The above formula is integrated as \( H \beta = \mathbf{t} \), where \( \beta = [\beta_1, \beta_2, \ldots, \beta_N] \), \( \mathbf{t} = [t_1, t_2, \ldots, t_N] \), and \( H \) can be expressed as follows:

\[
H = \begin{bmatrix}
    g(w_1 \cdot x_1 + b_1) & \cdots & g(w_N \cdot x_1 + b_N) \\
    \vdots & \ddots & \vdots \\
    g(w_1 \cdot x_M + b_1) & \cdots & g(w_N \cdot x_M + b_N)
\end{bmatrix}_{M \times N}, \tag{12}
\]

where \( h \) is the output matrix of the ELM hidden layer, \( \beta \) is the output weight matrix of the hidden layer, and \( \mathbf{t} \) is the expected output matrix. The least-square solution of hidden layer output weight can be obtained by directly solving linear equations:
\[ \hat{\beta} = H^* T, \]

where \( H^* \) is the generalized inverse matrix of \( h \), and it can be proved that the norm of \( \hat{\beta} \) estimation is the minimum and inimitable.

For the ELM algorithm, the parameters of hidden layer nodes do not need to be adjusted deliberately, and the generalization performance is good. Compared with other algorithms, the ELM algorithm also has the advantages of fast learning speed and fewer training parameters. Compared with the high-frequency components, the intermediate-frequency, and low-frequency components are relatively regular, so the prediction effect of the ELM model for lower frequency data may be better.

### 3.5. Evaluation Indicators of the Ensemble Forecasting Model

In this paper, \( R^2 \), mean absolute error (MAE) and root mean squared error (RMSE) are used to measure the prediction ability of the integrated model.

For the accuracy of prediction, we first use \( R^2 \) to evaluate the forecasting model:

\[
R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2},
\]

where \( R^2 \) reflects the degree that the independent variable can explain the dependent variable, and the higher the value, the better the prediction effect of the ensemble models.

MAE refers to the average value of the deviation between the Bitcoin price and the forecast price. The formula is as follows:

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|.
\]

RMSE refers to the arithmetic square root of the square expected value of the difference between the estimated value of the parameter and the actual value. RMSE is used to evaluate the degree of change of the data. The smaller the RMSE value, the higher the accuracy of the forecasting model in describing the experimental data. The calculation formula is as follows:

\[
RMSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2.
\]

The smaller the MAE and RMSE, the smaller the error of the forecasting model and the better the prediction effect. This study will use these two indicators to compare the ensemble prediction models with the benchmark model to get the relatively optimal Bitcoin price prediction model.


The empirical study of ensemble prediction models for Bitcoin currency price is explained as follows.

#### 4.1. Data Description and Processing

In Section 3, the integrated forecasting model of cryptocurrency prices is constructed, and the applicability and effectiveness of the model for the integrated forecasting of Bitcoin prices are explored theoretically. This section applies the ensemble models based on multiscale analysis and deep learning to the empirical research of daily closing price forecasting. According to the combination of different algorithms, the ensemble models are set as EMD-LSTM, VMD-LSTM, EMD-ELM, VMD-LSTM, EMD-LSTM-ELM, and VMD-LSTM-ELM. For EMD-LSTM-ELM and VMD-LSTM-ELM, LSTM is used to predict high-frequency components, and ELM predicts medium-frequency and low-frequency components. Considering that Bitcoin is the main currency in cryptocurrency and its market value exceeds 50% of cryptocurrency’s market value and referring to Kristjanpoller and Minutolo [47], Trucós [48], and Maghyereh and Abdoh [49], we take Bitcoin as the representative of cryptocurrency for empirical research. On December 1st, 2017, Bitcoin futures was officially approved for listing by the US Commodity Futures Trading Commission, which indicates that the financial attributes of Bitcoin are further enhanced, and its influence in the financial market is gradually expanding. Therefore, this study chooses the previous week of this launch time as the sample start time. The data sample interval is set from November 24th, 2017, to November 27th, 2020, with 1098 data samples. The trend of Bitcoin closing price is shown in Figure 6. The data show significant nonlinearity during the sample period. For example, in December 2017, the price soared to about $20,000 per Bitcoin. By November 2020, the price of Bitcoin experienced a bear market for three years, and it dropped significantly.

In our study, Bitcoin price series data are divided into the training set and test set, 80% of which is the training set. The training set is used to train and optimize the ensemble model, and the remaining 20% of the data is used to test the prediction ability of the ensemble models. Specifically, the data from November 24th, 2017, to April 21st, 2020, are the training set, including 878 trading days. The data from 2020 April 22nd to November 27th, 2020, are the test set, including 220 trading days.

