ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

Combined CEEMDAN-CNN-BiLSTM-ATT Model for Forex Forecasting

Xin Li[1,a], Guoqiang Tang[1,b,*], Yumei Ren[1,c], Xuchang Chen[1,d]

[1]*Faculty of Science, Guilin University of Technology, Guilin, Guangxi Zhuang Autonomous Region, China*
[a]*llixin0219@163.com*, [b]*tanggq@glut.edu.cn*, [c]*848126613@qq.com*, [d]*499179952 @qq.com*
**Corresponding author*

Abstract: *With the deepening of economic globalization and the expanding scale of the foreign exchange market, the study of the volatility characteristics and forecasting of foreign exchange has received widespread attention.In this paper, a new forex portfolio forecasting model is proposed, which uses fully integrated empirical modal adaptive noise decomposition (CEEMDAN) to decompose the original forex price into sub-layers of different frequencies; then, convolutional neural network, bi-directional long and short term memory and attention mechanism module (CNN-BiLSTM-Attention) are used to combine the forecasting of each sub-layer; finally, the forecasting results of different frequency sub-layers are summed Reconstructed as the final prediction value. The forex series of USD, EUR, JPY and HKD against CNY were selected for empirical analysis, and based on the results of both evaluation metrics and DM tests, it was concluded that the CEEMDAN-CNN-BiLSTM-Attention model performed better in the forex market and had higher accuracy in exchange rate forecasting compared to the single model and other combined models.*

Keywords: *Foreign exchange forecasting; CEEMDAN; CNN-BiLSTM-Attention; DM test*

1. Introduction

As by far the world's largest financial market, the foreign exchange market includes large banks, central banks, currency speculators, multinational corporations, governments and other financial markets and institutions, is the place where foreign exchange transactions, exchange and other businesses are conducted. Although the history of the foreign exchange market is much shorter than that of the stock, gold, futures and interest markets, the pace of development has been phenomenal, with daily trading volume already reaching US$6 trillion, hundreds of times the daily stock trading volume of the New York Stock Exchange. Therefore, the possibility of constructing stable and accurate forecasting models to predict changes in foreign exchange has become one of the hotly debated and researched topics in the financial field.

Current models and methods for forecasting foreign exchange fall into three broad categories: traditional econometric models, artificial intelligence methods, and integrated and comprehensive approaches. Early forecasts on foreign exchange mainly revolve around ARMA and GARCH models, with Minakhi Rout et al[1] using adaptive ARMA models and differential evolution-based training to forecast currency exchange rates,Yu You et al[2]using GARCH-MIDAS methods to forecast short-term exchange rate fluctuations using currency fundamentals, and Stavros et al [3] building an AR(1)-GARCH(1,1)-skT framework to study the fitness of day-to-day volatility to exchange rates, and Kevin B et al [4] construct a multivariate GARCH-M approach to exchange rate shocks and trade studies. These types of econometric models are highly practical but do not capture enough the non-linear characteristics of foreign exchange data.

To better capture the nonlinear characteristics of foreign exchange data, researchers have started to experiment with more effective artificial intelligence methods.Katsuki Ito et al [5] built an LSTM model and used limit singletons to predict foreign exchange rates, Zhen Fang et al [6] built an adaptive LSTM-BN network model for exchange rate prediction, Pradeepta Kumar Sarangi et al [7] used an ANN-GA hybrid approach for forecasting INR to USD currency exchange rates, Amit R et al [8] used a bidirectional GRU and LSTM model for multi-currency exchange rate forecasting, Wang Jingyang et al [9] proposed a CNN-TLSTM model for forecasting FX closing prices, and Abedin et al [10] combined BR regression and BiLSTM for forecasting 21 currencies against USD. Such methods can more fully

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

extract features such as the non-linearity and volatility complexity of the FX series, but the need for optimization of a large number of parameters in the forecasting process may cause the forecasts to fall into a local optimum or overfitting state.

To address the shortcomings in artificial intelligence algorithms, some scholars have started to adopt decomposition integration class approach in the foreign exchange market.Chiun-Sin Lin et al [11] developed EMD-LSSVR model to improve the accuracy of foreign exchange forecasting, Wei Guo et al [12] proposed a hybrid model of VMD and ARIMA-TEF to improve the accuracy of forecasting financial time series, Sun Shaolong et al [13] integrated VMD and SVNN to forecast four major foreign currencies, Yang Hengli et al [14] applied EMD and back propagation neural network to forecast exchange rates, Bhusana et al [15] used EMD-SVR model to forecast foreign currencies, and Nur et al [16] integrated EMD and LSSVM and used PDC to reconstruct the foreign exchange decomposition term before forecasting. Such methods solve the problems of overfitting and local optimality that tend to exist in single AI algorithms, but the disadvantages of the decomposition algorithm itself, such as modal conflation and poor adaptivity, limit its forecasting ability.

In summary, the existing decomposition integration class methods for forex forecasting articles mostly use EMD,EEMD,CEEMD decomposition sequences, whose processes mostly suffer from modal mixing, long computational time consuming and large reconstruction errors; most of the AI class algorithms used cannot learn long distance dependencies and lack the extraction of key information.

To this end, a combined model of CEEMDAN and CNN-BiLSTM-ATT is proposed in the paper for foreign exchange forecasting, the main contributions of this paper are listed as follows: 1.This work presents a new deep learning architecture for high frequency and financial time series forecasting. The architecture designs a decomposed integrated model incorporating an attention mechanism to forecast high-frequency and low-frequency data separately to effectively predict foreign exchange rate trends and characteristics.2.In this paper, the combined model CEEMDAN-CNN-BiLSTM-ATT, which uses bi-directional long and short-term memory (BiLSTM) and convolutional neural network (CNN) for deep feature extraction, can better improve the forex prediction accuracy compared with the traditional model. In four sets of numerical experiments on exchange rates, this paper compares the combined model with nine other models such as GRU and LSTM using evaluation metrics, and in general, the combined model has better generalization ability.3.In order to obtain more comprehensive and valuable information to evaluate the merits of the models, this paper further compares the rationality of the deep learning architectures using DM tests, and empirically verifies that the combined models have good predictive stability.

The remainder of the paper is structured as follows: Sect. 2 we introduce the details of the proposed method. In Sect. 3, describes and analyses the experimental results. Section 4 we summarize the paper and outline further work in future.

