



Cross Entropy Based Long Short Term Memory Recurrent Neural Network Model for Analyzing the Time Series on Stock Market Price

Dwaram Jayanarayana Reddy^{1*} **Akepogu David Donald¹** **Kothapalli Seshadri Ramana¹**
Karanam Srividya Lakshmi¹ **Pullala Cheruvu Sai Divya¹**

¹*G Pullaiah College of Engineering and Technology (Autonomous),
Department of Computer Science and Engineering, Kurnool, India*

* Corresponding author's Email: djnreddy@gmail.com

Abstract: In the financial stock market, a sequence of prices obtained from the share market with respect to the time series is usually examined. Generally, time series in finance, particularly shows importance in predicting investment in today's share market. Since there are too many factors such as public opinions, general economic conditions, or political events, vulnerability in the economy are directly or indirectly reflects on the evolution of financial time series. The desire of the investor is to predict the future stock prices neglecting whether the investor is a long-term investor or a day-trader. A major challenge is to develop and design an efficient predictive model that guides investors to make appropriate decisions. In this research work, Long Short Term Memory-Recurrent Neural Network (LSTM-RNN) is developed to overcome such disputes and contributing an efficient technique for predicting the future stock prices financially. In addition to the model, cross entropy is calculated using a Mutual Information feature selection model to minimize the optimization problems that create the time complexity in the system. The proposed LSTM-RNN has achieved best accuracy of 61.33% of prediction accuracy compared to state-of-the-art techniques.

Keywords: Cross entropy, Global financial crisis, Long short term model, Recurrent neural network, Stock market.

1. Introduction

The stock market is a place where the aggregation of both buyers and sellers happens in a single platform for offering shares to the general public. Meanwhile, the capital is raised on products needed for expansion of new operations [1]. During the past two decades, the stock market has been advanced as the main form of investment in numerous organizations as well as individuals for arranging huge investment funds [2]. As a result, many companies have been listed in stock markets around the world and investing a huge amount of their capital regularly. The time series analysis and topic modelling have been used in various applications such as the environment, the economy, health, and politics, even the social media. An overview of time series analysis of social media in different settings and focus areas were provided [3]. The traditional financial theory is a foundation for deciding the

effective market system that shows the investors are fully rational. The stock price in turn reflects all available information precisely at any time. The mood factors affect the judgment of investors and behaviour of investors that impact on the stock price in great demand. The investors sometimes overreact to good news when they are in a good mood or for bad news. Thus, investors tend to buy more or sell less stocks when they are in a good mood than they are in a bad mood, which causes the abnormal change in stock price [4]. In order to reduce the dimensionality of time series in univariate data, an Asynchronism Principal Component Analysis (APCA) was developed based on Dynamic Time Warping (DTW). ARIMA (Auto Regressive Integrated Moving Average) model was developed for the prediction of stock market movement. The Univariate time series models reduce the range of economical phenomena through the historical behaviour of a dependent variable. The accuracy

decrease as more feature causes more computing cost and thus achieves poor accuracy in predicting the price. The proposed RNN-LSTM handles financial time series and overcomes such limitation. The results showed that the RNN-LSTM handle financial time series data better than traditional time series prediction method. The advantage of the proposed model is that, the design of the model requires lowest computation cost and less computation, obtains highest predictive accuracy.

This research paper is arranged as follows. In section 2, many research papers on time data analysis using stock market data are reviewed. Detailed explanation about the proposed system is given in section 3. In addition, section 4 illustrates about the evaluation of the results and comparative analysis of the proposed system. The conclusion is made in section 5.

2. Literature works

Several methods have been developed by researchers in analysing the time series of stock market data, weather, etc. In this segment, a review of important contributions to the existing literature is presented for stock market data.

