Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 143 (2018) 947–953

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Advances in Computing and Communication (ICACC-2018)

Applicability of Deep Learning Models for Stock Price Forecasting An Empirical Study on BANKEX Data

A. Jayanth Balaji*, D. S. Harish Ram, Binoy B. Nair

SIERS Research Laboratory, Department of Electronics and Communication Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

Abstract

Stock price time series are extremely nonlinear in nature and hence, accurate stock price forecasting has been a challenge. Accurate prediction of stock prices and the direction of stock price movement is also essential for a stock trader/investor in order to trade profitably. A deep learning approach to stock price forecasting is presented in this study. A total of fourteen different deep learning models based on Long-Short Term Memory (LSTM), Gated Recurring Unit (GRU), Convolutional Neural Networks (CNN) and Extreme Learning Machines (ELM) are designed and empirically evaluated on all stocks in the S&P BSE-BANKEX index for their ability to generate one-step ahead and four-step ahead forecasts. Performance of the proposed systems is evaluated in terms of the Root Mean Squared Error (RMSE), Directional Accuracy (DA) and the Median Absolute Percentage Error (MdAPE). Results indicate that deep learning models proposed in this study are capable of generating highly accurate stock price forecasts.

Keywords: Deep Learning; CNN; ELM, GRU; LSTM; Financial Time-Series.

1. Introduction

Nonlinear time series such as stock prices have long been considered to be impossible to forecast. From the earliest studies reported, eg. In [1] and [2] to the Efficient Market Hypothesis [3, 4], there have been several studies that claim that stock prices tend to follow a random walk and consequently, cannot be predicted with any reasonable degree of accuracy. However, with the advent of artificial neural networks and other soft computing techniques, it has become easier to model (to a limited degree, at least) the nonlinear behavior exhibited by stock price time series as is exhibited in [5, 6, 7, 8]. Deep learning models have become very popular recently and have been yielding state-of-the-art performance for tasks such as image classification[9, 10]. However, the applicability of such deep learning

* Corresponding author. Tel.: +91-8056798764.
E-mail address: a_jayanthbalaji@cb.amrita.edu

10.1016/j.procs.2018.10.340

models for stock forecasting is still being explored. An effort has been made in this study to generate one-day-ahead and four-day-ahead stock price forecasts using lean deep learning models. A total of fourteen different deep learning models based on Long-Short Term Memory (LSTM), Gated Recurring Unit (GRU), Convolutional Neural Networks (CNN) and Extreme Learning Machines (ELM) have been evaluated on the entire set of stocks in the S&P BSE-BANKEX[11] index. The paper is structured as follows: section 2 describes the datasets and the preprocessing steps, section 3 describes the design of the forecasting systems evaluated in the present study, section 4 presents the results and analysis and section 5 presents the conclusions.

2. Data description and Preprocessing

The datasets considered in the present study are the closing price time series for the stocks in the S&P BSE-BANKEX index. Ten publicly traded banking stocks constitute the S&P BSE-BANKEX. The list of stocks that constitute the S&P BSE-BANKEX as on April 11, 2018, along with the considered time-frame and the number of samples considered for each stock is presented in Table 1.

Table 1. BANKEX Entities with symbols and Start and End date of data considered.

S. No	Symbols	Entity	Start Date	End Date	Number of Samples
1	AXIXBANK.BO	Axis Bank	12.07.2005	03.11.2017	3,057
2	BANKBARODA.BO	Bank of Baroda	12.07.2005	03.11.2017	3,057
3	FEDERALBNK.BO	Federal Bank	12.07.2005	03.11.2017	3,057
4	HDFCBANK.BO	HDFC Bank	12.07.2005	03.11.2017	3,057
5	ICICBANK.BO	ICICI Bank	12.07.2005	03.11.2017	3,057
6	INDUSINDBK.BO	Indus Ind Bank	12.07.2005	03.11.2017	3,057
7	KOTAKBANK.BO	Kotak Mahindra	12.07.2005	03.11.2017	3,057
8	PNB.BO	PNB	12.07.2005	03.11.2017	3,057
9	SBIN.BO	SBI	12.07.2005	03.11.2017	3,057
10	YESBANK.BO	Yes Bank	12.07.2005	03.11.2017	3,057

