Nat. Volatiles & Essent. Oils, 2021; 8(6): 5132-5151

Stock Market Prediction Using Principal Component Analysis, Iterative Filtering And Ensemble Deep NeuralNetworks

[1]Er. Amanpreet Singh , Prof. (Dr.) [2]R.K. Bathla

[1,2]department Of Computer Science and Engineering, Desh Bhagat University, India

Abstract: Stock market behavior is extremely volatile and complex in nature due to the randomness of the governing parameters. This makes attaining a high degree of accuracy in stock market forecasting extremely challenging. This paper presents a combination of discrete wavelet transform and gradient boosting approach for stock market prediction. A recursive wavelet decomposition approach is employed to filter the noisy data and subsequently a gradient boost approach is applied to different co-efficient values of the wavelet transform. Principal component Analysis has also been used for dimensional reduction. A composite predicted output them computed from the individual neural networks trained with different wavelet co-efficient. The day-wise data for shares has been considered with opening prices, closing prices, average price and volume as the temporal parameters. The performance evaluation has been done based on the mean absolute percentageerror, mean square error, accuracy of prediction and regression values. It has been shown that the proposed approach and attains an average accuracy of 88.97% and a regression value of 0.99 for multiple benchmark datasets. A comparative analysis of the proposed technique shows that the method outperforms existing baseline techniques.

Keywords—Time Series Forecasting, Stock Market Movement, Wavelet Transform, Multi-level Decomposition, Gradient Boost, Gradinet Boosting, Accuracy, Regression.

1. Introduction

Stock markets have been a vital area of research with respect to economy. The evaluation of stock prices and stock markets are important to understand the changing dynamics of market investments. It is crucial for the investors and policy makers to predict the stock prices and gaugethe movement of stock values. It is essential for investment planning and investment-based profits. Stock investment is a substantial financial activity that can lead to huge profits or losses. Stock market predictions based on the stock price prediction are a key aspect for share market analysisand planning Attigeri et al. (2015). The volatile nature of the share market is a major challenge that needs to be focused. Stock market prediction is of paramount importance in this context, which aimsat predicting the future stock values Baek et al. (2018). Accuracy in prediction of stock market prices can aid in creating profitable investment strategies. A robust prediction model can help in forecasting accurate stock values. There have been rapid developments in this area of research with the use of artificial intelligence and machine learning methods Bansal et al. (2019). The domain of artificial intelligence has helped in understanding and analyzing the financial markets in an effective manner. There has been a widespread increase in the use of machine learning backed methods for efficient evaluation of financial markets. The technical analysis in prediction of stock market investments is based on past data of share market can be done using efficient machine learning based methods Billah et al. (2016). This domain has been an active area of research for accurate prediction of stock values based on machine learning based methods. While different machine learning approaches have been explored to forecast stock market behavior accurately, the non- linearity, extremely high randomness and non-stationarity of the noisy data sets renders inaccuracies in forecasting for conventional machine learning algorithms used for time series forecasting problems Bouktif et al. (2020). The paper aims at devising a technique for stock market forecasting employing statistical data filtering coupled with machine learning. For that purpose, the wavelet transform is utilized, as a transform domain iterative filter along with principal component analysis for data optimization. The gradient descent approach is employed subsequently for training Cakra et al. (2015). The remaining section of the paper is organized as follows: Section II presents theconventional techniques in the domain of the proposed work. Section III discusses the data processing and training approach. Section IV discusses the results obtained using the proposed approach. Section V makes the concluding remarks regarding the findings.