W. Deng and G. Wang [5] analysed a framework on weakly Dissolve Oxygen time-series data collected from the river Yangtze of five intensive stations. The developed method utilized Two-Dimensional Normal Cloud Representation (2D-NCR) which is time-series instance dataset. To analyse the water quality using the time series, the process of clustering and classification were integrated into the framework. The contribution of the developed method included to generate a water quality time series data based on data mining. The similarity search and similarity matrix discovered the water quality by generating the pattern. The results obtained from the developed method showed that the frame work was feasible and an efficient method to mine the hidden time series data for predicting water quality. However, an efficient time-series data mining methods are needed into the framework as well as the extension of the application.

W. Deng, G. Wang, and J. Xu [6] developed a novel time series representation model such as 2D-NCR based on the cloud model theory. The developed method was evaluated for 20 time-series datasets using 2D-NCR. The developed method reflected the distribution of each segment with the variations of the time. The developed method also addresses the uncertainty in the series of time data. The experimental results were validated for 2D-NCR and classification was performed using clustering.

The new time series data mining showed better results compared to dimensionality reduction methods. The developed method showed a better performance and suggested improvement in the segmentation process. The developed model could have used six numerical characteristics instead of four numerical characteristics for better results.

M. Wen, P. Li, L. Zhang, and Y. Chen [7] developed a novel technique for predicting time series in the field of finance by reconstructing time series through higher order structures. The data set included in the proposed method are the Standard & Poor's 500 stock (S&P 500) index and the individual stocks like Google Inc. (GOOGL), International Business Machines Corporation (IBM), The Boeing Company (BA), etc. A new method was used to simplify noisy-filled financial temporal series through reconstructing the leveraging motifs (frequent patterns) and used Convolutional Neural Network (CNN) to release the time series of spatial structure. The developed model extracted useful raw financial series data and information from the developed CNN architecture that facilitates the prediction. Reconstructing of the time series used the motifs, where the important features were captured from the original data at the same time filters out low frequency noises. However, the window size was too large and in turn, calculating the time was too much.

D. T. Tran, A. Iosifidis, J. Kannianen, and M. Gabbouj [8] developed a new neural network layer type for multivariate time-series data analysis. The developed temporal attention mechanism improved the performance of the bilinear structure and enhanced interpretability. The developed layer leverages on the basis of bilinear projection and is augmented with a temporal attention mechanism. The theoretical analysis of the developed layer created complexity in comparison with existing attention mechanisms in the recurrent structure. The developed layer constituted lower memory and complexity in computation. A quantitative analysis of the attention patterns is generated during the training process for the patterns of interest.

F. Zhu, W. Quan, Z. Zheng, and S. Wan [9] developed a Bayesian Learning Method for Financial Time-Series Analysis. The datasets included the Standard & Poor's 500 stock (S&P 500) index. The developed model incorporates particle-filtering technology. The Markov Chain Monte Carlo (MCMC) algorithm sequentially updates the estimation of a commonly-used time-series asymmetric GARCH model. The sequential Bayesian analysis method analysed complex models, such as, interest rate term structure models, stochastic volatility models, and multiple factor models. The

variance terms became complex when the long memory characteristics characterized asymmetry data. The parameter estimation of a complex model using the Bayesian method needs further improvement.

T. Fischer, and C. Krauss [10] developed a deep learning with LSTM networks for financial market predictions. The developed model was applied to a financial market with large-scale using LSTM networks for the prediction task on the S&P. From the regression analysis, the developed LSTM was prior to transaction costs, obtained highest predictable accuracy and lowered the risk exposure compared to existing models such as Random Forest. However, LSTM has the maximum drawdown of 52.33 percent compared to all other models and the general market.

C. Stolojescu [11] developed a forecasting WiMAX BS Traffic by Statistical Processing in the Wavelet Domain. The developed model considered two particular MIB (Management Information Base), outgoing (downlink), objects, and incoming (uplink). The predictions obtained accurate estimates on the components and was provided with minimal computational time. However, the volume of the data need to be reduced for the prediction of traffic more accurately.