All the stock datasets considered in this study have been sourced from [12]. As a pre-processing step, '0-1'normalization has been performed on all the datasets. The '0-1'normalization is calculated as given in equation (2). Let the set of N samples of a stock price daily close time series be represented by:

$$Y = \{y(t),\ y(t-1),\cdots,\ y(t-(N-1))\} \tag{1}$$

Where, $y(t)$ represents the closing price at day t, $y(t-1)$ is the closing price on the immediate previous trading day, and so on.
Then, the'0-1'normalization, transforms the sample $y(t)$ to:

$$y_{Norm}(t) = \frac{y(t)-Max(Y)}{Max(Y)-Min(Y)} \tag{2}$$

Where, $Max(Y)$ and $Min(Y)$ corresponds to the maximum and minimum values respectively, in the dataset. The normalized dataset can now be represented as:

$$Y_{Norm} = \{y_{Norm}(t),\ y_{Norm}(t-1),\cdots,\ y_{Norm}(t-(N-1))\} \tag{3}$$

Once all the samples are normalized, the normalized data is fed into the deep learning models for forecasting. The following section presents a detailed description of deep learning-based forecasting systems considered in this study.

3. System Description

A total of fourteen different models based on four different deep learning techniques: Long-Short Term Memory (LSTM), Gated Recurring Unit (GRU), Convolutional Neural Network (CNN) and Extreme Learning Machines (ELM) are designed as depicted in Fig 1. Each input vector for one-step as well as four-step ahead forecasts consists of the normalized values of four most recent closing prices as represented in equation (4).