2. Model construction

2.1. CEEMDAN

For the EMD algorithm to decompose the problem of modal mixing, EEMD and CEEMD methods are proposed, but these two algorithms will leave a certain amount of white noise when decomposing the signal, which affects the subsequent signal decomposition and processing, CEEMDAN[17] is therefore proposed as a more effective solution to the problem of transferring white noise from high to low frequencies and is constructed as follows.:

(1)Construct a sequence with noise:

$$X^{j}(t)=X(t)+W^{j}(t) \quad (j=1,2,...,N) \tag{1}$$

Where $X(t)$ is the actual value of the exchange rate at the t moment and $W^{j}(t)$ is the j Gaussian white noise signal obeying the standard normal distribution.

(2)The new signal is obtained by adding white noise to the signal to be decomposed, the new signal is EMD decomposed, and then the overall average of the resulting N modal components is performed to obtain the first-order eigenmode components and residuals of the CEEMDAN decomposition.

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

$$\overline{imf_1(t)} = \frac{1}{N}\sum_{j=1}^{N} imf_1^j(t) \tag{2}$$

$$r_1(t) = X^j(t) - \overline{imf_1(t)} \tag{3}$$

(3)The new signal is obtained by adding positive and negative pairs of white noise to the first-order residuals, which is used as a vehicle for EMD decomposition to further obtain the second-order modal components and second-order residuals:

$$\overline{imf_2(t)} = \frac{1}{N} e_1\left\{r_1(t)+\sigma_1 e_1\left[W^j(t)\right]\right\} \tag{4}$$

$$r_2(t) = r_1(t) - \overline{imf_2(t)} \tag{5}$$

where $e_1(\cdot)$ is the first order IMF operator of the EMD decomposition and σ_1 is the initial standard deviation of the added Gaussian white noise.

(4)Repeat the above steps until the residual signal is obtained as a monotonic function that cannot be further decomposed, at which point the original signal is reconstructed as:

$$X(t) = \sum_{k=1}^{K} \overline{imf_k(t)} + r(t) \tag{6}$$

2.2. CNN

Convolutional neural networks (CNN)[18] are feed-forward neural networks, where each neuron weight connects all neurons in the upper layer, resulting in a large number of weights and increasing the data volume and complexity of the entire network. The CNN has two important features, "local awareness" and "weight sharing", which can effectively extract data features while reducing the number of weights. The convolutional network is mainly composed of a convolutional layer and a pooling layer, and its convolution operation is shown in Equation 7.

$$y_i^k = f\left(y^{k-1} * W_i^k + b_i^k\right) \tag{7}$$

Where y^{k-1} denotes the input of the k convolutional layer, * denotes the convolutional operation, W_i^k denotes the weight of the k convolutional kernel of the i convolutional layer, and b denotes the bias of the k convolutional layer.

The pooling layer only compresses and reduces the dimensionality of the acquired feature information, with the aim of reducing the computational complexity and highlighting the main features.The basic structure of a CNN is shown in Figure 1.

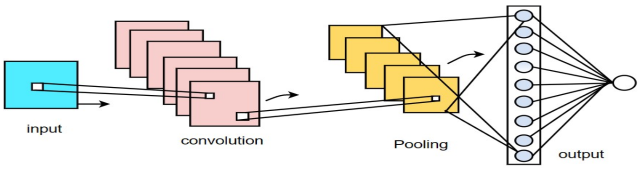


Figure 1: CNN structure.

2.3. BiLSTM

BiLSTM[19] is an improved model on the basis of LSTM, consisting of forward LSTM and backward LSTM, which takes into account the influence of the before and after data transformation while retaining the characteristics of LSTM in processing long time sequences, effectively improving the global nature of feature extraction. The sequence data enters the hidden layer through the input layer, and the forward and backward calculations are carried out respectively. The final output result is obtained by fusing the forward LSTM and backward LSTM output results in the output layer according to certain weights, and its output calculation formula is shown in equation (8)-equation (10):

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

$$h_q^f = lstm\left(y_q, h_{q-1}^f\right) \tag{8}$$

$$h_q^n = lstm\left(y_q, h_{q-1}^n\right) \tag{9}$$

$$h_q = \alpha h_q^f + \beta h_q^n \tag{10}$$

where equations 8 and 9 are the forward and backward hidden layer states at moment q. α and β represent the forward and backward hidden layer output weights, respectively. The structure of the BiLSTM is shown in Figure 2.

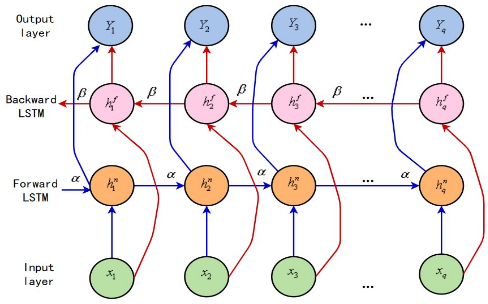


Figure 2: BiLSTM model structure.

2.4. Attention mechanism

Originally derived from the simulation of the attentional features of the human brain, Attention[20] was first applied to the field of image processing. In the field of deep learning, this mechanism can assign weights to different features according to their importance thus improving the efficiency of information processing. Its transformation process is shown in equations (11)-(14).

$$A_{qi} = V\tanh\left(Wy_q + Uy_i + b\right), \quad i = 1,2,3,\ldots,q-1 \tag{11}$$

$$a_{qi} = \frac{\exp\left(A_{qi}\right)}{\sum_{k=i}^{q}\exp\left(A_{qk}\right)}, i = 1,2,3,\ldots,q-1 \tag{12}$$

$$F = \sum_{i=1}^{q} a_{qi} \times y_i, \quad i = 1,2,3,\ldots,q-1 \tag{13}$$

$$y_q^{'} = f\left(F, y_q, h_q\right) \tag{14}$$

Where a_{qi} denotes the BLSTM hidden layer output value h_i for the current input attention weight , hq is the BiLSTM hidden layer state vector at the moment of q , V, W, U, b is the learning parameter of the model, which will be updated as the model is trained. The structure of the Attention unit is shown in Figure 3.

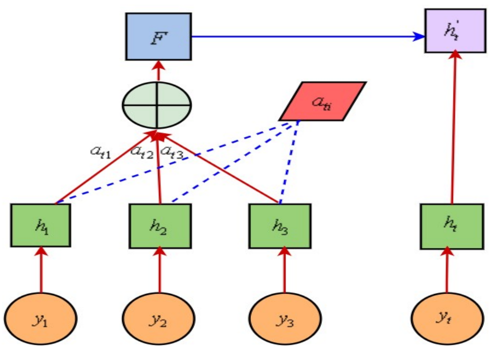


Figure 3: Structure of the attention mechanism module.

2.5. Constructing a portfolio model

To reduce the difficulty of modelling foreign exchange forecasts, this paper combines the temporal decomposition of CEEMDAN with CNN-BiLSTM temporal forecasting based on the attention mechanism to propose a high precision exchange rate forecasting method, and the forecasting method framework is shown in Fig 4.By performing CEEMDAN decomposition of the original data, CNN-BiLSTM-ATT modelling prediction of the eigenmode function IMF and summation and integration of the predicted values, the final high-precision sequence prediction values are obtained.

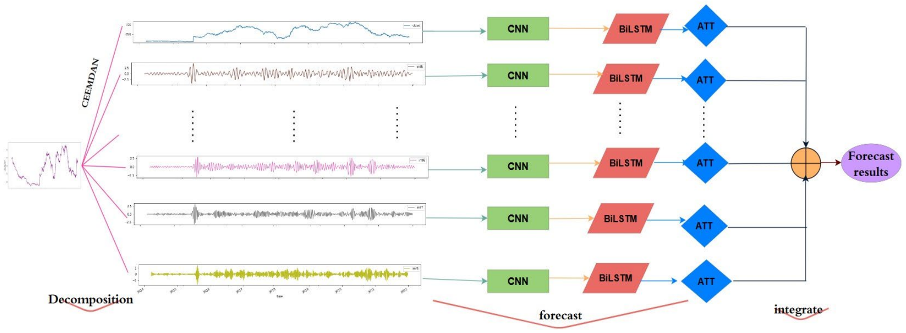


Figure 4: Structure of the predictive model.

Step 1: Decompose the foreign exchange data using CEEMDAN to decompose the RMB median exchange rate into IMFs and residuals arranged from high to low frequencies;

Step 2: The decomposed IMF components are normalized and the total data is divided into a training set and a test set according to a 4:1 ratio;

Step 3: Prediction of the pre-processed training and test sets using a CNN model consisting of a one-dimensional convolutional layer and a pooling layer to extract the local spatial state features of the data;

Step 4: Passing the predicted values from the CNN into the BiLSTM model and using BiLSTM modelling to learn the internal dynamic change patterns of the local features extracted by the CNN to extract temporal feature information;

Step 5: Using the features generated by the BiLSTM hidden layer as input to the Attention mechanism, the extracted temporal information is automatically weighted using the Attention mechanism to obtain a distinction of importance;

Step 6: The input layer of the Attention layer is used as the input of the fully connected layer, and the prediction results of each IMF are obtained and then summed up, and the final prediction results are obtained by inverse normalization;

In the prediction process, the Dropout technique was introduced to prevent overfitting of parts of the prediction model, the Mean Squared Error (MSE) was used for the loss function, and the Adam (Adaptive Moment Estimation) optimisation algorithm was used to update the network parameters of each layer in order to improve the generalisation capability of the model.

2.6. Model assessment

In order to make the evaluation of the model more objective and comprehensive, this paper chooses two methods, regression evaluation index and DM test, to assess both the accuracy of the prediction results and the degree of variation between prediction models.

2.6.1. Evaluation indicators

The difference between the predicted and actual values is often expressed in terms of the loss error, which is used to evaluate the predictive effectiveness of the model. In the paper, root mean square error $(RMSE)$, mean absolute error (MAE) and coefficient of determination (R^2) are used to assess the goodness of the model, calculated as follows.

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i-y_i\right)^2} \tag{15}$$

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{16}$$

$$R^2 = 1 - \frac{\sum_i\left(\hat{y}_i - y_i\right)^2}{\sum_i\left(\bar{y}_i - y_i\right)^2} \tag{17}$$

Where $\hat{y}_i$ is the predicted value, y_i is the actual value and $\bar{y}_i$ is the mean value of the data. The better the model is evaluated based on the value of R^2, the closer the result is to 1, the better the model fit is, while for the remaining two evaluation indicators, the closer the result is to 0, the smaller the model error is.

2.6.2. DM test

Although the above three evaluation metrics can compare the predictive ability of the models for a given data sample, the difference in such errors between two two models may not be significant, and therefore comparing only the predictive accuracy and loss functions of the models is not sufficiently convincing. Therefore, this paper uses the DM test[21]to further assess the strengths and weaknesses between two two models. The specific ideas and steps are as follows.

(1)From the values of the original data series, the predicted values of the different models are determined, and then the residuals of the two predicted series are calculated, and then the relative loss function is calculated based on this.

$$g_t = g\left(\varepsilon_{1,t}\right) - g\left(\varepsilon_{2,t}\right) \tag{18}$$

(3)With a relative loss function, a one-sided test is used with the following original and alternative hypotheses:

$$\begin{aligned} &H_0 \ : E\left(g_t\right) < 0 \\ &H_1 \ : E\left(g_t\right) \geq 0 \end{aligned} \tag{19}$$

The P value and significance level α are used to compare the model strengths and weaknesses. If $P < \alpha$ is used, the original hypothesis is rejected and Model 2 is better than Model 1, and vice versa, the original hypothesis is accepted and Model 2 is worse than Model 1. With the help of the DM test and the evaluation metrics above, the model prediction can be evaluated more comprehensively and objectively. With regard to the form of the loss function the mean squared error (MSE) is used here.

3. Empirical studies

3.1. Data description

The data used in this article are the exchange rate values of 100 units of foreign currencies against the RMB for a total of 2,915 days from 15 July 2010 to 15 July 2022 for four daily RMB mid-rates, namely USD, EUR, JPY and HKD, all from the State Administration of Foreign Exchange(http://www.safe.gov.cn/safe/rmbhlzjj/index.html).The first 80% of the data were taken as the training set, spanning from July 15, 2010 to February 21, 2020, and the second 20% were taken as the test set, spanning from February 24, 2020 to July 15, 2022.

Table 1: Results of descriptive statistics on exchange rates.

Name	Mean	Min	Max	T-value	Skewness	Peak	J-B	LB(20)
100USD/CNY	651.7	609.3	713.2	-1.837	0.323	2.003	171.603	0.0
100EUR/CNY	755.2	648.5	964.5	-1.495	0.629	3.223	110.095	0.0
100JPY/CNY	6.6	4.857	8.373	-0.906	0.766	2.605	303.803	0.0
100HKD/CNY	83.9	78.58	92.00	-1.550	0.384	2.081	174.091	0.0

The exchange rates of the four foreign currency assets against the RMB are shown in Figue 5, and the descriptive statistics are shown in Table 1. The ADF test for each of the four foreign currencies shows that the series is non-stationary at the 1% significance level; the Jarque-Bera test shows that all four foreign currencies do not obey a normal distribution; the ARCH effect of the series is tested using the Ljung-Box statistic, and the results indicate that the p-value is much less than 0.05 after a lag order of

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

more than 2, indicating that the foreign currency series has significant volatility aggregation. In summary, the forex series is non-stationary and contains a large amount of noise, and has long memory information, so compared to traditional models, this paper introduces CEEMDAN for adaptive decomposition, denoising and reconstruction of the series, and then uses a non-linear machine learning composite model to predict more reasonably.

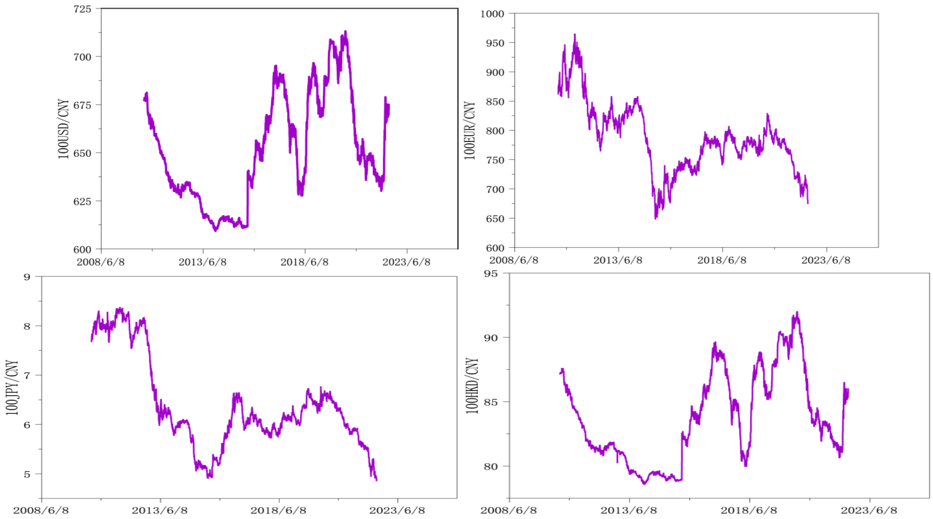


Figure 5: Four original sequences of exchange rates.

3.2. CEEMDAN decomposition and normalisation of data

Following the CEEMDAN decomposition process described in the previous section, an adaptive noise-complete ensemble empirical modal decomposition was performed on the four RMB daily mid-rate prices, and the results are shown in Figure 6. 100 USD/RMB, 100 EUR/RMB and 100 HKD/RMB were decomposed into 8 IMFs and 1 residual term in descending order, and 100 JPY/RMB was decomposed into 9 IMFs and 1 residual term, where the horizontal axis denotes the time series number and the vertical axis denotes the frequency of each IMF.

To improve the training speed of the model as well as the prediction accuracy of the model, the different spectra after decomposition are normalised according to equation (20).

$$X_n(t) = \frac{X_i(t) - X_{\min}(t)}{X_{\max}(t) - X_{\min}(t)} \tag{20}$$

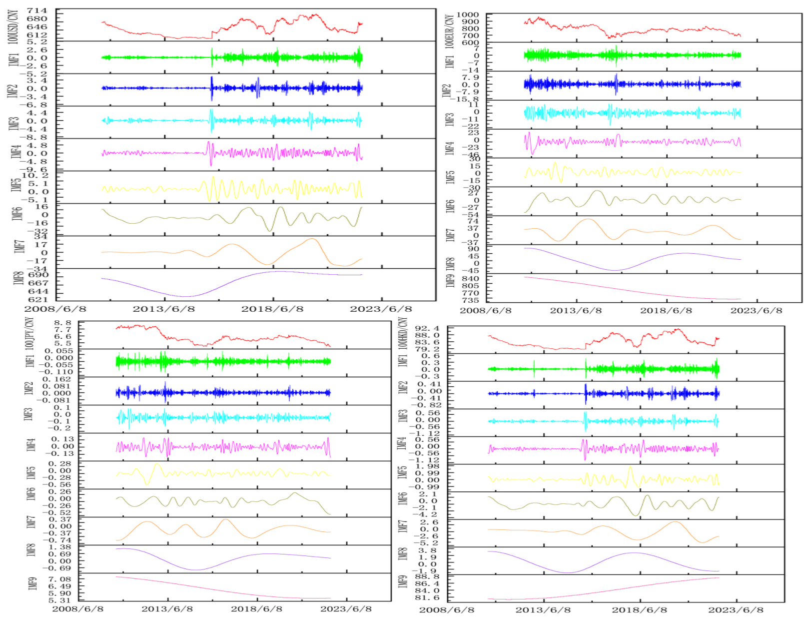


Figure 6: Decomposition results for the four exchange rates.

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

3.3. Time step setting

In order to get better prediction results, a suitable training time step needs to be found. Using USD/CNY 100 as an example, different time steps were set to compare the prediction results, and the results are shown in Figure 7. The experimental comparison found that the longer the step length, the longer the training time will be. When the step size is 3, the predicted value deviates from the actual result because the short time step does not take into account the influence of global factor changes. When the time step is set to 20, a larger time range is taken into account, but the effects of short time periods are relatively ignored and the predictions are not as good. In comparison, when the time step is set to 10, the error is minimised and the accuracy is higher, so 10 is used as the time step for subsequent model predictions.

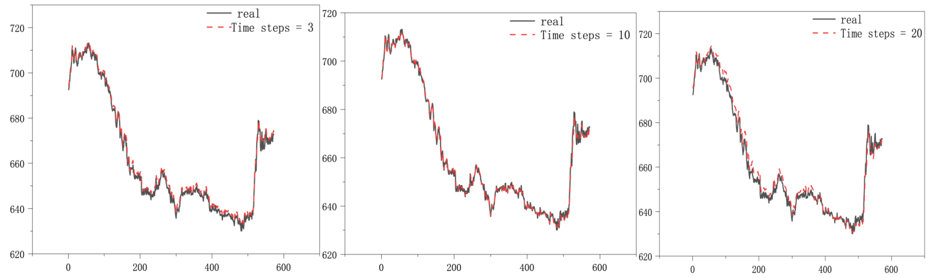


Figure 7: Predicted results with different step sizes.

3.4. Predicted results

In order to verify whether the proposed model is accurate and reliable in forecasting foreign exchange, the following section proceeds to test the model performance in two parts. The first part uses evaluation metrics to assess the performance strengths and weaknesses of this and other models; the second part assesses the inter-model goodness based on the DM test.

3.4.1. Comparison of evaluation indicators

Table 2: Test results for evaluation indicators.

Datasets	100USD/CNY			100EUR/CNY			100JPY/CNY			100HKD/CNY		
Indicators	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
LSTM	3.548	1.536	0.989	6.113	2.131	0.986	0.074	0.231	0.988	0.45	0.542	0.99
GRU	3.35	1.514	0.99	4.818	1.896	0.991	0.095	0.266	0.981	0.446	0.555	0.991
BiLSTM	3.623	1.584	0.989	4.5	1.828	0.992	0.066	0.223	0.99	0.447	0.551	0.991
CNNLSTM	4.212	1.933	0.985	5.355	2.037	0.989	0.063	0.215	0.992	0.437	0.541	0.991
CNNBiLSTM	3.197	1.461	0.992	4.545	1.85	0.992	0.063	0.214	0.992	0.411	0.521	0.992
CNN-BiLSTM-ATT	2.979	1.444	0.993	4.405	1.834	0.993	0.055	0.206	0.993	0.359	0.502	0.994
C-LSTM	2.457	1.338	0.995	3.509	1.664	0.995	0.068	0.252	0.991	0.324	0.485	0.995
C-GRU	1.863	1.839	0.997	2.371	1.394	0.997	0.064	0.249	0.992	0.285	0.456	0.996
C-BiLSTM	2.267	1.298	0.996	2.652	1.455	0.997	0.067	0.254	0.991	0.297	0.482	0.996
C-CNNLSTM	1.776	1.175	0.997	2.334	1.361	0.998	0.06	0.239	0.993	0.25	0.444	0.997
C-CNNBLSTM	1.687	1.118	0.998	2.296	1.369	0.998	0.049	0.217	0.995	0.203	0.383	0.998
C-CNN-BiLSTM-ATT	1.022	0.858	0.999	2.18	1.309	0.998	0.026	0.153	0.998	0.181	0.371	0.998

Note: C in the table stands for CEEMDAN (Fully Integrated Empirical Modal Adaptive Noise Decomposition).

Commonly used machine learning algorithms and combinatorial algorithms were evaluated using three evaluation metrics, and the results are shown in Table 2. In order to exclude the effects caused by environmental changes and parameter settings and to compare the performance of the models themselves only, the BiLSTM, LSTM and GRU were all set to 256 hidden layers and the CNN was set to 128 hidden layers in the prediction study.From the prediction results, the CNN-BiLSTM-ATT model has a R^2 value of 0.993, which is greater than the other single models, and RMSE and MAE values of 2.979 and 1.444, which are smaller than the other single models. In 100EUR/CNY, the LSTM has an RMSE value of 6.113 and a R^2 value of 0.986. Changing it to the BiLSTM model, the RMSE value is 4.5 and the R^2 value is

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

0.992, which is 1.613 less than the LSTM model RMSE value and the R^2 value is improved, also for the mid-price of JPY and HKD, indicating that the bi-directional LSTM forecasts This indicates that the two-way LSTM predictions are more accurate.The RMSE values of all four foreign currencies were further reduced and worthy of improvement after the inclusion of CNN models, reflecting the effectiveness of the inclusion of CNN models, and the improved generalisation ability of the models after the introduction of the attention mechanism on this basis, indicating the reasonableness of all components of the forecasting models proposed in this paper. Some benchmark models perform better in one type of foreign exchange forecasting, but not significantly in other foreign exchange forecasting, indicating that the forecasting level of such models is not stable. The introduction of the CEEMDAN algorithm significantly improves the forecasting accuracy of all models compared. In the 100USD/CNY series, the RMSE and MAE of the combined model were 1.022 and 0.858 respectively, which were smaller than the other forecasting models, and the R^2 value was 0.999, which ranked first among the forecasting models. Similarly, among the other three foreign exchange forecasts, the models that incorporate the decomposition algorithm have a much higher forecasting ability than the other models. This indicates that the CEEMDAN-CNN-BiLSTM-ATT model performs better in forex forecasting based on the results of the evaluation metrics.

3.4.2. DM test

Table 3: DM test results (1).

		GPU	BiLSTM	CNNLSTM	CNNBLSTM	CNN-BiLSTM-ATT
100USD/CNY	LSTM	3.524 (0.000)	3.251 (0.001)	-3.103 (0.002)	7.461 (0.000)	3.925 (0.000)
	GPU		-3.251 (0.001)	-4.912 (0.000)	3.543 (0.000)	2.675 (0.008)
	BiLSTM			-3.103 (0.002)	7.461 (0.000)	3.925 (0.000)
	CNNLSTM				5.817 (0.000)	3.925 (0.000)
	CNNBLSTM					2.568 (0.010)
100EUR/CNY	LSTM	8.237 (0.000)	8.886 (0.000)	5.432 (0.000)	9.413 (0.000)	8.595 (0.000)
	GPU		5.595 (0.000)	-4.057 (0.000)	2.908 (0.004)	2.664 (0.008)
	BiLSTM			-7.361 (0.000)	4.080 (0.000)	8.739 (0.000)
	CNNLSTM				7.032 (0.000)	11.869 (0.000)
	CNNBLSTM					1.969 (0.050)
100JPY/CNY	LSTM	-10.094 (0.000)	4.336 (0.000)	7.056 (0.000)	6.199 (0.000)	10.223 (0.000)
	GPU		10.460 (0.000)	10.086 (0.000)	11.231 (0.000)	12.808 (0.000)
	BiLSTM			2.746 (0.006)	2.910 (0.004)	11.244 (0.000)
	CNNLSTM				0.249 (0.000)	3.418 (0.007)
	CNNBLSTM					3.295 (0.010)
100HKD/CNY	LSTM	0.368 (0.713)	0.360 (0.719)	2.987 (0.003)	7.923 (0.000)	3.678 (0.002)
	GPU		-0.010 (0.921)	1.392 (0.165)	4.417 (0.000)	4.180 (0.000)
	BiLSTM			1.354 (0.176)	5.945 (0.000)	4.366 (0.000)
	CNNLSTM				8.366 (0.000)	4.366 (0.000)
	CNNBLSTM					3.143 (0.002)

(1)DM test for CNN-BiLSTM-ATT

Before using the combined model, the local prediction model (CNN-BiLSTM-ATT) was first DM-

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

tested against commonly used deep learning and machine learning methods to verify the effectiveness of the chosen algorithm. Here the loss function is taken to be MSE and the results are shown in Table 3. The p-value of the DM test is shown in parentheses, and the test is only meaningful if the value is less than 0.05. The positive or negative DM value is used to assess how good the model is, if the value is less than 0, the model in the row performs better, and if the value is greater than 0, the model corresponding to the column is better. The results show that the LSTM and GRU and BiLSTM models do not perform consistently well for the four RMB mid-rates, and there is some risk in using these benchmark models for forecasting, compared to the models with CNNs, which perform more robustly, with the CNNBiLSTM model having the highest forecasting accuracy, with DM test values greater than 0 compared to the other three benchmark models. The CNN-BiLSTM-ATT model performs the best among the three intermediate price forecasts, with DM test values greater than zero, except for 100USD/CNY where the p-value is greater than 0.05. Therefore, on balance, this model has better forecasting ability.

(2)DM test for CEEMDAN-CNN-BiLSTM-ATT

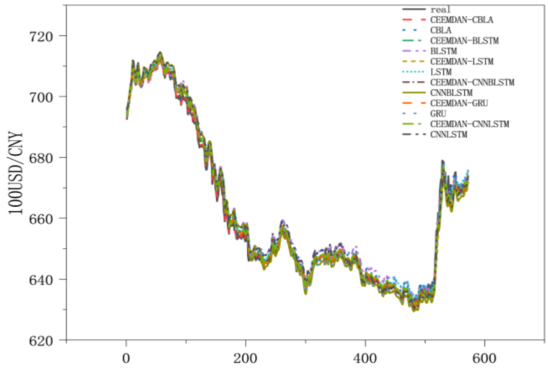
In the DM test, this combined model was tested against other models containing decomposition algorithms to verify the effectiveness of the decomposition part of the combined model. The detailed results are shown in Table 4. In USD/RMB100, the prediction ability of the combined CNN-BiLSTM-ATT model is stronger than the rest of the models, and the prediction ability of the basic models LSTM and GRU is relatively poor; in general, the prediction model proposed in this paper performs better and has stronger prediction ability compared with the other models. Figures 8-11 show the fit of each model to the true values on the test set.

Table 4: DM test results (2).

		C-GPU	C-BiLSTM	C-CNNLSTM	C-CNNBLSTM	C-CBLA
100USD/CNY	C-LSTM	8.3225 (0.000)	3.7992 (0.000)	7.9451 (0.0000)	9.1719 (0.0000)	11.1281(0.0000)
	C-GPU		-8.3060 (0.000)	3.9987 (0.0000)	9.2010 (0.0000)	11.6869(0.0000)
	C-BiLSTM			7.6916 (0.0000)	9.6783 (0.0000)	11.0369(0.0000)
	C-CNNLSTM				4.4887(0.0000)	13.0513(0.0000)
	C-CNNBLSTM					10.4895(0.000)
100EUR/CNY	C-LSTM	10.7724 (0.000)	8.9263 (0.000)	10.7723 (0.000)	10.2913 (0.000)	9.2509 (0.0000)
	C-GPU		-10.2595 (0.000)	5.0798 (0.000)	1.9196 (0.055)	1.4548 (0.146)
	C-BiLSTM			10.2594 (0.000)	9.3587 (0.000)	5.2839 (0.000)
	C-CNNLSTM				1.3082 (0.191)	1.9197 (0.055)
	C-CNNBLSTM					1.4547 (0.146)
100JPY/CNY	C-LSTM	7.6793 (0.000)	0.5164 (0.606)	21.8323 (0.000)	18.7392 (0.000)	37.7977(0.000)
	C-GPU		-2.3581 (0.019)	13.2994 (0.000)	27.3950 (0.000)	57.0044(0.000)
	C-BiLSTM			5.2690 (0.000)	15.8863 (0.000)	32.5596(0.000)
	C-CNNLSTM				14.4937 (0.000)	42.4029(0.000)
	C-CNNBLSTM					49.7954(0.000)
100HKD/CNY	C-LSTM	4.0075 (0.000)	4.3444 (0.000)	7.1177 (0.000)	8.9705 (0.000)	9.2862 (0.000)
	C-GPU		-1.4622 (0.144)	3.6883 (0.002)	9.9949 (0.000)	8.8483 (0.000)
	C-BiLSTM			8.3296 (0.000)	9.2514 (0.000)	10.2528(0.000)
	C-CNNLSTM				5.9494 (0.000)	8.7988 (0.000)
	C-CNNBLSTM					3.5395 (0.000)

Note: C in the model stands for CEEMDAN and CBLA stands for CNN-BiLSTM-ATT.

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713



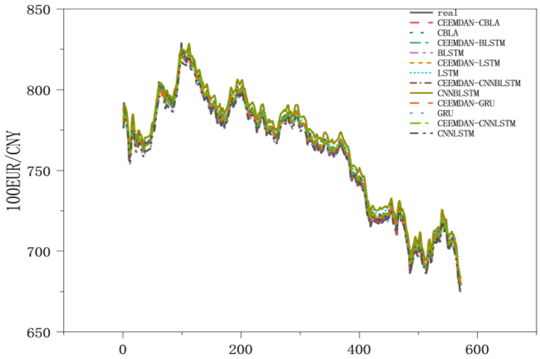
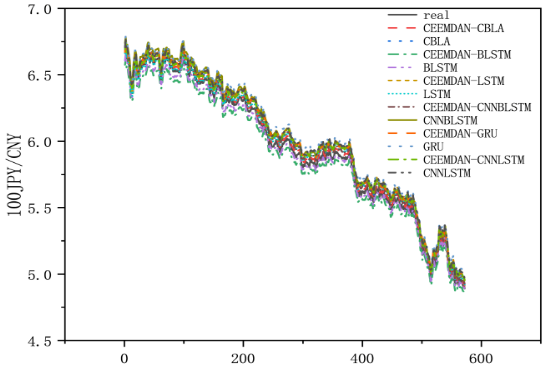


Figure 8: Forecast results for 100USD/CNY. Figure 9: Forecast results for 100EUR/CNY.



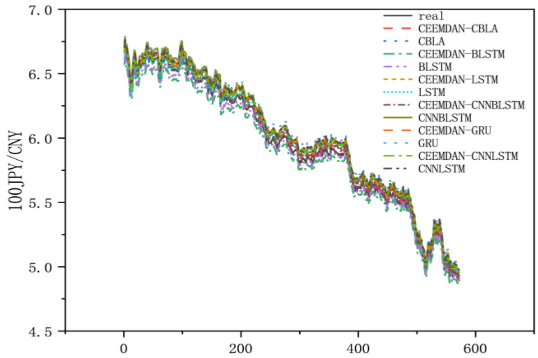


Figure 10: Forecast results for 100JPY/CNY. Figure 11: Forecast results for 100HKD/CNY.

From the above results, it can be inferred that the introduction of CEEMDAN in the hybrid algorithm can reduce the loss error, which plays a crucial role in prediction. In prediction, the introduction of a two-way LSTM model performs better, and the inclusion of an attention mechanism can improve prediction accuracy. The combined model reflects both long-term and short-term dependencies and also extracts key information with high forecasting accuracy. This section therefore illustrates that the CEEMDAN-CNN-BiLSTM-ATT model has better performance capability in forecasting foreign exchange.

3.5. Discussion

From the results of both evaluation metrics and DM tests, the model proposed in this paper performs well in forex forecasting, and its accuracy is significantly improved compared to the models used by previous authors.The LSTM model[23], GRU model[24]and BiLSTM model[25] as basic machine learning models have large deviations in forecasting foreign exchange and do not predict the changes in the trend of foreign exchange rates more accurately. The combined CNNLSTM model [26] and CNNBiLSTM model[27] and the prediction model after adding the attention mechanism have improved the prediction accuracy compared with the individual models, but there are still errors, and the prediction cannot be more accurate. Therefore, we introduce EMD[28], EEMD[29], and CEEMDAN algorithms to decompose, predict, and integrate the prediction process, and the results show that the combination of CEEMDAN decomposition algorithm and machine learning model has higher prediction accuracy, which indicates that the algorithm used in this paper has better prediction effect.

4. Conclusion

Effective and timely forecasting of foreign exchange rates is of great importance to those involved, investors and financial practitioners alike. Researchers have also been trying to introduce new methods to analyse and forecast foreign exchange rates in order to improve the accuracy of forecasting. In machine learning, CNN can extract effective features from data, BiLSTM is suitable for prediction of time series data, and attention mechanism can improve the network feature extraction ability. These two deep learning models and attention mechanism algorithms perform well in the forecasting of foreign exchange rates. And CEEMDAN, as a decomposition algorithm that can solve the problem of transferring white noise from high to low frequencies, has also dealt with many problems in time series prediction. Therefore, this paper constructs a new combinatorial model CEEMDAN-CNN-BiLSTM-ATT to forecast

ISSN 2616-5775 Vol. 6, Issue 7: 88-100, DOI: 10.25236/AJCIS.2023.060713

these four RMB exchange rate median prices. The empirical RMSE, MAE, and R^2 results show that the model performs well in forecasting the four foreign exchange rates, and the DM test results also indicate that CEEMDAN-CNN-BiLSTM-ATT has better forecasting results compared with several other models. In summary, the CEEMDAN-CNN-BiLSTM-ATT model proposed in this paper can play a certain positive role in the foreign exchange market.

Finally, this paper only analyses the univariate time series and does not consider other factors that influence foreign exchange volatility. If these factors were introduced into the model, perhaps the forecasting performance would be better. In future work, it remains to be explored what influencing factors could be introduced and improve the forecasting performance of the model, and hopefully there will be further research on this to provide valuable reference for investors.

Acknowledgments

We are very grateful for the research data provided by the State Administration of Foreign Exchange. We are very grateful to the National Natural Science Foundation of China (62166015) and General Program of Guangxi Natural Science Foundation (2022GXNSFAA035499) for their support of thesis creation work.

References

[1] Minakhi Rout, Babita Majhi, Ritanjali Majhi, Ganapati Panda, Forecasting ofcurrency excha-nge rates using an adaptive ARMA model with differential evolution based training, Journal of King Saud University-Computer and Information Sciences, Volume 26, Issue1, 2014, 7-18. http://go. microsoft. com/ fwlink/p/?LinkId=255141

[2] Yu You, Xiaochun Liu. Forecasting short-run exchange rate volatility with monetary fundame-ntals: A GARCH-MIDAS approachJ, Bank. Financ. (2020). https://doi. org/10. 1016/j. jbankfin. 2020. 105849.

[3] Stavros D, Artemis P. Multiple-Days-Ahead Value-at-Risk and Expected Shortfall Forecasting for Stock Indices, Commodities and Exchange Rates:Inter-Day versus Intra-Day Data[J], International Review of Financial Analysis, Volume 49, 2017:176-190. https://doi. org/10. 1016/j. irfa. 2016. 10. 008.

[4] Kevin B, Aaron D. Exchange Rate Shocks and Trade:A Multivariate GARCH-M Approach[J], Journal of International Money and Finance, Volume 37, 2013: 282-305, https://doi. org/10. 1016/ j. jimonfin. 2013. 05. 010.

[5] Katsuki Ito, Hitoshi Iima, Yoshihiro Kitamura, LSTM forecasting foreign exchange rates usin-g limit order book, Finance Research Letters, Volume 47, Part A, 2022. https://doi. org/10. 1016/j. frl. 2021. 102517.

[6] Zhen Fang, Xu Ma, Huifeng Pan, Guangbing Yang, Gonzalo R. Arce, Movement forecasting of financial time series based on adaptive LSTM-BN network, Expert Systems with Applications, Volume 213, Part C, 2023, 119207, https://doi. org/10. 1016/j. eswa. 2022. 119207.

[7] Pradeepta Kumar Sarangi, Muskaan Chawla, Pinaki Ghosh, Sunny Singh, P. K. Singh, FOREX tre-nd analysis using machine learning techniques: INR vs USD currenyexchange rate using ANN-GA hybrid approach, Materials Today: Proceedings, Volume 49, Part 8, 2022, Pages 3170-3176. https://doi. org/ 10. 1016/j. matpr. 2020. 10. 960.

[8] Amit R, Avinash J. Using Bidirectional GRU and LSTM Neural Network Methods for Multi-C-urrency Exchange Rates Prediction[J]. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(7). https://kns. cnki. net/kcms/detail/detail. aspx?FileName= SQFH1140BBCCB F6E3CA99B750A5BEA2511A9&DbName=GARJ2021_1

[9] Wang J Y, Wang X X, Li J Z, et al. A Prediction Model of CNN-TLSTM for USD/CNY Excha-nge Rate Prediction[J]. 2021, 9.

[10] Abedin MZ, Moon M H, Hassan M K, et al. Deep Learning-Based Exchange Rate Prediction During the COVID-19 Pandemic. [J]. Annals of Operations Research, 2021. https://doi. org/10. 1007/S10479-021-04420-6.

[11] Chiun-Sin Lin, Sheng-Hsiung Chiu, Tzu-Yu Lin, Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting, Economic Modelling, Volume 29, Issue 6, 2012, Pages 2583-2590. https://doi. org/10. 1016/j. econmod. 2012. 07. 018.

[12] Wei Guo, Qingfu Liu, Zhidan Luo, Yiuman Tse, Forecasts for international financial series with VMD algorithms, Journal of Asian Economics, Volume 80, 2022, 101458. https://doi. org/10. 1016 /j. asieco. 2022. 101458.

[13] Shaolong Sun, Shouyang Wang, Yunjie Wei, A new multiscale decomposition ensemble appr-oach

for forecasting exchange rates, Economic Modelling, Volume 81, 2019, Pages 49-58. https://doi. org/ 10. 1016/j. econmod. 2018. 12. 013.

[14] Yang H L, Lin H Z. Applying EMD-Based Neural Network to Forecast NTD/USD Exchange Rate[C]. The 7th International Conference on Networked Computing and Advanced Information Management, 2011, 352-357. https://ieeexplore. ieee. org/document/5967574/ citations? tabFilter =papers#citations

[15] Bhusana P, Chris T. Improving Prediction of Exchange Rates Using Differential EMD[J]. Ex-pert Systems with Applications, 2013, P377-384, https://doi. org/10. 1016/j. eswa. 2012. 07. 048.

[16] Nur I A R, Ani S, Ruhaidah S. New Hybridization of Empirical Mode Decomposition and Le-ast Squares Support Vector Machine Model in Forecasting Malaysia Exchange Rates[J]. 2017 International Conference on Robotics, Automation and Sciences(ICORAS), 2017, pp. 1-5. https://doi. org/10. 1109/ ICORAS. 2017. 8308069

[17] Maria E T, Marcelo A C, Gaston S, et al. A Complete Ensemble Empirical Mode Decomposit-ion with Adaptive Noise[C]. Speech and Signal Processing. 2011:4144-4147. https://doi. org/10. 1109/ ICASSP. 2011. 5947265.

[18] Lecun Y, Bottou L, Bengio Y, et al. Gradient-Based Learning Applied to Document Recognitio-n[J]. Proceedings of the IEEE, 1998, 86(11). https://doi. org/10. 1109/5. 726791.

[19] Li W J, Qi F, Tang M, et al. Bidirectional LSTM with Self-Attention Mechanism and Multi-Channel Features for Sentiment Classification[J]. Neurocomputing, 2020, 387. https://doi. org/ 10. 1016/j. neucom. 2020. 01. 006.

[20] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]. arXiv, 2017.

[21] Diebold F X, Mariano R S. Comparing Predictive Accuracy[J]. Journal of Business & Econo-mic Statistics, 2002, 20(1).

[22] Liang Y H, Lin Y, Lu Q. Forecasting Gold Price Using A Novel Hybrid Model with ICEEM-DAN and LSTM-CNN-CBAM[J]. Expert Systems With Applications, 2022, 206. https://doi. org/10. 1016/ j. eswa. 2022. 117847.

[23] Hum Nath Bhandari, Binod Rimal, Nawa Raj Pokhrel, Ramchandra Rimal, Keshab R. Da-hal, RajendraK. C. Khatri, Predicting stock market index using LSTM, Machine Learning with Applications, Volume 9, 2022, 100320, ISSN 2666-8270, https://doi. org/10. 1016/j. mlwa. 2022. 100320.

[24] Umang Gupta, Vandana Bhattacharjee, Partha Sarathi Bishnu, StockNet—GRU based stock index prediction, Expert Systems with Applications, Volume 207, 2022, 117986, ISSN 0957-4174, https://doi. org/10. 1016/j. eswa. 2022. 117986.

[25] Mo Yang, Jing Wang, Adaptability of Financial Time Series Prediction Based on BiLSTM, Procedia Computer Science, Volume 199, 2022, Pages 18-25, ISSN 1877-0509, https://doi. org/10. 1016/ j. procs. 2022. 01. 003.

[26] Furkan Elmaz, Reinout Eyckerman, Wim Casteels, Steven Latré, Peter Hellinckx, CNN-LST-M architecture for predictive indoor temperature modeling, Building and Environment, Volume 206, 2021, 108327, ISSN 0360-1323, https://doi. org/10. 1016/j. buildenv. 2021. 108327.

[27] Reuben Varghese Joseph, Anshuman Mohanty, Soumyae Tyagi, Shruti Mishra, Sandeep Ku-mar Satapathy, Sachi Nandan Mohanty, A hybrid deep learning framework with CNN and Bi-directional LSTM for store item demand forecasting, Computers and Electrical Engineering, Volume 103, 2022, 108358, ISSN 0045-7906, https://doi. org/10. 1016/j. compeleceng. 2022. 108358.

[28] Huang N. E. , Shen Z. , Long S. , Wu M. , Shih H. , Zheng Q. , Liu H. (1998). Theem-pirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proceedings of the Royal SocietyA-Mathematical Physical and Engineering Sciences, 454(1971), 903–995. https://doi. org/10. 1098/ rspa. 1998. 0193.

[29] Wu Z., & Huang N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. Adv. Adapt. Data Anal. , 1, 1–41. https://doi. org/10. 1142/ S1793536909000047.