2. Related Work

The concept of transfer entropy-based machine learning was proposed and implemented by Kim et al. (2020) which tried to use the time varying effective transfer entropy (ETE) in conjugation with different machine learning algorithms. The approach used the ETE as an exogenous input for the logistic regression, LSTM, and MLP algor ithms. It was shown that the ETE acted as an effective feature for time series forecasting. Bouktif et al. (2020) proposed the use of texture-based features along with temporal data for stock market forecasting. This approach tries to incorporate public opinion regarding stock prices and market movement. Sentiment analysis and its possible impacton stock market movement was studied by Li et al. (2020). It as shown that the correlation between the sentiment analysis and stock market movement can be used for forecasting closing values of stock prices. Porshnev et al. (2013) utilized Twitter data for estimating public sentiments pertaining to stock prices and coupled it with the historical data to generate the training vector. A similar approach was proposed by Smailović et al. (2013) to utilize the knowledge discovery process for stock market forecasting.The long short term memory (LSTM) neural networks have beenextensively used for time series forecasting problems. Eapen et al. (2019) have used a bidirectional and conventional LSTM network respectively for stock market forecasting, It as shown by Baek etal. (2018) that the LSTM can be prone to overfitting which may result in the network being stuck in local minima, which needs to be avoided. It was shown that the LSTM was able to track the more recent trends in the data as compared to other deep learning techniques such as the convolutional neural networks (CNNs). Support Vector Regression (SVR) has also been used extensively for stock market forecasting. SVR has been employed by Guo et al. (2018). An adaptive SVR-Wavelet model was proposed by Raimundo et al. (2018) where the Wavelet transform is used as a data processing tool. Recurrent Deep Neural Networks (RDNNs) were used by Yoshihara et al. (2014). A comparative analysis of RNN, LSTM and Sliding CNN was proposed by Selvin et al. (2017) Fuzzy logicand expert systems were employed for stock market forecasting by Sadaei et al. (2016). Adaptive Neuro Fuzzy Inference Systems (ANFIS) was proposed and implemented by Lincy et al. (2016) for time series stock market forecasting. Another approach which was extensively used was the linear regression, logistic regression and Support Vector Regression for stock market movement estimation and forecasting problems by Devi et al. (2015). The use of sentiment analysis and opinion mining was proposed by Cakara et al. (2015) and Li et al. (2015). The use of hybridtechniques for forecasting was proposed by Patel et al. (2015). Neural networks along with its different variants were used for stock market prediction as discussed by Sim et al. (2013). Recurrent Networks were used for prediction by Yoshihara et al. (2014). The common performance metrics for the evaluation of the system were the mean square error (MSE), mean absolute error (MAE), normalized root mean squared error (NRMSE), regression (R) and accuracy.

3. Methodology

The proposed methodology tries to employ techniques for data pre-processing to remove noise effects from the data as well as feed the machine learning model with an exogenous input in terms of recent moving average so as to facilitate in recent pattern analysis. Mostly large stock datasets are inherently noisy in nature due to the near random movement of the prices based on several financial and socio-economic factors. In the technique developed, a major focus has been on processing the data to alleviate the effects of noisy fluctuations in the data which makes it extremely challenging for machine learning models to find the intrinsic patterns in the data. The data processing has been proposed employing a two-fold approach of noise filtering and dimensional reduction. For the purpose, the wavelet transform and the principal component analysis have been planned to be employed and fine-tuned for the data model. Such an approach helps to understand the retention of the actual data patterns in the data over a long span while removing the noisy part of the data. This is a novel approach which leads to better training of the employed machine learning model. The extensive statistical analysis helps to gain insight into the correctness of the data being used to train the machine learning model. A moving average is then proposed to be an exogenous input to the learning model. The moving average would allow the machine learning algorithm to find recent patterns in the data along with overall historical data. The final data vector would thus contain the DWT coefficients of the decomposition along with the moving average exogenous input. Gradient Boosting is then proposed to compute the final prediction output as a summation of the outputs from the individual learning models. The mathematical modelling of the proposed approach is described subsequently. To attain high accuracy of forecasting, a combination of PCA, Wavelet Transform and Gradient Descent (GD) based approaches has been proposed.

