Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques

Aytaç ALTAN[a], Seçkin KARASU[a], Stelios Bekiros[b,c,*]

[a]Department of Electrical-Electronics Engineering, Zonguldak Bülent Ecevit University, 67100 Zonguldak, Turkey

{aytacaltan, seckin.karasu}@beun.edu.tr

[b]European University Institute, Department of Economics, Villa La Fonte, Via delle Fontanelle, 18, I-50014, Florence, Italy

[c]Rimini Centre for Economic Analysis (RCEA), LH3079, Wilfrid Laurier University, 75 University Ave W., ON, N2L3C5, Waterloo, Canada

*Corresponding author e-mail: aytacaltan @ beun.edu.tr

Abstract

The price forecasting of the digital currencies in the financial market is of great importance, especially after the recent global economic crises. Due to the nonlinear dynamics, which is including inherent fractality and chaoticity of the digital currencies, it is understood from the research conducted by many researchers that a single model is not sufficient in the forecasting of the digital currencies with very high accuracy. Since the single models used in the forecasting of digital currencies have weaknesses as well as their own strengths, they might not grant the best forecasting achievement in all situations for all the time. A new hybrid-forecasting framework has been proposed in digital currency time-series to minimize this negative situation and increase forecasting achievement. In this study, a novel hybrid forecasting model based on long short-term memory (LSTM) neural network and empirical wavelet transform (EWT) decomposition along with cuckoo search (CS) algorithm is developed for digital currency time series. The model is obtained by combining the LSTM neural network and EWT decomposition technique, and optimizing the intrinsic mode

function (IMF) estimated outputs with CS. The price of the four most traded digital currencies such as BTC, XRP, DASH and LTC, is estimated by the proposed model and the performance of the model has been tested. The experimental results show that the hybrid model proposed for digital currency forecasting can capture nonlinear properties of digital currency time series.

Keywords: Cryptocurrency, Hybrid forecasting model, Fractality, Long short-term memory (LSTM), Empirical wavelet transform (EWT), Cuckoo search algorithm.

1. Introduction

The evolution of the money used in the exchange of goods and services is rapidly progressing from primitive exchange methods to commodity money, gold and silver, valuable paper which has gold provision, trustworthy money which has no gold provision and digital/crypto currency, respectively. The cryptocurrency, which is sub-set of digital currencies, is defined as a digital cash money designed to operate as a cryptographic medium of exchange tool to control the formation of additional units of currency, secure its own transactions, and confirm the transfer [1]. Interest in crypto currencies, whose total market value has reached hundreds of billions of US dollars and affects people's investment and transaction behavior, has increased rapidly over the past few years, this situation attracts the attention of both academics and practitioners [2]. Bitcoin (BTC), which is presented to the financial market based on its unique protocol and the systematic structural feature of Nakamoto [3], is known as an accomplished cipher currency with the largest market capitalization in the digital currency market [4]. As of the end of 2018, four cryptocurrencies with the highest market value in the digital currency market are Ethereum (ETH), Ripple (XRP), Litecoin (LTC), EOS, except BTC, respectively. In addition, four crypto currencies with the highest total volume, except BTC, are Tether (USDT), ETH, LTC, EOS, respectively. In this study, we propose a new hybrid framework consisting of bio-inspired signal processing for digital

2

currency time-series forecasting. The performance of the proposed model has been empirically tested on Digital Cash (DASH), LTC, which have high liquidity and data availability of at least one thousand observations with BTC and XRP, which have both the highest market value and highest total volume in the digital currency market.

The cryptocurrencies have a decentralized and distributed structure, as opposed to the central electronic currency generated by central banks or banking systems. The control of this distributed structure is carried out with blockchain technology. Introducing of new cryptocurrencies and verification of transactions by decoding of crypto-puzzles are realized by blockchain technology [5]. The infrastructure of the cryptocurrencies in the financial market is based on the block chain structure. Thanks to the peer-to-peer network in the blockchain structure [3], electronic payments can be made from one side to the other without a center, and transaction records are kept as local copies in the databases on each node. As the security of cryptocurrencies is provided by the blockchain structure, it plays an important role in increasing the transaction volume and total volume in the exchange transactions. Recently, the price forecasting of cryptocurrencies with the highest market value, transaction volume or total volume are made by many researchers. The price predictability of digital currencies can be determined by examining their intrinsic nonlinear dynamics [6-8], including fractality [9-11] and intrinsic chaoticity [11-13]. In this study; firstly, the predictability of digital currencies has been evaluated by examining their inherent nonlinear dynamics including inherent chaoticity and fractality. Secondly, a new hybrid digital currency forecasting model has been proposed based on the deep learning and decomposition method with the bio-inspired algorithm by considering the nonlinear dynamics of the digital currencies. The performance of the proposed digital currency forecasting model is tested on BTC, which has the highest both market value and total volume in the financial market, XRP that has high market value, and DASH and LTC that are used in forecasting studies in literature.

3

In [14], the financial time series estimation is performed by support vector machine (SVM) method. The obtained model in the study is compared with case-based reasoning (CBR) and back propagation neural network (BPNN) models. SVM models are created for various kernel parameters and upper limit values, and the identified model is indicated to perform satisfactorily compared to considered models in predicting financial time series. In [15], linear regression (LR), linear SVM (L-SVM) and polynomial SVM (P-SVM) models are used to estimate BTC price for the next one-step ahead. The models used in the study consist of two inputs and one output. An accumulation/distribution (A/D) oscillator is applied to the model input together with one of moving average (MA) or weighted moving average (WMA) filters. Closing, maximum or lowest price time series are used for the calculation of these technical indicators. It is specified that the forecasting model with the least error is obtained by testing with various parameter combinations such as SVM with polynomial and linear kernel functions. Patel et al. [16] have proposed a hybrid model of two-stage approach based on machine learning methods to estimate the future values of the stock market index. In the first stage of the proposed model, estimation of the future statistical parameter values is made by support vector regression (SVR) method. In the second stage of the model, the index estimation is performed for the various future step values of two stock exchanges using artificial neural network (ANN), random forest (RF) and SVR methods, which use the predicted parameter values as input. The value of the stock index is estimated by using ten technical indicators at the input of the proposed hybrid model. In [17], in the estimation of the BTC price trend for the next day; the ANN-based prediction model is formed by using the price and transaction volume data of the previous days. It is stated that the identified single-layer feed forward network model performance is higher than the online complex models. In [18], an experimental study is performed on the estimation and modeling of BTC prices using Bayesian neural network (BNN) based on blockchain information related to BTC supply and

4

demand. The performance of the model is measured by comparing it with other linear and nonlinear models. In [19], machine learning ensemble algorithms, which is the combination of tree classifier and recurrent neural networks (RNNs), with the help of feature selection techniques, achieve the best results, have been suggested in estimating the BTC price trend. The SVM and ANN have been utilized for regression of the minimum, maximum and closing prices of the BTC. In [20], the capabilities of the various alternative univariate and multivariate models used to estimate four crypto currencies consisting of BTC, ETH, XRP and LTC are compared. A set of crypto-estimators have been used to combine a large set of univariate dynamic linear model sets and multivariate vector autoregressive models with different forms of time variation. It is specified that there are statistically significant improvements are found in point estimation when using density estimation and combinations of univariate models based on the selection of multivariate models. In [21], the predictability of the most traded digital currencies in the financial market has been determined by investigating the inherent nonlinear dynamics including the inherent fractality and chaoticity. The time series of all digital currencies are reported to show self-similarity, long memory and fractal dynamics. The price of the four most commonly traded digital currencies such as BTC, XRP, DASH and LTC has been estimated with long short-term memory (LSTM) from deep learning techniques. The forecasting performance of the LSTM model and the generalized regression neural networks (GRNN) model performance are compared. It is indicated that LSTM model performance is significantly higher than GRNN model performance.

Neural network, machine learning and deep learning methods can be used as a single forecasting model in the price [22], volatility [4] and trend [23] forecasting of cryptocurrencies in the financial market as well as they can be also used as a combined forecasting model together with decomposition or optimization algorithms. In [24], a stock market time series prediction model is developed based on the combination of empirical mode

decomposition (EMD) and adaptive network-based fuzzy inference system (ANFIS). In the first phase of the developed model, the intrinsic mode function (IMF) components are decomposed by the EMD method applied to the stock market time series. The decomposed components are applied to the ANFIS model input and the stock market time series are estimated for the next one-step ahead. It is expressed that the hybrid model has a higher performance when compared to Chen's model [25], Yu's model [26], auto-regressive (AR), SVR and ANFIS. In [27], the BTC price estimation is performed by using multi-layer perceptron (MLP) based on nonlinear auto-regressive with exogenous (NARX) model. It is used the opening, closing, minimum and maximum past price data together with MA technical indicators as the model input. Model parameters such as input lag, output lag and hidden units are optimized with particle swarm optimization (PSO) algorithm. In this study, the decomposition method and the meta-heuristic based optimization technique are combined with the deep learning method.