3. Proposed method

An overview of the proposed method is shown in Fig. 1. The block diagram consists of collection of stock data, technical indicators, LSTM model, evaluation of the model.

3.1 Dataset

The Yahoo search engine data (Apple, Microsoft) of 1 year are collected for stock market predictions in

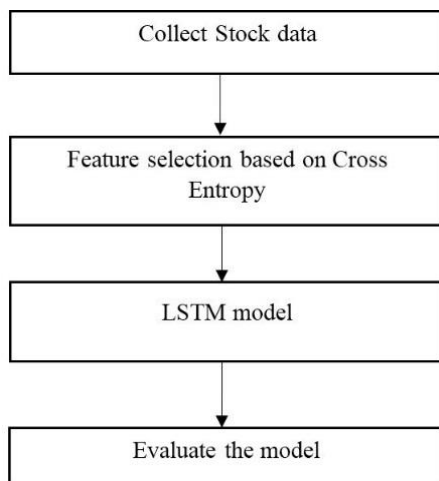


Figure.1 The block diagram of the proposed method

the proposed method. The daily information of each sample includes high price, low price, close price, opening price, and trading volume. The stock price of apple generated each week from the first week of September 2007 to the end of August 2012 is extracted from the Yahoo finance website [12]. This data set contains the high, open, close, low and adjusted close prices of apple stock on every Monday throughout (2007-2012). The dataset contains trading volume values as well. To achieve consistency, close prices are used as a general measure of stock price of Apple. The Yahoo finance news data starts at a point and set to one month earlier than the stock price’s starting point to study the relationship between news at one time and stock price at a later time.

3.2 Feature selection based on cross entropy

In this work, a novel algorithm is presented to overcome the optimization problem, which is the foundation for Mutual Information feature selection method. The proposed method automatically estimated the value of k . Mutual Information feature selection algorithm with the Cross Entropy feature is used to filter out the set of features among the currently available features. Through a stochastic method, selection of efficiently minimal feature set $U \subseteq X$ minimizes the conditional entropy between U and the class attribute y are utilized. The quality of the method is presented when the promising results are obtained on standard real datasets. All the estimated variables are calculated through the optimized algorithms and are not approximated. The candidate feature was not evaluated with every individual and the pair wise was evaluated with respect to the selection of feature subset. Finally, the algorithm autonomously estimated the dimension of the set of selected features.

The optimization problem is represented in Eq. (1) as follows,

$$\min S(x): X \in R^n \rightarrow R \tag{1}$$

The above equation is estimated as probability distribution problem. Which is randomized according to a family of probability density functions defined on, and the auxiliary problem is obtained and represented in Eq. (2).

$$l(\gamma) = P_u (S(X) \leq \gamma) = E_u [I_{\{S(X) \leq \gamma\}}] \tag{2}$$

Where E is expectant operator and in certain level γ , I is the indicator function, whose value is one if $S(X) \leq \gamma$ and zero for otherwise. In order to reduce the number of samples, cross entropy uses the

importance sampling method, which is represented in Eq. (3).

$$l(\gamma) = \frac{1}{N} \sum_{i=1}^N I_{\{S(x^i \leq \gamma)\}} \log f(x^i; v) \quad (3)$$

Where x^i is a random sample from $f(x; v)$ with sampling density $g(x)$. The optimal importance sampling density and the Kullback-Leibler distance is introduced to Calculate the distance between two densities, the cross entropy, and minimize the cross entropy as represented in Eq. (4),

$$g(x^*) = \min \frac{1}{N} \sum_{i=1}^N I_{\{S(x^i \leq \gamma)\}} \log f(x^i; v) \quad (4)$$

Consequently, optimal density $g(x^*)$ is obtained by solving the Eq. (4).

The optimized values are obtained from the cross entropy and then utilized for the classification to predict the prices for the stock market.

