International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2000-000X
Vol. 3 Issue 1, January – 2019, Pages: 1-13

Empirical Mode Decomposition Based Ensemble Random Forest Model for Financial Time Series Forecasting

Utpala Nanda Chowdhury*, Sanjoy Kumar Chakravarty, Md. Tanvir Hossain, and Shamim Ahmad
Department of Computer Science and Engineering
University of Rajshahi, Rajshahi, Bangladesh
*unchowdhury@gmail.com

Abstract—*Financial time series (TS) forecasting has gained profound research interest in the financial sector and various models have been proposed. The accuracy of such a forecasting model may offer investors an opportunity to maximize their profits. But, the inherent non-linearity and non-stationary characteristics of the stock market and financial TS has made this task immensely challenging. To address this problem, an ensemble method constituting Empirical Mode Decomposition (EMD) and Random Forest (RF) algorithm is proposed in this paper for future stock price prediction. At first, several intrinsic mode functions (IMFs) and one residual component were decomposed from the historical stock price TS. Then RF model was built to predict the future price for the target stock. 16 years' historical data of three prominent stocks from three different sectors listed in Dhaka Stock Exchange (DSE), Bangladesh are used to test the effectiveness of the proposed EMD-RF method. Empirical results demonstrated efficacy of the proposed method compared with six other forecasting methods.*

Keywords—Empirical mode decomposition; Random forest; Financial time series; Forecasting; Dhaka Stock Exchange

1. INTRODUCTION

Financial time series (TS) forecasting has become very essential for the financial sector and hence has gained extreme attention from both the researchers and financial practitioners. An accurate prediction can govern the investment strategy as well as the company policy to mitigate the risk and gain high investment return. But this prediction task is very challenging due to the inherent nonlinearity and non-stationary characteristics of the stock market and financial TS [1]. Moreover the stock market is influenced by various aspects like general economic conditions, government policies, company performance, investor's interest etc.

Numerous methods and proposals with different levels of success have been published in literature so far. Among them, autoregressive integrated moving average (ARIMA), the traditional statistical models have been very popular and are widely chosen because of their forecasting performance [2]. With the recent development of machine-learning algorithms plentiful approaches have been introduced, which includes Artificial Neural Networks (ANN), genetic algorithm, rough set (RS) theory, fuzzy logic and others [3] [4]. Most of these approaches suffer from different problems like over-fitting or under-fitting, initializing large number of control parameters, finding the optimum solutions etc. To resolve most of these shortcomings, support vector regression (SVR), the regression model of support vector machine (SVM) [5], has been widely used in various nonlinear regression tasks [6]. This is largely because; SVR uses the structural risk minimization principal for function estimation while the traditional methods implement empirical risk minimization principal [7]. Recent studies have found Random Forest (RF) to perform better for TS forecasting [8] RF is an ensemble learning method that generates many regression trees (CART) and aggregates their results.

Feature extraction (transforming the original features into new ones) and feature selection (choosing the most influential set of features) are very crucial in developing an efficient SVR forecasting model. Various feature extraction and selection methods have been integrated with basic forecasting models to ameliorate the performance which includes principal component analysis (PCA) [9] [10], kernel PCA [11], linear [12] and nonlinear [13] independent component analysis (ICA), and both PCA and ICA [14]. These approaches fail to address the inherent nonlinear and non-stationary problem properly and hence degrading forecasting performance. Over the last few decades, empirical mode decomposition (EMD) have been incorporated to overcome the limitations of financial TS forecasting [15]. The EMD method decomposes the signal into a finite set of nearly orthogonal oscillating components, called intrinsic mode functions (IMFs), which directly addresses the non-linearity issue. EMD is integrated with various base forecasting models like ARIMA [16], SVR [17], least squares SVR [18], υ-SVR [19] etc. for financial TS forecasting. But, EMD is not still ensemble with RF to investigate its effectiveness. Therefore, in this work, integration of EMD with RF is mainly focused.

In this paper, we integrated EMD and RF to form an ensemble model for financial TS forecasting. At first, EMD decomposes the historical stock price TS into several IMFs and one residual component. An RF model is built to predict the desired stock price using corresponding IMFs and residual component. The attractiveness of the proposed method is demonstrated on 16 years' historical data of three prominent stocks from three different sectors listed in Dhaka Stock Exchange (DSE), Bangladesh. Predictions are made for 1, 5, 10, 15, 20, 25 and 30 days in advance targeting the close price of stocks. The empirical results were compared with six benchmark learning algorithms: single RF, single SVR, EMD-SVR, PCA-SVR, ICA-SVR and PCA-ICA-SVR.

The reminder of this paper is organized into 6 sections. Section 2 provides a brief overview of the methodologies used in this study which includes EMD and RF. Section 3 introduces the proposed method followed by the experimental setup in section 4 which comprises research data and performance criteria. Section 5 reports the experimental results obtained from the study. Finally section 6 contains the concluding remarks.

2. METHODOLOGY

2.1 Empirical Mode Decomposition

EMD was proposed by Huang et al. [15] as a fundamental part of the Hilbert-Huang transformation (HHT). It is a signal transformation method that decomposes a non-stationary and nonlinear TS into several intrinsic mode functions (IMFs) and a residual component. According to Huang et al. each IMF must satisfy two conditions: firstly, the number of extreme values and zero-crossings either are equal or differ at the most by one; and secondly, the mean value of the envelope constructed by the local maxima and minima is zero at any point.

The shifting process which EMD uses to decompose the signal into IMFs is described as follows:

- For a time series signal $x(t)$, let m_1 be the mean of its upper and lower envelopes as determined by a cubic-spline interpolation of local maxima and minima.
- The first component h_1 is computed by subtracting the mean from the original time series: $h_1 = x(t) - m_1$.
- In the second shifting process, $h_1$is treated as the data, and $m_{11}$is the mean of h_1's upper and lower envelopes: $h_{11} = h_1 - m_{11}$.
- This shifting procedure is repeated k times until one of the following stop criterion is satisfied: i) m_{1k}approaches zero, ii) the numbers of zero-crossings and extrema of h_{1k} differs at most by one, or iii) the predefined maximum iteration is reached. h_{1k} can be treated as an IMF in this case and computed by: $h_{1k} = h_{1(k-1)} - m_{1k}$.
- Then it is designated as $c_1 = h_{1k}$, the first IMF component from the data, which contains the shortest period component of the signal. We separate it from the rest of the data: $x(t) - c_1 = r_1$. The procedure is repeated on r_j: $r_1 - c_2 = r_2, \ldots, r_{(n-1)} - c_n = r_n$.