This study is the first study that uses both the decomposition method and the optimization algorithm in combination with deep learning to predict digital currency price. Also, unlike other studies on this issue, predicts the digital currency price with higher performance than other studies in the literature with the hybrid model, which include LSTM, empirical wavelet transform (EWT) and cuckoo search (CS) optimization algorithm. The remainder of this paper is organized as follows. In Section 2, the distinctive methodology is introduced in detail, including the normalization method, the optimization algorithm and the developed hybrid model. The experimental results of the proposed model and other models, which is compared to the proposed model, are given Section 3. In Section 3, in order to further prove the availability and accuracy of the proposed model, the serious and tangible discussions are also presented. Finally, the conclusions are highlighted in Section 4.

2. Digital currency time-series forecasting framework

In this section, firstly, the largest Lyapunov exponent (LLE) and a detrended fluctuation analysis (DFA) based on the extracted Hurst exponent (HE) of the time series used to detect chaos and/or fractal characteristics of the underlying digital currencies are described. Then, our approach to forecasting digital currency price time-series which has higher performance than the current models is explained. In the proposed EWT-LSTM-CS digital currency price-forecasting model, an estimation model is identified with the LSTM neural network for each IMF obtained by the EWT decomposition technique and the weight coefficients of each output are optimized with CS algorithm. The methods used in our approach are explained in detail in the following sub-sections.

2.1 *Largest Lyapunov exponent*

The sensitivity of the digital currency time series to changes in initial conditions is quantified by Lyapunov exponent (LE). The convergence or divergence rate of the two initially close points along their trajectory over time is measured by LE. This rate, measured by an exponential function, is used to examined the local stability of linear or nonlinear systems. In this test, the positive values of the exponents show exponential divergence of the digital currency time series, high sensitivity to initial conditions, and hence the chaotic process. In many applications the calculation of only the LLE is sufficient for the presence of chaos [11]. In this study, the approach of Rosenstein et al. [12] is used to estimate the LLE. The LLE algorithm which is given in [21] directly tests the exponential divergence of nearby trajectories. As a result of the test, we are able to decide whether it really makes sense to calculate a Lyapunov exponent for a given digital currency time series.

2.2 *Detrended fluctuation analysis based on the extracted Hurst exponent*

The DFA is very useful method to reveal the extent of long-range correlations in time series. Since it is less dependent on noisy data and non-stationary assumptions, it is appropriate to

quantify nonlinear dynamics and complexity in the time series [9]. The computational steps of the DFA method for any signal are defined in [21]. This computational steps are repeated over all time scales to characterize the relationship between detrended time series function, the average fluctuation, and the time scale. The fluctuations can be characterized by a scaling exponent. Hurst exponent is estimated by running a regression of logarithm of the overall fluctation upon the logarithm of time scale. In this study, it is determined by HE whether the four active cryptocurrencies used have long-memory characteristics.

2.3 *Normalization process*

Normalization is the process of including attributes that are valued at different cluster ranges into the same set. The importance of this phase is to ensure that all attributes have an equal effect on the result during the classification process. The normalization method, which rescales attributes or outputs from a range of values to a new predefined range of values and provides linear transformation in the original data range, is called min-max normalization. The attribute selection is the process of obtaining the best attributes for classifying by eliminating unnecessary or unrelated attributes that reduce classifier performance [28]. In this study, the attributes are transferred to the $[0, 1]$ range by min-max normalization method. In this normalization method, rescaling is usually performed using a linear interpretation formula such as

$$d' = (f_{max}^{new} - f_{min}^{new}) \frac{(d - f_{min})}{(f_{max} - f_{min})} + f_{min}^{new} \tag{1}$$

where d is the range of original data, f_{min} is minimum value of attribute, f_{max} is maximum value of attribute, f_{min}^{new} and f_{max}^{new} are predefined boundaries. When min-max normalization is applied to the original raw data, each attribute will be in the new range of values and will remain the same. Also, this normalization method has the advantage that all relationships in the data are fully protected.

2.4 *Empirical wavelet transform*

The EWT is proposed by Jérôme Gilles [29] to extract different intrinsic modes of a time series by building adaptive wavelets. The proposed EWT for non-stationary signal analysis is an adaptive signal decomposition method based on the information content of the signal. In the EWT decomposition algorithm consisting mainly of three steps, the frequency components of the signal applied by using fast Fourier transform (FFT) are firstly determined. Afterwards, different intrinsic modes, subband signals, are extracted by obtaining the appropriate segmentation of the Fourier spectrum. Finally, scaling and wavelet functions corresponding to each detected segment are applied.

The empirical wavelets can be defined as bandpass filters on each Λ_n, where Λ_n represents each segment $\Lambda_n = [\Omega_{n-1}, \Omega_n]$. Ω_n is the limits between each segments, where $\Omega_0 = 0$ and $\Omega_n = \pi$, then $\cup_{n=1}^{N} \Lambda_n = [0, \pi]$. $\forall n > 0$, the empirical wavelet function and the empirical scaling function can be defined by the Eq. (2) and (3), respectively.

$$\hat{\psi}_n(\Omega) = \begin{cases} 1 & if\ (1+\rho)\Omega_n \leq |\Omega| \leq (1-\rho)\Omega_{n+1} \\ \cos\left[\frac{\pi}{2}\alpha\left(\frac{1}{2\tau_{n+1}}(|\Omega| - \Omega_{n+1} + \tau_{n+1})\right)\right] & if\ (1-\rho)\Omega_{n+1} \leq |\Omega| \leq (1+\rho)\Omega_{n+1} \\ \sin\left[\frac{\pi}{2}\alpha\left(\frac{1}{2\tau_n}(|\Omega| - \Omega_n + \tau_n)\right)\right] & if\ (1-\rho)\Omega_n \leq |\Omega| \leq (1+\rho)\Omega_n \\ 0 & otherwise \end{cases} \quad (2)$$

$$\hat{\phi}_n(\Omega) = \begin{cases} 1 & if\ |\Omega| \leq (1-\rho)\Omega_n \\ \cos\left[\frac{\pi}{2}\alpha\left(\frac{1}{2\tau_n}(|\Omega| - \Omega_n + \tau_n)\right)\right] & if\ (1-\rho)\Omega_n \leq |\Omega| \leq (1+\rho)\Omega_n \\ 0 & otherwise \end{cases} \quad (3)$$

The function $\alpha(x)$ is an arbitrary function defined as

$$\alpha(x) = \begin{cases} 0 & if\ x \leq 0 \\ and\ \alpha(x) + \alpha(1-x) = 1 & if\ \forall x \in [0,1] \\ 1 & if\ x \geq 1 \end{cases} \quad (4)$$

Note that the related function $\alpha(x)$ is mostly used in the literature [30] as

$$\alpha(x) = x^4(35 - 84x + 70x^2 - 20x^3) \tag{5}$$

The condition of tight frame can be described as [29]

$$\rho < min_n \left(\frac{\Omega_{n+1} - \Omega_n}{\Omega_{n+1} + \Omega_n} \right) \tag{6}$$

The detail and approximation coefficients are obtained from the inner product of the applied signal $y(t)$ with the empirical wavelet and the empirical scaling functions.

2.5 *Long short term memory neural network*

The LSTM neural network, a special type of RNNs, which has a strong ability to handle long-term and short-term dependency problems with its success in processing nonlinear sequential data, cannot only keep adjacent ad-hoc information, but also control long-term information. The core of the LSTM neural network is the memory cell that replaces the hidden layers of conventional neurons [31]. The LSTM can protect previous information that can help significantly improve its ability to learn signal sequences and inherent nonlinear patterns. The LSTM memory cell can remember or forget any cell state depending on the inputs. The cell is supported by the *input gate*, *output gate*, and *forget gate*. The LSTM neural network can add or remove information from these gates to the memory cell state [32]. This structure provides the LSTM can detect which cells suppress which cells are excited based on the previous state, current input and current memory. So, the LSTM neural network structure shown in Fig. 1 effectively overcomes the vanishing gradient problem and the exploding gradient problem, thus ensuring that the neural networks can remember information from a long range [33]. In LSTM neural network structure, in order to identify new information that can be accumulated to the cell, the input data is multiplied by the output of the input gate. To calculate the information that can be spread to the network, the output data for the network is multiplied by the activation of the output gate. To determine whether the last state of the cell should be

forgotten, the cell states of the previous time are multiplied by the activation of the forget gate [34]. The procedure of the LSTM [33] is as follows:

- the stage of deciding what information to be discarded from the cell state: The value of x_t and h_{t-1} will be obtained, and determine whether to discard through a sigmoid function.

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

- the stage of determining which new information is stored in the cell state: It is decided by a sigmoid layer which information will be stored in the cell state. Then, the values obtained from the $\tanh(\cdot)$ layer by x_t and h_{t-1} are taken as a new candidate value $\tilde{C}_t$.