To alleviate the influence of units of input variables, the standardization of component series \( x_1, x_2, \ldots, x_t \) is set as follows:

\[
y_i = \frac{(x_i - \bar{x})}{s},
\]

\[ i = 1, 2, \ldots, t, \]

where \( \bar{x} \) is the mean value of the Bitcoin price series and \( s \) is the standard deviation of the Bitcoin price series.

#### 4.2. Sequence Decomposition and IMF Reconstruction

EMD and VMD are used to decompose Bitcoin closing price, respectively, and six IMF components and one residual term can be obtained. Each series component shows Bitcoin closing price fluctuation characteristics on different time scales from high frequency to low frequency, as shown in Figure 7.
For different IMFs, there exist differences and similarities. Under EMD decomposition, as shown in Figure 7(a), the frequency of IMF1-IMF2 components is higher, and the volatility amplitude is smaller, reflecting the short-term change trend of the Bitcoin closing price. The frequency and volatility amplitude of the component IMF3-IMF6 is moderate, which shows the medium-and long-term trend of Bitcoin closing price. The frequency fluctuation of the RES component is low, and the long-term trend is manifested.

To reduce the error and complexity of the prediction, we calculate the entropy of the decomposed components. The entropy values of each component are shown in Figure 8.

The entropy value evaluates the degree of autocorrelation and complexity of each IMFs (as shown in Figure 8), and the IMFs with the nearest entropy value are considered to combine a new component. By IMFs combination, we can decrease the number of IMFs and reduce the complexity of the prediction models. To avoid arbitrary selection, we calculate the difference in entropy value of each IMFs and classify them with the intergroup difference larger than 0.01, and the intragroup difference less than 0.01, thus ensuring the similarity of the reconstructed IMFs. Based on the criteria, we combine the IMF1 and IMF2 as the high-frequency component for EMD, IMF3–IMF5 as the medium-frequency component, and the rest IMFs as reconstructed as the low-frequency components. The reconstruction results for EMD and VMD IMFs are shown in Table 1.

According to the component composition in Table 1, the high-, medium-, and low-frequency components are reconstructed based on EMD and VMD and are shown in Figure 9.

We calculate the correlation coefficients of IMFs. Judging by the results in Figure 9, there are different patterns for high-, medium-, and low-frequency components. The high-frequency components have more fluctuations, while the low-frequency ones show a noticeable long-term trend. For the reconstruction shown in Figure 10, the heatmap is as follows.

As shown in Figure 10, the correlation coefficients of reconstructed IMFs are relatively small, indicating that the classification of original IMFs is effective and can be used for further prediction. We use the K–S test to compare the differences among low-, medium-, and high-frequency components to further compare whether there are significant differences among components. We take the low- and medium-frequency sequence as an example. The null hypothesis \( H_0 \) is set as follows: low-frequency and medium-frequency components obey the same distribution. We suppose the low- and medium-frequency sample size is \( n_1 \) and \( n_2 \). \( F_1(x) \) and \( F_2(x) \) are the cumulative empirical distribution, and \( D_i = F_1(x) - F_2(x) \); thus, the K–S test statistics is expressed as follows:

\[
Z = \max_i |D_i| \sqrt{\frac{n_1 n_2}{n_1 + n_2}}
\]  

(18)

According to the K–S test, given \( D < D_{\text{crit}} \) (\( D_{\text{crit}} \) is the critical value at the significance level \( \alpha \)), there is no significant difference between the low-frequency and medium components. When \( D \geq D_{\text{crit}} \), it is considered that there is a significant difference between low- and medium-component series. Table 2 shows the K–S test results based on different matchups of components.

The above results show that the price series at different frequencies are significantly different and have different fluctuation characteristics. Therefore, ignoring these characteristics on different frequencies may lower the prediction models’ performance. In addition, by comparing EMD and VMD reconstruction sequences, it can be found that the changing trend of price fluctuation sequences at the same frequency is the same. However, it is also worth noting that the high-frequency and medium-frequency sequences based on VMD decomposition have a slight fluctuation amplitude at the beginning of the sample period, which reflects that VMD can avoid modal mixing to some degree.

4.3. Empirical Results and Discussion. As discussed in Section 3, LSTM is selected as the primary prediction method for high-frequency data, and the ELM algorithm is chosen for low-frequency and medium-frequency components. The loss function of the LSTM algorithm is the mean square error loss function. The optimizer is Adam since its hyperparameter has the advantages of better interpretability, high calculation efficiency, and automatic learning rate adjustment. According to the reconstructed subsequence in Figure 9, the high-frequency subsequence is input into the LSTM model for training and prediction. The medium-frequency and low-frequency subsequences are input into the ELM model, respectively. Finally, the prediction results of each part are integrated to obtain the outcome of ensemble prediction models.

