A Hybrid Model Using Deep Learning to Predict Stock Price Index

Mohammad Reza Shahraki[1]

Abstract–Predicting the stock price is a demanding task since multiple factors affect it. To enhance the stock price index prediction accuracy, the current study hybridizes variational mode decomposition (VMD) with the CNN-LSTM model. The proposed model, VMD-CNN-LSTM, works based on the decomposition-and-ensemble framework. To do this, VMD and CNN-LSTM were used to deal with the nonstationary and nonlinear nature of the stock price data. The former was first applied to the decomposition of time-series data into a number of components. Then, CNN-LSTM was applied to the prediction of the components. To end with, all the components' prediction results were summed up to attain the final prediction result. To verify the effectiveness of the proposed model in terms of predicting the stock price index, its performance was compared to some single models as well as some VMD- and EMD-based hybrid models. The results not only confirmed the superiority of the hybrid models over the single ones, but also showed the higher effectiveness of VMD-based models compared to EMD-based ones regarding the prediction accuracy.

Keywords: Stock price, Prediction, Deep learning

1. Introduction

Stock markets, in their nature, are known as non-linear, non-parametric, and noisy deterministic chaotic systems. All these qualities cause the effective and efficient prediction of the future stock price to be a highly challenging task. The stock price fluctuates in an uncertain atmosphere; a behavior that is affected by different interconnected factors. These factors include, but not limited to, massive globalized economic data, changes to the unemployment rate, monetary policies taken by powerful countries, natural calamities, immigration policies, and public health conditions. It is widely acknowledged that all of the stock market stakeholders are in a continuous attempt to augment their profits and, on the other hand, decrease the risks from the thorough market evaluation. The huge challenge in this domain is how to collect the relevant multifaceted information in one basket and then create a model capable of accurately predicting the stock price. The stock/equity market involves several stock exchanges that occur across the world. People in this market buy and sell shares of high volatility in their price, which is attributed to the law of supply and demand. A share or stock denotes the holder's ownership in the firm or corporation. Buyers attempt to buy a share at the minimum possible price, while sellers attempt to sell it at the maximum possible price. [1]

Stock market is widely acknowledged as a key money-raising platform along with debt markets that are more imposing but do not trade publicly. Highly liquidity is one of the stock exchange's qualities, which eases the selling or buying of securities. An important characteristic of any upcoming economy is the increase of the number of people involved in the stock market and also the upward trend of this market. The movements in stock market considerably affect both the people and the economy. If share prices are collapsed, the economy will become dysfunctional. The stock market crash of 1929, for instance, triggered the Great Depression of the 1930s. With high stock prices, more companies tend to issue Initial Public Offering (IPO) with the aim of raising capital through transfer ownership. In a bull market, mergers and acquisitions are also widespread. An outstanding economic growth can happen due to this increased investment. In the current literature, a cutting-edge model called Prescriptive Analytics has attracted much attention. This model, which could be used by both

1* Corresponding Author: Department of Accounting, Azadshahr Branch, Islamic Azad University, Azadshahr, Iran.
Email:shahrakimohammadreza80@gmail.com
Received: 2023.08.03; Accepted: 2023.09.19

investors and the general public, can suggest a profitable mode of action.[2]

A question is what would happen if investors knew the exact time of the increase or decrease of a certain stock price. It is imaginable that they would invest all their money in it and maximize their profit. This is impossible; however, they can make informed estimations on the basis of the past and present information accessible regarding certain shares.

By showing the utility of deep learning in the stock market, the current study attempts to show how its evolution is beneficial to society by reducing human efforts and providing optimum results.

2. Related work

Many researchers have already focused on proposing proper models applicable to the prediction of the stock market, and, more specifically, the literature consists of several studies reviewing the deep learning methods developed in the literature. As they have concentrated on the deep learning methods applications, the stock market prediction could be only one instance among numerous financial problems. This section lists them chronologically and discusses the motivation and unique perspectives of the current paper.