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{8}$$

$$\tilde{C}_t = tanh(w_C \cdot [h_{t-1}, x_t] + b_C) \tag{9}$$

- the step of updating the previous cell state C_{t-1} to the new cell state C_t: The cell state C_{t-1} is multiplied by f_t to forget what information we decided to forget. Then, to obtain a new cell state C_t, i_t is multiplied by $\tilde{C}_t$ and new cell state C_t is determined by adding the term obtained to the previous term.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{10}$$

- the stage of deciding what information will be output: It is decided by a sigmoid layer which information will be output in the cell state.

$$s_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{11}$$

$$h_t = s_t * tanh(C_t) \tag{12}$$

where $w_f \, w_i \, w_C \, w_o$ denotes the weight matrices, $b_f \, b_i \, b_C \, b_o$ represents the bias vectors, respectively. $\sigma(\cdot)$ is the logistic sigmoid function which is utilized as the gate activation function as

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{13}$$

and $tanh(\cdot)$ is a hyperbolic tangent function; it is used as activation function of input and output block as

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{14}$$

2.6 *Meta-heuristic optimization algorithm*

Cuckoo search (CS) is a meta-heuristic algorithm developed based on the obligate brood parasitic behavior of some cuckoo species along with the Lévy flight behavior [35] of fruit flies and some birds [36]. The CS algorithm not only offers effective global search (explorative random walk) and local search (exploitative random walk) capabilities but also has the ability to find the local optimum with Lévy flight. These two search capabilities are controlled by a switching/discovery probability (p_a). The meta-heuristic CS algorithm can be implemented through three ideal rules [37]. The first rule is that each cuckoo lays one egg at a time and dump its egg into a randomly selected nest. The second rule is that the best nests that contain high quality eggs are moved to next generations. The third rule is that the number of available host nests is fixed at the location where the egg discarded by a cuckoo is discovered by the host bird with a measured fraction probability, $p_a \in [0,1]$. The main iteration equation of the new generation of $x^{(k+1)}$ for the i th cuckoo is

$$x_i^{(k+1)} = x_i^k + \gamma \otimes L\acute{e}vy(\lambda) \tag{15}$$

where x_i^k refers to the current location, which is the only way to determine the next location $x_i^{(k+1)}$, $\gamma > 0$ is the step size related to the dimension of the problem, $\otimes$ is the entry-wise multiplication, and $L\acute{e}vy(\lambda)$ is Lévy flight representing the most effective characteristics of cuckoo search to generate new eggs by a random walk. Since the step length is much longer in the long run, the random walk over the Lévy flight is more effective in exploring the search space than the other optimization methods such as genetic algorithm (GA) and PSO [38]. An explorative random walk by using Lévy flights can be defined as

$$L\acute{e}vy \sim u = k^{-\lambda}, \qquad (1 < \lambda \le 3) \tag{16}$$

where k is the iteration number, and λ is a parameter that is the mean or expectation of the occurrence of the event during a unit interval. Also, Eq. (16) has an infinite mean and an infinite variance.

In the CS algorithm, each nest represents a solution and a nest' population is used to find the best solution to the problem. The main steps of CS algorithm are expressed in follow [39]:

- **Initialization step**

N_{pop} host nests' population is defined as $x = [x_1 \quad x_2, \cdots, x_{N_{pop}}]^T$

where each nest $x_i = [P_{i1} \quad P_{i2}, \cdots, P_{ij}, \cdots, P_{iN_{pop}}]$ indicates power output of units. This output is initialized by

$$P_{ij} = P_j^{min} + randn_\omega * (P_j^{max} - P_j^{min}), \qquad randn_\omega \in [0,1] \tag{17}$$

- **Generation step of new solution via Lévy flights**

The new solution for each nest is determined by the previous best nests via Lévy flight. The new solution for each nest is calculated as

$$x_i^{new} = x_i^{best} + \gamma. randn_\varepsilon. \Delta x_i^{new} \tag{18}$$

where $randn_\varepsilon$ is a normally distributed stochastic number, and $\gamma > 0$ the updated step size. Δx_i^{new} is defined as

$$\Delta x_i^{new} = \left[\frac{s(\beta). randn_\mu}{randn_\xi} \right]^{\frac{1}{\beta}} . (x_i^{best} - x_i), \qquad 1 \leq \beta \leq 3 \tag{19}$$

where $randn_\mu$ and $randn_\xi$ are two normally distributed stochastic variables, and these variables are random value between zero and one $randn_\mu, randn_\xi \in [0,1]$. Also, the standard deviation $s(\beta)$ is determined as

$$s(\beta) = \left(\frac{\Gamma(1+\beta). \sin(\pi. \beta/2)}{\Gamma\left(\left(\frac{1+\beta}{2} \right). \beta. 2^{\left(\frac{\beta-1}{2} \right)} \right)} \right)^{\frac{1}{\beta}} \tag{20}$$

where $\Gamma(\cdot)$ is gamma distribution function, and (β) is the distribution factor $0.3 \leq \beta \leq 1.99$.

- ***Foreign egg discovery and randomization steps***

The action of exploration of an foreign egg in nest of a host bird with the probability of p_a creates a new solution for the problem similar to Lévy flights, and the new solution can be calculated as follow:

$$x_i^{exp} = x_i^{best} + M.\Delta x_i^{exp} \tag{21}$$

where M is the updated coefficient determined based on the probability of a host bird to explore a foreign egg in its nest and it is defined as

$$M = \begin{cases} 1 & if \ randn_\varphi < p_a \\ 0 & else \end{cases} \tag{22}$$

Also, Δx_i^{exp} is determined by

$$\Delta x_i^{exp} = randn_\varphi.\left[randp_1\left(x_i^{best}\right) - randp_2\left(x_i^{best}\right)\right] \tag{23}$$

where $randn_\varphi$ is the distributed random number in $[0,1]$, $randp_1\left(x_i^{best}\right)$ and $randp_2\left(x_i^{best}\right)$ are the random perturbation for locations of nests in x_i^{best}.

- ***Stopping criteria step***

The algorithm is stopped when the number of iterations reaches the predefined value.

2.7 *Framework of proposed hybrid EWT-LSTM-CS digital currency forecasting model*

In this study, the contents of the proposed a novel hybrid digital currency price forecasting framework consist of data pre-processing stage, a forecasting model identification stage by combining the EWT and LSTM, the stage of optimizing IMF predicted outputs with CS, and model evaluation stage. The whole process of the proposed hybrid EWT-LSTM-CS digital currency forecasting model is shown in Fig. 1. In the data pre-processing stage, the digital currency time-series data are normalized by min-max normalization method before model identification stage. In the model identification stage, IMFs are obtained by EWT, which is one of the decomposition methods. To eliminate noise signals and stochastic volatility, EWT

decomposes the original series and is used to reconstruct the time series. For each output of the IMFs, an estimation model consisting of the LSTM neural network and EWT is identified. At the optimization stage, the weighted coefficients of each IMF output are optimized for the identification of the best forecasting model with the CS algorithm. The objective function is done with MSE, which is one of the criteria used to measure the performance of the models.
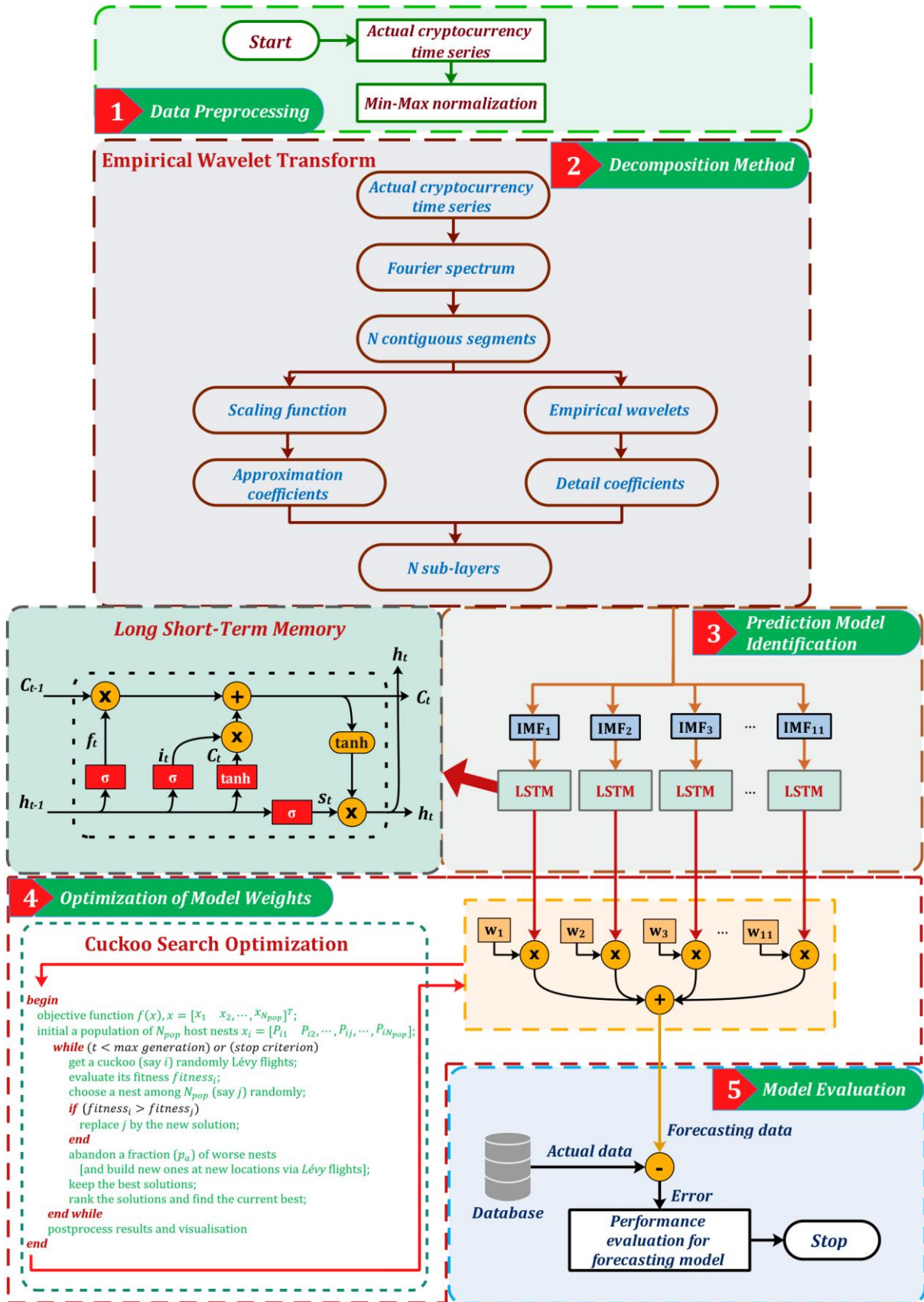
Fig. 1. The overall framework of hybrid EWT-LSTM-CS digital currency forecasting model.