4.3.1. Performance of the Ensemble Prediction Models. Figure 11 shows the prediction effect of the ensemble prediction models based on multiscale analysis and deep learning methods. It shows that the prediction value of the ensemble prediction model has a small error compared with the actual price. Further calculation of \( R^2 \) of each prediction model shows that the prediction accuracy of benchmark models ARIMA, BP, SVR, LSTM, and ELM are 83.52%, 80.23%, 85.76%, 89.32%, and 89.43%, respectively. The
average prediction accuracy for benchmark models is 85.65%. The prediction accuracy of EMD-LSTM, VMD-LSTM, EMD-ELM, VMD-LSTM, EMD-LSTM-ELM, and VMD-LSTM-ELM are 90.32%, 92.51%, 94.33%, 95.45%, 98.85%, and 99.12%, respectively. The average accuracy of ensemble models is 95.12%, higher than that of benchmark models. From the MAE and RMSE indicators (shown in Table 3), the error of the ensemble prediction model for Bitcoin price is relatively small, indicating that the ensemble models can achieve high prediction accuracy. It is worth noting that the ensemble models can predict the price decline to a certain extent. For example, on May 9th, 2020, the price of Bitcoin dropped from 9935.62 US dollars the previous day to 9683.06 dollars, and it fell for three consecutive days, with a drop rate of 10.81%. The VMD-LSTM-ELM model predicted the downward trend of the price on May 8th. On September 2nd, 2020, the price dropped from 12035.56 US dollars the previous day to 11326.8 US dollars and dropped for eight consecutive days, with a drop rate of 14.83%. The VMD-LSTM-ELM model also predicted this downward trend on August 31st. Therefore, the ensemble prediction model can better forecast Bitcoin prices.
4.3.2. Further Comparison of Ensemble Models and Benchmark Models. We conduct a model comparison in this section to further test the performance of ensemble models and other soft computing techniques. Since the methods of ARIMA, SVR, BP, RNN, KNN, GBDT, RF, XGBoost, LSTM, and ELM are widely used in different fields (Ibrahim et al. [3], McNally et al. [6], Basher and Sadorsky [11], Dong et al. [14], Zhang et al. [20], Ma et al. [21], Rtyai and Enneya [23], Liu et al. [24], Lee et al. [29], and Ghosh et al. [50]), we select these models as the benchmark models. At the same time, the EMD and VMD methods are combined with different types of intelligent algorithms to form EMD-LSTM, VMD-LSTM, EMD-ELM, VMD-LSTM, and EMD-LSTM-ELM. The MAE and RMSE value of each model is calculated and shown in Table 3.

From Table 3, we can find that the MAE and RMSE of the benchmark model are higher than those of the ensemble models (except LSTM and ELM models), and the stability of the ensemble models is relatively high.

Furthermore, this study defines the prediction deviation as the ratio of the difference between the predicted value and the actual value to the real value and compares the deviation of different models. The result is shown in Figure 12. The time-varying of deviation is relatively small for ensemble models, which indicates that the ensemble models have better prediction performance.

According to RMSE, MAE, and other evaluation indicators, the effect of the ensemble forecasting model is better than that of a single method. In addition, although ensemble models such as the EMD-LSTM model have relatively good generalization ability, their performance is not superior to VMD-LSTM-ELM and EMD-LSTM-ELM models. This result shows that the time-series data after multiscale decomposition need to choose an appropriate deep learning algorithm to improve the models’ prediction performance and generalization ability. Because multiscale analysis and deep learning can consider the multiscale attributes and nonlinear characteristics of time series data, matching components with different frequencies with different deep learning algorithms can improve the prediction accuracy of the prediction models.

4.3.3. Influence of Multiscale Decomposition on the Ensemble Prediction Model. To examine the influence of multiscale analysis on ensemble prediction, we compare the prediction performance of EMD and VMD with different machine learning algorithms. The first pair of models are EMD-LSTM and VMD-LSTM, which are denoted by I-VMD and I-EMD, respectively. The second pair is EMD-ELM and VMD-ELM, and the third pair is EMD-LSTM-ELM and VMD-LSTM-ELM. These four models are represented as II-EMD, II-VMD, III-EMD, and III-VMD. The deviation of the model for the test samples is counted, and the results are shown in Table 4.