3.1 Data processing

The data utilized for the present experimental study is the SBI, Infosys and Reliance stock prices listed in the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE), India. The data preparation and processing techniques bolster the training algorithms for stock market forecasting. The techniques used in this paper are the Principal Component Analysis (PCA) and the Discrete Wavelet Transform (DWT) Walden (2001). The PCA has been used as data optimization and dimensional reduction tool for time series forecasting problems. In the context of this paper, it is particularly useful for segregate the fluctuating component of the data from the predictable component of the data, which is effective to filter out noisy data samples from the data samples of predictable coherent nature, Sim et al. (2013) The application of the PCA on the raw data allows to dimensionally reduce the data into a more coherent and compact data vector with lesser dimensions. This allows to lessening the perturbations in the data employing truncation of the dimensions which exhibit relatively lesser coherence Meesad et al. (2013). The data representationof the target vector can be given by:

$$T = T_{(t),}T_{(t-k),}..........T_{(t-NK)} \quad (1)$$

Here,

T is the composite target vector

t is the time variable

k is the delay or lag variable

N is the number of lags

The target vector is generally influenced by several governing parameters or features which are often non-linear and non-stationary in nature. The PCA can be used to generate a dimensionallyreduced set of the data vector and in the process removes the uncorrelated data variables inthe noise spectrum of the data. The filtered data can be expressed as:

$$T_t = P_r^{n,m} S_r^{n,m} \Delta_r \quad (2)$$

Here,

T_t is the truncated and dimensionally reduced data
$P_r^{n,m}$ vectoris the principal components
$S_r^{n,m}$ is the reduced singular
Δ_r matrixis the PCA matrix

The data is truncated and filtered to contain closely correlated data points and discard the uncorrelated data samples thereby rendering dynamic closeness to the data under consideration. Thus, the PCA helps to produce a time series data set by projecting the data statespace into matrix of singular vectors into a reduced principal component state space, given by:

$$T_n \overrightarrow{P_r^{n,m}} T_r^m \tag{3}$$

Here,
T_n is the n-dimensional data vector
T_r^m is the m dimensional reduced vector
$\overrightarrow{P_r^{n,m}}$ is the principal components for employing PCA

The next step is the decomposition of the PCA reduced data using the discrete wavelet transform. The discrete wavelet transform is a sampled version of the continuous wavelet transform which can be used for filtering time series data. The DWT acts as a multi-level filter, with even length scaling and shifting operations acting as high and low pass filtering operations. On contrary to the Fourier based methods, which can be applied to smooth signals and data patterns due to constraints of Dirichlet's conditions Raimundo et al. (2018). However, due to the non-smooth kernel functions, the wavelet transform is useful to analysis of non -smooth, non-stationary and abruptly changing signals and data [26]. The scaling operation satisfies:

$$\sum_{n=0}^{L-1} k_i^2 = 1 \tag{4}$$

$$\sum_{n=0}^{L-1} k_i^2 + 2n = \sum_{i=-\infty}^{\infty} k_i^2 + 2n = 0 \tag{5}$$

Here,
k_i^2 :l = 0………L-1 represents the kernel of the transform n represents the set of non-zero integers. The scaling co-efficient values of the DWT with a multi-level pyramidal filtering structure for level 'i' is given by:

$$S_{i,t} = \sum_{l=0}^{L-1} h_l X_{i-1,(2t+1-l) \bmod N_{i-1}} \tag{6}$$

The wavelet co-efficient values of the DWT with a multi-level pyramidal filtering structure for level 'i' is given by:

$$W_{i,t} = \sum_{l=0}^{L-1} g_l X_{i-1,(2t+1-l) \bmod N_{i-1}} \tag{7}$$

Here,
g_l and h_l are related as:

$$h_l = (-1)^l g_{L-l-1} \tag{8}$$

And

$$g_l = (-1)^{l+!} h_{L-l-1} \tag{9}$$

For l=0, ………, L-1

t denotes the time variable

X denotes the time dependent sequent to be transformed.

