ICIC Express Letters ICIC International ©2020 ISSN 1881-803X
Volume **14**, Number **10**, October **2020** pp. **1017–1024**

DEVELOPING THE HYBRID FORECASTING MODEL ON THE SHORT-TERM TIME SERIES

Van-Chung Pham[1] and Ngoc-Canh Pham[2]

[1]Faculty of Information Technology
Industrial University of Ho Chi Minh City
No. 12, Nguyen Van Bao, Ward 4, Go Vap District, Ho Chi Minh City, Vietnam
pchung@iuh.edu.vn

[2]Faculty of Information Technology
Ho Chi Minh City Open University
No. 97, Vo Van Tan, Ward 6, District 3, Ho Chi Minh City, Vietnam
canhpn.178i@ou.edu.vn

Received October 2019; accepted January 2020

Abstract. *The problem of time series data forecast has many practical benefits, thus getting more and more attention from many people. Doing forecast on short-term time series is a practical problem today. There are two commonly used methods: the exponential smoothing and artificial neuron network (ANN). However, many studies show that for trend and seasonal data, the quality of the ANN is not high. Similarly, for non-linear data, the quality of the exponential smoothing is not good. In this paper, we proposed a new model to forecast short-term time series data by developing a hybrid model between the ANN and exponential smoothing. Empirically, this hybrid model can forecast on different data types and give good forecast results.*
Keywords: Time series, Exponential smoothing, Artificial neuron network, Forecast

1. **Introduction.** In the real world, time series has many different types, and by observing their values, they have the main components such as:

- Long-term trend component: their values measured over the long-term are increasing or decreasing. The graph shows whether it can be a smooth curve or a straight line.
- Seasonal component: their values increase or decrease at a specific period, such as a month, a quarter or a year. For example, the amount of electricity consumed usually increases during the summer and often decreases in the spring.
- Cycle component: their values change in a certain cycle (cycle can be weekly or monthly or yearly).
- Abnormal component: abnormal variable values are not cyclical. Therefore, past data are incapable of predicting these abnormal values.
- In terms of characteristics, they also can be linear time series (data with certain rules variation) or non-linear (non-rules variation) or can contain both.

Currently, there have been many works to solve the prediction problem on time series data such as [1-9]. However, time series data have many types, as mentioned above. Most forecasting methods have forecasted on a specific data type but not on different data types. According to Hyndman and Athanasopoulos [10], the combination of forecasting methods will lead to increased efficiency. Therefore, to increase the efficiency of forecasts on multi-component data as well as many characteristics, a hybrid model should be developed between existing processing models [11-14]. This paper develops a new model based on existing models. This new model is a hybrid between the artificial neuron network (ANN)

DOI: 10.24507/icicel.14.10.1017

and the triple exponential smoothing (ES3), which is called the ANN-ES3 hybrid model. The ANN-ES3 model is capable of forecasting short-term data of many components and characteristics. To verify the quality and execution time, we experimented on ten different multi-component datasets on three existing models: ANN, ES3, and ANN-ES1 (single exponential smoothing) to compare with the ANN-ES3 model of us. Experimental results show that this new model gives better forecasting on time series data with many components.

The rest of the paper is organized as follows. Section 2 presents much background knowledge and related works for forecast data on time series. Section 3 introduces the ANN-ES3 model that we propose. Section 4 reports on the experiments of ANN-ES3 model in the comparison model ANN [2,9], ES3 [1,3] and ANN-ES1 [15]. Finally, Section 5 gives some conclusions and future work.

2. **The Literature Review.** In this section, we introduce much background knowledge as well as some related works in recent years to forecast time series data.

2.1. **Exponential smoothing (ES).** This model was proposed in the late 1950s (Brown, 1959; Holt, 1958; Winters, 1960) [16], and had motivated some of the most successful forecasting methods. In this model, older observations are weighed down exponentially. That is, the more new observations, the greater the assigned weight values compared to the older observations. ES can forecast seasonal and trend time series and it is simple, easy to implement, and low-cost. However, ES is only a class of linear model and thus it cannot capture the nonlinear feature of time series [15]. Overview of this model is as follows.