In the model evaluation stage, the performance of the proposed hybrid model is measured by taking into account the mean absolute error (MAE), root mean square error (RMSE), and the mean absolute percentage error (MAPE). Note that, for these three metrics, the smaller the index value, the better the model performance. The performance of the digital currency forecasting models is measured by considering MAE, RMSE and MAPE values. The mathematical formulas of these error criteria are given in Table 1. The validity and reliability of the proposed hybrid EWT-LSTM-CS digital currency forecasting model has been tested by these indexes.

Table 1

Error metrics for evaluation of proposed digital currency forecasting model.

Metric	Definition	Formula
MAE	Mean absolute error	$MAE = \dfrac{1}{N}\sum_{j=1}^{N}\left|p_{estimated}^{j} - p_{real}^{j}\right|$
RMSE	Root mean square error	$RMSE = \sqrt{\dfrac{1}{N}\sum_{j=1}^{N}(p_{estimated}^{j} - p_{real}^{j})^2}$
MAPE	Mean absolute percentage error	$MAPE = \dfrac{1}{N}\sum_{j=1}^{N}\left|\dfrac{p_{estimated}^{j} - p_{real}^{j}}{p_{real}^{j}}\right| \times 100\%$

3. Experimental results and discussion

In this section, the studies performed with BTC, XRP, DASH and LTC cryptocurrencies data have been presented to prove the forecasting effectiveness of the proposed hybrid EWT-LSTM-CS digital currency forecasting model. The obtained results have been compared and discussed with the results of the current studies in literature. Also, the effect of CS optimization algorithm on the digital currency forecasting model has been investigated.

3.1 *Digital currency data sets*

In this study, data sets containing daily prices in US dollars have been used for BTC, XRP, DASH and LTC cryptocurrencies. In the selection of these four cryptocurrencies, the

availability of data consisting of at least one thousand observations or high liquidity played an important role. In addition, these four cryptocurrencies, which are also used in current studies in literature, have been selected in order to compare the experimental results of our digital currency forecasting model. The time series containing daily prices in US dollars have been included data from 18 July 2010 to 28 March 2019 for the BTC, 22 January 2015 to 28 March 2019 for the XRP, 14 February 2014 to 28 March 2019 for DASH, and 24 August 2016 to 28 March 2019 for LTC. Thus, the time series length consists of 3176 samples for BTC, 1527 samples for XRP, 1869 samples for DASH, and 947 samples for LTC. Each cryptocurrency time series data set used in the study is separated into two sets as training and test sets. The first 85% of the cryptocurrency time series data is used for training purposes and the remaining 15% is used for testing purposes. The LLE estimation and DFA based on the extract HE calculation are used both on training and testing sub-samples, to investigate fractality, inherent chaoticity and any other nonlinear characteristics throughout all time periods. Some statistical values for data sampling of four cryptocurrencies data sets, including minimum, maximum, mean, and standard deviation are presented in Table 2.

Table 2

Statistical values of the four datasets used in study for cryptocurrency data sampling.

Dataset	Samples	Numbers	Statistical values (USD)			
			Min	Max	Mean	Std.
BTC	All samples	3176	0.10	19346	1609.60	2960.90
	Training	2700	0.10	16850	611.03	1333.50
	Testing	476	3228.70	19346	7273.50	3268.20
XRP	All samples	1527	0.0036	2.78	0.2247	0.3524
	Training	1298	0.0036	2.78	0.1993	0.3750
	Testing	229	0.2630	0.5827	0.3690	0.0819
DASH	All samples	1869	0.3149	1550.80	120.92	216.09
	Training	1589	0.3149	1550.80	117.59	232.81
	Testing	280	59.69	265.53	139.84	60.83
LTC	All samples	947	3.50	359.40	64.52	66.00
	Training	805	3.69	359.40	75.22	66.04
	Testing	142	3.50	4.69	3.90	0.22

The estimated values of the LLE and DFA based on the extract HE for the training and testing sub-samples of each digital currency are given in Table 3.

Table 3

Estimated LLE and HE values.

	LLE		HE	
	Training sub-sample	Testing sub-sample	Training sub-sample	Testing sub-sample
BTC	0.1180	-6.1681	1.0103	0.8924
XRP	0.8741	-0.0103	1.0317	0.7983
DASH	0.3623	-9.6873	0.9213	0.9911
LTC	0.1962	-3.0027	1.0271	0.8227

The LLE values for the training and test sub-samples of the four cryptocurrencies time series used in the study are positive and negative, respectively. The price sample used in the training phase of the cryptocurrency time series exhibits chaotic dynamics, while the test phase reveals

19

that there is convergence with classical attractors. Further, all estimated values of HE indicate that four active digital currencies have long-memory characteristics.

3.2 *Empirical results and analysis*

In this study, the proposed hybrid digital currency forecasting model consists of EWT, LSTM, and CS. The proposed hybrid EWT-LSTM-CS cryptocurrency forecasting model is obtained in three stages after the preprocessing stage. In preprocessing step, the time series data have been transferred to the [0,1] interval by using min-max normalization method to provide that all attributes have an equal effect on the result during the classification process. In the first stage of the hybrid model, BTC, XRP, DASH and LTC cryptocurrencies time series data have been decomposed into the components of IMF by the EWT decomposition method to eliminate noise signals and stochastic volatility. In the study, BTC, XRP, DASH and LTC cryptocurrencies time series data have decomposed into 22, 12, 17 and 12 subband signals by EWT, respectively. The first eleven IMFs of the EWT, which is used as the decomposition method in the identification of the proposed hybrid EWT-LSTM-CS cryptocurrency forecasting model, are presented in Figs. 2-5 for the BTC, XRP, DASH and LTC cryptocurrencies time series, respectively. In the second stage of the proposed model, a combined forecasting model is identified with EWT and LSTM neural network for each output of the IMF. In the training phase of the LSTM neural network model, data sets allocated for training of BTC, XRP, DASH and LTC cryptocurrencies time series have used. Each output of the IMF obtained for BTC, XRP, DASH and LTC has been trained by LSTM neural network. In this study, 63 IMF in total, where BTC, XRP, DASH and LTC have 22, 12, 17 and 12 IMF, respectively, have been trained in LSTM neural network model. In LSTM model, the number of hidden units is set to 500 and the maximum epoch is set to 300. All experiments for BTC, XRP, DASH and LTC cryptocurrencies time series are applied on a PC device which has Intel Xeon W-2155, 10 core, 4.50 GHz Turbo, 32 GB RAM. All codes are

20

compiled with MATLAB 2018b. The training time of the 63 IMF in total by using LSTM neural network has been carried out in 113 minutes and 42 seconds. The cryptocurrency time series data obtained at this phase are used in the third stage, which is CS optimization process. In the third stage of the proposed hybrid model, the weighted coefficients of each IMF output are optimized to form the best cryptocurrency forecasting model with the CS algorithm instead of accumulation of estimated IMF outputs. The MSE is selected as the objective function called also fitness function.