As a result, the original time series signal is decomposed as a set of functions:

$$x(t) = \sum_{n}^{i=1} c_i + r_n \tag{1}$$

where the number of functions n in the set depends on the original signal.

2.2 Random Forest

Random forests (RF) is an ensemble of both classification and regression learning methods [20]. RF is generated by combining a set of unpruned decision trees (Breiman's CART - Classification And Regression Trees [21]) that are grown to maximum size by selecting random subspaces of the feature space. Each tree is constructed using different bootstrap sample taken from the training data and a subset of randomly chosen features (input variables or predictors) at each node. After the forest is formed, the final decision is obtained by aggregating over the ensemble, i.e. by averaging the output for regression or by majority voting for classification. This ensemble method called bagging improves the stability and accuracy of the model, reduces variance and helps to avoid overfitting [22].

The random forests algorithm (for both classification and regression) is as follows [23]:

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2000-000X
Vol. 3 Issue 1, January – 2019, Pages: 3-13

- From the training data of n samples draw n_{tree} bootstrap samples.
- For each of the bootstrap samples, grow classification or regression tree with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m_{try} of the predictors and choose the best split from among those variables. The tree is grown to the maximum size and pruning in skipped. Bagging can be thought of as the special case of random forests obtained when $m_{try} = p$ the number of predictors.
- Predict new data by aggregating the predictions of the n_{tree} trees (i.e., majority votes for classification, average for regression).

The RF generalization error is estimated by an out-of-bag (OOB) error, i.e. the error for training points which are not contained in the bootstrap training sets (about one-third of the points are left out in each bootstrap training set). An OOB error estimate is almost identical to that obtained by N-fold cross-validation. The large advantage of RFs is that they can be fitted in one sequence, with cross-validation being performed along the way. The training can be terminated when the OOB error stabilizes.

RF requires three parameters to be tuned: number of trees (n_{tree}) number of descriptors randomly sampled as candidates for splitting at each node (m_{try}) and minimum node size. When the forest is growing, random features are selected at random out of the all features in the training data. The number of features employed in splitting each node for each tree is the primary tuning parameter (m_{try}). To improve the performance of RF, this parameter should be optimized. The number of trees (n_{tree}) should only be chosen to be sufficiently large so that the OOB error has stabilized. In many cases, 500 trees are sufficient (more are needed if descriptor importance or molecular proximity is desired). There is no penalty for having *too many* trees, other than waste in computational resources, in contrast to other algorithms which require a stopping rule. Another parameter, minimum node size, determines the minimum size of nodes below which no split will be attempted. This parameter has some effect on the size of the trees grown. In RF, for classification, the default is 1, ensuring that trees are grown to their maximum size and for regression, the default is 5.

3. Proposed Ensemble Method

In this paper, EMD based ensemble RF (EMD-RF) model is proposed where the original financial TS is decomposed into a series of sub-datasets until they are simple enough to be analyzed. At first, EMD decomposes the stock price data into several IMFs and one residual component. Then a RF model is trained using the IMFs and residual component to generate the final prediction result. Fig. 1 shows the schematic diagram of this proposed ensemble method and the procedures can be described as:

1. The original TS is decomposed by EMD into several IMFs and one residual components.
2. These IMFs and residual component are used to train a RF model.
3. The trained RF model produces an ensemble prediction result for the TS forecasting.

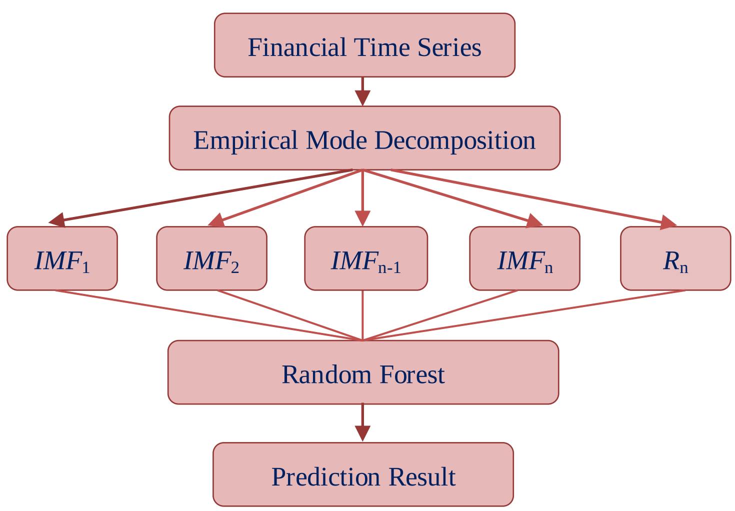

Fig. 1. *The proposed EMD-RF forecasting model for financial TS.*

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2000-000X
Vol. 3 Issue 1, January – 2019, Pages: 4-13

4. EXPERIMENT SETUP

4.1 Datasets

To conduct the study and evaluate the performance of the proposed approach, the 16 years' historical data of daily transaction for the time period from January 2000 to December 2015 are collected from Dhaka Stock Exchange, Bangladesh (https://www.dsebd.org/). This data covers 3600 trading days and each data comprises five attributes: open price, high price, low price, close price and trade volume. We have considered three companies from three different sectors: *Square Pharmaceuticals Limited*, *AB Bank Limited*, and *Bangladesh Lamps Limited* as these are the most prominent stocks in DSE. The first one is a leading company in pharmaceuticals sector, the second leads the banking sector and the last one belongs to the engineering sector. 70% of the total sample points (around 2520 trading days) are used as the training sample and the remaining 30% of the total sample points (around 1080 trading days) are holdout to be used as the testing sample.