Table 4 shows that the integrated forecasting method based on the VMD method is robust. The predicted deviation of the VMD method is relatively small, and the distance of its outliers from the median is smaller. From the average and standard deviation of the error, the VMD model also has better performance, and the Q1, Q2, and Q3 values of the deviation of VMD models are smaller than those of EMD models, which shows that the prediction effect under VMD decomposition is relatively better to some extent.
Figure 9: Reconstructed components based on EMD and VMD. Note: the x-axis in each figure denotes the time daily, and the y-axis is the price of each component of Bitcoin in the unit of USD dollar. (a) Reconstruction based on EMD. (b) Reconstruction based on VMD.

Figure 10: Heatmap of reconstruction of IMFs based on EMD and VMD methods. (a) Correlation coefficient based on EMD IMFs. (b) Correlation coefficient based on VMD IMFs.
In addition, the prediction error of a specific date from the VMD method is also small. For example, on April 24th, 2020, the closing price of Bitcoin on the day was 7502.76 US dollars, the prediction value of the integrated model under EMD decomposition was 7653.58 US dollars, with an error of 2.01%, and the prediction value generated by the VMD-LSTM-ELM model was 7503.98 US dollars, with an error of 0.016%. On June 17th, 2020, the closing price was 9409.91 USD, and the predicted value of VMD-LSTM-ELM was 9409.925 USD, with an error of only 0.00015%, while the error of the EMD-LSTM-ELM model was 1.42%. From the overall trend of the forecast, when the closing price suddenly changes, the prediction model based on VMD decomposition is more stable, and VMD-LSTM-ELM responds better to sudden or extreme financial events. The relative advantages of VMD models could be attributed to that the VMD could better avoid overenveloping and modal mixing.

5. Robust Analysis Based on Different Market Conditions

To investigate the influence of market conditions on the performance of ensemble prediction models, we select the interval with high volatility to test the prediction accuracy when the market is in a boom or bust period. The subperiods are shown in Table 5.

Overall, the ensemble prediction model has good robustness regarding the different market conditions. Table 5 shows that the prediction accuracy is mostly above 80% during extreme market conditions. The small sample size could explain the relatively low accuracy in the subperiod (2020/05/10–2020/05/17).

To further test the robustness of the model, we divided the sample period into different market conditions based on the conditional volatility of Bitcoin price by using GARCH models. We suppose $P_t$ is the closing price on the $T$ day, $P_{t-1}$ is the closing price on the $t-1$ day, and the yield $R_t = \ln P_t - \ln P_{t-1}$. The GARCH (1,1) model is set as follows:

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where $\omega > 0, \alpha \geq 0, \beta_1 \geq 0$ is a constant parameter, $\sigma_t^2$ is the volatility of Bitcoin’s closing price in the period of $t$, and $\epsilon$ is the independent random error term with the same distribution.
The GARCH model has a good fitting effect, reflecting the volatility clustering and persistence, as shown in Figure 13. According to the changing trend of conditional volatility, the forecast intervals are divided into high volatility intervals of 2020/4/22–2020/6/4, 2020/7/28–2020/9/10, and 2020/10/28–2020/11/27, which are denoted as interval I, interval III, and interval V. The low volatility intervals 2020/6/4–2020/7/28 and 2020/9/10–2020/10/28, respectively, are interval II and interval IV in the figure. The sample period is divided, as shown in Figure 13.

Through calculation, the prediction accuracy of the VMD-LSTM-ELM model is 93.97% and 94.69% in two low volatility intervals and 93.28%, 90.04%, and 95.25% in three high volatility intervals, respectively. When the volatility is high, the prediction performance of the ensemble models still keeps at a relatively satisfactory level. Therefore, the ensemble prediction models are robust in different market conditions of Bitcoin markets.

### Table 4: Performance comparison of ensemble prediction models.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>I-EMD</th>
<th>I-VMD</th>
<th>II-EMD</th>
<th>II-VMD</th>
<th>III-EMD</th>
<th>III-VMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum value</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.0959</td>
<td>0.1179</td>
<td>0.0682</td>
<td>0.1413</td>
<td>0.0648</td>
<td>0.0704</td>
</tr>
<tr>
<td>Average value</td>
<td>0.0182</td>
<td>0.0160</td>
<td>0.0177</td>
<td>0.0170</td>
<td>0.0144</td>
<td>0.0129</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.0170</td>
<td>0.0160</td>
<td>0.0179</td>
<td>0.0123</td>
<td>0.0121</td>
<td>0.0119</td>
</tr>
<tr>
<td>Q1</td>
<td>0.0058</td>
<td>0.0051</td>
<td>0.0074</td>
<td>0.0059</td>
<td>0.0051</td>
<td>0.0034</td>
</tr>
<tr>
<td>Q2</td>
<td>0.0130</td>
<td>0.0124</td>
<td>0.0145</td>
<td>0.0120</td>
<td>0.0111</td>
<td>0.0092</td>
</tr>
<tr>
<td>Q3</td>
<td>0.0251</td>
<td>0.0207</td>
<td>0.0239</td>
<td>0.0239</td>
<td>0.0210</td>
<td>0.0193</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.6658</td>
<td>2.8382</td>
<td>1.1722</td>
<td>2.8688</td>
<td>1.3357</td>
<td>1.3871</td>
</tr>
</tbody>
</table>