Another way of looking at the DWT is a recursive pyramidal filter which breaks down the function to be transformed into the approximate and detailed co-efficient values. While the detailed co-efficient values would contain the noisy disturbances, the approximate co-efficient values would contain the intrinsic patterns in the data except the noisy component. Thus, iteratively applying the DWT, discarding the detailed co-efficient values and retaining the approximate coefficient values would allow to filter out the noisy component of the data. While some data loss may be incurred in the process, but it would outweigh the detrimental effects of the disturbances in the data. Mathematically, it can be summed up as:

$$X_{(t)} \underrightarrow{S,W} Co-efficients(Approx, Detailed) \tag{10}$$

Here,

Approx. represents the approximate co-efficient values

Detailed represents the detailed co-efficient values

S and W are the scaling and wavelet filters of the DWT

X is the time domain samples of the data

3.2 Training

The raw data is structured and processed using the PCA and the DWT which yields the actual training data. The data features used in this study are date, previous day closing price, present day opening price, volume (swing), highest and lowest price of the day. The training algorithm employed here is the back propagation-gradient descent. Following a standard convention, 70% of the data is utilized for training the neural network and 30% is used for testing. The back propagation mechanism is chosen since it is found to yield relatively good results for temporal prediction problems. In this approach, the data vector is applied to the network to be train ed, and the error or objective function is computed. Recursive trainings or iterations are performed based on the following constraints:

3.2.1 The objective function is minimized and the values becomes stationary for multipleiterations consecutively.

3.2.2 The objective function doesn't attain stationarity but the maximum number of iterations defined have been reached.

Satisfying either of the two conditions results in the termination of the training process.The back propagation fundamentally utilizes the chain rule given by:

$$W_{k+1} = W_k - \alpha \frac{\partial^2 e}{\partial w} \tag{11}$$

Here,

W_{k+1} is the weight of the next iteration

W_k is the weight of the present

e iteratione is the error

α is the learning rate

$$\frac{\partial^2 e}{\partial w} = \frac{\partial^2 e}{\partial y}, \frac{\partial^2 y}{\partial w} \tag{12}$$

The chain rule in relation (12) can be used for computing the error gradient. Summarizingthe back propagation algorithm employed, the following steps are to be employed:

$$T^{l} = P \qquad (13)$$

Here,
P is the initial training vectorT is the feed-forward vector
The superscript ‘i’ represents the initial state of the network

Next, the propagation of the training vector through the network would yield:

$$T^{m+1} = f^{m+1}(W^{m+1}a^{m} + b^{m} + 1)\forall m \in 0,1......M-1 \qquad (14)$$

At each step, the following vector is updated:

$$a=a^{m} \qquad (15)$$

Here,
m is the iteration number f is the activation function
W is the weight of the networkb is the bias of the network

Subsequently, the sensitivities of the network are to be fed back to the next iteration as a composite training vector, given by [27]:

$$S^{M} = 2F^{M}n^{M}(t-a) \qquad (16)$$

Here,
S denotes the sensitivity
t denotes the time metric
a denotes the delay or lag after which the sensitivity is computed in the network.
Typically, the error of iteration ‘m’ along with the error gradients are utilized as the variables for the Sensitivity vector in feedback i.e.

$$e_{m}\frac{\partial^{2} y}{\partial x} \in S \qquad (17)$$

The values of the weight a bias is generally updated as:

$$W_{k} = W_{k-1} - \alpha S^{m} a^{m-1^{T}} \qquad (18)$$

$$b_{k}^{m} = b_{k-1}^{m} - \alpha S^{m} \qquad (19)$$

Here,
α denotes the learning rate
k denotes the subsequent iteration
k-1 denotes the previous iteration

In general, the steepest descent is the simplest to implement, but is slow in convergence. Hence, we employ the scaled conjugate variant of the steepest descent to speed up the convergence of the objective function. This needs the computation of a search vector for each iteration ‘k’ which would point to a scaled gradient with the maximum error change rate with respect to weights. It is given by:

$$g_k = \max\{g\}\forall W, \text{ e:at iteration k} \qquad (20)$$

A parallel search for the learning rate is carried out to minimize the objective function along the direction of search of the error gradient in each iteration. Mathematically, it is given by:

$$R_k = R_{k-1} = \alpha P_k \qquad (21)$$

Here,
R is the scaling step
α is the learning rate
P_k is the search vector for iteration 'k'