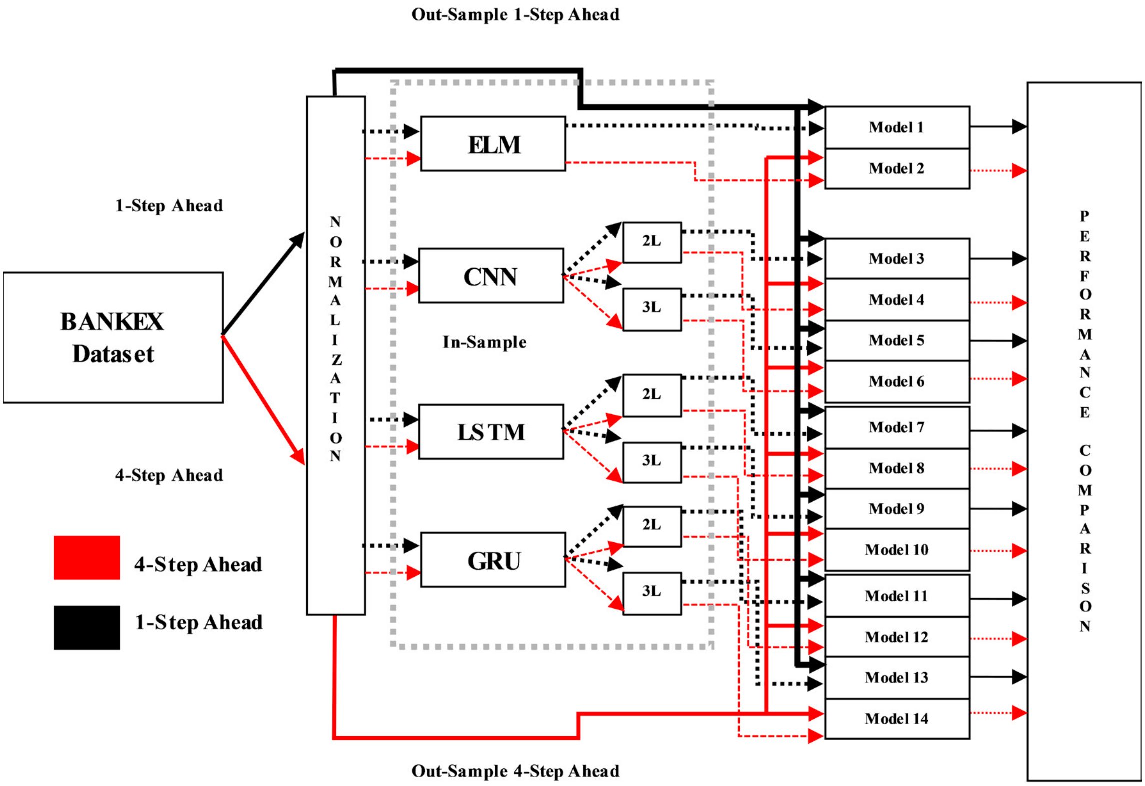

Fig. 1. Block Diagram of Proposed System

$$Y(t) = \begin{Bmatrix} y_{Norm1}(t) & y_{Norm_1}(t-1) & y_{Norm_1}(t-2) & y_{Norm_1}(t-3) \\ y_{Norm2}(t) & y_{Norm2}(t-1) & y_{Norm2}(t-2) & y_{Norm2}(t-3) \\ \vdots & \vdots & \vdots & \vdots \\ y_{Norm10}(t) & y_{Norm10}(t-1) & y_{Norm10}(t-2) & y_{Norm10}(t-3) \end{Bmatrix} \tag{4}$$

Where, $y_{Norm_i}(t)$ represents the normalized closing price for stock i, $1 \leq i \leq 10$
The corresponding target vector for one-step ahead forecasts is as represented in equation (5) and for the four step ahead forecasts, is represented in equation (6).

$$D_1(t) = \begin{Bmatrix} y_{Norm1}(t+1) \\ y_{Norm2}(t+1) \\ \vdots \\ y_{Norm10}(t+1) \end{Bmatrix} \tag{5}$$

and

$$D_4(t) = \begin{Bmatrix} y_{Norm1}(t+1) & y_{Norm1}(t+2) & y_{Norm1}(t+3) & y_{Norm1}(t+4) \\ y_{Norm2}(t+1) & y_{Norm2}(t+2) & y_{Norm2}(t+3) & y_{Norm2}(t+4) \\ \vdots & \vdots & \vdots & \vdots \\ y_{Norm10}(t+1) & y_{Norm10}(t+2) & y_{Norm10}(t+3) & y_{Norm10}(t+4) \end{Bmatrix} \tag{6}$$

Hence, the entire input training dataset can be represented as $X = \{Y(t), Y(t-1), \cdots, Y(t-N+4)\}$ and the corresponding target datasets for one-step ahead and four step ahead forecasts can be represented as
$D_1 = \{D_1(t), D_1(t-1), \cdots, D_1(t-N+5)\}$ and $D_4 = \{D_4(t), D_4(t-1), \cdots, D_4(t-N+5)\}$, respectively. The deep

learning models are trained to learn patterns from these datasets. A detailed description of the models used and the training parameters is presented below.

3.1. Long Short-Term Memory (LSTM)

A Recurrent Neural Network (RNN) comprises of series of repeating neural network with simpler '*tanh*'blocks throughout their structures. LSTM [13] is a kind of RNN with capability to learn longer dependencies in data. LSTM has been effectively used in anomaly detection[14], event forecasting[15], and time series forecasting as demonstrated in [16, 17].
Two types of LSTM models are evaluated, one with two hidden layers and the other with three hidden layers. Each of these configurations is trained to generate one-step ahead and four-step ahead forecasts, resulting in a total of four trained models. All the hidden layers in all of the four models described above, have 64 neurons each, selected using trial-and-error. Linear activation functions are used in all layers.

3.2. Gated Recurrent Unit (GRU) [13]

GRU is a less complex RNN when compared with computations associated with LSTM. GRU is composed only of two gates, namely, a reset gate and an update gate. It doesnt possess any internal memory and output gate like LSTM. There have been several experimental evaluations of both LSTM and GRU with respect to their performance, as in [13].
Similar to the LSTM configuration, two types of GRU models are evaluated, first one with two hidden layers and the second with three. Each of these configurations is trained to generate one-step ahead and four-step ahead forecasts, thus a total of four trained models are generated. All the hidden layers in all of the four models described above, have 64 neurons each, selected using trial-and-error. Linear activation functions, similar to the ones used for LSTM models, are used in all layers.