3.3 Classification using LSTM-RNN classifier

In an existing online algorithm, an optimization process is stopped in the middle of a learning model. To overcome the problem in the existing method, the novel LSTM-RNN classifier model is proposed. This is particularly useful for very large data sets like stock price data and the convergence is measured and learned quite early. The online learning model is ideally used for stochastic learning paradigm as the model and is trained on every data point. Each parameter update uses a randomly chosen data point, and the oscillations observed with the weights converge on the same optimal value. This section presents an LSTM-RNN classifier to address the limitations of long-term dependencies, a linear self-loop memory cell that showed gradient values through the long sequences were allowed. The memory cell is moderated through the amount of information flow in and out from the cell. Instead of a simple RNN unit, an LSTM unit has a memory cell that has state $c(t) \in R^K$ at time t . Through the memory cell, the information is flowed and controlled by three gates such as an input gate secondly a forget gate and finally an output gate. The input gate $x(t) \in R^K$ controls the flow of inputs into the cell. Forgetting memory cell is controlled by the $f(t) \in R^K$ forget gate and the output gate $o(t) \in R^K$ modifies the output flow from the memory cell. The element wise sigmoid function of a vector by σ and the element wise product of two vectors by the three gated are all the sigmoid units that sets each and every element gate value from 0 and 1.

A special subset of RNNs captured a context-specific temporal dependency for long terms of period. Each memory cell in the LSTM neuron stored other information that maintains cell state of its own. Whereas, the neurons in normal RNNs merely consider their previous hidden state and the present output input state to a new hidden state. The LSTM neuron considers its old cell state and the outputs of its new cell state. Block diagram of the proposed LSTM is shown in Fig. 2. The LSTM consists of Output gate, input gate and Forget gate.

A. Forget gate: The forget gate decides when specific portions of the cell state are to be replaced with more recent information. It's outputs values, close to 1 of the cell state should be retained, and zero for values that should be neglected.

B. Input gate: Based on the input output $o(t - 1)$, input $x(t)$, and previous cell state $c(t - 1)$, the network learns the conditions under any information that should be stored or updated in the cell state.

C. Output gate: Depending on the input and cell state, output gate decides which information is propagated forward where the output $o(t)$ and cell state $c(t)$ to the next node in the network. Thus, for exploring the variations in one stock's price, an ideal LSTM RNN affects the prices of other stocks over a long term of time. The dynamic fashion for long information about specific past trends in stock price movement needs to retain property in order to predict the future scenarios in the variation of stock prices. The results obtained for the proposed LSTM-RNN model with a calculation of Cross Entropy are evaluated and discussed in the next section.

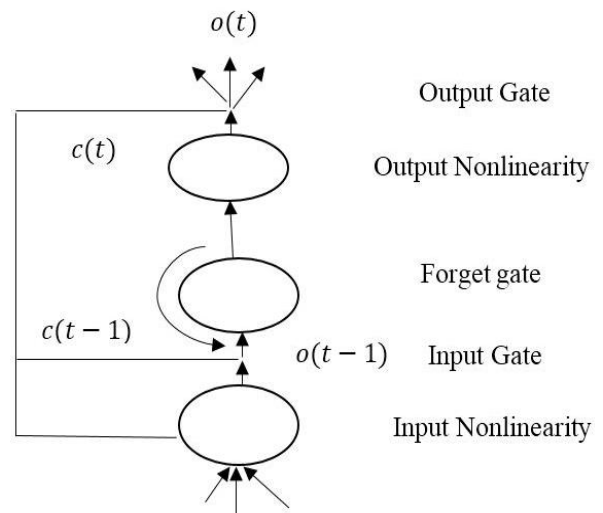


Figure.2 Block diagram of the proposed LSTM –RNN method

These three layers plays an important role in activation function. The activation function decides whether the neuron should be activated or deactivated. The activation function does the non-linear transformation to the input making it capable to perform more complex tasks. The activation function is expressed in the following Eq. (5) and (6)

$$o(1) = W(1)c + b(1) \tag{5}$$

$$a(1) = z(1) \tag{6}$$

Where,
 $z(1)$ is the vectorized output of the input layer,
 $W(1)$ be the vectorized weights assigned to neurons of hidden layer i.e. w_1, w_2, w_3 and w_4
 X be the vectorized input features
 b is the vectorized bias assigned to neurons in hidden layer like b_1, b_2 etc.,
 $a(1)$ is the vectorized form of any linear function

Thus the activation function is capable to solve linear and non-linear transformations and perform complex tasks. The Bidirectional LSTM architecture is shown in Table 1. Based on the mentioned parameters RNN is designed.