2.1.1. *Simple exponential smoothing method (ES1).* This method is suitable for forecasting data with no clear trend or seasonal component. The past values of time series are smoothed out (average) by exponentially reducing. The forecast value A_t (value after smoothing) at time t is calculated as follows:

$$A_t = \alpha Y_{t-1} + (1-\alpha)A_{t-1} \text{ or } A_t = A_{t-1} + \alpha(Y_{t-1} - A_{t-1}),$$

where Y_{t-1}, the actual value at time $t-1$ and α, smoothing parameter. The smoothing parameter ($0 \leq \alpha \leq 1$) indicates the percentage of error of the prediction, α is usually selected within the range $[0.1, 0.5]$, choosing a small value will be forecasted to be stable, vice versa when fast forecasting is needed.

2.1.2. *Triple exponential smoothing methods (ES3).* It is suitable for forecasting when data has trends and seasons [7]. There are two models that can be used in ES3: the additive model and the multiplicative model.

- ***Additive model.*** The seasonal component is expressed as a mass, which can be added or subtracted from the average of the time series to unify the seasonality as follows [1,10]:

$$\begin{aligned}
Y_{t+p} &= (L_t + pT_t) + S_{t+p-s}, \\
S_t &= \gamma(Y_t - L_t) + (1-\gamma)S_{t-s}, \\
T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}, \\
L_t &= \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1}).
\end{aligned}$$

- ***Multiplicative model.*** It also has four equations [1]:

$$\begin{aligned}
Y_{t+p} &= (L_t + pT_n)S_{t+p-s}, \\
S_t &= \gamma Y_t/L_t + (1-\gamma)S_{t-s}, \\
L_t &= \alpha Y_t/S_{t-s} + (1-\alpha)(L_{t-1} + T_{t-1}), \\
T_t &= \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1},
\end{aligned}$$

where Y_{t+p}, the value of forecast at time t gives p future cycles; T_t, estimating trend at time t; S_t, estimating seasonality at time t; L_t, the level at the time t; s, the length of the season (the number of seasons in a year) and α (the smoothing constant for level), β (the smoothing constant for seasonality) and γ (smoothing constant for the trend) are in the range $[0, 1]$ which play the role of smoothing control.

2.2. **Artificial neuron network.** Some recent works [6,7,15] have shown that the artificial neuron network (ANN) is a model that can approximate various nonlinearity very well in time series and cost low. This paper uses the ANN model [17] with two algorithms: back propagation [18] and RPROP [19], and then chooses the best forecasting result between these two algorithms.

2.2.1. *Combining back propagation algorithm with momentum.* This algorithm may not find global extremes, but only local extremes can be found. If the learning rate η is too large, the error function will oscillate between the extreme, preventing the error from falling to a certain value. Conversely, if the coefficient of learning is too small, many steps must be taken (each step costs a one-time calculation of output results and correction of network weights). In order to meet acceptable results, it leads to a longer convergence time. That is, the time required to reach the convergence threshold depends heavily on η. The Momentum concept [18] has overcome these limitations, $0 \leq \mu < 1$ is a constant called for weight adjustment process to reduce the possibility of falling into local minimum points, reducing training time. It is based on the value of the weights in previous iterations t to get the weight value during the second training $t + 1$ with the following formula.

$$\Delta w_{ij}(t) = -\eta \partial E / \partial w_{ij}(t) + \mu \Delta w_{ij}(t + 1).$$

2.2.2. *Resilient propagation (RPROP).* RPROP [19] performs a direct adaptation of the weight step based on local gradient information. Δw_{ij} determines the size of the weight-updating, this adaptive update value evolves during the learning process based on its local sight on the error function E, according to the following learning rule.

Let $A = \partial E(t-1)/\partial w_{ij} * \partial E(t)/\partial w_{ij}$, then

$$\Delta w_{ij}(t) = \begin{cases} \eta^{+} * \Delta w_{ij}(t-1), & \text{if } A > 0 \\ \eta^{-} * \Delta w_{ij}(t-1), & \text{if } A < 0 \\ \Delta w_{ij}(t-1), & \text{else} \end{cases} \tag{1}$$

where $0 < \eta^{-} < 1 < \eta^{+}$, fixed coefficients of the learning process, if the derivative is positive (increasing error), the weight is decreased by its update-value, if the derivative is negative then the update-value is added:

$$\Delta w_{ij}(t) = \begin{cases} -\Delta w_{ij}(t), & \text{if } \partial E/\partial w_{ij}(t) > 0 \\ +\Delta w_{ij}(t), & \text{if } \partial E/\partial w_{ij}(t) < 0 \\ 0, & \text{else} \end{cases} \tag{2}$$

$w_{ij} = w_{ij}(t) + \Delta w_{ij}(t)$.

3. **ANN-ES3 Hybrid Model.** This section presents the content of the ANN-ES3 model. As in Section 2.2, ANN is capable of capturing nonlinear components. The exponential smoothing technique is a class of linear models, which can capture linear components in the time series. ES3 can be forecasted on data with seasonality and trend [1], which allows for the forecast on short-term time series. In [5,15], it is shown that the hybrid model between ANN and ES results in better forecasting. In this study, we developed a hybrid predictive model between ES3 and ANN because it is simple, low cost and especially suitable for short-term time series. Moreover, it allows forecasts on data that are nonlinear, trending and seasonal. In ANN, we implemented both algorithms: back propagation connecting Momentum and RPROP algorithm; in ES3 we also did both: additive and

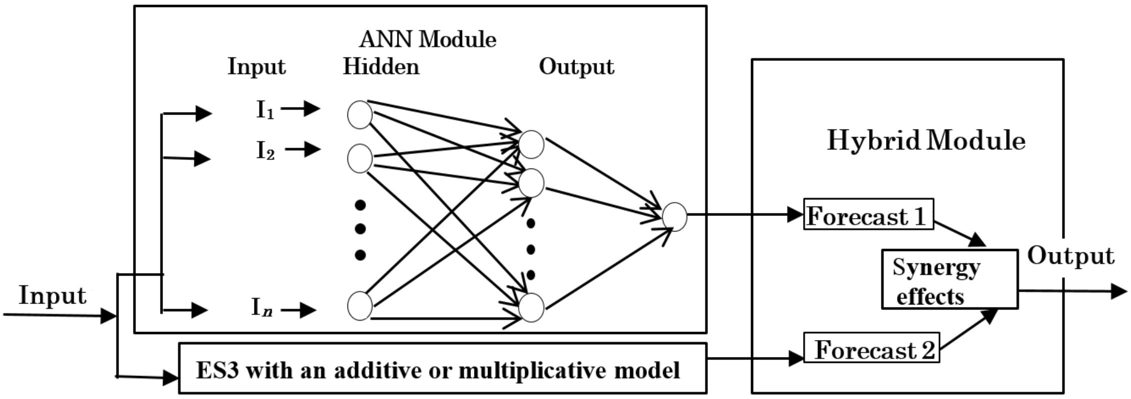

FIGURE 1. ANN-ES3 hybrid forecast model

multiplicative model (Section 2.1). Figure 1 shows the general diagram of the ANN-ES3 hybrid model.

3.1. **ANN module.** It provides a predictable result (Forecast 1). Its contents include:

- Users can select several units in the input layer and the hidden layer depending on the characteristics of the time series.
- The input value is the short-term time series to forecast.
- The sigmoid function is used to activate the function in the hidden layer and the output layer.
- Momentum and RPROP are network training algorithms. Users can choose one of these two algorithms.
- The error function on the weight vector w during network training is used:

$$E(\overrightarrow{w}) \equiv \frac{1}{2}\sum_{d\in D}(t_d - o_d)^2, \tag{3}$$

  where D, training data set; d, a sample in D; t_d, target value of d; o_d, the output value of the network.
- The conditional learning process is stopped by two parameters: epoch number (number of the training process of the network, selected by the user) and forecast error. MAPE, MAE, and MSE are used to calculate forecast errors as shown in Table 1. Maximum epoch value and minimum deviation λ user selected (we choose epoch = 1000, λ = 0.001). The learning process will stop when the difference value of $E(\overrightarrow{w})$ two consecutive learning epochs is less than or equal to λ or epoch has reached its maximum epoch value. If the network has just finished learning, the MAPE is quite

TABLE 1. Standard statistical errors measures

(n, the length of forecast series; y_t, the actual value and $\hat{y}_t$, the forecast value)

Name	Formulas	Usage
Mean Absolute Percentage Error	$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\lvert\hat{y}_t - y_t\rvert}{y_t}100$	*MAPE* shows the difference between the actual value and forecast value.
Mean Square Error	$MSE = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2$	*MSE* is often used to detect large forecast errors. For example, a method has several errors that are larger than abnormal in many small errors.
Mean Absolute Error	$MAE = \frac{1}{n}\sum_{t=1}^{n}\lvert\hat{y}_t - y_t\rvert$	*MAE* is often used to measure forecast errors in the same unit as the original data.

large, the user can increase the number of epochs and reduce the error and then continue studying until the desired MAPE is reached (to reduce the cost of learning, possibly based on the experience of experts).