3.3. Convolutional Neural Network (CNN) [18]

Convolutional Neural Networks (CNNs) have been widely reported to be capable of generating state-of-the art performance in fields such as image classification and natural language processing. In general, CNN is composed of an input layer, multiple hidden layers and an output layer. Each hidden layer is made up of the convolutional layer and pooling layer with batch normalization taking place after the convolution layer. The output layer is a fully connected layer with the number of neurons being equal to the number of classes (or outputs), as defined in the training data. CNNs has been used for applications such as wind forecasting [19] etc.. A detailed analysis of CNN architecture can be found in [18]. In the present study, two types of CNN models are evaluated, first one with two hidden layers and the second with three. Each of these configurations is trained to generate one-step ahead and four-step ahead forecasts, resulting in a total of four trained models. All the hidden layers in all of the four models described above, have 64 neurons each, selected using trial-and-error. However, unlike the deep learning models described earlier, Rectifier Linear Unit (ReLU) activation functions, are used in all layers.

3.4. Extreme Learning Machines (ELMs)

ELM a kind of feed-forward neural network [20] which is composed of a single hidden layer, where the weights between input and hidden layer are assigned randomly following a normal distribution and the weights between hidden and output layers are learnt in single step by a pseudo-inverse technique making the network the fastest learner when compared with other techniques considered in the present study. ELMs are vast and the performance has been claimed to be comparable to many state-of-art deep learning models, as in [20],[21],[22],[23] and [24]. The forecast output from the ELM can be calculated as in equation (7).

$$\widehat{D}_i = W_h \sigma_A \left(W_p X \right) \tag{7}$$

where W_p and W_h are the matrix of weights between input-hidden layer and hidden-output layer, $\sigma_A(.)$ is the activation function and $\hat{D}_i$ is i-th step ahead forecast, $i \in \{0, 4\}$. The activation function for ELM models is sigmoid and since, conventionally, the hidden layer neuron count is fixed to be twice as the size of inputs, the two models considered (one each for one-step ahead and four step ahead) have the hidden neurons as 80.

4. Results and analysis

All the models are trained using a Nvidia 1080 Ti GPU. The training terminates if any of these two stopping criteria are met: (a) max. iterations:1000 or (b) minimum MSE: 10^{-6}. For all the models, 92% of the data is used for training and the remaining 8% data is used for testing. Since the number of samples is same for all the stocks, this results in 2,806 samples being used for training and 251 samples for testing. The models are evaluated for their ability to generate 1 - day ahead and 4 day ahead forecasts using three different performance measures, namely, the Root Mean Squared Error (RMSE), Median Absolute Percentage Error (MdAPE)[25]. Since the direction of stock price movements is also of critical importance for a stock trader, the third measure used for evaluating the effectiveness of the models is Directional Accuracy (DA)[26].
The forecasting results are tabulated in the tables 2-5. The Table 2 and 3 gives the DA for different forecast horizons of the models under consideration for BANKEX entities, Table 4 and Table 5 present the MdAPE values and the RMSE is presented in Tables 6 and 7.

Table 2. Directional Accuracy Single Step Ahead prediction for NSE - BANKEX

Technique	Layers	Axis Bank	Bank of Baroda	Federal Bank	HDFC Bank	ICICI Bank	Indus Ind Bank	Kotak Mahindra	PNB	SBI	Yes Bank
CNN	2	59.35	59.35	58.94	55.69	61.38	**58.13**	**56.5**	52.85	54.47	55.28
	3	59.76	56.1	54.47	51.63	57.32	55.69	53.66	55.69	54.07	52.44
GRU	2	**71.95**	60.98	58.54	53.66	58.94	57.72	54.47	58.94	**65.85**	51.63
	3	64.63	58.94	60.16	**58.54**	60.16	49.19	50.41	**63.41**	63.41	48.78
LSTM	2	64.63	**61.38**	**61.38**	57.32	61.38	50.41	56.1	57.32	56.1	51.63
	3	63.82	58.94	53.25	50.41	56.5	48.37	54.47	54.47	57.72	48.37
ELM	1	61.79	59.35	54.07	55.69	**65.04**	48.37	55.28	59.35	60.57	**63.82**