The scaled version of the steepest descent converges in 'n' iterations for a quadratic objective function typically the mean squared error. However, after the 'n' iterations are over, a repetitive cycle of scaled gradients is employed.
As data is fed to a neural network for pattern recognition, the weights keep updating. However, ithas been found that in case of time series problems, the latest data sample have the maximum impact on the latest output. Hence it is logical to calculate a moving average of latest (previous) dataand apply it to the neural network. This is also called a moving average. Mathematically,

$$I_k = X_{1,k}, Mean(X)_{k,k-n}, Y_k \qquad (22)$$

Here,
I_k is the kth input sample to the neural network
$X_{1,k}$ are the data samples from the first to the kth sample
$Mean(X)_{k,k-n}$ is the mean of the data samples form k-n to k, i.e. it is a moving average depending on the value of k
Y_k is the target

Thus, a moving average of the C_A *and* C_A values can be computed after the after the application of the PCA. The next step would be creating a new training vectorcomposing of the following variables:

$$T_r = [X1_{CA,CD}, X2_{CA,CD} ....Xn-1_{CA,CD}\, Xn_{CA,CD}, Avg_{n-k}, Y] \qquad (23)$$

Here,
T_r is the training vector,
Y is the target vector.
$X1_{CA,CD}, X2_{CA,CD} ....Xn-1_{CA,CD}\, Xn_{CA,CD}$ are the individual decomposed values of the features using the DWT iteratively.
Avg_{n-k} is the, moving average of the variables.

The moving average would allow the machine learning algorithm to find recent patterns in the data along with overall historical data. The final data vector would thus contain the DWT co-efficient of the decomposition along with the moving average exogenous input. Gradient Boosting is then proposed to compute the final prediction output as a summation of the outputs from the individual learning models.

The performance evaluation of the proposed model is done based on the evaluationof the following parameters:

1. Mean Absolute Percentage Error (MAPE)
2. Mean square error (MSE)

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\frac{|v_t - \hat{v}_t|}{v_t} \quad (24)$$

$$MSE = \frac{1}{N}\sum_{t=1}^{N} e_t^2 \quad (25)$$

Here,
N is the number of predicted samples
v is the predicted value
$\hat{v}_t$ is the actual value
e is the error value

The next section discusses the obtained results.

4. Experimental Results

The data has been extracted from https://in.finance.yahoo.com/quote. Three data sets have been used for the analysis of the proposed algorithm, which are the SBI, Infosys and Reliance. The feature selected are the date, opening price, closing price, mean price of the day, maximum price and minimum price. The day wise parameters are modelled to affect and gov ern the final share prices. The data processing and prediction methodologies are explained subsequently using the obtained results.