3.2. **ES3 module.** It produces a different predictive result (Forecast 2 in Figure 1). This module performs the functions of ES3 (including the addition and multiplication models) and chooses the best forecasting result among them. Using R(D)COM software in R (available at: http://rcom.univie.ac.at/) estimates the smoothing parameters α, β and γ (as in Section 2.1). These parameters are returned by the Holtwinters function [18] and standard statistical errors (as shown in Table 1) calculate forecasting errors. This module allows for forecasting on seasonal and trending datasets.

3.3. **Hybrid module.** It combines Forecast 1 and Forecast 2 results to produce synergy effects which are output forecasting results based on the formula [15]

$$\hat{y}_t^{Hybrid} = \xi \hat{y}_t^{ANN} + (1-\xi)\hat{y}_t^{ES3}$$

where $\hat{y}_t^{Hybrid}$, forecast results of hybrid module; $\hat{y}_t^{ANN}$, forecast results of ANN module; $\hat{y}_t^{ES3}$, forecast results of ES3 module and $0 < \xi < 1$.

The hybrid weight ξ between ES3 and ANN needs to be determined, and it will be calculated based on MSE:

$$MSE = \sum \left(Y_t - \left[\xi \hat{y}_t^{ANN} + (1-\xi)\hat{y}_t^{ES3}\right]\right)^2$$

where Y_t, the string used to train; $\hat{y}_t^{ANN}$, the forecasting series of Y_t after passing ANN module and $\hat{y}_t^{ES3}$, the forecasting series of Y_t after passing ES3 module. The best value ξ is calculated using the formula [12] (if $\xi < 0$, choose $\xi = 0$, if $\xi > 1$, choose $\xi = 1$).

$$\xi = \sum \left(\hat{y}_t^{ANN} - \hat{y}_t^{ES3}\right)\left(Y_t - \hat{y}_t^{ES3}\right) \Big/ \sum \left(\hat{y}_t^{ANN} - \hat{y}_t^{ES3}\right)^2$$

This paper chooses the ANN-ES3 model to make forecasts with the following purposes.

a) The ES3 model is capable of forecasting on a short-term time series of trends and seasons, and it is possible to use R software to estimate the parameters for it. Moreover, users can choose the forecast result between addition and multiplication methods.

b) In the ANN model, users can choose the forecast result between the back propagation algorithm and RPROP (as in Section 2.2).

c) Hybridization of ANN and ES3 achieves forecast results of both a) and b).

4. **Experimental Evaluation.** We implemented four models including ANN, ES3, ANN-ES1, and ANN-ES3 with Microsoft Visual C # and conducted experiments on an Intel®Core i7, 16GB RAM, Quad-core 1.8GHz. We experimented on these models to compare forecast results between them on ten available datasets of different areas. These datasets have indexes: 1, 2, 3, 5, 6, 7, 8 and 9 (see Table 2) which are obtained from the webpage https://github.com/FinYang/tsdl (R. Hyndman, Monash University, Australia). The other two datasets are from the https://timeseries.weebly.com/dataset.html (Auckland, New Zealand University).

We use R software to estimate α parameter of ES1 and α, β and γ parameters of ES3. The program derives results from R through the R(D)COM 3.0.1.5 intermediary software available at www.rcom.univie.ac.at/. Use the Holtwinters function [10,20] in R to get the output, and analyze the output to get the parameters α, β, and γ (If they have a negative value then assign it a value of 0, if they are greater than 1 then assign them a value of 1). The Holtwinters function relies on minimizing the value of MSE (see Table 1) to derive parameters α, β, and γ. The forecast results obtained from the experiment on the ten datasets are shown in Table 2.

TABLE 2. Compare MAPE of four models on ten datasets

(a, forecast accuracy; e, execution time (milliseconds))

Index	**Dataset description**	**ES3**	**ANN** $\boldsymbol{a/e}$	**ANN-ES1** $\boldsymbol{a/e}$	**ANN-ES3** $\boldsymbol{a/e}$
1	Basic monthly iron production in Australia (linear, seasonal and trend, length: 479)	2.524%/9506	3.387%/555	11.167%/9968	**2.498%**/10063
2	Monthly M2 financial data in USA (linear, seasonal and trend, length: 398)	1.245%/9922	5.986%/1893	1.512%/11929	**1.222%**/11125
3	Monthly Chocolate consumption in Australia (linear, seasonal and trend, length: 458)	**6.195%**/15675	13.569%/15656	15.556%/31332	6.874%/31463
10	Number of monthly migrants in Australia (linear, seasonal and trend, length: 1094)	5.641%/9675	6.974%/11510	7.023%/21128	**4.133%**/21186
5	The monthly closing price of the Dow-Jones industrial index in USA (abnormal, length: 291)	2.599%/9597	**2.060%**/30	3.779%/9585	2.571%/9628
8	The average monthly water level of Lake Erie in USA (abnormal, length: 600)	3.854%/15749	**1.539%**/3647	3.204%/19486	1.543%/19023
4	Monthly temperatures in Paris, France (abnormal, seasonal and trend, length: 154)	20.514%/10362	17.265%/613	16.979%/9975	**14.385%**/9688
6	The number of women over the age of 20 is unemployed every month in USA (abnormal, seasonal and trend, length: 408)	5.220%/9729	11.451%/4615	7.165%/14155	**5.078%**/14046
7	Quarterly cement sales in Australia (abnormal, seasonal and trend, length: 155)	**7.490%**/9688	12.277%/57	7.716%/9426	7.613%/9745
9	Quarterly electricity output in Australia (abnormal, seasonal and trend, length: 155)	1.940%/9360	3.004%/31	**1.748%**/9556	1.801%/9698

In Table 2, we have the following comments.

- In the group of datasets with the seasonal and trend characteristics, ANN-ES3 has the highest forecasting results compared to the other three models.
- In the group of abnormal datasets: ANN-ES3 has a forecasting result that is close to ANN and higher than ES3 and ANN-ES1.
- In datasets 4 and 9 with seasonal, trending and abnormal characteristics, ANN-ES3 has higher forecasting results than the other three models in the dataset 4 and close to ANN-ES1 in dataset 9 (ANN-ES1 has the highest forecasting results in dataset 9).
- Through all ten empirical datasets, ANN-ES3 produced stable forecasting results when compared to the remaining models, it had the highest results on five datasets 1, 2, 10, 4 and 6. The remaining datasets, ANN-ES3, yield close results to the models with the highest forecasting results and deviations from its forecasting results compared to the models with the highest forecasting results ranging from 0.004 to 0.175%.

For each model, we also have commented:

- ES3 is very well forecasted for short-term seasonal data and trend as for dataset 3, or nearly equal to the ANN-ES3 in datasets 1, 2, 6, 7. However, the forecast is not as good for abnormal datasets as in datasets 5, 8 and datasets contain all three (seasonal, trend and abnormal).
- ANN only forecast well for abnormal short-term datasets in datasets 5 and 8, but less accurate for seasonal and trend datasets.
- ANN-ES1 is only strong for short-linear data with high linearity in dataset 9, this also confirms the correctness of Lai et al. [15].
- ANN-ES3 has a higher or near-higher predictive quality than ES3 in short-term data with seasonal and trend components, almost similar to ANN in abnormal data. Finally, in the data of season, trend and abnormal components, ANN-ES3 showed higher results than the other three models.

For these ten datasets alone, Figure 2 shows the average accuracy of the forecast across four models.

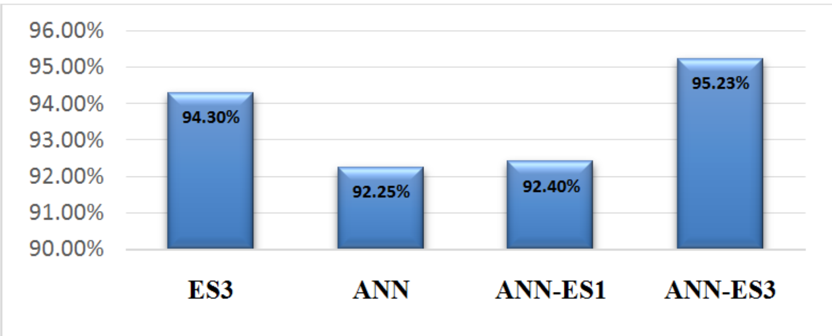

Figure 2. Comparison of the accuracy of the forecast

5. **Conclusions.** Forecast time series data is a matter of great interest in many years. We have developed a new ANN-ES3 hybrid model to do forecasting work on short time series. Experimental results on ten datasets belonging to different areas show that our proposed model has good efficiency in forecasting when compared with models ANN, ES3, and ANN-ES1.

The execution time of our model is better than the ANN-ES1 model and obviously, it is not as good as single model ES3 or ANN. This is reasonably acceptable. The ES3 model is a popular and effective approach to predicting seasonal data, but the results also depend on smoothing results when the data is too short and has a lot of noise.

As for future work, we plan to compare our model with ANN-SARIMA [4] using the classification input data of ANN and apply them to building a recommendation system on the school datasets.

Acknowledgment. This research is funded by Industrial University of Ho Chi Minh City in the scientific research project of 2020.

REFERENCES

[1] A. Rusyana, Nurhasanah, M. Oktaviana and Amiruddin, Forecasting passenger by using Holt's exponential smoothing and Winter's exponential smoothing, *The 23rd Seminar and Annual Meeting (Semirata) for Public Universities in Western Indonesia on Mathematics and Natural Sciences*, Universitas Sriwijaya, Palembang, 2016.

[2] E. Akarslan and F. Hocaoglu, Electricity demand forecasting of a micro grid using ANN, *The 9th International Renewable Energy Congress (IREC)*, 2018.

[3] C. Monica, S. Kalaivani and G. Jayalalitha, Forecasting the annual temperature of India by using exponential smoothing, *Journal of Advanced Research in Dynamical and Control Systems*, vol.10, no.6, 2018.

[4] P. Chen, A. Niu, D. Liu, W. Jiang and B. Ma, Time series forecasting of temperatures using SARIMA: An example from Nanjing, *IOP Conference Series Materials Science and Engineering*, vol.394, no.5, 2018.

[5] W. Sulandari, Subanar, Suhartono and H. Utami, Forecasting electricity load demand using hybrid exponential smoothing-artificial neural network model, *International Journal of Advances in Intelligent Informatics*, vol.2, no.3, pp.131-139, 2016.

[6] B. M. Al-Maqaleh, A. A. Al-Mansoub and F. N. Al-Badani, Forecasting using artificial neuron network and statistics models, *International Journal of Education and Management Engineering*, vol.6, no.3, pp.20-32, 2016.

[7] M. Khairalla, X. Ning and N. T. Al-Jallad, Hybrid forecasting scheme for financial time-series data using neural network and statistical methods, *International Journal of Advanced Computer Science and Applications*, vol.8, no.9, 2017.

[8] K. Zang, H.-X. Wang, H.-F. Wang and Z. Li, The time series prediction algorithm of enrollment based on the hesitancy degree correction rule of vague sets, *ICIC Express Letters*, vol.9, no.5, pp.1311-1318, 2015.

[9] A. H. Manurung, W. Budiharto and H. Prabowo, Algorithm and modelling of stock prices forecasting based on long short-term memory (LSTM), *ICIC Express Letters*, vol.12, no.12, pp.1277-1283, 2018.

[10] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd Edition, Monash University, Australia, 2018.

[11] W. J. Lee and J. Hong, A hybrid dynamic and fuzzy time series model for mid-term power load forecasting, *International Journal of Electrical Power & Energy Systems*, vol.64, pp.1057-1062, 2015.

[12] I. Khandelwal, R. Adhikari and G. Verma, Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition, *Procedia Computer Science*, vol.48, pp.173-179, 2015.

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

[14] S. Rathod, K. Naveena, S. Singh, S. Rathod and A. Singh, Hybrid ARIMA-ANN modelling for forecasting the price of robusta coffee in India, *International Journal of Current Microbiology and Applied Sciences*, vol.6, pp.1721-1726, 2017.

[15] K. K. Lai, L. Yu, S. Wang and W. Huang, Hybridizing exponential smoothing and neural network for financial time series predication, in *Computational Science – ICCS 2006. Lecture Notes in Computer Science*, V. N. Alexandrov, G. D. van Albada, P. M. A. Sloot and J. Dongarra (eds.), Berlin, Heidelberg, Springer, 2006.

[16] J. G. De Gooijer and R. J. Hyndman, 25 years of time series forecasting, *International Journal of Forecasting*, vol.22, pp.443-473, 2006.

[17] R. H. Nielscn, Theory of the backpropagation neuron network, *Proc. of International Joint Conference on Neuron Networks*, pp.593-611, 1987.

[18] T. M. Mitchell, *Machine Learning*, McGraw-Hill Science/Engineering/Math, 1997.

[19] M. Riedmiller and H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, *Proc. of the IEEE International Conference on Neural Networks*, pp.586-591, 1993.

[20] J. D. Cryer and K. S. Chan, *Time Series Analysis with Applications in R*, 2nd Edition, Springer, 2008.