4. Results and discussion

To explore whether it is possible to learn useful information from the first day’s price series before predicting the up and down compared to the market price on last day, an extensive value is tabulated on the financial datasets and compare the proposed method with several baseline methods. In this resultant section, the proposed system is simulated by using python 3, Intel i5 processor, and 4 GB RAM. The proposed system uses the Stock Market datasets for evaluating the results.

Table 1. Bi-directional LSTM architecture

Type	Output shape	Number of parameters
Input	100,1	0
LSTM(forward)	100,400	161600
LSTM(backward)	100,400	161600
Global ID max pooling	400	0
Fully connected Rectified Linear Unit (ReLU)	50	20050
Dropout	50	0
Fully connected (Sigmoid)	1	51

4.1 Experimental data

The datasets include the Standard & Poor’s 500 stock (S&P 500) index and its individual stocks such as: Google Inc. (GOOGL), International Business Machines Corporation (IBM), The Walt Disney Company (DIS), The Goodyear Tire & Rubber Company (GT), Apple Inc. (APPL). Each financial time series features such as a number of variables, Open, High, Low, Close, Adjusted Close, and Volume is evaluated. The experimental results showed that close variables are the research object. For each stock price series, the true label for each day’s up and down is marked according to the following rule is shown in the Eq. (7).

$$y_i = \begin{cases} 1 & X_i \leq X_{i+1} \\ 0 & X_i > X_{i+1} \end{cases} \tag{7}$$

Where y_i denotes the up or down, X_i is Closed value of each stock on i^{th} day and X_{i+1} is Closed value of each stock on $i + 1^{th}$ day. The confusion matrix evaluated for classification predictions are shown in Table 2.

4.2 Evaluation

Generally, the prediction of the trend of stock prices can be regarded as a two-category problem. Therefore, the prediction results fall into one of four cases based on the consistency of their real class labels and the predicted labels as True Positive, False Positive, True Negative, Or False Negative. These measures are denoted as TP, FP, TN, and FN the number of corresponding samples, respectively, the confusion matrix for the classification result has the following form. The standard performance measures are Accuracy, Recall, Precision, F1 score to evaluate the performance of individual stock prediction. These scores are calculated as follows:

Accuracy: The accuracy is defined as the degree of closeness of a calculated or measure value to its original value. Accuracy is the most important performance measure as it is simply a ratio of properly predicted observation to the total number of observations. The expression for accuracy, is given in Eq. (8).

Table 2. Confusion matrix of classification results

True Value	Forecast Result	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Table 3. Comparison of the existing and the proposed method for the performance measures Accuracy(Acc) and Recall (Rec)

	Data	S&P 500		APPL		GOOGL		IBM		DIS		GT	
	Metrics	Acc	Rec	Acc	Rec	Acc	Rec	Acc	Rec	Acc	Rec	Acc	Rec
METH ODS	Wavelet	48.77	49.52	51.83	54.48	51.80	53.49	51.18	52.26	51.43	53.49	49.67	49.27
	LSTM	55.88	71.91	54.34	77.08	56.25	84.21	62.85	63.34	56.11	27.53	52.50	39.21
	LSTM-RNN	61.33	74.56	58.9	78.97	58.65	86.43	68.76	65.76	59.65	54.43	56.75	53.51

Table 4. Comparison of the existing and the proposed method for the performance measures Precision (Pre) and F1 score (F1)