The forecasting accuracy of LSTM, EWT-LSTM and EWT-LSTM-CS models have been calculated for BTC, XRP, DASH and LTC cryptocurrencies time series and presented in Table 4-6. The performance of the proposed hybrid model has been evaluated by considering MAE, RMSE and MAPE values. The proposed model has the MAE values of 500.16, 0.0064, 2.0746 and 1.1066, the RMSE values of 623.41, 0.0088, 2.7776 and 1.7989, and the MAPE values of 3.55%, 1.72%, 1.47% and 2.77% for BTC, XRP, DASH and LTC cryptocurrencies, respectively. The performance of the proposed EWT-LSTM-CS hybrid cryptocurrency forecasting model has been compared to the performance of single LSTM and EWT-LSTM cryptocurrency forecasting models. The single LSTM model has the MAE values of 862.34, 0.0171, 7.0954 and 7.5919, the RMSE values of 1474.2, 0.0301, 10.831 and 8.1470, and the MAPE values of 9.59%, 4.38%, 4.99% and 13.05% for BTC, XRP, DASH and LTC cryptocurrencies, respectively. The EWT-LSTM model has the MAE values of 556.78, 0.0067, 2.1671 and 4.6509, the RMSE values of 776.74, 0.0092, 2.9343 and 4.2760, and the MAPE values of 6.14%, 1.81%, 1.52% and 5.63% for BTC, XRP, DASH and LTC cryptocurrencies, respectively. The forecasting performance of BTC, XRP, DASH and LTC cryptocurrencies time series for single LSTM, EWT-LSTM, and EWT-LSTM-CS models are presented in Figs. 6-8. The effect of EWT decomposition method on single LSTM model has been shown in Fig. 7, and the effect of CS optimizer on EWT-LSTM combined model has

been indicated in Fig. 8. The Fig. 8 and Table 6 show that the hybrid EWT-LSTM-CS model is the best cryptocurrency forecasting model compared to the single LSTM model and EWT-LSTM combined model.

When the performance of EWT-LSTM combined model is compared to the performance of single LSTM model, it is observed that 35.43%, 60.81%, 69.45% and 38.73% improvement in MAE values, 47.31%, 69.43%, 72.91% and 47.51% improvement in RMSE values, and 35.97%, 58.67%, 69.53% and 56.85% improvement in MAPE values for BTC, XRP, DASH and LTC, respectively. When the performance of proposed hybrid EWT-LSTM-CS model is compared to the performance of EWT-LSTM combined model, it is observed that 10.16%, 4.47%, 4.26% and 76.20% improvement in MAE values, 19.74%, 4.34%, 5.34% and 57.93% improvement in RMSE values, and 42.18%, 4.97%, 3.28% and 50.79% improvement in MAPE values for BTC, XRP, DASH and LTC, respectively.

Moreover, it is observed that the EWT-LSTM combined forecasting model significantly eliminates the noise signals and stochastic volatility. It has been found that the CS optimizer increased the forecasting accuracy of the cryptocurrency forecasting model by at least 4%. The proposed cryptocurrency forecasting model has been proven to be robust to nonlinear dynamics involving inherent fractality and chaoticity of cryptocurrencies.

To the best of our knowledge, this study is the first study to estimate the price of digital currency using deep learning together with decomposition method and optimization algorithm. When this study compared with [21], it is seen that the RMSE values have been improved by 77.33%, 82.36% and 85.60% for BTC, XRP and DASH, respectively, with the proposed hybrid cryptocurrency forecasting model. It is clear that the achievement of the proposed cryptocurrency forecasting model is quite satisfactory when compared to studies carried out with single models by using deep learning methods. It is also supported by the