Most of forecasting methods proposed in the early studies conducted on time series were focused on conventional econometric models such as GARCH model and ARIMA model [3,4]. In [5], the authors used ARIMA to predict Nifty-Midcap. The findings confirmed the high flexibility of this model when applied to time series. In another study[6], the authors employed several GARCH models in order to forecast the volatility of the composite indices of Shenzhen and Shanghai. Among all, GARCH (1,1) was found more successful in describing the stock index volatility. Econometric models have shown some successes in terms of forecasting tasks; though, many empirical studies have reported that the linear assumption disables the models to have a comprehensive reflection on the real distribution of data [7]; this hinders the achievement of accurate forecast outcomes on datasets of higher complexity. Likewise, Liu and Morley [8] argued that applying only the conventional econometric models to the modeling process results in serious deviations in prediction.

Currently, artificial intelligence (AI) techniques, in addition to econometric models, are extensively applied to time series; two well-recognized instances are the artificial neural network (ANN) [9,10,11] and support vector machine model [12,13,14]. The conventional ANNs have been found more successful when applied to nonlinear data [15]. On the other hand, they simply fall into overfitting and local extremum, which limit their application [16]. Moreover, stock price is auto-correlative; therefore, the historical data can have impact on the current stock price. Therefore, although changes occurring to the stock price do not meet the assumption of random walk, they could have proper impacts on neural networks. If the prediction process is done with the use of only the current information, the pieces of information the early data carry may get lost and, consequently, dealing with complex time series problems will be difficult. The literature introduces the recurrent neural networks (RNNs) as a response to such difficulties. These networks are able to keep memory of recent events by making a full connection among the hidden neurons existing within the networks and learning from internal feedback mechanism. However, RNNs suffer from the gradient disappearance problem. The long short-term memory (LSTM) model was proposed in the literature based on RNN for the aim of capturing both long- and short-term information. LSTM is provided with the special gate structure that filters the information of higher usefulness and significance from the training data; this characteristic causes LSTM to be more applicable to the time series problem. This model was used by [17] to forecast the future trends of the stock price of companies; the outcomes confirmed the acceptable accuracy of LSTM in this application. In another study [18], the authors used an LSTM-based model to forecast two indices. Their obtained results showed the superiority of LSTM in this regard over the other models to which it was compared. Though, the valid information could not attain more weights to show its significance in LSTM, which requires further improvement.

The integration of AI techniques into data preprocessing methods can lead to a considerable enhancement in predictions accuracy since if AI is used in this regard, consistent and stable results cannot be obtained. The predictive model complexity could be lessened using signal analysis methods. In these methods, the available data are initially decomposed and then the sub-sequences are modelled. The literature contains frequent use of the EMD applied to decompose the signals. EMD is applicable to transforming non-stationary signals into stationary ones; additionally, it has a desired decomposition impact on nonlinear data. If the original signal is decomposed into multiple intrinsic mode functions (IMFs) and trend items, the information hidden in the data will be mined properly. In the study of Wang [19], EMD and stochastic time neural network were efficaciously used to forecast financial time series. Though, an inevitable problem has still remained with the use of EMD-modal aliasing. Put simply, different IMFs consist of similar features. To solve this problem, VMD was developed in the literature as a non-recursive signal analysis method, on the basis of EMD. In [20], the authors used the effective VMD to decompose the stock price series in order to capture the key information. Their results confirmed the effectiveness of VMD in feature selection applications. The present study also used VMD as the decomposition algorithm since it is capable of integrating the decomposed signals into the real signals and also separating identical frequencies.

Accordingly, the current study proposes a decomposition ensemble model applicable to predicting the stock market price, combining the benefits of VMD and CNN-LSTM model.

Briefly, the literature contains many studies carried out on how to forecast the stock market as accurately as possible. In this realm, several researchers have preferred to be focused on complex statistical or machine learning techniques regardless of the type of attributable variables. Some other researchers have concentrated on the use of fundamental data regardless of other variables affecting the stock market prediction. Though, the literature still lacks a model integrating simplicity with the features of the stock market variables. Thus, the aim of this study is to develop a model without further complicating the model architecture and, at the same time, preserving a well-balanced set of variables to capture the stock market behavior from several dimensions.

3. Methodology

The present paper contributes to the body of knowledge by hybridizing two models into one in order to predict the stock prices with the use of advanced deep learning techniques. The proposed algorithm as shown in Fig.1 consisted of two steps, where VMD was employed to decompose the data into several components (i.e. IMFs), and CNN_LSTM to predict each component. Ultimately, the final result was computed by summing up the forecasting values of all the sub-components.