Table 3. Directional Accuracy Four Step Ahead prediction for NSE - BANKEX

Technique	Layers	Axis Bank	Bank of Baroda	Federal Bank	HDFC Bank	ICICI Bank	Indus Ind Bank	Kotak Mahindra	PNB	SBI	Yes Bank
CNN	2	58.95	57.1	55.04	**58.44**	63.07	**59.36**	**57.72**	53.29	55.14	56.17
	3	57.2	53.19	53.5	51.13	59.88	54.22	56.17	55.76	59.16	55.76
GRU	2	51.95	55.04	49.69	49.59	54.42	55.56	47.84	53.91	54.32	55.45
	3	53.19	53.19	49.9	55.45	53.91	51.13	50.72	53.4	53.09	50.93
LSTM	2	58.54	**59.05**	52.57	51.65	54.32	56.48	48.87	52.67	57	50.93
	3	57.51	54.53	**55.86**	48.35	55.97	47.53	48.25	54.73	53.7	52.78
ELM	1	**60.19**	58.44	54.42	52.88	**63.17**	48.15	54.73	**57.41**	**61.11**	**65.12**

Table 4. MdAPE Single Step ahead prediction for NSE - BANKEX

Technique	Layers	Axis Bank	Bank of Baroda	Federal Bank	HDFC Bank	ICICI Bank	Indus Ind Bank	Kotak Mahindra	PNB	SBI	Yes Bank
CNN	2	18.33	29.82	16.82	43.56	16.79	23.23	**11.22**	49.62	30.76	63.37
	3	25.51	52.45	27.08	42.43	25.59	15.88	16.77	42.44	26.89	59.83
GRU	2	**12.87**	**19.87**	12.11	23.01	16.43	**13.57**	19.78	18.5	**13.43**	17.44
	3	22.77	24.34	**11.14**	**17.9**	17.9	21.78	23.67	**16.81**	16.44	17.24
LSTM	2	15.62	21.09	12.19	22.69	**15.94**	18.77	21.15	29.78	19.58	27.2
	3	20.49	30.81	22.8	28.89	29.27	29.39	24.61	32.74	18.17	31.31
ELM	1	18.76	27.39	20.36	23.3	16.88	26.64	24.73	28.76	17.99	**7.169**

It is observed that the of all the fourteen models considered, the ELM, LSTM model with 2 hidden layers (LSTM 2 Layer, CNN model with three hidden layers (CNN 3 Layer) and GRU model with three hidden layers (GRU 3 Layer) are capable of generating better forecasts compared to other models considered. For illustration of their effectiveness, the out-sample forecasting results of these three models and the actual time series is plotted in Fig. 2.

Table 5. MdAPE Four Step ahead prediction for NSE - BANKEX

Technique	Layers	Axis Bank	Bank of Baroda	Federal Bank	HDFC Bank	ICICI Bank	Indus Ind Bank	Kotak Mahindra	PNB	SBI	Yes Bank
CNN	2	19.47	31.04	**19.92**	34.71	20.59	28.75	24.63	56.21	28.09	66.4
	3	**18.48**	44.46	23.32	41.64	23.13	35.08	**13.76**	66.15	32.3	66.26
GRU	2	42.61	51.94	41.51	44.06	46.93	**14**	55	48.36	36.16	17.64
	3	37.41	37.37	66.77	52.99	38.88	23.39	64.71	46.84	41.26	28.86
LSTM	2	21.86	**27.19**	27.18	**21.39**	32.73	22.77	56.47	58.52	38.14	32.51
	3	39.17	46.23	55.69	47.25	45.99	32.63	37.99	48.68	47.53	42.79
ELM	1	20.06	29.96	20.43	23.74	**18.45**	26.02	24.45	**32.77**	**19.93**	**7.02**

Table 6. RMSE Single Step ahead prediction for NSE - BANKEX

Technique	Layers	Axis Bank	Bank of Baroda	Federal Bank	HDFC Bank	ICICI Bank	Indus Ind Bank	Kotak Mahindra	PNB	SBI	Yes Bank
CNN	2	0.1496	0.3092	0.1385	0.2791	0.1441	0.1285	**0.0758**	0.1713	0.1791	0.4453
	3	0.1764	0.2909	0.1914	0.2582	0.1472	0.1401	0.1133	0.1773	0.1516	0.4044
GRU	2	**0.1097**	**0.1262**	**0.082**	**0.1553**	**0.0943**	**0.105**	0.1219	0.09	**0.0957**	0.143
	3	0.1726	0.1482	0.1162	0.1368	0.1429	0.1409	0.1334	**0.0859**	0.1174	0.1859
LSTM	2	0.1291	0.1434	0.1361	0.1661	0.1011	0.1769	0.1182	0.1183	0.1126	0.2104
	3	0.1496	0.175	0.1789	0.1815	0.2137	0.1904	0.1641	0.13	0.119	0.2634
ELM	1	0.1561	0.2615	0.2621	0.1995	0.1164	0.1685	0.152	0.1538	0.12	**0.1053**