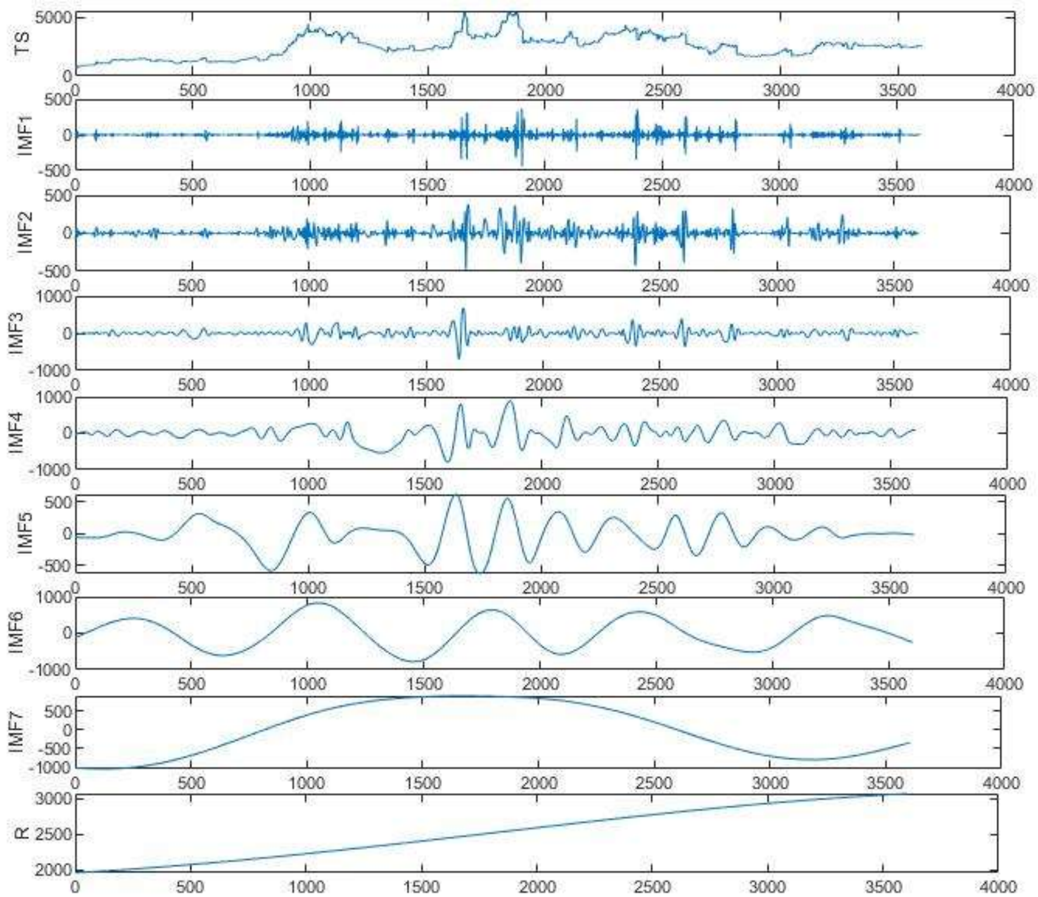

Fig. 2. *The decomposition result for Square Pharmaceuticals Ltd. from EMD.*

4.2 Performance Measure

In this paper, Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), relative Root Mean Squared Error (rRMSE) and Mean Squared Error (MSE) are used to evaluate the performance of the proposed model [24]. Formulas of these evaluation measures are shown in (2), (3), (4) and (5). These are the measures of deviation between actual and predicted prices. The smaller the values of these measures, the closer the predicted prices are to actual prices. They can be used to evaluate the predictive performance of any forecasting model.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|A_t - F_t|}{|A_t|} \times 100 \quad (2)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\frac{|A_t - F_t|}{|A_t|} \quad (3)$$

$$rRMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{A_t - F_t}{A_t}\right)^2} \times 100 \quad (4)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(A_t - F_t)^2 \quad (5)$$

where A_t is the actual value, F_t represent the predicted value and n is the total number of data points.

5. Results and Discussion

The modeling steps of the proposed EMD-RF are shown in Section 3. While EMD approach is applied the close price TS of three different stocks are decomposed into seven IMFs and one residual component. These decomposition results are shown in Fig. 2 to Fig. 4. This divide and conquer strategy may uplift the forecasting model's performance. The independent IMFs and residual components decomposed earlier are then used to train the RF forecasting model.