### Table 5: Prediction accuracy of different fluctuation intervals.

<table>
<thead>
<tr>
<th>Subperiods</th>
<th>Percentage of volatility (%)</th>
<th>Prediction accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020/04/22–2020/05/08</td>
<td>39.55</td>
<td>92.67</td>
</tr>
<tr>
<td>2020/05/10–2020/05/17</td>
<td>13.23</td>
<td>68.85</td>
</tr>
<tr>
<td>2020/05/26–2020/06/04</td>
<td>12.61</td>
<td>81.87</td>
</tr>
<tr>
<td>2020/07/20–2020/08/17</td>
<td>35.10</td>
<td>97.41</td>
</tr>
<tr>
<td>2020/09/08–2020/11/24</td>
<td>91.82</td>
<td>99.06</td>
</tr>
<tr>
<td>2020/05/17–2020/05/26</td>
<td>−11.14</td>
<td>86.43</td>
</tr>
<tr>
<td>2020/06/04–2020/07/20</td>
<td>−6.45</td>
<td>80.07</td>
</tr>
</tbody>
</table>

### 6. Conclusions and Recommendations

This study constructs the ensemble model for Bitcoin price prediction using multiscale analysis and the deep learning method. The empirical study shows that the ensemble model has good forecasting performance. The average prediction accuracy is 95.12%, and its average error is smaller than that of the benchmark model. Meanwhile, the ensemble model has good robustness in different market conditions. There are two possible explanations for the better performance of ensemble models.

On the one hand, the decomposed components based on EMD and VMD can better capture the features of Bitcoin price on different time scales. On the other hand, the corresponding deep learning algorithms can be adapted to various decomposed components. Thus, the prediction performance is improved. These results have practical values for investment and supervision. First, the practitioners could use ensemble models to forecast the cryptocurrency price and Bitcoin price in particular. Second, the decomposed component of the original price has different characteristics and impacts price forecasting. Thus, the investors and supervisors could make more satisfying decisions based on multiscale analysis.

The implications for future research are as follows: first, for the decomposed components, the impactors and their influence on cryptocurrency price have not been investigated in this study, while for better performance and interpretability of the ensemble model, future research could construct the ensemble models by incorporating the various input features, such as trading information, hyperfrequency price, supply and demand factors, and energy consumption of Bitcoin mining, into multiscale analysis and deep learning. Second, as the prediction models are ensemble by different machine learning algorithms, the parameter
uncertainty is also a limitation of the proposed model. Additionally, for obtaining a more precise or stable prediction of Bitcoin price, a possible way is to construct the ensemble models by incorporating trading information, hyperfrequency price, supply and demand factors, the energy consumption of Bitcoin mining, and other input features, rather than only price series, the increased number of features may also exacerbate the problem of parameter uncertainty. For solving this problem, a reliable, stable, and convergent prediction model with high prediction accuracy could be obtained by using metaheuristic algorithms, such as swarm-based algorithms, human-based algorithms, physics-based algorithms, and other algorithms as [21, 22] suggest. Third, although the reconstruction of IMFs reduces computation complexity, the ensemble model's overall complexity is increased by decomposition and matching different components with various machine learning algorithms. Future research could try dimension reduction techniques or adopt more domain knowledge to achieve a better-off balance between prediction accuracy, model complexity, and robustness.

Data Availability

The data used to train the models of this study can be obtained from the corresponding author upon request.

Disclosure

The manuscript was already published as a preprint based on the link https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4092347 [51].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Changqing Luo and Lurun Pan designed the study, performed the research, analysed the data, and wrote the paper; Binwei Chen and Huiru Xu corrected the grammatical errors and retouched the sentences.

Acknowledgments

This paper was supported by the Hunan Natural Science Foundation (no. 2020JJ4255), Key Research Project of the Ministry of Science and Technology of China (2022YFC3320904), and Postgraduate Scientific Research Program of Hunan Province (CX20221127 and CX20211129).

References