Table 7. RMSE Four Step ahead prediction for NSE - BANKEX

Technique	Layers	Axis Bank	Bank of Baroda	Federal Bank	HDFC Bank	ICICI Bank	Indus Ind Bank	Kotak Mahindra	PNB	SBI	Yes Bank
CNN	2	**0.1667**	0.3865	**0.2035**	0.2333	**0.1303**	0.1388	0.1363	0.1921	0.2061	0.4604
	3	0.1683	0.3336	0.2041	0.285	0.1452	0.2307	**0.126**	0.2401	0.1846	0.4582
GRU	2	0.316	0.2761	0.2668	0.2633	0.2833	**0.1098**	0.312	0.2197	0.2187	0.1793
	3	0.3476	0.2412	0.5195	0.2933	0.2608	0.2002	0.4488	0.1977	0.2087	0.2559
LSTM	2	0.1734	**0.194**	0.218	**0.1684**	0.1785	0.1375	0.277	0.2373	0.2129	0.2557
	3	0.2998	0.2362	0.3533	0.2954	0.252	0.1971	0.2435	0.1927	0.2513	0.3086
ELM	1	0.1789	0.2609	0.2571	0.2095	0.1321	0.1697	0.1526	**0.1769**	**0.1499**	**0.1255**

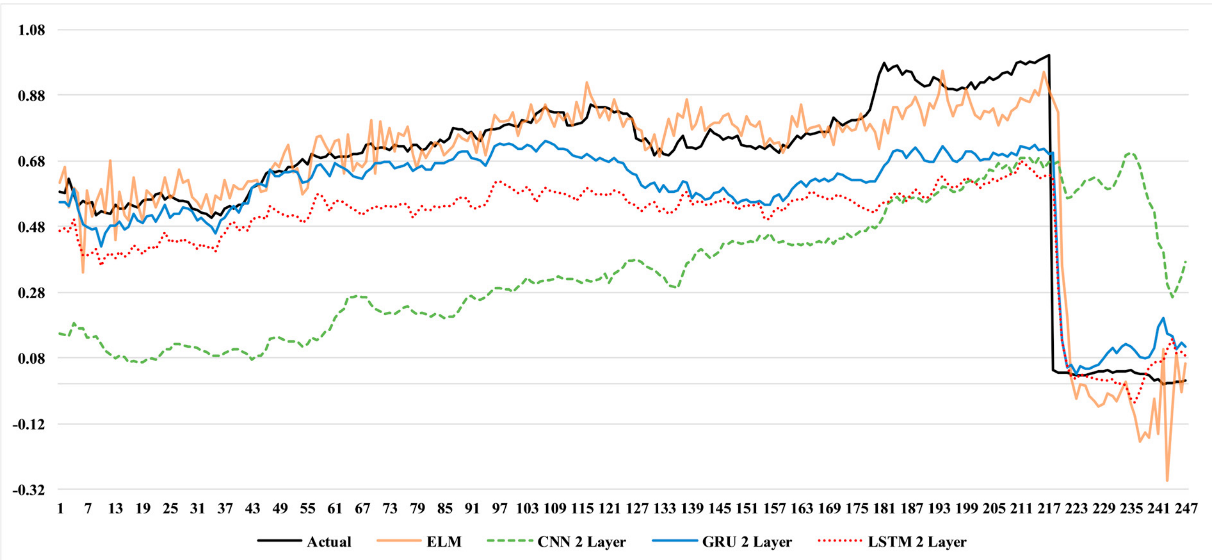

Fig. 2. One-day ahead out-sample forecasts for Yes Bank dataset

5. Conclusions

From the obtained results, it is observed that all the deep learning models are able to generate good forecast accuracy. The MdAPE values as low as 7.02% is observed. The proposed systems are also seen to be capable of generating correct forecast of the direction of stock price movements up to a maximum of 71.95% of the time. It is found that GRU based models provides better DA for shorter forecast horizon and ELM based models for longer forecast horizon. Similarly, GRU based models give the lowest MdAPE values for shorter forecast horizon and ELM based models give the least MdAPE values for longer forecast horizon. Hence, it can be stated that deep learning-based models are well suited for generating accurate forecasts for financial time series.