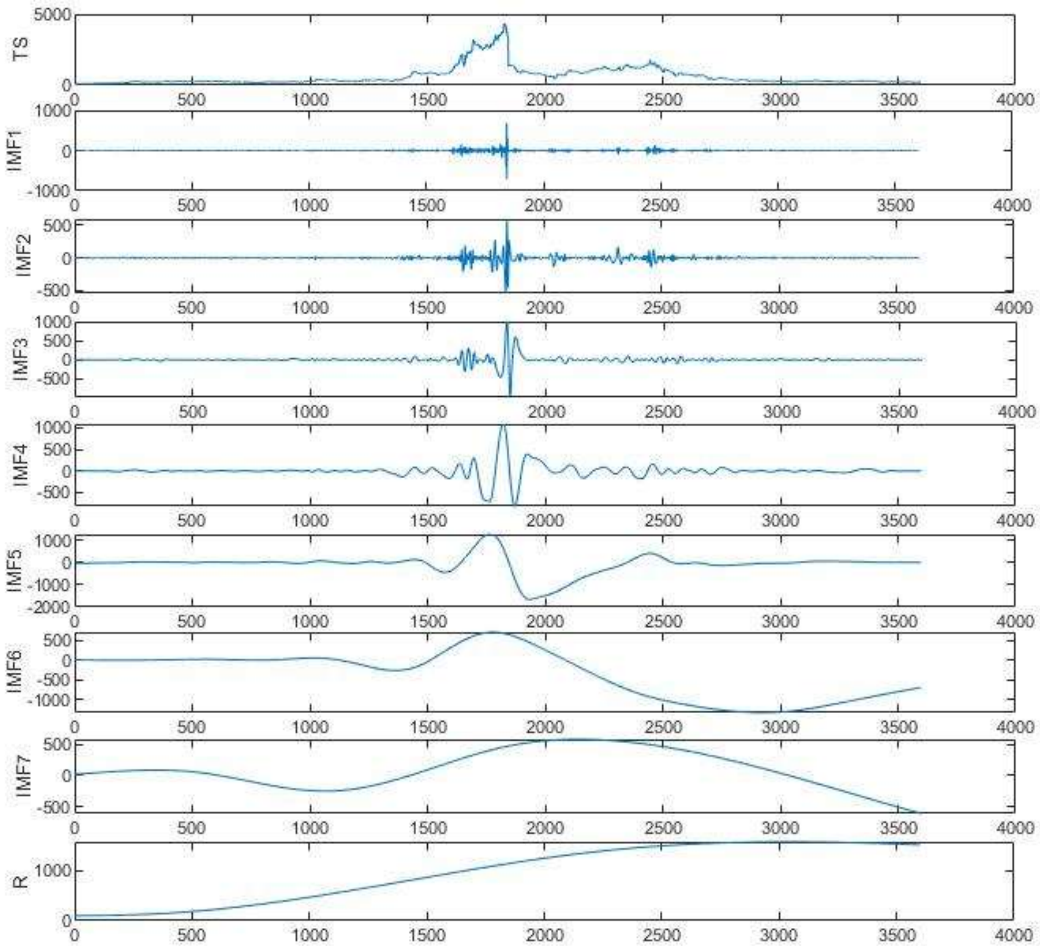

Fig. 3. *The decomposition result for AB Bank Ltd. from EMD.*

5.1 Parameter Tuning for RF

As mentioned in Section 2.2, parameter tuning is essential for RF model construction. For training an RF, the n_{tree}, $minLeaf$ and m_{try} need to be prescribed. Fig. 5 (a) shows the OOB mean squared error of RF as the number of trees increases for *Square Pharmaceuticals Ltd*. The figure shows that the OOB error drops to a steady rate when the once there are a sufficient number of trees (around 30). To determine the minimum node size, we have compared the OOB error rate with varying the number of node size using n_{tree}=50 and $m_{try} = 5$, as shown in Fig. 5 (b). The plot shows that default value 5 (five) for regression gives the lowest OOB error rate. Then, we further optimize the parameter m_{try} (which is much sensitive to the performance of RF compared with other parameters). More specifically, we gradually vary the value of m_{try} from 1 to 7 with a step size of 1 and the one which minimizes the OOB error is chosen. As a result, we find $m_{try} = 5$ corresponds to the minimal value of OOB error.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2000-000X
Vol. 3 Issue 1, January – 2019, Pages: 6-13