results presented in Table 6 and Fig. 8, where the proposed hybrid cryptocurrency forecasting model performs well from considered models according to the MAE, RMSE and MAPE performance indexes.
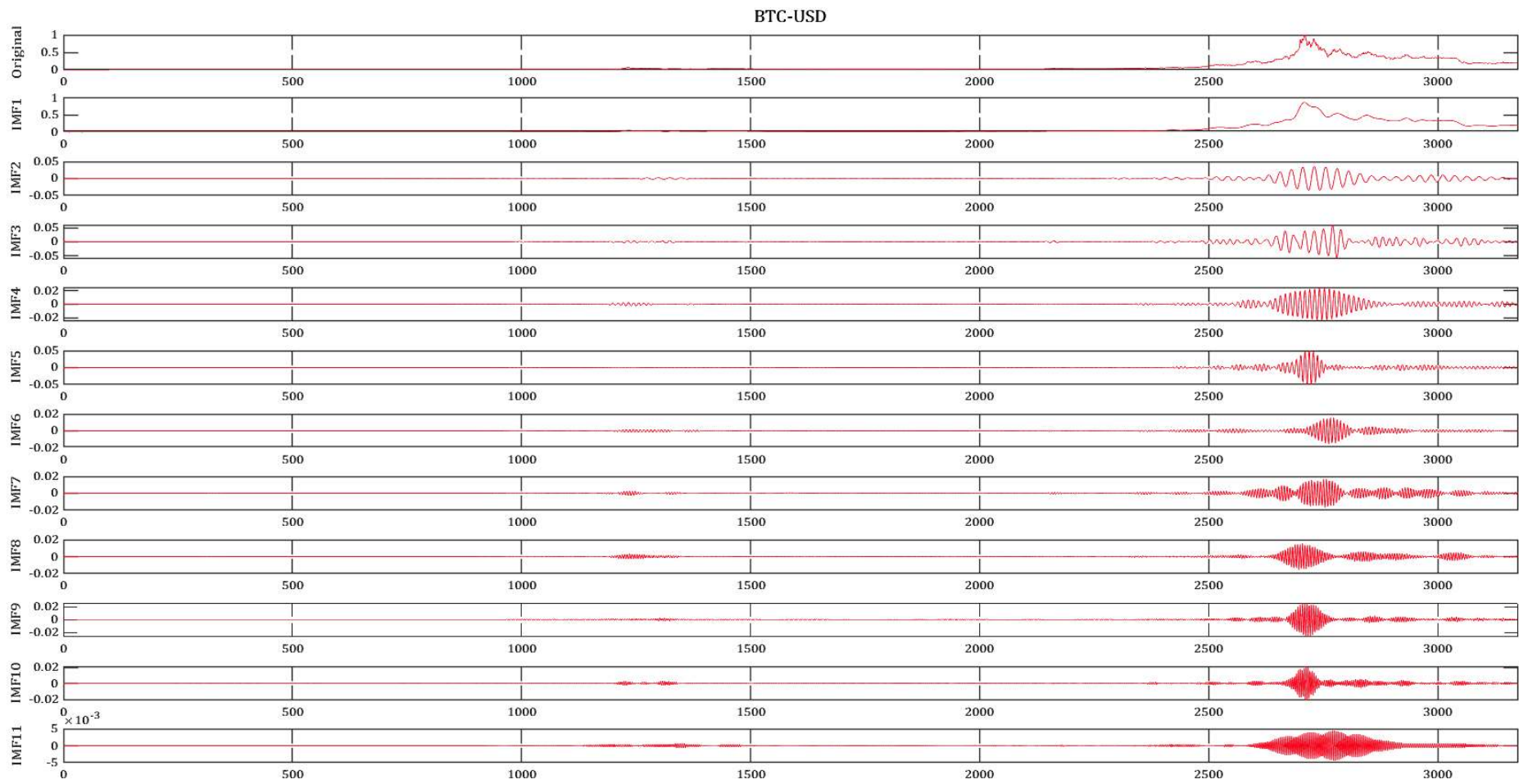
Fig. 2. IMFs of EWT for BTC-USD cryptocurrency time series.
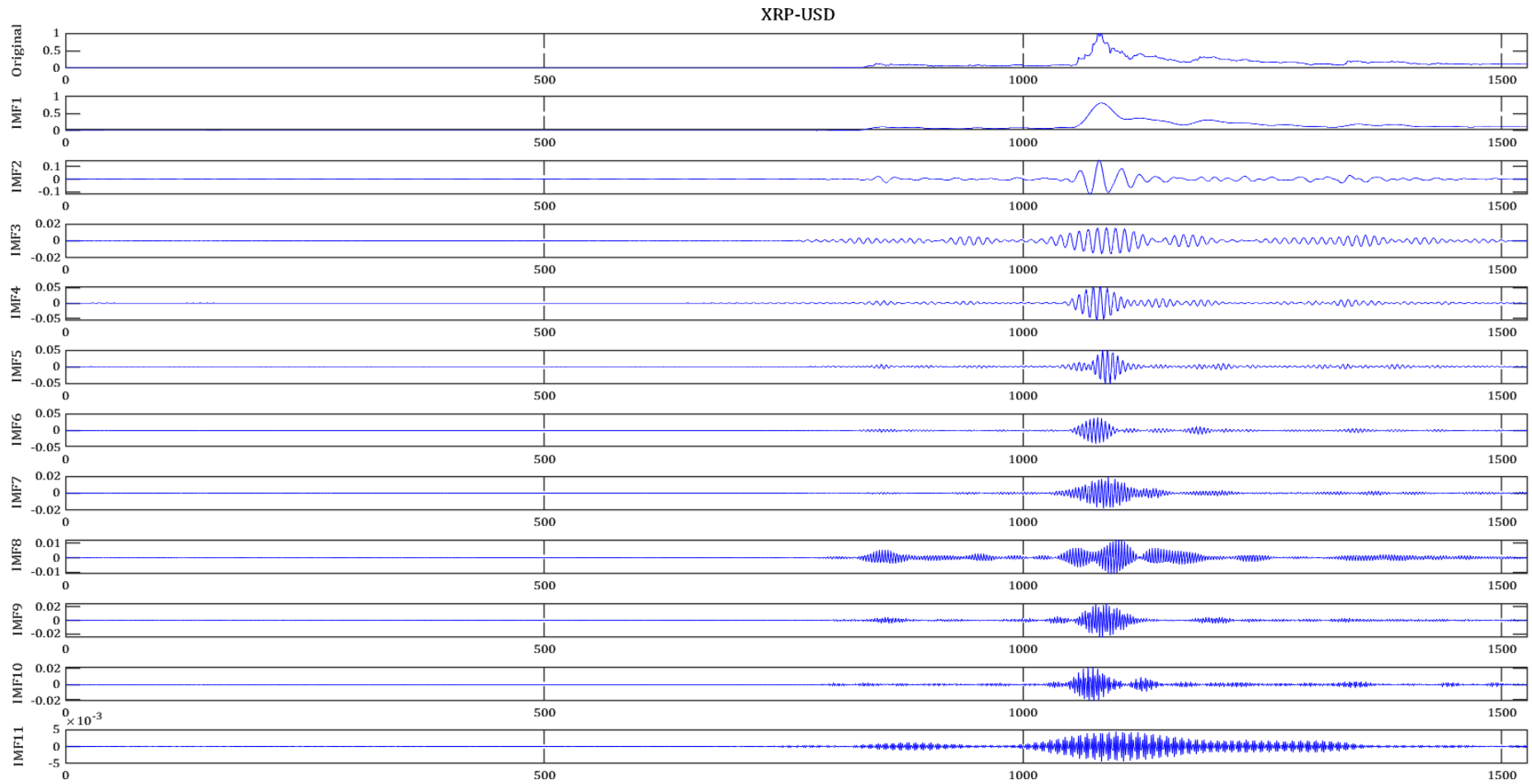
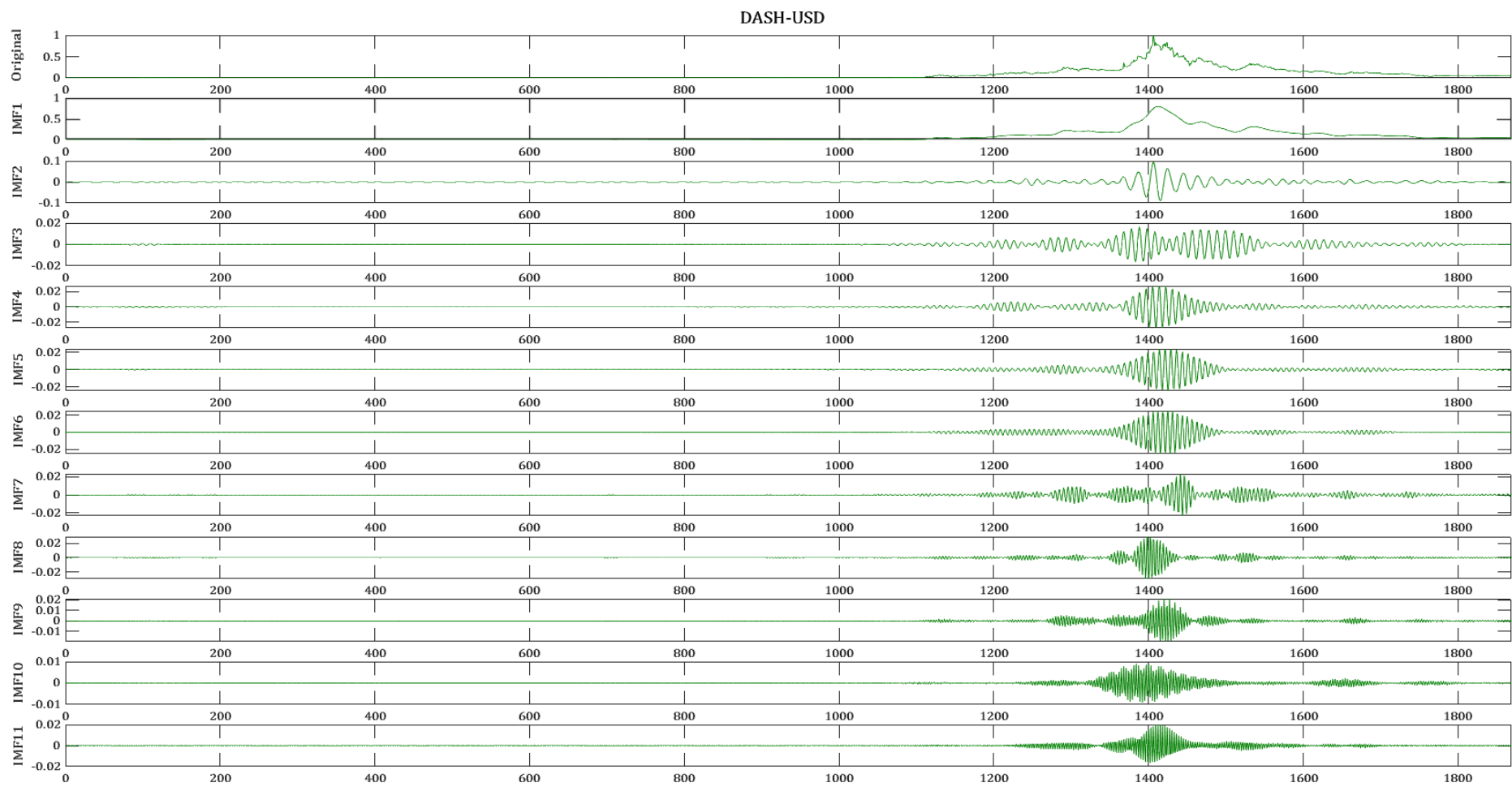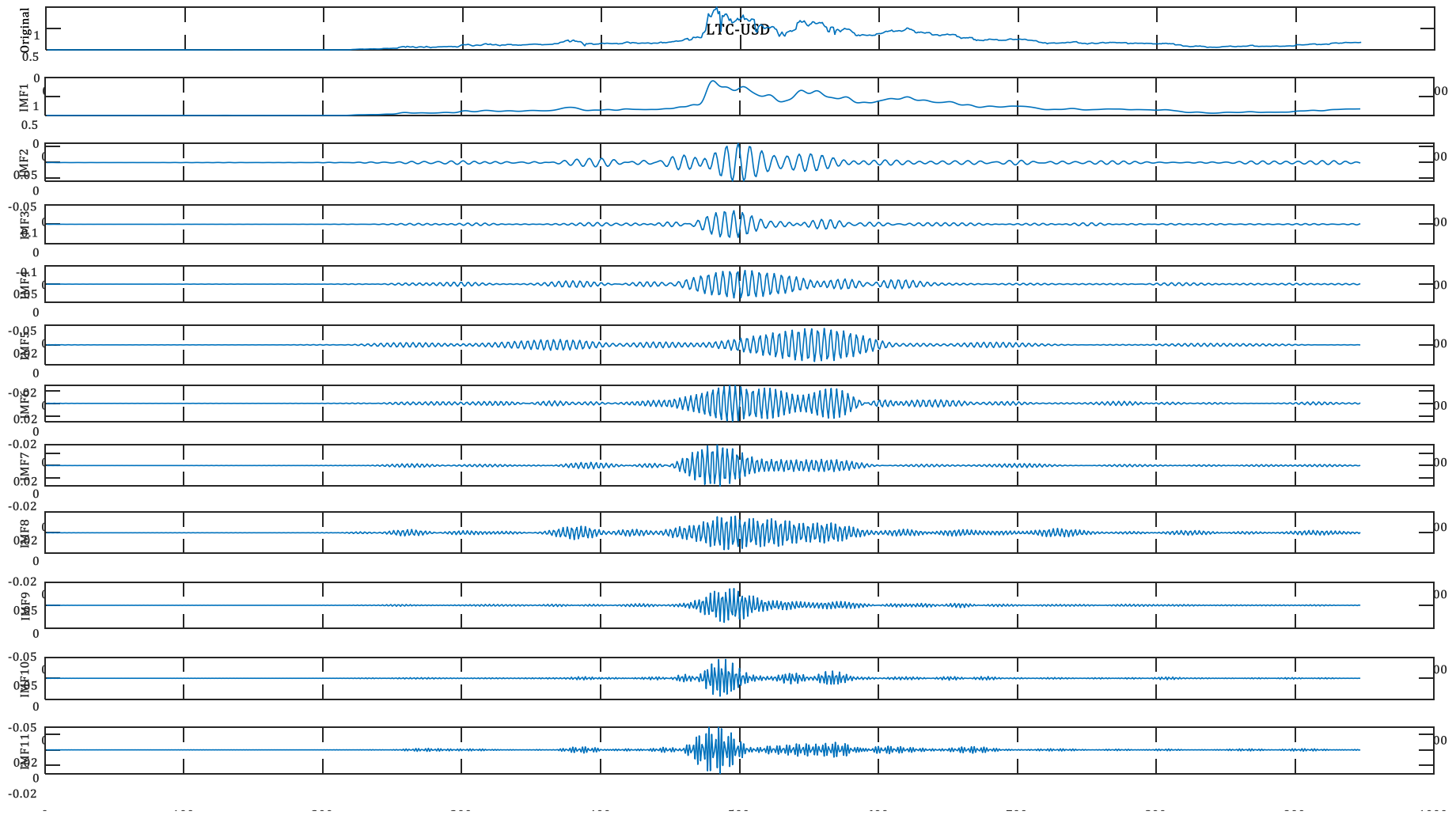**Fig. 3.** IMFs of EWT for XRP-USD cryptocurrency time series.