References

[1] A. Cowles 3rd, Can stock market forecasters forecast?, Econometrica: Journal of the Econometric Society (1933) 309–324.
[2] A. Cowles, Stock market forecasting, Econometrica, Journal of the Econometric Society (1944) 206–214.
[3] E. F. Fama, The behavior of stock-market prices, The journal of Business 38 (1) (1965) 34–105.
[4] E. F. Fama, Random walks in stock market prices, Financial analysts journal 51 (1) (1995) 75–80.
[5] B. B. Nair, P. S. Kumar, N. Sakthivel, U. Vipin, Clustering stock price time series data to generate stock trading recommendations: An empirical study, Expert Systems with Applications 70 (2017) 20–36.
[6] B. B. Nair, V. Mohandas, Artificial intelligence applications in financial forecasting–a survey and some empirical results, Intelligent Decision Technologies 9 (2) (2015) 99–140.
[7] B. B. Nair, V. Mohandas, An intelligent recommender system for stock trading, Intelligent Decision Technologies 9 (3) (2015) 243–269.
[8] B. B. Nair, V. Mohandas, N. Nayanar, E. Teja, S. Vigneshwari, K. Teja, A stock trading recommender system based on temporal association rule mining, SAGE Open 5 (2) (2015) 2158244015579941.
[9] G. Huang, Z. Liu, K. Q. Weinberger, L. van der Maaten, Densely connected convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, Vol. 1, 2017, p. 3.
[10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[11] Index reach - s&p bse bankex, accessed : April 12, 2018.
URL https://www.bseindia.com/indices/indiceswatch.aspx?index_Code=53&iname=BANKEX#about
[12] Yahoo finance - business finance, stock market, quotes, news, accessed : April 12, 2018.
URL https://in.finance.yahoo.com/
[13] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555.
[14] P. Malhotra, L. Vig, G. Shroff, P. Agarwal, Long short term memory networks for anomaly detection in time series, in: Proceedings, Presses universitaires de Louvain, 2015, p. 89.
[15] N. Laptev, J. Yosinski, L. E. Li, S. Smyl, Time-series extreme event forecasting with neural networks at uber, in: International Conference on Machine Learning, 2017.
[16] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, Recurrent neural networks for multivariate time series with missing values, Scientific reports 8 (1) (2018) 6085.
[17] K. Bandara, C. Bergmeir, S. Smyl, Forecasting across time series databases using long short-term memory networks on groups of similar series, arXiv preprint arXiv:1710.03222.
[18] D. Suryani, Convolutional neural network.
[19] M. Dalto, J. Matuško, M. Vašak, Deep neural networks for ultra-short-term wind forecasting, in: Industrial Technology (ICIT), 2015 IEEE International Conference on, IEEE, 2015, pp. 1657–1663.
[20] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, Vol. 2, IEEE, 2004, pp. 985–990.
[21] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1-3) (2006) 489–501.
[22] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42 (2) (2012) 513–529.
[23] G.-B. Huang, An insight into extreme learning machines: random neurons, random features and kernels, Cognitive Computation 6 (3) (2014) 376–390.
[24] A. J. Balaji, D. H. Ram, B. B. Nair, Machine learning approaches to electricity consumption forecasting in automated metering infrastructure (ami) systems: An empirical study, in: Computer Science On-line Conference, Springer, 2017, pp. 254–263.
[25] J. G. De Gooijer, R. J. Hyndman, 25 years of time series forecasting, International journal of forecasting 22 (3) (2006) 443–473.
[26] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, S.-P. Guo, Stock index forecasting based on a hybrid model, Omega 40 (6) (2012) 758–766.