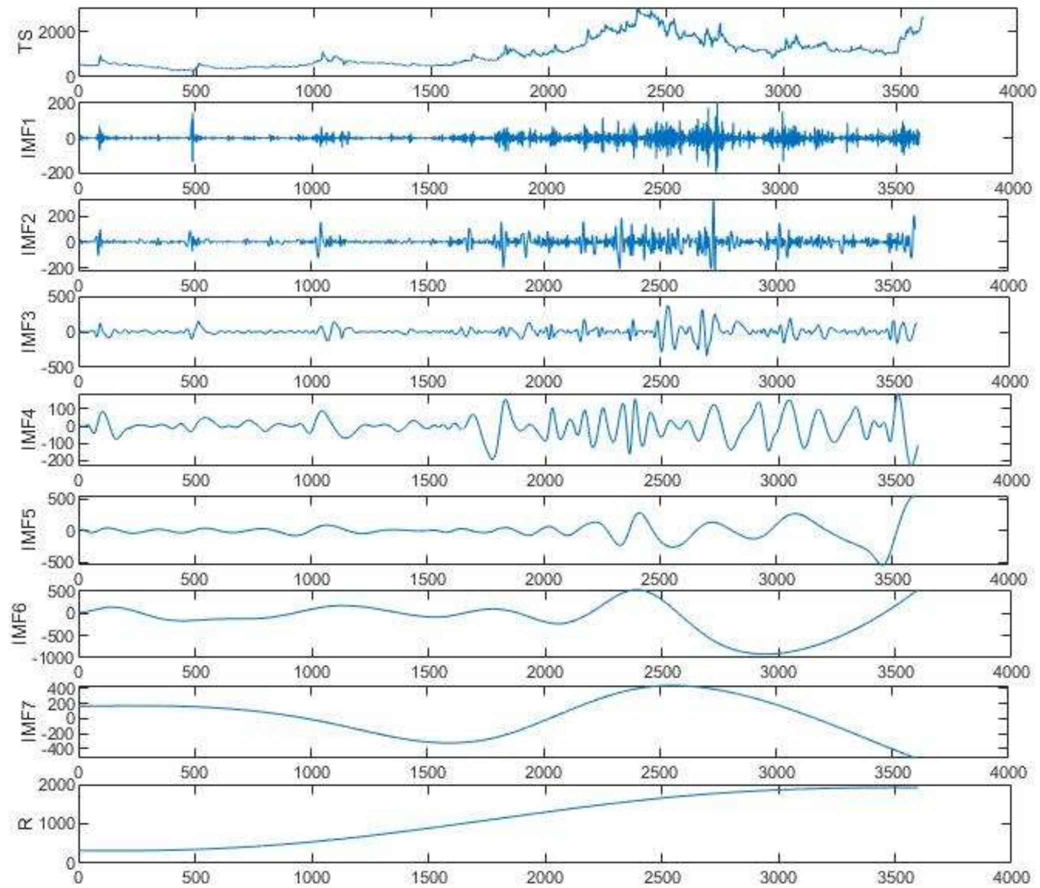

Fig. 4. *The decomposition result for BD Lamps Ltd. from EMD.*

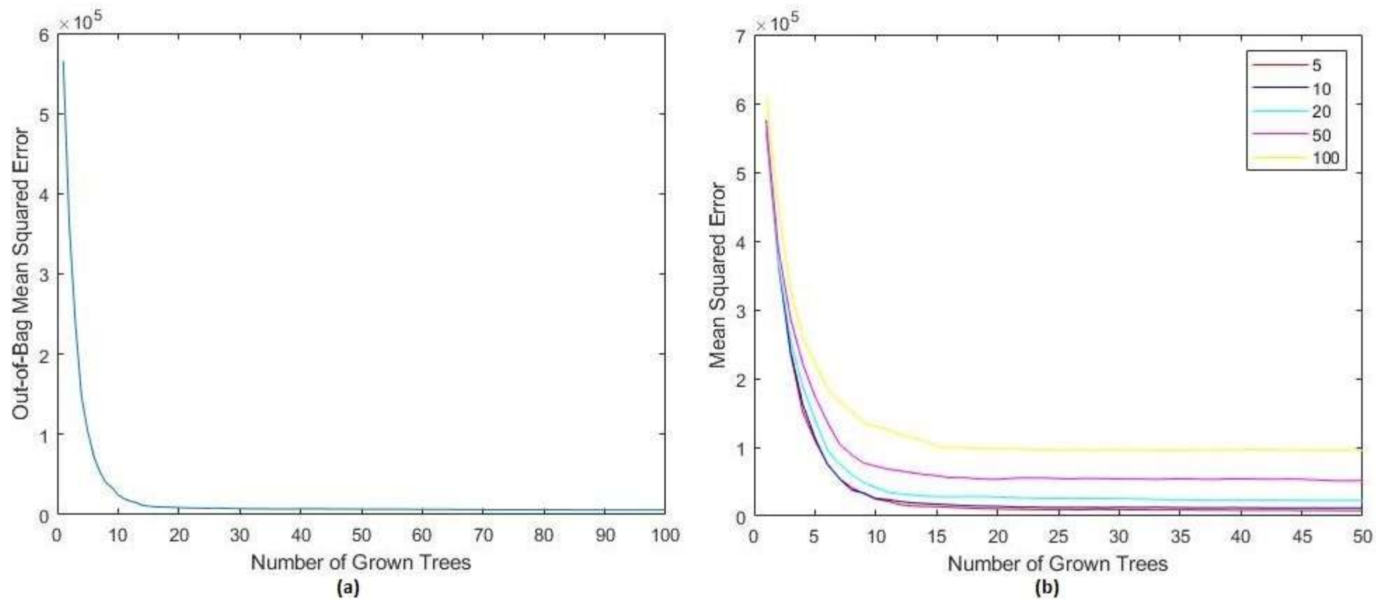

Fig. 5. *(a) OOB error rates of RF as the number of trees increases for Square Pharmaceutical Ltd.*
(b) OOB error rates of RF as the number of minimum node size increases for Square Pharmaceuticals Ltd.

5.2 Comparison of Forecasting Results

To verify the advantages of EMD based ensemble methods, we implemented the single structure models of SVR and RF for price forecasting, and conducted a comparison with their EMD based ensemble methods. Moreover, the comparison was carried out with previously proposed PCA-SVR, ICA-SVR and PCA-ICA-SVR models. The same EMD-based methodology steps of EMD-RF are followed to construct EMD-SVR model. For the remaining five models 30 technical indicators were used in the same way as in [14]. In all the cases, closing prices of the target stocks are predicted for 1, 5, 10, 15, 20, 25 and 30 days in advance. Prediction results for *Square Pharmaceuticals Limited* are listed in Table 1 to Table 4. Table 5 to Table 8 illustrate the comparative results of price forecasting for *AB Bank Limited*. Table 9 to Table 12 compare the performance of price forecasting for *Bangladesh Lamps Limited.*

It is evident from all the results that, the proposed EMD-RF model has produced lower MAPE (%), MAE, MSE and rRMSE for all three target stocks. EMD can bring significant improvements in the forecasting models especially in the long term stock price forecasting but the proposed EMD-RF model outperforms other five compared methods. This corroborates that the proposed EMD-RF approach can generate lower prediction errors than other five compared approaches. Again, it could be noticed from the results that, where the forecasting performance of all other approaches decreases as the predictions are made for more and more number of days in advance, the performance of the proposed EMD-RF model does not degrade in that ratio. This also means that the proposed method has more advantage compared with the benchmark models.

Table 1: MAPE(%) comparison of proposed EMD-RF with other models of *Square Pharmaceuticals Limited.*

No. of Days	**Forecasting Models**						
	SVR	***EMD-SVR***	***PCA-SVR***	***ICA-SVR***	***PCA-ICA-SVR***	***RF***	***EMD-RF***
1	2.35	3.14	2.27	2.36	1.95	**1.38**	1.42
5	3.77	3.10	3.72	3.61	3.50	2.30	**1.49**
10	5.49	3.09	5.51	5.13	5.21	2.91	**1.59**
15	7.04	3.00	7.12	6.18	6.57	3.19	**1.60**
20	7.93	3.11	7.97	6.99	7.37	2.95	**1.47**
25	8.89	3.27	8.89	7.86	8.34	3.01	**1.49**
30	9.81	3.57	9.77	8.59	9.17	2.91	**1.48**

Table 2: MAE comparison of proposed EMD-RF with other models of *Square Pharmaceuticals Limited.*

No. of Days	**Forecasting Models**						
	SVR	***EMD-SVR***	***PCA-SVR***	***ICA-SVR***	***PCA-ICA-SVR***	***RF***	***EMD-RF***
1	0.024	0.031	0.023	0.024	0.02	0.014	**0.014**
5	0.038	0.031	0.037	0.036	0.035	0.023	**0.015**
10	0.055	0.031	0.055	0.051	0.052	0.029	**0.016**
15	0.07	0.03	0.071	0.062	0.066	0.032	**0.016**
20	0.079	0.031	0.08	0.07	0.074	0.029	**0.015**
25	0.089	0.033	0.089	0.079	0.083	0.03	**0.015**
30	0.098	0.036	0.098	0.086	0.092	0.029	**0.015**

Table 3: MSE comparison of proposed EMD-RF with other models of *Square Pharmaceuticals Limited.*

No. of Days	**Forecasting Models**						
	SVR	***EMD-SVR***	***PCA-SVR***	***ICA-SVR***	***PCA-ICA-SVR***	***RF***	***EMD-RF***
1	7,245.60	9,814.87	7,277.29	9,555.73	6,892.71	6,010.60	**4,890.89**
5	23,759.40	9,781.71	23,958.27	19,667.32	21,497.89	11,334.21	**4,886.44**
10	59,123.94	10,930.92	59,702.18	41,336.03	53,083.03	20,841.20	**8,512.95**
15	96,325.49	9,430.56	96,957.13	62,528.79	84,125.77	23,899.74	**6,435.60**
20	116,961.23	9,359.73	118,096.24	79,061.48	100,531.52	16,772.51	**4,734.76**
25	133,204.18	10,265.13	134,236.35	93,265.20	115,086.34	19,022.94	**4,665.20**