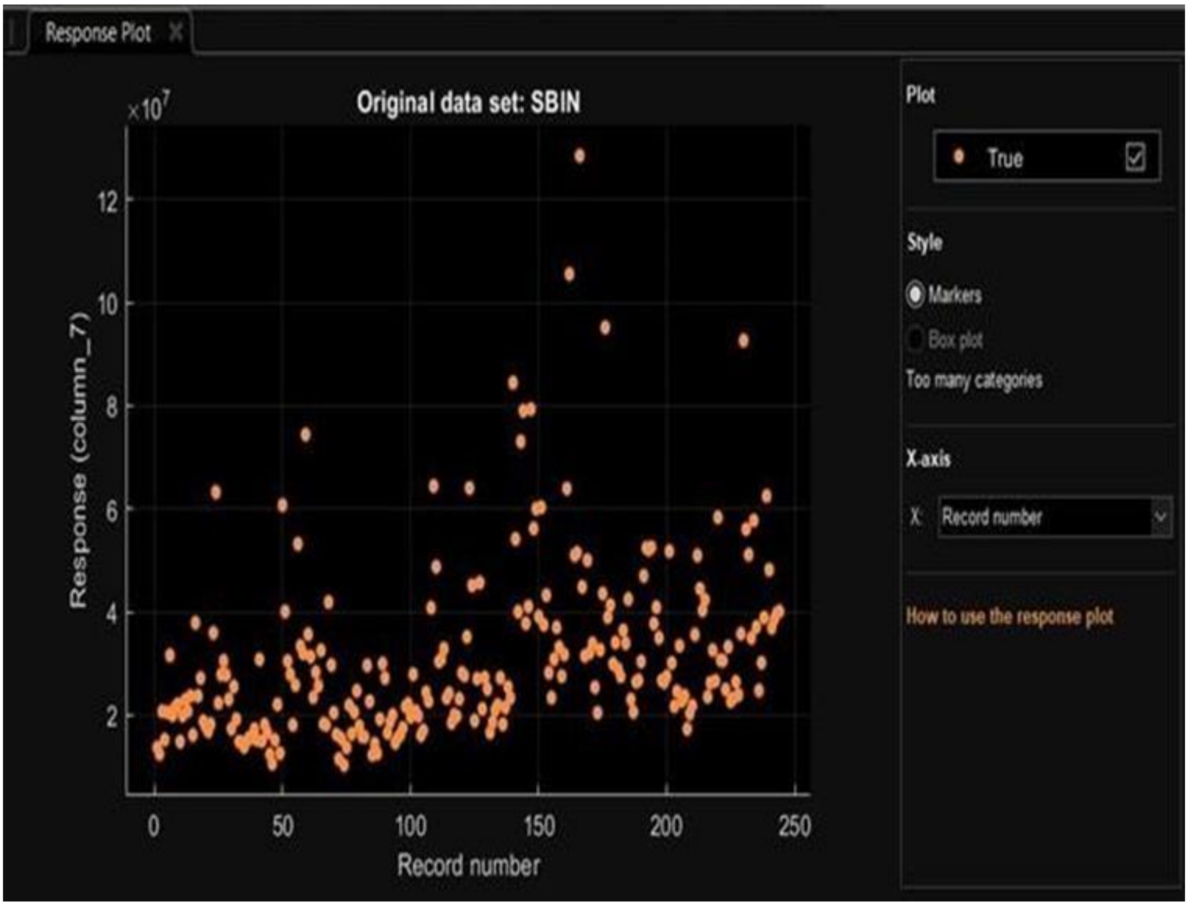

Fig.1 The scatter plot of the raw unfiltered data samples (SBI dataset)

Subsequently, the PCA is applied to the data for dimensional reduction and converting it into amore coherent and compact data vector with lesser dimensions.

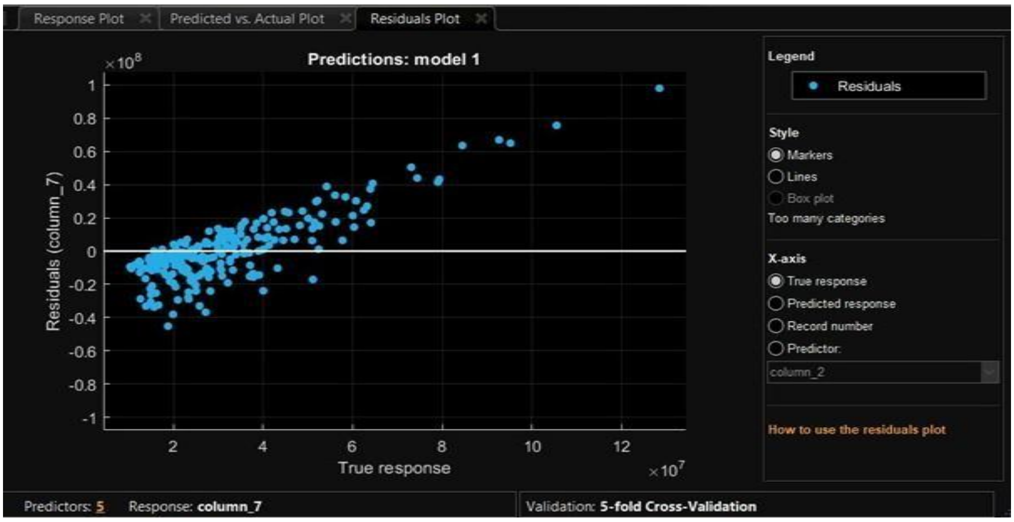

Fig.2 The scatter plot of the data after application of PCA (SBIdataset)

It can be observed that the local noise floor of the data can be removed by the application ofPCA which is the objective of pre-processing the data so as to facilitate the pattern recognition for the neural network. After the dimensional reduction of the data and removal of the base noise floor, the DWT is applied iteratively to obtain further filtering of the data.