Fig. 4. IMFs of EWT for DASH-USD cryptocurrency time series.

Fig. 5. IMFs of EWT for LTC-USD cryptocurrency time series.

27

Table 4

Error metrics of four cryptocurrencies for hybrid LSTM model.

Cryptocurrencies	MAE	RMSE	MAPE (%)
BTC	862.34	1474.2	9.59
XRP	0.0171	0.0301	4.38
DASH	7.0954	10.831	4.99
LTC	7.5919	8.1470	13.05

Table 5

Error metrics of four cryptocurrencies for hybrid EWT-LSTM model.

Cryptocurrencies	MAE	RMSE	MAPE (%)
BTC	556.78	776.74	6.14
XRP	0.0067	0.0092	1.81
DASH	2.1671	2.9343	1.52
LTC	4.6509	4.2760	5.63

Table 6

Error metrics of four cryptocurrencies for hybrid EWT-LSTM-CS model.

Cryptocurrencies	MAE	RMSE	MAPE (%)
BTC	500.16	623.41	3.55
XRP	0.0064	0.0088	1.72
DASH	2.0746	2.7776	1.47
LTC	1.1066	1.7989	2.77
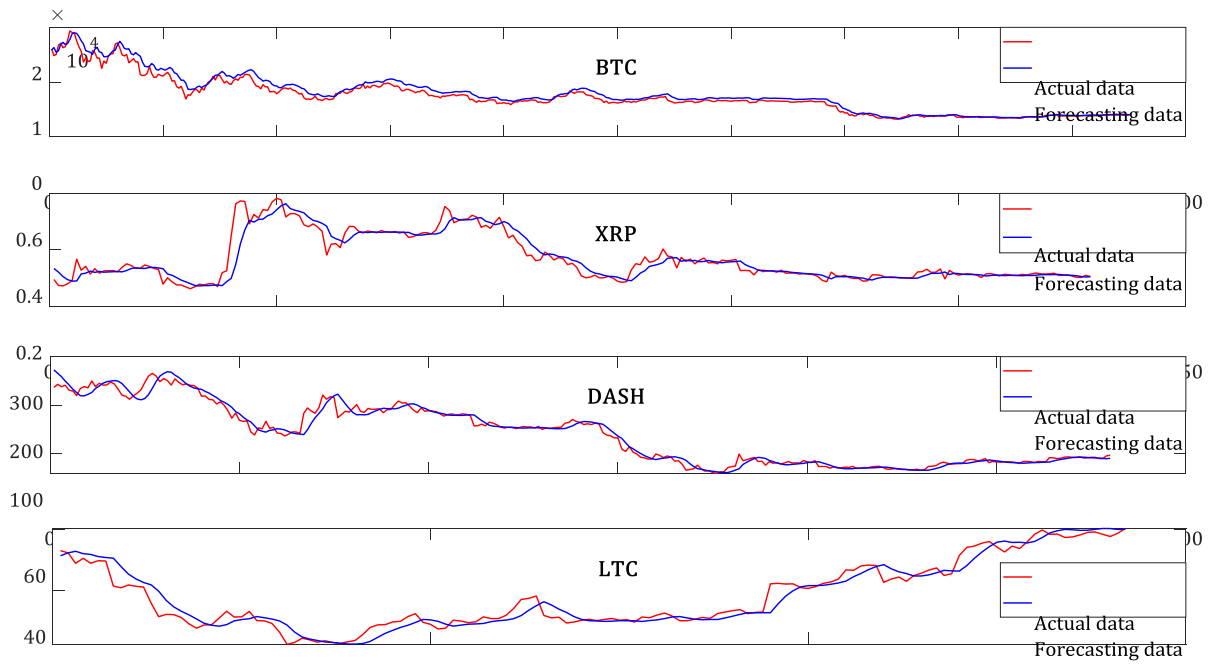
Fig. 6. The forecasting and actual values of BTC, XRP, DASH and LTC cryptocurrencies time series for single LSTM model. The time horizon is represented by the x-axis, and the price values are indicated on the y-axis in USD.
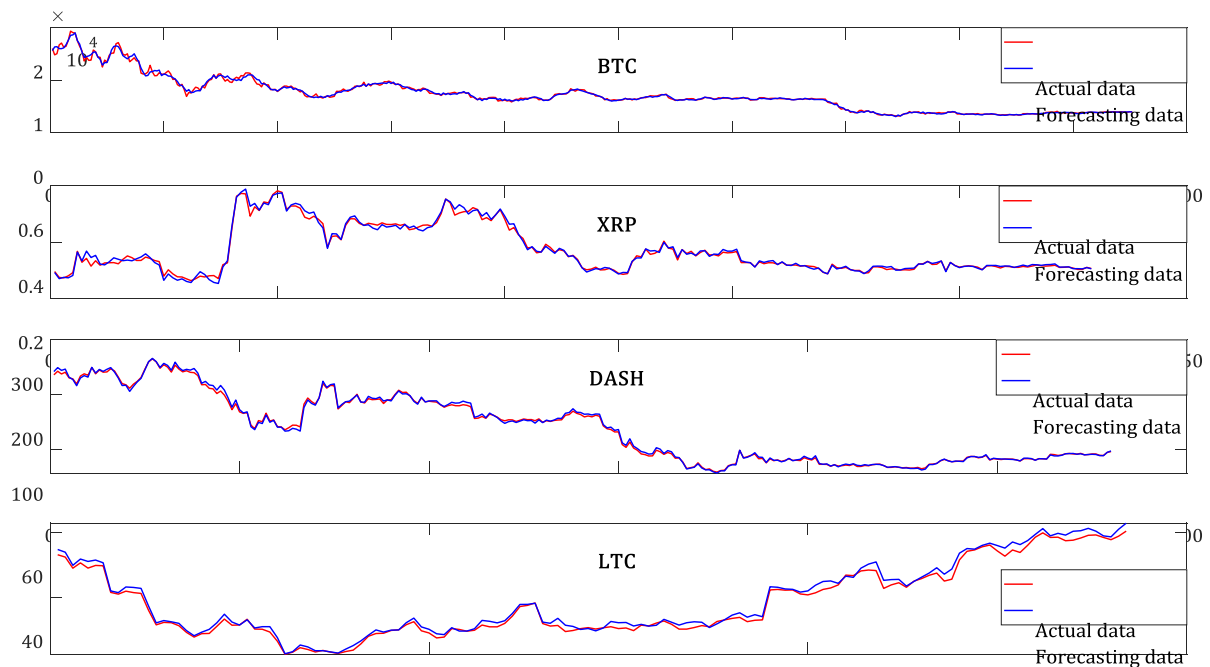
Fig. 7. The forecasting and actual values of BTC, XRP, DASH and LTC cryptocurrencies time series for EWT-LSTM combined model. The time horizon is represented by the x-axis, and the price values are indicated on the y-axis in USD.
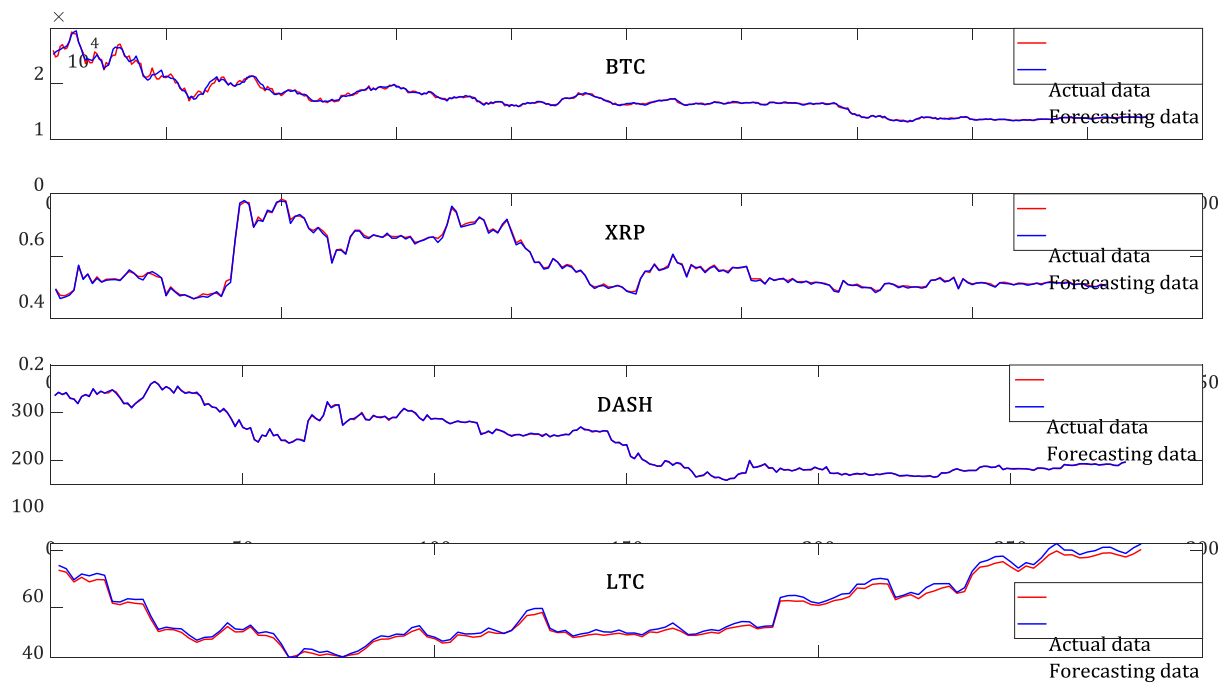
Fig. 8. The forecasting and actual values of BTC, XRP, DASH and LTC cryptocurrencies time series for hybrid EWT-LSTM-CS model. The time horizon is represented by the x-axis, and the price values are indicated on the y-axis in USD.