30	153,708.54	11,304.45	154,924.18	112,216.41	135,480.39	17,116.62	**4,885.29**

Table 4: rRMSE comparison of proposed EMD-RF with other models of *Square Pharmaceuticals Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	0.492	0.319	0.429	0.271	0.145	**0.033**	0.091
5	0.486	0.153	0.458	0.215	0.286	0.027	**0.022**
10	0.446	0.158	0.45	0.222	0.325	0.079	**0.078**
15	0.501	0.194	0.556	0.237	0.311	0.124	**0.084**
20	0.536	0.152	0.528	0.257	0.298	0.129	**0.038**
25	0.577	0.185	0.542	0.36	0.464	0.139	**0.018**
30	0.729	0.313	0.649	0.503	0.556	0.206	**0.075**

Table 5: MAPE(%) comparison of proposed EMD-RF with other models of *AB Bank Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	5.29	10.47	5.40	4.97	4.56	**2.12**	2.30
5	7.87	11.04	7.94	7.28	6.92	3.74	**2.40**
10	9.77	10.25	9.75	8.67	8.59	4.27	**2.54**
15	11.64	9.91	11.47	10.69	10.55	4.40	**2.48**
20	13.22	11.07	13.06	11.71	11.87	4.12	**2.46**
25	14.59	9.69	14.72	12.73	12.82	3.83	**2.42**
30	15.73	11.07	15.79	13.22	13.76	3.61	**2.57**

Table 6: MAE comparison of proposed EMD-RF with other models of *AB Bank Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	0.053	0.105	0.054	0.050	0.046	**0.021**	0.023
5	0.079	0.110	0.079	0.073	0.069	0.037	**0.024**
10	0.098	0.102	0.097	0.087	0.086	0.043	**0.025**
15	0.116	0.099	0.115	0.107	0.105	0.044	**0.025**
20	0.132	0.111	0.131	0.117	0.119	0.041	**0.025**
25	0.146	0.097	0.147	0.127	0.128	0.038	**0.024**
30	0.157	0.111	0.158	0.132	0.138	0.036	**0.026**

Table 7: MSE comparison of proposed EMD-RF with other models of *AB Bank Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	1,874.72	78,502.66	1,998.75	2,062.05	1,389.85	**1,240.77**	4,311.32
5	14,235.50	99,562.30	14,417.60	8,732.63	10,702.26	3,662.13	**2,436.77**
10	31,896.96	76,308.07	32,183.37	15,957.62	25,461.77	6,472.95	**6,315.30**
15	56,815.80	68,013.35	57,166.17	38,082.74	47,132.45	7,104.85	**2,277.13**
20	60,728.37	104,923.29	61,248.34	39,327.08	49,519.63	4,067.53	**4,517.40**
25	81,052.34	63,274.07	81,264.13	58,959.43	64,624.05	5,169.81	**2,161.65**
30	89,603.37	108,849.00	89,608.15	62,653.87	68,441.05	6,397.28	**5,285.19**

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2000-000X
Vol. 3 Issue 1, January – 2019, Pages: 9-13

Table 8: rRMSE comparison of proposed EMD-RF with other models of *AB Bank Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	0.240	1.101	0.252	0.561	0.343	**0.021**	0.072
5	0.124	0.966	0.057	0.328	0.050	0.017	**0.010**
10	0.061	1.071	0.291	0.141	0.380	0.114	**0.026**
15	0.269	1.266	0.165	0.362	0.410	0.232	**0.099**
20	0.247	1.036	0.075	0.271	0.476	0.121	**0.069**
25	0.164	1.086	0.245	0.508	0.402	0.059	**0.015**
30	0.182	1.115	0.204	0.335	0.526	0.007	**0.031**

Table 9: MAPE(%) comparison of proposed EMD-RF with other models of *Bangladesh Lamps Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	5.89	8.64	5.85	5.88	5.22	4.03	**3.68**
5	6.07	5.55	6.06	5.76	5.49	3.73	**2.27**
10	7.12	5.21	7.20	6.81	6.64	3.58	**2.07**
15	10.18	7.14	10.30	9.55	9.48	4.96	**3.48**
20	9.75	5.69	9.73	8.85	9.15	3.73	**2.21**
25	11.85	7.00	11.92	11.03	11.22	4.87	**3.34**
30	11.72	5.87	11.68	10.63	11.02	3.78	**2.48**

Table 10: MAE comparison of proposed EMD-RF with other models of *Bangladesh Lamps Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	0.059	0.086	0.059	0.059	0.052	0.040	**0.037**
5	0.061	0.056	0.061	0.058	0.055	0.037	**0.023**
10	0.071	0.052	0.072	0.068	0.066	0.036	**0.021**
15	0.102	0.071	0.103	0.095	0.095	0.050	**0.035**
20	0.097	0.057	0.097	0.088	0.091	0.037	**0.022**
25	0.119	0.070	0.119	0.110	0.112	0.049	**0.033**
30	0.117	0.059	0.117	0.106	0.110	0.038	**0.025**

Table 11: MSE comparison of proposed EMD-RF with other models of *Bangladesh Lamps Limited.*

No. of Days	Forecasting Models						
	SVR	*EMD-SVR*	*PCA-SVR*	*ICA-SVR*	*PCA-ICA-SVR*	*RF*	*EMD-RF*
1	3,202.23	6,274.35	3,229.37	2,694.80	2,631.77	2,620.65	**1,613.10**
5	9,602.82	6,845.43	9,692.84	7,788.14	8,509.78	5,672.10	**2,058.00**
10	14,460.58	4,552.39	14,616.84	12,425.01	13,030.18	5,156.76	**1,378.18**
15	19,901.87	4,761.16	20,016.32	16,738.83	17,712.53	4,490.28	**1,875.40**
20	25,247.56	6,057.36	25,492.69	21,162.09	23,032.29	4,174.57	**1,683.56**
25	28,757.42	5,421.76	29,003.58	23,645.94	26,315.89	3,980.77	**1,483.36**
30	34,146.43	5,938.42	34,446.62	27,374.06	31,225.98	3,874.87	**1,650.48**

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2000-000X
Vol. 3 Issue 1, January – 2019, Pages: 10-13

Table 12: rRMSE comparison of proposed EMD-RF with other models of *Bangladesh Lamps Limited.*

No. of Days	Forecasting Models						
	SVR	***EMD-SVR***	***PCA-SVR***	***ICA-SVR***	***PCA-ICA-SVR***	***RF***	***EMD-RF***
1	1.435	1.861	1.402	1.481	1.196	0.659	**0.565**
5	0.797	0.753	0.754	0.778	0.681	0.089	**0.070**
10	0.620	0.569	0.612	0.659	0.488	0.107	**0.371**
15	1.097	1.251	1.147	1.111	0.949	0.498	**0.379**
20	0.542	0.804	0.494	0.570	0.472	0.077	**0.029**
25	0.976	1.260	0.961	1.103	0.876	0.597	**0.484**
30	0.650	0.934	0.570	0.655	0.492	0.210	**0.192**

5.3 Robustness

Evaluation of the proposed EMD-RF model's robustness is performed by comparing its performance with other six methods. Five relative ratios, 50%, 60%, 70%, 80%, and 90% of training sample size with respect to the complete dataset size, and two performance measures, MAPE (%) and rRMSE, are used for the comparison. Closing price of the target stock is predicted for 15 days ahead. Table 13 to Table 15 illustrate the prediction results for *Square Pharmaceuticals Limited*, *AB Bank Limited* and *Bangladesh Lamps Limited* respectively. The results showcase that the proposed EMD-RF method outperforms other methods under all five different relative ratios for all three target stocks. It therefore concludes that EMD-RF approach clearly produces less forecasting error than other six approaches. This demonstrates the effectiveness of our proposal.

Table 13: Robustness evaluation of proposed EMD-RF with other forecasting models with different relative ratios of *Square Pharmaceuticals Limited.*

Relative Ratio	Performance Measure	Forecasting Models						
		SVR	***EMD-SVR***	***PCA-SVR***	***ICA-SVR***	***PCA-ICA-SVR***	***RF***	***EMD-RF***
50	**MAPE(%)**	6.764	3.403	6.810	6.073	6.323	3.384	**1.808**
	rRMSE	0.884	0.484	0.896	0.520	0.612	0.302	**0.083**
60	**MAPE(%)**	6.429	3.078	6.423	5.949	6.072	3.077	**1.671**
	rRMSE	0.192	0.250	0.177	0.065	0.047	0.034	**0.027**
70	**MAPE(%)**	7.041	3.004	7.122	6.177	6.570	3.193	**1.603**
	rRMSE	0.501	0.194	0.556	0.237	0.311	0.124	**0.084**
80	**MAPE(%)**	6.855	3.390	6.923	6.051	6.550	2.691	**1.373**
	rRMSE	0.399	0.343	0.399	0.274	0.272	0.069	**0.041**
90	**MAPE(%)**	7.107	3.138	7.178	6.201	6.558	2.792	**1.348**
	rRMSE	0.345	0.165	0.356	0.194	0.233	0.062	**0.025**

Table 14: Robustness evaluation of EMD-RF with other forecasting models with different relative ratios of *AB Bank Ltd.*

Relative Ratio	Performance Measure	Forecasting Models						
		SVR	***EMD-SVR***	***PCA-SVR***	***ICA-SVR***	***PCA-ICA-SVR***	***RF***	***EMD-RF***
50	**MAPE(%)**	11.713	11.220	11.770	10.666	10.460	5.208	**3.209**
	rRMSE	0.853	1.707	0.828	0.943	0.790	0.347	**0.125**
60	**MAPE(%)**	11.455	10.190	11.295	11.093	10.310	4.534	**2.846**
	rRMSE	0.179	1.434	0.196	1.092	0.731	0.146	**0.097**
70	**MAPE(%)**	11.644	9.910	11.474	10.689	10.545	4.405	**2.475**
	rRMSE	0.269	1.266	0.165	0.362	0.410	0.232	**0.099**
80	**MAPE(%)**	11.185	11.585	11.188	9.828	9.876	3.975	**2.361**
	rRMSE	0.232	1.207	0.317	0.199	0.419	0.158	**0.025**
90	**MAPE(%)**	11.622	8.191	11.632	10.712	10.570	3.804	**2.405**

	rRMSE	0.385	0.788	0.331	0.374	0.499	0.088	**0.035**

Table 15: Robustness evaluation of proposed EMD-RF with other forecasting models with different relative ratios of *Bangladesh Lamps Limited.*

Relative Ratio	Performance Measure	Forecasting Models						
		SVR	***EMD-SVR***	***PCA-SVR***	***ICA-SVR***	***PCA-ICA-SVR***	***RF***	***EMD-RF***
50	**MAPE(%)**	9.975	6.972	10.018	9.319	9.290	5.230	**3.683**
	rRMSE	1.017	1.477	0.978	0.949	0.830	0.511	**0.492**
60	**MAPE(%)**	8.844	5.534	8.859	8.174	8.293	4.177	**2.395**
	rRMSE	0.638	0.646	0.620	0.697	0.564	0.139	**0.059**
70	**MAPE(%)**	10.177	7.143	10.298	9.546	9.483	4.961	**3.482**
	rRMSE	1.097	1.251	1.147	1.111	0.949	0.498	**0.379**
80	**MAPE(%)**	8.636	5.544	8.633	7.994	7.933	3.958	**2.597**
	rRMSE	0.467	0.778	0.409	0.375	0.246	0.160	**0.095**
90	**MAPE(%)**	10.000	7.477	10.149	9.451	9.354	5.549	**3.657**
	rRMSE	0.946	1.472	1.061	0.884	0.773	0.492	**0.398**

6. Summary and Conclusions

In this paper, we proposed an ensemble RF model composing EMD for financial TS forecasting. EMD first decomposed the stock price TS into several IMFs and a residual component which were then fed to an RF model to predict future stock prices. Close prices for three leading stocks from three different sectors listed in Dhaka Stock Exchange, Bangladesh were used to validate the effectiveness of the proposed method. The performance was compared with six other methods. The empirical results show that EMD based hybrid models like EMD-SVR and the proposed EMD-RF achieve better forecasting performance compared to the corresponding single structure models. Again, as the predictions are made for more and more number of days in advance, the performance of most the methods degrades significantly. But, the proposed EMD-RF model yields insignificant degradation of forecasting performance. Hence, the proposed EMD-RF approach outperforms single RF, single SVR, EMD-SVR, PCA-SVR, ICA-SVR and PCA-ICA-SVR on several criteria. Thus, the proposed EMD-RF model might be used as an effective tool for financial TS forecasting.

In this research work, only the price related historical data is used here to predict future prices. But, it is well known that various other aspects like general economic conditions, government policies, company performance, investor's interest etc. also play vital roles in stock market. In future, these aspects can also be incorporated as input features for prediction which may buttress the accurate prediction. Moreover, the proposed EMD-RF model can also be tested on other TS data, such as renewable energy data and weather data to evaluate the performance in generic situations.

7. References

[1] Shian-Chang Huang, Pei-Ju Chuang, Cheng-Feng Wu, and Hiuen-Jiun Lai. *Chaos-based support vector regressions for exchange rate forecasting*. Expert Systems with Applications, 37(12):8590-8598, 2010.

[2] Mehdi Khashei, Mehdi Bijari, and Gholam Ali Raissi Ardali. *Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (anns)*. Neurocomputing, 72(4-6):956-967, 2009.

[3] Kyoung-jae Kim and Ingoo Han. *Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index*. Expert systems with Applications,19(2):125-132, 2000.

[4] JingTao Yao and Joseph P Herbert. *Financial time-series analysis with rough sets.* Applied Soft Computing, 9(3):1000-1007, 2009.

[5] Corinna Cortes and Vladimir Vapnik. *Support-vector networks*. Machine learning, 20(3):273-297, 1995.

[6] Jian-Yi Lin, Chun-Tian Cheng, and Kwok-Wing Chau. *Using support vector machines for long-term discharge prediction*. Hydrological Sciences Journal, 51(4):599-612, 2006.

[7] Wen-Qi Duan and H Eugene Stanley. *Cross-correlation and the predictability of financial return series*. Physica A: Statistical Mechanics and its Applications, 390(2):290-296, 2011.

[8] Grzegorz Dudek. *Short-term load forecasting using random forests*. In Intelligent Systems' 2014, pages 821-828. Springer, 2015.

[9] Huanhuan Yu, Rongda Chen, and Guoping Zhang. *A svm stock selection model within pca*. Procedia computer science, 31:406-412, 2014.

[10] Utpala Nanda Chowdhury, Md Abu Rayhan, Sanjoy Kumar Chakravarty, and Md Tanvir Hossain. *Integration of principal component analysis and support vector regression for financial time series forecasting*. International Journal of Computer Science and Information Security (IJCSIS), 15, 2017.

[11] Anass Nahil and Abdelouahid Lyhyaoui. *Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of casablanca stock exchange*. Procedia Computer Science, 127:161-169, 2018.

[12] Chi-Jie Lu, Tian-Shyug Lee, and Chih-Chou Chiu. *Financial time series forecasting using independent component analysis and support vector regression*. Decision Support Systems, 47(2):115-125, 2009.

[13] Ling-Jing Kao, Chih-Chou Chiu, Chi-Jie Lu, and Jung-Li Yang. *Integration of nonlinear independent component analysis and support vector regression for stock price forecasting*. Neurocomputing, 99:534-542, 2013.

[14] Utpala Nanda Chowdhury, Sanjoy Kumar Chakravarty, and Md Tanvir Hossain. *Short-term financial time series forecasting integrating principal component analysis and independent component analysis with support vector regression*. Journal of Computer and Communications, 6(03):51, 2018.

[15] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. *The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis*. In Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences, volume 454, pages 903-995. The Royal Society, 1998.

[16] Md Islam, Md Rashed-Al-Mahfuz, Shamim Ahmad, Md Molla, Khademul Islam, et al. *Multiband prediction model for financial time series with multivariate empirical mode decomposition*. Discrete Dynamics in Nature and Society, 2012, 2012.

[17] Noemi Nava, Tiziana Di Matteo, and Tomaso Aste. *Financial time series forecasting using empirical mode decomposition and support vector regression*. Risks, 6(1):7, 2018.

[18] Chiun-Sin Lin, Sheng-Hsiung Chiu, and Tzu-Yu Lin. *Empirical mode decomposition-based least squares support vector regression for foreign exchange rate forecasting*. Economic Modelling, 29(6):2583-2590, 2012.

[19] Xueheng Qiu, Huilin Zhu, PN Suganthan, and Gehan AJ Amaratunga. *Stock price forecasting with empirical mode decomposition based ensemble υ-support vector regression model*. In International Conference on Computational Intelligence, Communications, and Business Analytics, pages 22-34. Springer, 2017.

[20] Leo Breiman. *Random forests*. Machine learning, 45(1):5-32, 2001.

[21] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[22] Subrata Pramanik, Utpala Nanda Chowdhury, Bimal Kumar Pramanik, and Nazmul Huda. *A comparative study of bagging, boosting and c4. 5: The recent improvements in decision tree learning algorithm*. *Asian J. Inf. Tech*, 9(6):300-306, 2010.

[23] Andy Liaw, Matthew Wiener, et al. *Classification and regression by randomforest*. R news, 2(3):18-22, 2002.

[24] Jigar Patel, Sahil Shah, Priyank Thakkar, and Ketan Kotecha. *Predicting stock market index using fusion of machine learning techniques*. Expert Systems with Applications, 42(4):2162-2172, 2015.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2000-000X
Vol. 3 Issue 1, January – 2019, Pages: 13-13

Authors

Utpala Nanda Chowdhury received his B.Sc. (Hons) and M.Sc. from the department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh. Currently he is an assistant professor in the department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh. Prior to his academic career he had been in Investment Corporation of Bangladesh. His research interest includes Data mining, Machine learning, Bioinformatics and Data analytics.

Sanjoy Kumar Chakravarty completed B.Sc. and M.Sc. from the department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh. He has been working as an assistant professor in the same department. He is currently pursuing his Ph.D. degree from the department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh. His research interest includes Digital signal processing, Brain computer interfacing, and Machine learning.

Md. Tanvir Hossain received his B.Sc. and M.Sc. from the department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh. Currently he is a lecturer in the department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh. His research interest includes Image Processing, Computer Graphics, Virtual Reality, Augmented Reality, Computer Vision, Mixed Reality, Data Analysis, and Machine Learning.

Shamim Ahmad completed B.Sc. and M.Sc. from University of Rajshahi, Rajshahi, Bangldesh and then completed D.Engg. from Chubu University, Japan. Currently he is a professor in the department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh. His research interest includes Embedded Systems, Bioinformatics, Digital Signal Processing, and Machine learning.