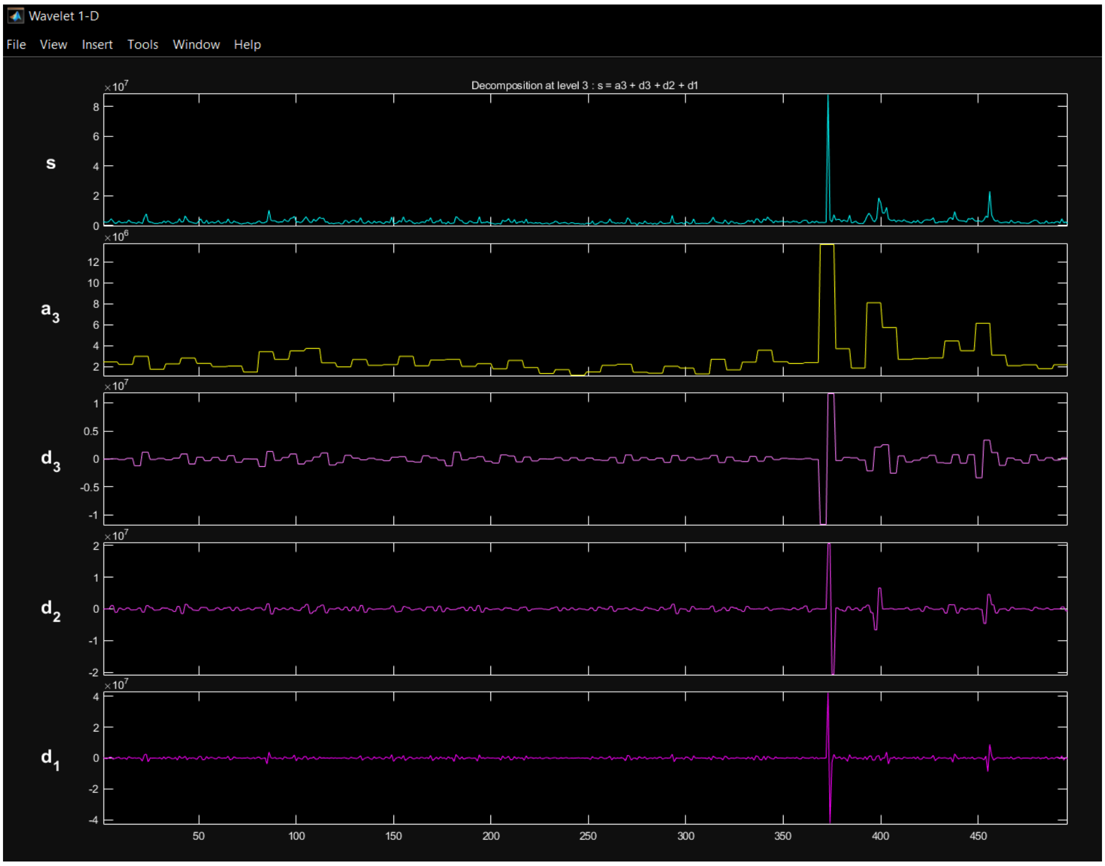

Fig.3 Wavelet Decomposition of the data at level 3 using Haar family

The data is decomposed into the approximate co-efficient and the 3 level detailed co-efficient values. Here‘s’ represents the approximate co-efficient value while d1, d2, d3 and d4 represents the detailed co-efficient values. The noise spectrum can be analyzed using the histogram analysisof the data.