	Data	S&P 500		APPL		GOOGL		IBM		DIS		GT	
	Metrics	Pre	F1	Pre	F1	Pre	F1	Pre	F1	Pre	F1	Pre	F1
METH ODS	Wavelet	52.41	50.92	53.55	54.01	49.46	51.39	53.09	52.67	55.67	54.56	55.14	52.03
	LSTM	56.14	63.05	54.41	63.79	59.25	69.56	68.80	58.60	63.33	38.38	43.47	41.23
	LSTM-RNN	61.56	65.64	61.32	66.63	65.41	73.43	72.2	59.76	67.65	59.34	59.32	54.89

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{8}$$

Precision: The precision is a measure of closeness of two or more measurements of each other. Precision is the ratio of True Positive to the sum of the False Negative and True Positive. The expression of precision is given in Eq. (9).

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

Recall: The recall is the ability of the classifier to determine all the positive samples. The worst value is 0 and best value is 1. Recall can be defined as the ratio of true positive to the sum of True Positive and False Negative, which represented in Eq. (10).

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

F1 score: F1 Score is the weighted average of Recall and Precision. Therefore, the F1 score takes both False Negatives and False Positives into account. The expression for F1 score is derived in Eq. (11).

$$F1\ Score = \frac{2TP}{2TP+FP+FN} \tag{11}$$

Where TP is the number of samples in which the positive class sample is predicted to be a positive class, and TN is the number of samples in which the negative class sample is predicted to be a negative class, the denominator is the number of all test sample.

True Positive: True positive is a result where the model correctly predicts the positive class.

True Negative: True negative is a result where the model correctly predicts the negative class.

False Positive: False positive is a result where the model incorrectly predicts the positive class.

False Negative: False negative is a result where the model incorrectly predicts the negative class.

4.3 Comparative analysis

By extracting trend features from the reconstructed sequence, the developed method improved the accuracy of the prediction results on almost all the datasets. From Tables 3 and 4, LSTM is a competitive baseline considering higher Recall and Precision scores on some datasets. However, LSTM is essentially time consuming than LSTM-RNN in the training period due to the intrinsic vanishing gradient problem. It is also shown in both Tables 3 and 4 that, traditional time series analysis method and signal processing method like Wavelet method was inferior to machine learning based methods LSTM which are 4% - 6% less in prediction accuracy and more less than the deep learning model. This reveals the limitations of traditional financial time series prediction methods on nonlinear time series modelling. Overall, the proposed LSTM-RNN model is superior to the baseline methods. With respect to Accuracy and Recall traditional signal processing methods, the proposed LSTM-RNN attained 7% to 8% higher values respectively. Although F1 scores do not achieve the best performance on all datasets, they are close to the best results. Table 2 shows the comparison results of the existing and the proposed method for the performance measure's accuracy and recall.

Similarly, Table 3 shows the comparison of the existing and the proposed method comparison values like precision and F1 score.

Fig. 3 and Fig. 4 shows the comparison of the graphs for the proposed method and the existing methods. The values obtained from T. Fischer and C. Krauss [11] for LSTM and Wavelet techniques evaluated by C. Stojescu, A. Cusnir, S. Moga, and A. Isar [12] are compared with the proposed LSTM-RNN model, with various financial markets. The

developed models used a number of features that caused gradient vanishing and data redundancy. The accuracy decrease as more feature causes more computing cost. The proposed model overcomes such limitation and achieves best prediction accuracy compared to the existing models. The results showed that the Recurrent Neuron Network with LSTM (Long Short-term Memory) handle financial time series data better than traditional time series prediction method.

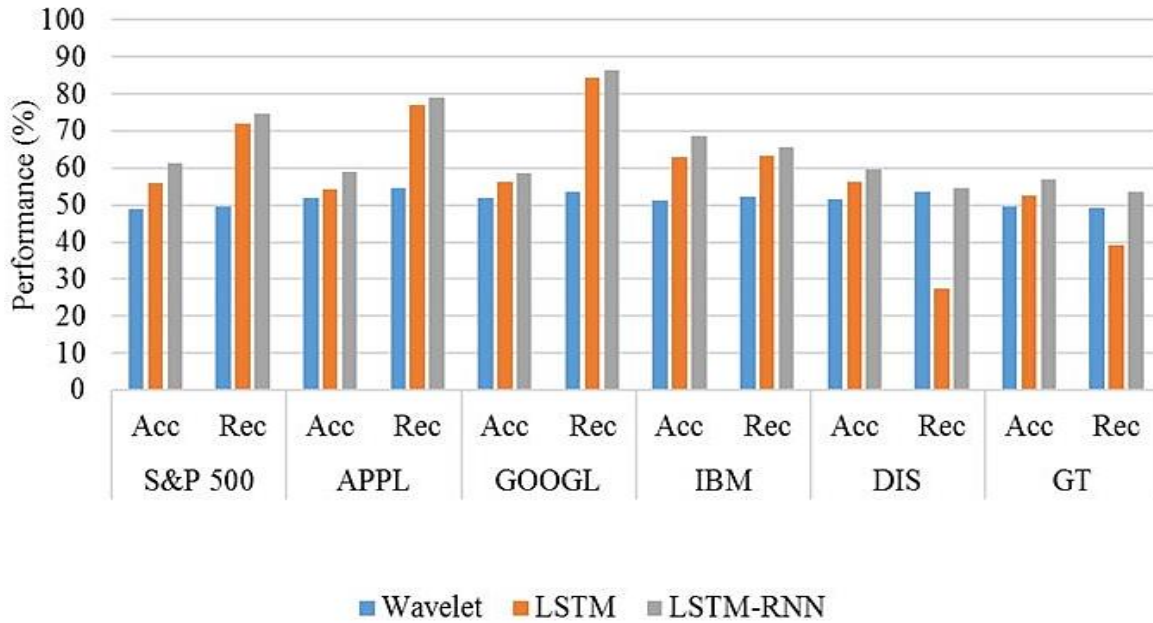


Figure.3 Comparison graph of the existing and the proposed method for the performance measures Accuracy (Acc) and Recall (Rec)

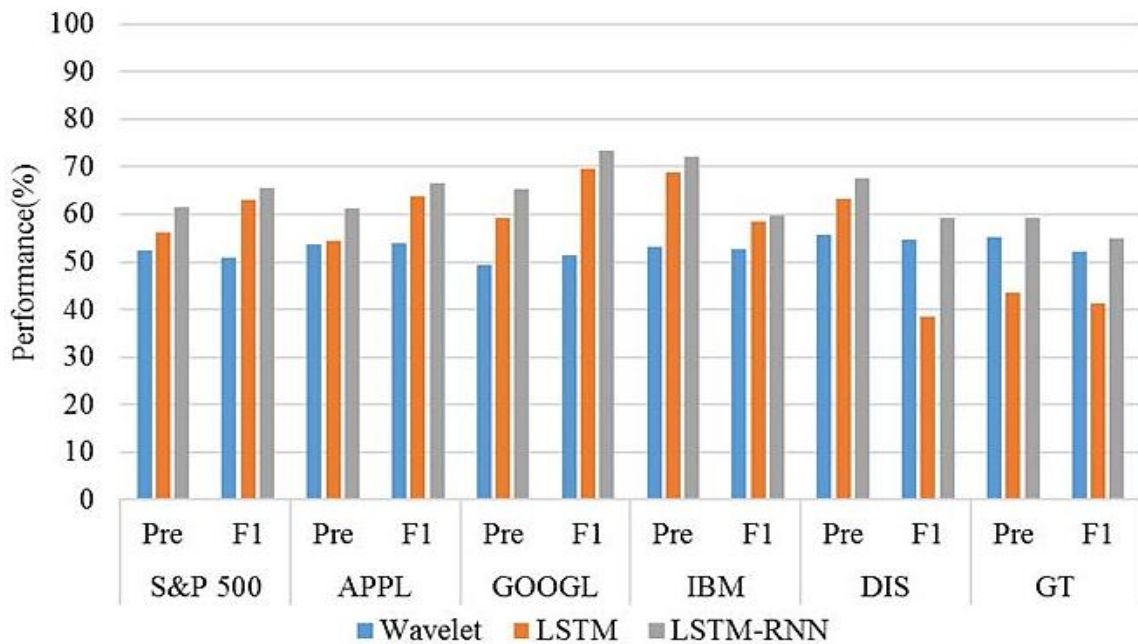


Figure.4 Comparison graph of the existing and the proposed method for the performance measures Precision (Pre) and F1 score (F1)

5. Conclusion

In this research work, a novel technique for financial time series trend prediction by using a modified LSTM-RNN model is proposed. The RNNs are used to predict the ups and downs in the sale to collect useful information on the sale. The existing methods are compared with the LSTM, which utilizes RNNs as a sequential model and remarkably more demanding in terms of complexity in the computation. The developed models used a number of features that caused gradient vanishing and data redundancy. On the other hand, the proposed LSTM-RNN overcame the computation problem, thereby evaluating results on real financial time series datasets that validated the effectiveness of capturing trend information of stock shares. The proposed Mutual Information feature selection model calculates the cross entropy values to minimize the optimization problems that reduce the time complexity in the system. The LSTM-RNN model obtained 61.33% of predictive accuracy in the financial time series and provided a novel solution for price prediction for a certain range of days. In the future, this research can be extended with more technical stock indicators for a different organization's stock analysis.

References

- [1] D. J. Reddy, R. Varaprasad, and N. Parashuram, "Secure Data Deduplication Using Improved Elliptic Cryptography with Associative Support and Load Balancing", *Journal of Advanced Research in Dynamical and Control Systems*, Vol.9, pp.553-563, 2009
- [2] N. F. Ibrahim and X. Wang, "Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media", *Computers in Human Behavior*, Vol.96, pp.32-45, 2019.
- [3] Q. Ruan, M. Zhang, D. Lv, and H. Yang, "SAD and stock returns revisited: Nonlinear analysis based on MF-DCCA and Granger test", *Physica A: Statistical Mechanics and its Applications*, Vol.509, pp.1009-1022, 2018.
- [4] S. M. Idrees, M. A. Alam, and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data", *IEEE Access*, Vol.7, pp.17287-17298, 2019.
- [5] W. Deng and G. Wang, "A novel water quality data analysis framework based on time-series data mining", *Journal of environmental management*, Vol.196, pp.365-375, 2017.
- [6] W. Deng, G. Wang, and J. Xu, "Piecewise two-dimensional normal cloud representation for time-series data mining", *Information Sciences*, Vol.374, pp.32-50, 2016.
- [7] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series", *IEEE Access*, Vol.7, pp.28299-28308, 2019.
- [8] D. T. Tran, A. Iosifidis, J. Kannianen, and M. Gabbouj, "Temporal attention-augmented bilinear network for financial time-series data analysis", *IEEE Transactions on Neural Networks and Learning Systems*, Vol.30, No.5, pp.1407-1418, 2018.
- [9] F. Zhu, W. Quan, Z. Zheng, and S. Wan, "A bayesian learning method for financial time-series analysis", *IEEE Access*, Vol.6, pp.38959-38966, 2018.
- [10] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions", *European Journal of Operational Research*, Vol.270, No.2, pp.654-669, 2018.
- [11] C. Stojescu, A. Cusnir, S. Moga, and A. Isar, "Forecasting WiMAX BS traffic by statistical processing in the wavelet domain", *IEEE International Symposium on Signals, Circuits and Systems*, pp.1-4, 2009.
- [12] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series", *IEEE Access*, Vol.7, pp.28299-28308, 2019.