4. Conclusions

It is known that high accuracy and precision cryptocurrency forecasting have become very important especially after the global economic crises in the last decade. Because of the nonlinear dynamics, which is including inherent fractality and chaoticity of cryptocurrencies, it is seen that single models are not sufficient for estimating cryptocurrency with high accuracy. The hybrid models come to the forefront in digital currency estimation in order to deal with the non-stationary and uncertainty of the digital currency time series. In this paper, we have proposed a hybrid cryptocurrency forecasting model consisting of decomposition method, deep learning and meta-heuristic based optimization technique for estimation of high accuracy digital currencies. This study is the first study to estimate the price of digital currency using deep learning together with decomposition method and meta-heuristic based optimization algorithm. In the proposed model, LSTM neural network and EWT decomposition technique along with CS optimization algorithm have been used as hybrid. The proposed hybrid EWT-LSTM-CS digital currency forecasting model has been developed by combining the LSTM neural network method stage and the decomposition technique stage, and optimizing estimated IMF outputs of EWT with the CS optimization algorithm. In the model evaluation stage, both the price of the four most traded digital currencies such as BTC, XRP, DASH and LTC has been estimated by proposed model and the performance of the model has been tested. The obtained experimental results show that the proposed model for digital currency forecasting can capture nonlinear characteristics of digital currency time series. Furthermore, it is seen from the experimental results that the forecasting performance of the proposed model is clearly superior compared to all considered models according to statistical error criteria. It is thought that this study will significantly contribute to the forecasting of the volatility of digital currencies with high performance.

31

References

[1] Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). Bitcoin and cryptocurrency technologies: a comprehensive introduction. Princeton University Press.

[2] Zhang, W., Wang, P., Li, X., & Shen, D. (2018). Some stylized facts of the cryptocurrency market. Applied Economics, 1-16.

[3] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

[4] Kristjanpoller, W., & Minutolo, M. C. (2018). A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. Expert Systems with Applications, 109, 1-11.

[5] Symitsi, E., & Chalvatzis, K. J. (2018). Return, volatility and shock spillovers of Bitcoin with energy and technology companies. Economics Letters, 170, 127-130.

[6] Lahmiri, S., & Bekiros, S. (2018). Chaos, randomness and multi-fractality in Bitcoin market. Chaos, Solitons & Fractals, 106, 28-34.

[7] Takaishi, T. (2018). Statistical properties and multifractality of Bitcoin. Physica A: statistical mechanics and its applications, 506, 507-519.

[8] Lahmiri, S., Uddin, G. S., & Bekiros, S. (2017). Nonlinear dynamics of equity, currency and commodity markets in the aftermath of the global financial crisis. Chaos, Solitons & Fractals, 103, 342-346.

[9] Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. Physical Review e, 49(2), 1685.

[10] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. Physica A: Statistical Mechanics and its Applications, 316(1-4), 87-114.

[11] Gencer, M., & Unal, G. (2016). Testing Non-Linear Dynamics, Long Memory and Chaotic Behaviour of Energy Commodities.

[12] Rosenstein, M. T., Collins, J. J., & De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets. Physica D: Nonlinear Phenomena, 65(1-2), 117-134.

[13] Chenaghlu, M. A., Jamali, S., & Khasmakhi, N. N. (2016). A novel keyed parallel hashing scheme based on a new chaotic system. Chaos, Solitons & Fractals, 87, 216-225.

[14] Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1-2), 307-319.

[15] Karasu, S., Altan, A., Saraç, Z., & Hacioğlu, R. (2018, May). Prediction of Bitcoin prices with machine learning methods using time series data. In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[16] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications, 42(4), 2162-2172.

[17] Almeida, J., Tata, S., Moser, A., & Smit, V. (2015). Bitcoin prediciton using ANN. Neural Networks, 1-12.

[18] Jang, H., & Lee, J. (2018). An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. IEEE Access, 6, 5427-5437.

[19] Mallqui, D. C., & Fernandes, R. A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. Applied Soft Computing, 75, 596-606.

[20] Catania, L., Grassi, S., & Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. International Journal of Forecasting, 35(2), 485-501.

[21] Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons & Fractals, 118, 35-40.

[22] Atsalakis, G. S., Atsalaki, I. G., Pasiouras, F., & Zopounidis, C. (2019). Bitcoin price forecasting with neuro-fuzzy techniques. European Journal of Operational Research.

[23] Sun, X., Liu, M., & Sima, Z. (2018). A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Research Letters.

[24] Wei, L. Y. (2016). A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. Applied Soft Computing, 42, 368-376.

[25] Chen, S. M., & Chung, N. Y. (2006). Forecasting enrollments using high- order fuzzy time series and genetic algorithms. International Journal of Intelligent Systems, 21(5), 485-501.

[26] Yu, H. K. (2005). Weighted fuzzy time series models for TAIEX forecasting. Physica A: Statistical Mechanics and its Applications, 349(3-4), 609-624.

[27] Indera, N. I., Yassin, I. M., Zabidi, A., & Rizman, Z. I. (2017). Non-linear autoregressive with exogeneous input (NARX) Bitcoin price prediction model using PSO-optimized parameters and moving average technical indicators. Journal of Fundamental and Applied Sciences, 9(3S), 791-808.

[28] Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical normalization and back propagation for classification. International Journal of Computer Theory and Engineering, 3(1), 1793-8201.

[29] Gilles, J. (2013). Empirical wavelet transform. IEEE Transactions on Signal Processing, 61(16), 3999-4010.

[30] I. Daubechies,Ten Lectures on Wavelets, Society for Industrial and Applied Mathematics, Philadelphia,1992.

[31] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[32] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5-6), 602-610.

[33] He. F.. Zhou. J.. Feng. Z. K.. Liu. G.. & Yang. Y. (2019). A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. Applied Energy. 237. 103-116.

[34] Li. Y.. Wu. H.. & Liu. H. (2018). Multi-step wind speed forecasting using EWT decomposition. LSTM principal computing. RELM subordinate computing and IEWT reconstruction. Energy Conversion and Management. 167. 203-219.

[35] Pavlyukevich, I. (2007). Lévy flights, non-local search and simulated annealing. Journal of Computational Physics, 226(2), 1830-1844.

[36] Yang, X. S., & Deb, S. (2009, December). Cuckoo search via Lévy flights. In 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC) (pp. 210-214). IEEE.

[37] Berrazouane, S., & Mohammedi, K. (2014). Parameter optimization via cuckoo optimization algorithm of fuzzy controller for energy management of a hybrid power system. Energy Conversion and Management, 78, 652-660.

[38] Rajabioun, R. (2011). Cuckoo optimization algorithm. Applied Soft Computing, 11(8), 5508-5518.

[39] Basu, M., & Chowdhury, A. (2013). Cuckoo search algorithm for economic dispatch. Energy, 60, 99-108.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: