HUIT Journal of Science ***25*** *(6) (2025) 142-149*

DOI: https://doi.org/10.62985/j.huit_ojs.vol25.no6.370

FORECASTING VN-INDEX TIME SERIES BASED ON CEEMDAN DECOMPOSITION AND DEEP LEARNING

Nguyen Hoang An, Ha Binh Minh, Huynh Do Bao Chau, Phan Dinh Phung*

Ho Chi Minh University of Banking

*Email: *phungpd@hub.edu.vn*

Received: 21 May 2024; Revised: 16 February 2025; Accepted: 5 March 2025

ABSTRACT

In this note we propose a model of prediction for VN-Index time series by combining three deep learning techniques including Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), Temporal convolutional networks (TCN) and Gated recurrent unit neural networks (GRUNN) together. The aim of our research is to reduce the number of components in CEEMDAN decomposition so that it decreases time consumption for training models. Instead of dividing the components in CEEMDAN decomposition into many parts, our proposed model grouping these components into two main groups, one with high permutation entropy and other with low permutation entropy. Empirical results indicate the better performance of the model for predicting a financial time series VN-Index than those of other models.

Keywords: Forecasting VN-Index time series, Temporal convolutional networks, Gated recurrent unit neural networks, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise.

1. INTRODUCTION

The Stock Exchange market plays an important role to measure the flow of capital in the economics of a country. It is the market place where stocks, bonds, and other financial instruments are traded within many parties involved such as investors, investment banks, companies, hedge funds, governments, and other financial institutions. In Vietnam, the most important indicator of the Stock Exchange market is VN-Index.

Predicting the prices of VN-Index in various period of time, long or short, is important for investors to make their right decisions. It needs to be smart for investors and other institutions to make decision in the stock exchange market due to the fluctuation of VN-Index in every day. Considering VN-Index as a financial time series, the problem of predicting VN-Index becomes the one of forecasting time series which has been of much interest recently (see [1, 2] and the references therein). The problem can be divided into two main branches based on the techniques underlying: statistical techniques and deep learning techniques. In recent years, some certain deep learning models are applied in predicting time series such as Deep Feed Forward Neural Network (DFFNN), Recurrent Neural Network (RNN), Elman RNN, Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), and a variant of CNN, Temporal Convolutional Networks (TCN) [2, 3]. For improving the accuracy and reducing the noise in prediction, one often employs Empirical Mode Decomposition (EMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise model (CEEMDAN) and their variants or combinations as the powerful tools to achieve the higher quality of forecasting accuracy [4-13]. For instance, the CEEMDAN and LSTM models are combined to apply in [6-8], or the CEEMDAM is associated with other predicting algorithms, for example, the flower pollination one or chaotic local search (CLSFPA) in [9]. In [10], a trio combination of EMD-PE-ANN is used in which permutation entropy (PE) is for reconstructing the components of IMF; LSTM model for multi-step prediction of chaotic time series based on dilated convolution network is applied in [11]. Especially, in [12-13] the authors use the TCN interval prediction model for wind speed forecasting or combine with the GRU to predict electricity and energy data.

In this article we propose the other combination of prediction models so-called CEEMDAN-TCN-GRU model, in which we combine three well-known techniques of deep learning. The advantages of our prediction models is as follows. (1) The technique of CEEMDAN decomposition is used firstly in this kind of prediction models; (2) By grouping the components of CEEMDAN decomposition into two groups based on the value of permutation entropy, we reduce the time consumption for training models; (3) Each group is trained by using deep learning model whose architecture is the combination of Temporal convolutional networks (TCN) and Gated recurrent unit neural networks (GRUNN); (4) Finally, the predicted results of two group are summed together in order to assess the overall result.

The rest of paper is organized as follows. Section 2 is to present the proposed models. Section 3 is devoted to perform empirical results with real financial VN-Index time series data need to be predicted. Section 4 summarizes the paper with conclusion.

2. PROPOSED MODEL: FORECASTING VN-INDEX TIME SERIES

In this section we introduce some deep learning techniques to build our proposed model.

2.1. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAM)

CEEMDAN is an advanced technique of Empirical mode decomposition (EMD). Proposed by Huang et al. [14], EMD is an adaptive time-frequency signal processing method which decomposes the signal according to the time scale feature of the data without presetting of any basis function. It decomposes the time series into a finite number of components which is called Intrinsic Mode Functions (IMFs). Each IMF component represents different feature of the original signal at different time scales.

2.2. Temporal Convolutional Networks (TCNs)

The architecture of TCNs is shown in *Figure 2* where the activation values for each layer are computed by the values of previous layer. Specifically, the values of the neurons from the previous layer will contribute to those of the neurons in the next layer by using dilated convolution, both local and temporal information are captured, and that is the reason the method called Temporal Convolutional Networks.

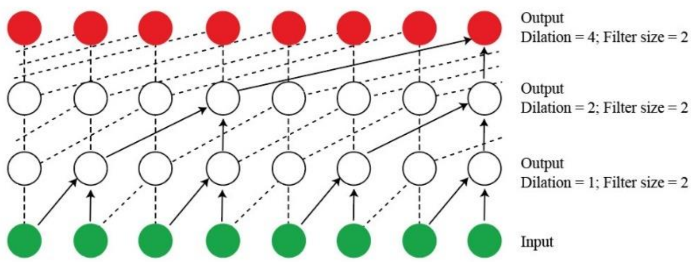

Figure 1. Visualization of Temporal Convolutional Network

2.3. Gated recurrent unit neural network (GRUNN)

The architecture of GRUNN is shown in *Figure 2*. GRUNN has only two gate control units (reset gate and update gate) which are simple and flexible structure [15]. As shown in *Figure 2*, reset gate in GRUNN is the merger of the input gate and forget gate in LSTM. The gate is closed when its value is near zero, which forbids the outflow of information. Furthermore, reset gate is used to erase prior memory cell states and to capture medium and short-term dependence relations. Correspondingly, update gate is employed to update current memory cell states via candidate.

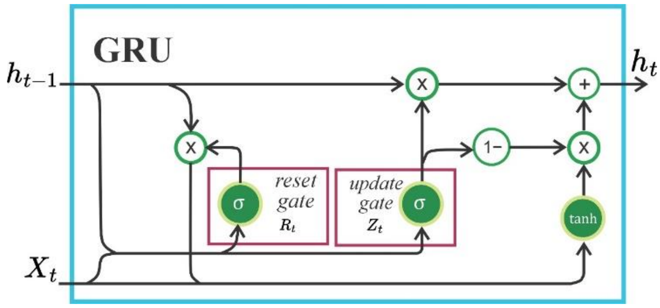

Figure 2. Structure of GRU network

2.4. Proposed models

To establish forecasting models for VN-Index, our models use CEEMDAN decomposition and the architecture of TCN-GRU structure as follows. First the original VN-Index time series is decomposed into several IMF sequences. Each IMF has different complexity and requires different optimal parameters to train. To simplify it we divide IMF sequences into two groups based on the analysis of PE: one group with higher PE and the other group with lower PE. The rest is to determine two set of optimal hyper-parameter for two sets of IMF, which has the benefit that reduces rapidly the time consumption for searching hyper-parameter. In detail, our proposed model is summarized in Algorithm 1 as follows.

Algorithm 1 Predicting VN-Index time series by CEEMDAN-TCN-GRU model

Input: The training dataset of VN-Index time series

Output: The predicted VN-Index time series

Step 1: Decompose time series VN-Index into several components IMF by CEEMDAN decomposition.

Step 2: This is the critical step in our proposed model. By choosing threshold *PE=0.5*, the component IMF with *PE > 0.5* can be considered as high complexity, and the rest can be considered as low complexity

Step 3: In each group we choose one component IMF having the highest PE to optimize hyper-parameters.

Step 4: Each IMF is trained by using the architecture of TCN-GRU structure shown in Figure 3 and the hyper-parameters in Step 3.

Step 5: Assess the prediction results obtained from proposed model by using three error measures as follows: mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). These measures are given by

$$\text{MAE} = \frac{1}{\text{N}}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

$$\text{MAPE} = \frac{\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i}}{N}$$

where y_i is the actual value at time i, and $\hat{y}_i$ is the predicted value.

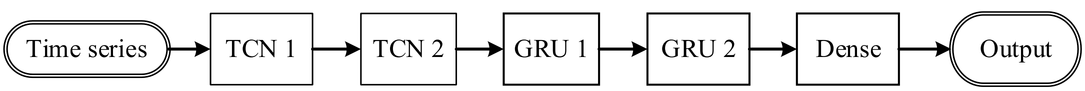

Figure 3. The proposed TCN-GRU model structure

Furthermore, we measure the performance of each forecasting model, regarding the classification problem of predicting whether VN-Index would increase or decrease on the next trading day. Comparing with the variation of actual data VN-Index, we can construct the confusion matrices, which help to determine the accuracy (ACC) of proposed models.

3. EMPIRICAL RESULTS

3.1. Research data

For testing the performance of the forecasting model, the daily closing price of VN-Index are selected from Ho Chi Minh Stock Exchange. Fig. 4 presents the data of the time series which is taken from January 21, 2016 to June 05, 2023, including 1839 data points. The statistical analysis of the time series data is presented in Table 1. We select 80% data for training set, and the latter 20% data for the test set.

Table 1. Descriptive statistics for daily closing price of VN-Index

Year	Count	Mean	Std	Min	Max
2016	238	629.78	45.73	521.88	688.89
2017	250	780.17	76.90	672.01	984.24
2018	248	1,008.22	81.60	888.69	1,204.33
2019	250	971.54	28.14	878.22	1,024.91
2020	252	889.48	90.03	659.21	1,103.87
2021	250	1,311.77	111.57	1,023.94	1,500.81
2022	249	1,252.68	179.37	911.90	1,528.57
2023	102	1,059.62	19.08	1,021.25	1,117.10
Total	**1,839**	**984.20**	**239.08**	**521.88**	**1,528.57**

Figure 4. VN Ho Chi Minh Stock Index from Jan-2016 to Jun-2023
Source: Ho Chi Minh Stock Exchange

3.2. Numerical results

In this note we perform the models on the virtual computer with 8 vCPU, 32 GB RAM and use Jupyter Lab and Tensorflow Framework.

The training process begins with input size 8 and the hyper-parameters for search space as shown in Table 1. In these empirical models, TCN layers have their dilation obtained values as 1, 2, 4, ReLU with kernel size 2 is the transfer function, the GRU layer activation function chosen is LeakyReLU, and transfer function of the output layer is Tanh. The epochs, the number of training, is 1000, and the Loss function is chosen as Mean Squared Error.

To obtain the better prediction result, the optimal hyper-parameters are selected through tuning hyper-parameter optimization by using the method called Tree-structured Parzen Estimator (TPE) [16] with search space shown in Table 2. The optimal hyper-parameters are presented in Table 3.

Table 2. Hyperparameter search space

Hyperparameter	Search space
Number of filters in TCN layer	*nb_filter:* 2, 4, 8, 16
Number of neural units in GRU layer	*neural units:* 64, 128, 256
Optimization algorithm	Stochastic Gradient Descent (SGD) Adaptive Moment Estimation (Adam) Root Mean Square Propagation (RMSprop)

Table 3. Optimal hyperparameter of two groups of model

	CEEMDAM-TCN-GRU-HF	CEEMDAM-TCN-GRU-LF
TCN layer 1	*nb_filters:* 16	*nb_filters:* 8
TCN layer 2	*nb_filters:* 4	*nb_filters:* 2
GRU layer 1	*neural units:* 128	*neural units:* 128
GRU layer 2	*neural units:* 64	*neural units:* 64
Optimization algorithm	SGD	Adam

In the first step, the original VN-Index time series is decomposed into eight IMFs in Figure 5, varying from high frequency to low frequency. The high-frequency components are more difficult to predict than the low-frequency ones, but the latter have an important role in prediction.

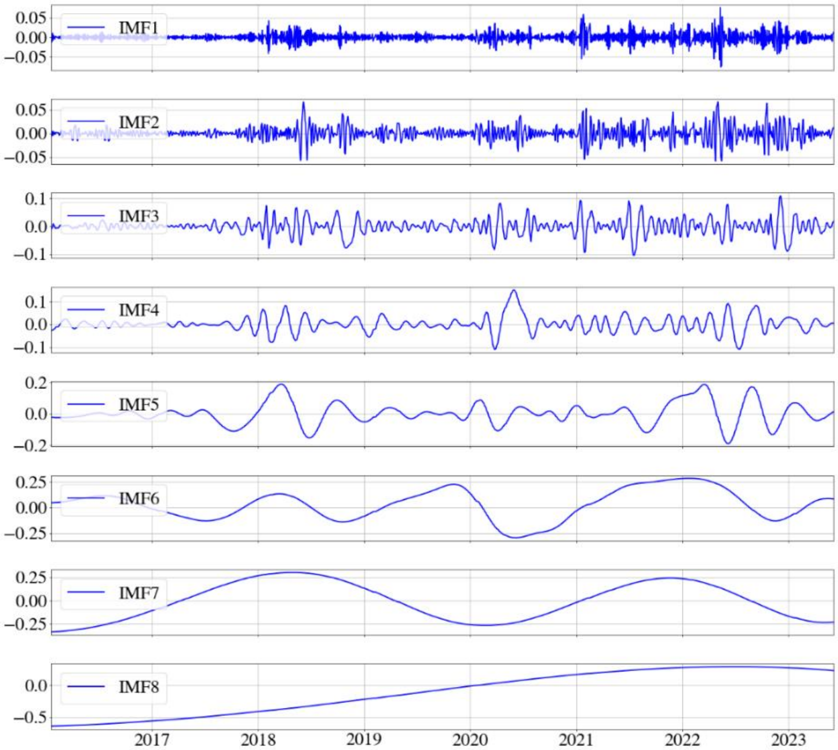

Figure 5. Eight IMFs by applying CEEMDAN analysis

Proposed models numbered from 0 to 4 are set up in Table 4, and they will be compared with the first model TCN+GRU, this model is used to forecast the entire time series without decomposing it into components. All models are reconstructed from two groups of different number of IMF components. They all use VN-Index time series to train and to forecast.

Table 4. Proposed models reconstructed from two groups of different number of IMF components

Model	IMFs	ΣIMFs
TCN+GRU	-	-
Model 0	1, 2, 3, 4, 5, 6, 7, 8	-
Model 1	1	$\sum_2^8 \mathrm{IMF}_i$
Model 2	1, 2	$\sum_3^8 \mathrm{IMF}_i$
Model 3	1, 2, 3	$\sum_4^8 \mathrm{IMF}_i$
Model 4	1, 2, 3, 4	$\sum_5^8 \mathrm{IMF}_i$

As shown in Table 5, model 1 beats other models by performing the best result of prediction the original VN-Index time series. All three error measures RMSE, MAE, and MPAE values of model 1 are 8.131, 5.538 and 0.589 respectively, which is the smallest among that of other models. Model 2 and model 3 also have good performance compared to others except model 1. The results are compared with the SVM model. The accuracy of three models 1, 2, 3 ranges from 71.65% to 75.59% which show the good results in predicting and indicate that the predicting value is closer to the original one.

Table 5. The prediction error and the accuracy of models

	Model	RMSE	MAE	MPAE	ACC
1	TCN+GRU	24.965	19.010	1.607	51.91%
2	Model 0	10.426	7.573	0.636	71.87%
3	Model 1	14.619	10.838	0.909	71.12%
4	Model 2	16.970	12.649	1.057	57.75%
5	Model 3	20.366	15.022	1.270	49.77%
6	Model 4	22.117	15.890	1.333	50.63%
7	SVM	17.150	12.105	1.024	52.22%

Source: Calculated results from JupyterLab 3.0

To deeper implement the analysis of forecasting models, Table 6 and Table 7 present the confusion matrices of models. It shows that model 1, 2, 3, and 4 have the values of true-positive and true-negative rate better than other models.

Table 6. Confusion matrices of TCN+GRU, model 0, and model 1

TCN+GRU			Model 0			Model 1		
	Up	Down		Up	Down		Up	Down
Up	95	91	Up	134	52	Up	130	54
Down	83	92	Down	50	125	Down	50	125

Table 7. Confusion matrices of model 2, model 3, and model 4

Model 2			Model 3			Model 4		
	Up	Down		Up	Down		Up	Down
Up	104	81	Up	92	93	Up	93	92
Down	71	104	Down	88	87	Down	86	89

For better visualization, *Figure 6* in the following shows the actual VN-Index and its predicted values of model 1.

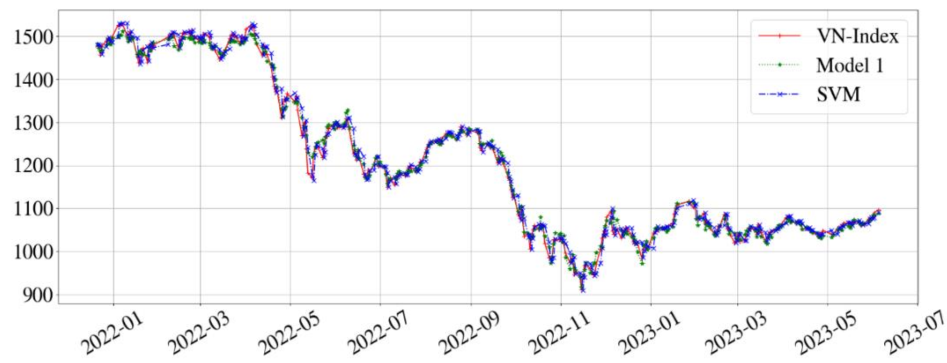

Figure 6. Actual and predicted values of all models

4. CONCLUSION

The class of CEEMDAN-TCN-GRU model developed in the paper has demonstrated its capability to predict the financial VN-Index data. The advantages when applying this proposed model are the following: (1) reduce the computational cost during the process of training due to the fact that our model needs only two time series (with PE > 0.5 and the other) to be trained instead of all time series of IMF components; (2) show a good performance in the forecasting results which has less error than other prediction models by using error analysis (MAE, RMSE, MAPE). It can be improved our prediction models discussed in the paper to be more accurate and robust. Our research in the future should be concerning to other financial time series, for example, high, low, open prices and trading volume as the input of model.

REFERENCES

1. Makridakis S., Wheelwright S. C. - Hyndman RJ. Forecasting methods and applications. USA: John Wiley and Sons (2008).
2. Torres J. F., Hadjout D., Sebaa A., Martínez-Álvarez F., Troncoso A. - Deep Learning for Time Series Forecasting: A Survey. Big Data **9** (1) (2021) 3-21. https://doi.org/10.1089/big.2020.0159
3. Zhang Q., Yang L. T., Chen Z., Li P. - A survey on deep learning for big data. Information Fusion 42 (2018) 146-157. https://doi.org/10.1016/j.inffus.2017.10.006
4. Minh H. B., An N. H., Tuan N. M. - Multi-step-ahead time series forecasting based on CEEMDAN decomposition and temporal convolutional networks. 2022 International Conference on Advanced Computing and Analytics (2022) 54-59. doi:10.1109/ACOMPA57018.2022.00015
5. An N. H., Minh H. B., Lam M. D, Phung P. D., Minh C. B. - Multi-Head Time Series Prediction by Decomposition and LSTM–Self-Attention–LSTM Model. 2022 International Conference on Advanced Computing and Analytics (2023) 19-25. doi:10.1109/ACOMPA61072.2023.00013
6. Anh T. T. T. - Dự báo chỉ số chứng khoán VN-Index bằng phương pháp phân rã trạng thái thực nghiệm CEEMDAN kết hợp với kỹ thuật LSTM. Tạp chí Kinh tế và Ngân hàng Châu Á **164** (1) (2019) 5-14. https://doi.org/10.63065/ajeb.vn.2019.164.44468
7. Cao J., Li Z., Li J. - Financial time series forecasting model based on CEEMDAN and LSTM. Physica A: Statistical mechanics and its applications **519** (2019) 127–139. https://doi.org/10.1016/j.physa.2018.11.061.
8. Chacon H. D., Kesici E., Najafirad P. - Improving financial time series prediction accuracy using ensemble empirical mode decomposition and recurrent neural networks. IEEE Access **8** (2020) 117133–117145. https://doi.org/10.1109/ACCESS.2020.2996981.
9. Zhang W., Qu Z., Zhang K., Mao W., Ma Y., Fan X. - A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy Conversion and Management **136** (2017) 439–451. https://doi.org/10.1016/j.enconman.2017.01.022.

10. Ruiz-Aguilar J. J., Turias I., González-Enrique J., Urda D., Elizondo D. - A permutation entropy-based EMD–ANN forecasting ensemble approach for wind speed prediction. Neural Computing and Applications **33** (2021) 2369-2391. https://doi.org/10.1007/s00521-020-05141-w.

11. Wang R., Peng C., Gao J., Gao Z., Jiang H. - A dilated convolution network-based LSTM model for multi-step prediction of chaotic time series. Computational and Applied Mathematics **39** (30) (2020). https://doi.org/10.1007/s40314-019-1006-2

12. Gan Z., Li C., Zhou J., Tang G. - Temporal convolutional networks interval prediction model for wind speed forecasting. Electric Power Systems Research **191** (2020), 106865. https://doi.org/10.1016/j.epsr.2020.106865

13. Shi H., Wang L. E. I., Scherer R., Woźniak M., Zhang P., Wei W. - Short-Term Load Forecasting Based on Adabelief Optimized Temporal Convolutional Network and Gated Recurrent Unit Hybrid Neural Network. IEEE Access **9** (2021) 66965–66981. https://doi.org/10.1109/ACCESS.2021.3076313

14. Huang N. E., Shen Z., Long S. R., Wu M. C., Snin H. H., Zheng Q., Yen N. C., Tung C. C., Liu H. H. - The empirical mode decomposition and the Hibert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **454** (1971) (1998) 903–995. https://doi.org/10.1098/rspa.1998.0193

15. Cho K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. - Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar (2014) 1724–1734. https://doi.org/10.3115/v1/D14-1179

16. Bergstra J., Bardenet R., Bengio Y., Kégl B. - Algorithms for hyper-parameter optimization. Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011 (2011) 1–9.

TÓM TẮT

DỰ BÁO CHUỖI THỜI GIAN VN-INDEX DỰA TRÊN PHÂN TÍCH CEEMDAN VÀ HỌC SÂU

Nguyễn Hoàng Ân, Hà Bình Minh, Huỳnh Đỗ Bảo Châu, Phan Đình Phùng*

Trường Đại học Ngân hàng Thành phố Hồ Chí Minh

*Email: *phungpd@hub.edu.vn*

Trong bài báo này tác giả đề xuất mô hình dự báo cho chuỗi thời gian VN-Index bằng việc sử dụng kết hợp ba mô hình học sâu gồm Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), mạng tích chập theo thời gian (TCN) và mạng nơ ron hồi tiếp có cổng (GRUNN). Ý tưởng chính trong mô hình đề xuất là phân chia các thành phần IMF thành hai nhóm, một nhóm có hệ số PE cao, một nhóm có hệ số PE thấp. Điều này giúp giảm đáng kể thời gian tìm các siêu tham số tối ưu cho các mô hình huấn luyện. Kết quả tính toán số cho thấy mô hình đề xuất này cho kết quả tốt hơn một số mô hình khác.

Từ khóa: Dự báo chuỗi thời gian VN-Index, Mạng tích chập theo thời gian, Mạng nơ ron hồi tiếp có cổng, CEEMDAN.