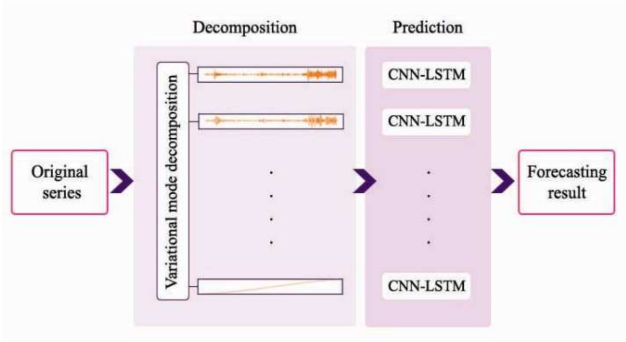

Fig. 1. Proposed hybrid model

3.1Variational Mode Decomposition Algorithm (VMD)

In methods working based on EMD, accuracy depends on at least three factors: the method used to explore the local extrema, the interpolation of local extrema to form carrier envelopes, and the stopping criteria set for IMFs. Thus, EMD-based methods are not well theorized from a mathematical point of view, and they operate based on recursive sifting, which causes them not to apply the backward error correction. The literature has proposed VMD in an attempt to address properly the above-mentioned limitations. VMD works by determining the relevant bands adaptively and, simultaneously, extracting the corresponding modes. Thus, error correction is feasible between them. According to Yang et al. [21], compared to EMD, the VMD method is able to achieve decomposition results of higher desirability.

In VMD, a real-valued time series $x(t)$ is decomposed into predefined sub-signals/modes $u_k(t)$ with sparsity properties, where each mode has a limited spectral bandwidth $B(u_k(t))$ compacted around a central frequency w_k .This bandwidth is determined along with the input signal decomposition. Then, the Hilbert transform of the associated analytic signal is computed for each mode in order to determine the bandwidth of each mode. This is done essentially to achieve unilateral frequency spectrum. Afterwards, the frequency spectrum for each mode u_k is shifted to baseband by mixing with the exponential tuned to the respective estimated center frequency wk. The bandwidth $B(u_k(t))$ is finally calculated with taking into account the H1 Guassian smoothness of the demodulated signal (For more details see[21]).

The number of modes K in VMD decomposition must be defined beforehand. In such a condition, the optimal K-value has not been achieved yet and still needs further research to be determined. The proper K-value may be dependent upon the number of IMFS that exist within the to-be-decomposed signal. In this regard, if the K-value is too large, redundant information may be created, while if it is too small, the mode-mixing problem may come to exist.

Fig.2displays the obtained results indicating that the original data were decomposed into six components.

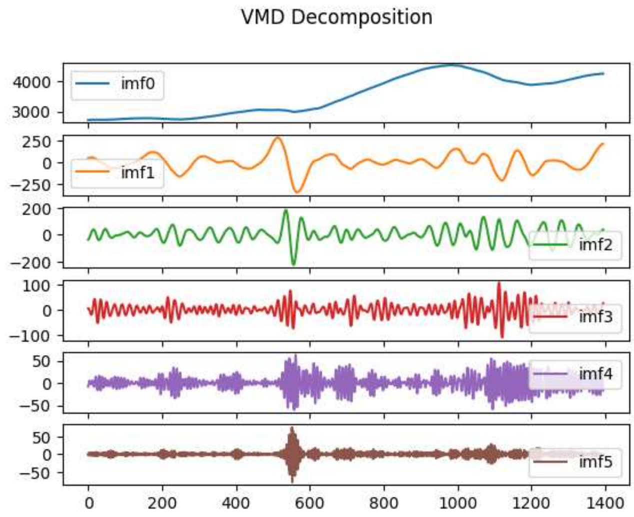

Fig. 2. Decomposition result using VMD

3.2 Prediction model

Convolutional layers can extract valuable knowledge and learn the internal representation of time-series data. LSTM, on the other hand, is able to determine both short- and long-term dependencies. The model developed in this study, CNN–LSTM, works based on the efficient combination of the benefits of these deep leaning techniques. CNN–LSTM comprises two k components: the first one contains convolutional and pooling layers where complicated mathematical operations are executed for the development of the input data features, whereas the second component uses the features produced by LSTM and dense layers. In the following, the convolutional, pooling, and LSTM layers (which form the core of CNN–LSTM) are described briefly.

-Convolutional and pooling layers

The convolutional layers are responsible for applying the convolution operation between the raw input data and convolution kernels to produce fresh feature values. It is important for the input data to have structured matrix form since this technique was originally developed for the purpose of extracting features from image datasets [23]. When compared to the input matrix, the convolution kernel (filter) is regarded as a tiny window presenting coefficient values in the form of a matrix. The window 'slides' throughout the input matrix and applies convolution operation on each sub-region (patch) that is 'met' across the input matrix. This process results in a convolved matrix that represents a feature value specified by the coefficient values and the dimension size of the applied filter. A number of convolved features can be generated by applying different convolution kernels to the input data. In general, these features are more useful compared to the original initial features of the input data. This enhances the model's performance quality.

-LSTM layers

LSTM neural networks [24], which are actually a sub-group of the recurrent neural networks (RNNs), can learn long-term dependencies by using feedback connections.

Conventional RNNs are focused on solving the feedforward neural networks problem, i.e., the lack of memory; this problem causes usually causes poor performance on sequences and time-series problems. Cyclic connections are used on the hidden layer of these models to attain short term memory and capture information from time-series and sequences data. On the other hand, RNNs generally have the problem of vanishing gradient. This problem confines the model to learning long-range dependencies. As a result, LSTM was proposed as a solution to this problem; it stores valuable information on memory cells while eliminating the needless one. This way, it can generally outperform the conventional RNN.

In each unit of LSTM, there are one memory cell and three main gates, i.e., input, output, and forget. This way, LSTM creates an information flow under a full control through determining which information must be subjected to 'forget' and which one to 'remember'. It results in effectively managing to learn long-term dependencies. More specifically, the input gate i_t along with a second gate c_t^*, controls the new information which is stored into the memory state c_t at time t. The forget gate, f_t, is responsible for controlling the past information that must be eliminated or preserved in the memory cell at time $t-1$, and, on the other hand, the output gate, o_t, is responsible for controlling the pieces of information that could be employed as the memory cell's output. Equations (1)–(5) express the operations done by a unit of LSTM.

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i), \quad (1)$$
$$f_t = \sigma(U_g x_t + W_g h_{t-1} + b_g), \quad (2)$$
$$c_t^* = \tanh(U_c x_t + W_c h_{t-1} + b_c), \quad (3)$$
$$c_t = g_t \circ c_{t-1} + i_t \circ c_t^*, \quad (4)$$
$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o), \quad (5)$$

where x_t stands for the input, W and U denote weight matrices, b_* signifies the vectors of bias term, σ represents the sigmoid function, and the operator ○ is the component-wise multiplication. To end with, Eq (6) is used to compute the hidden state, h_t, constituting the memory cell output, as follows:

$$h_t = o_t \circ \tanh(c_t). \quad (6)$$

When multiple LSTM layers are stacked together, both the memory state c_t and the hidden state h_t of each LSTM layer are forwarded as inputs to the following LSTM layer.

4. Experimental results

In this section, we explain the experiments and also the relevant analyses performed to assess the quality of the proposed hybrid model in forecasting stock price. The experiments were executed by means of the Python language using the Google Colab platform. For the purpose of this paper, the required data were gathered from www.yahoo.com for 1397trading days (the daily closing prices of the S&P) ranging from Jan 02, 2018 to July21, 2023. The collected data were separated first into two sets: training (80% of data) and testing (20%). The data in the former were applied to determining the weights and threshold values of the prediction models, whereas those in the latter to assessing the developed models' performance. For the verification of the models' efficiency, the closing price of the S&P given in Fig.3was considered, assuming that the stock price of one day is continuous and related to the its prices during the last 30 days.

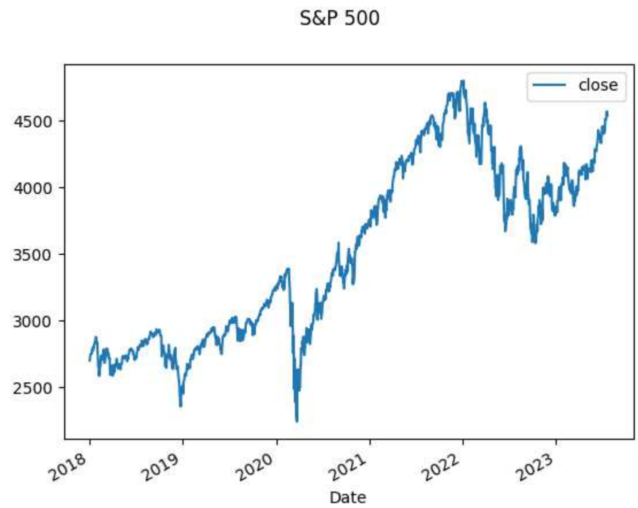

Fig. 3. Daily closing price of the S&P

Considering both the existing literature and the experiments carried out so far, the CNN-LSTM forecasting model (which has shown an acceptable performance) was selected in this paper. Based on the experiment results, the optimum prediction accuracy was attained with setting the time step to 30 days. Various hidden neurons for deep learning models were taken into account, with candidate values of 16, 32, 64, and 128. The results showed the models with 128 hidden neurons as the optimal condition. As a result, a one-layer deep learning models containing 128 hidden neurons was used in this study. In addition, Adam was used as optimizer due to its high computational efficiency and effectiveness. To establish a rapid computation ,the batch size and the epochs were fixed at 16 and 100, respectively. Moreover, the learning rate was fixed at 0.001. Each experiment was run for 25 times and the value obtained in average was noted as the final result.

4.1Experimental Performance Evaluation

The current study selects four performance measures, i.e., the root mean absolute error (RMAE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE). These measures were taken into account to assess the proposed models' predictive capacity. RMAE measures the bias between the predictive and practical models. RMSE reveals relatively large prediction errors. MAPE allows us to make a judge about the performance of predictions in statistics. Moreover, the loss of training set of each epoch was set in order to reflect on the proposed VMD-CNN-LSTM's performance under the impact of different parameters. The loss is assessed using the Mean square error (MSE).

$$RMAE = \sum_{t=1}^{N} \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (x_t - \hat{x}_t)^2} \quad (8)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{x_t - \hat{x}_t}{x_t} \right| \quad (9)$$

$$MSE = \frac{1}{N} \sum_{t=1}^{N} (\hat{x}_t - x_t)^2 \quad (10)$$

Table 1. The forecasting results using different methods

Methods	R2	RMSE	MAE	MAPE
ANN	0.90	62.19	50.37	1.26
LR	0.94	49.23	37.77	0.94
SVR	0.83	84.08	63.54	1.60
RF	0.88	70.91	56.71	1.42
LSTM	0.93	54.17	41.87	1.05
GRU	0.93	51.18	39.10	0.98
EMD-RF	0.92	55.75	42.51	1.06
EMD-LSTM	0.95	41.32	33.11	0.83
EMD-GRU	0.96	38.53	29.47	0.74
VMD-RF	0.96	37.52	28.83	0.71
VMD-LSTM	0.96	37.66	28.35	0.71
VMD-GRU	0.95	41.89	33.84	0.82
VMD-CNN-LSTM	0.96	37.02	28.04	0.71

The proposed hybrid model was comprehensively compared to several single models, i.e., the ANN, LR, SVR, RF, LSTM, and GRU as well as several hybrid models, i.e., the EMD-RF, EMD-LSTM, EMD-GRU, VMD-RF, VMD-LSTM and VMD-GRU. The finding results of the comparative experiments revealed that, by integrating VMD and EMD into single models, the distance between the predicted and the actual values dropped considerably, implying the higher effectiveness of the hybrid models over the single ones. In addition, the ANN models hybridized with VMD were found more accurate than those hybridized with EMD (for instance, VMD-LSTM > EMD-LSTM > LSTM). This confirmed the superiority of VMD over EMD in terms of predicting financial time series. Moreover, VMD-CNN-LSTM demonstrated a higher prediction capability compared to VMD-LSTM and EMD-LSTM in case of the stock data series.

5. Conclusion

The present paper developed a hybrid model, called VMD-CNN-LSTM, to enhance the accuracy of stock price indices prediction. The main challenge in these indices is that they are nonlinear, non-stationary, and uncertain, which is known as a complex problem. In the hybridized model, VMD is responsible for decomposing the original signal into a confined number of components, and CNN-LSTM for constructing a training and predicting model for each component. The concluding prediction result is achieved through aggregating all the components forecasted. VMD, in comparison with EMD, is of higher capacities such as data series decomposition with high accuracy, fast convergence, and strong noise robustness.

References

[1] Chhajer, P., Shah, M. and Kshirsagar, A., 2022. The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. *Decision Analytics Journal*, *2*, p.100015.

[2] Gandhmal, D.P. and Kumar, K., 2019. Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, *34*, p.100190.

[3] Jakubowski, A. and Szewczak, Z.S., 2021. Truncated moments of perpetuities and a new central limit theorem for GARCH processes without Kesten's regularity. *Stochastic Processes and their Applications*, *131*, pp.151-171.

[4] Ding, Y., 2019, February. Empirical Analysis of Logarithmic Return Rate of China's Financial Stocks—based on the ARMA-GARCH Model. In *2018 International Symposium on Social Science and Management Innovation (SSMI 2018)* (pp. 286-290). Atlantis Press.

[5] Devi, B.U., Sundar, D. and Alli, P., 2013. An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process*, *3*(1), p.65.

[6] Gu, F.J. and Cen, Z.D., 2011. Study on the volatility of Chinese Shanghai and Shenzhen stock markets with GARCH and SV Models. *J Math Pract Theory*, *1*, pp.4-22.

[7] Xu, X.S. and Chen, Y.B., 2001. Empirical study on nonlinearity in China stock market. *Quantitative and Technical Economics*, *18*, pp.110-113.

[8] Liu, W. and Morley, B., 2009. Volatility forecasting in the hang seng index using the GARCH approach. *Asia-Pacific Financial Markets*, *16*, pp.51-63.

[9] Yuan, S.L., Lu, H.M., Wang, P.P., Tian, C.G. and Gao, Z.J., 2021. Boron removal from metallurgical grade silicon by slag refining based on GA-BP neural network. *Rare Metals*, *40*, pp.237-242.

[10] Aras, S. and Kocakoç, İ.D., 2016. A new model selection strategy in time series forecasting with artificial neural networks: IHTS. *Neurocomputing*, *174*, pp.974-987.

[11] Raut, R.D., Priyadarshinee, P., Gardas, B.B. and Jha, M.K., 2018. Analyzing the factors influencing cloud computing adoption using three stage hybrid SEM-ANN-ISM (SEANIS) approach. *Technological Forecasting and Social Change*, *134*, pp.98-123.

[12] Richhariya, B. and Tanveer, M., 2018. EEG signal classification using universum support vector machine. *Expert Systems with Applications*, *106*, pp.169-182.

[13] Atmaja, B.T. and Akagi, M., 2021. Two-stage dimensional emotion recognition by fusing predictions of acoustic and text networks using SVM. *Speech Communication*, *126*, pp.9-21.

[14] Fan, L., Pan, S., Li, Z. and Li, H., 2016. An ICA-based support vector regression scheme for forecasting crude oil prices. *Technological Forecasting and Social Change*, *112*, pp.245-253.

[15] Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, pp.159-175.

[16] Chen, Y. and Hao, Y., 2018. Integrating principle component analysis and weighted support vector machine for stock trading signals prediction. *Neurocomputing*, *321*, pp.381-402.

[17] Ghosh, A., Bose, S., Maji, G., Debnath, N. and Sen, S., 2019, September. Stock price prediction using LSTM on Indian Share Market. In *Proceedings of 32nd international conference on* (Vol. 63, pp. 101-110).

[18] Baek, Y. and Kim, H.Y., 2018. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Systems with Applications*, *113*, pp.457-480.

[19] Wang, J. and Wang, J., 2017. Forecasting stochastic neural network based on financial empirical mode decomposition. *Neural Networks*, *90*, pp.8-20.

[20] Bisoi, R., Dash, P.K. and Parida, A.K., 2019. Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis. *Applied Soft Computing*, *74*, pp.652-678.

[21] Yang, W., Peng, Z., Wei, K., Shi, P. and Tian, W., 2017. Superiorities of variational mode decomposition over empirical mode decomposition particularly in time–frequency feature extraction and wind turbine condition monitoring. *IET Renewable Power Generation*, *11*(4), pp.443-452.

[22] Rawat, W. and Wang, Z., 2017. Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), pp.2352-2449.

[23] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[24] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, *9*(8), pp.1735-1780.