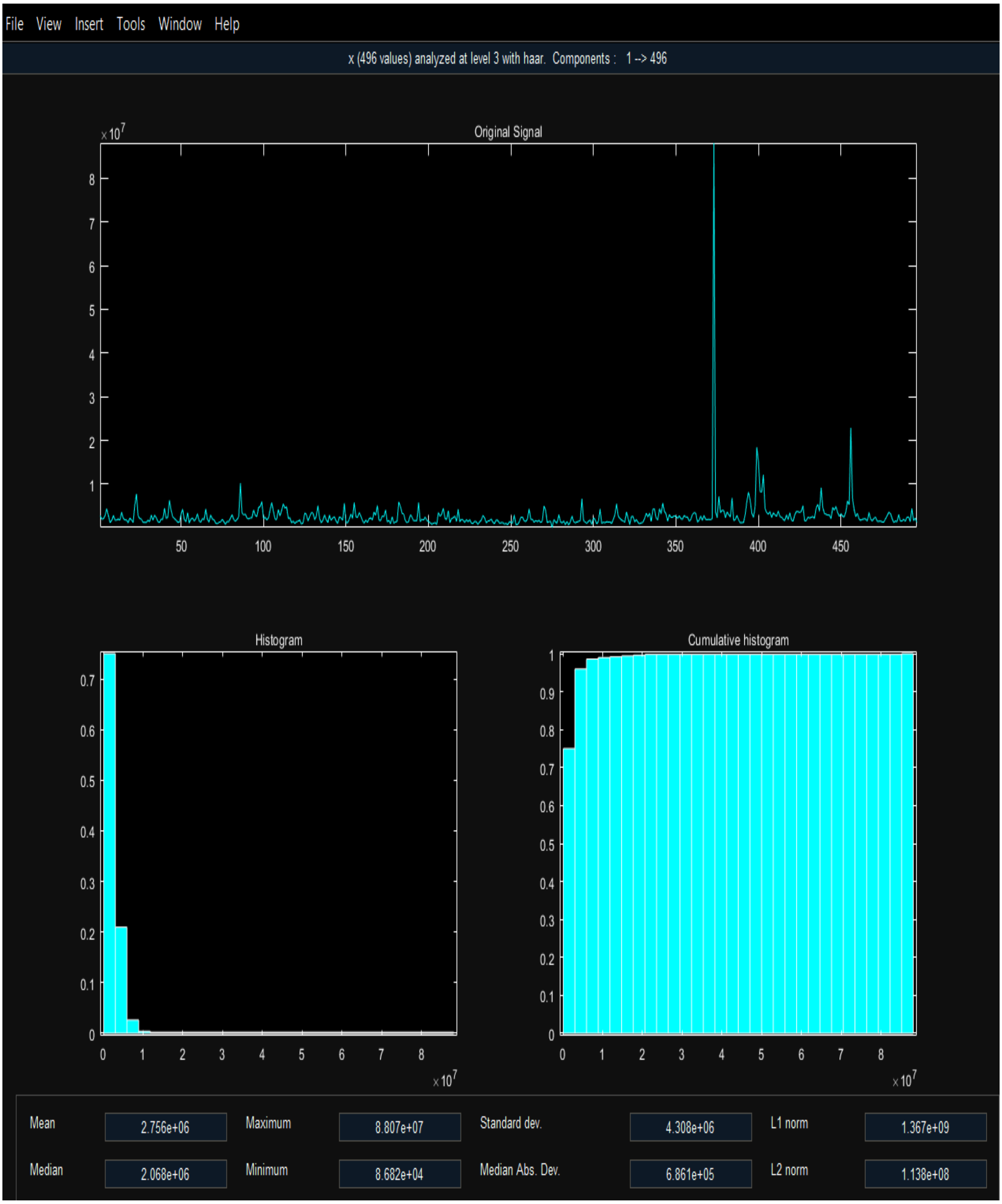

Fig.4 Histogram Analysis of original dataset (SBI)

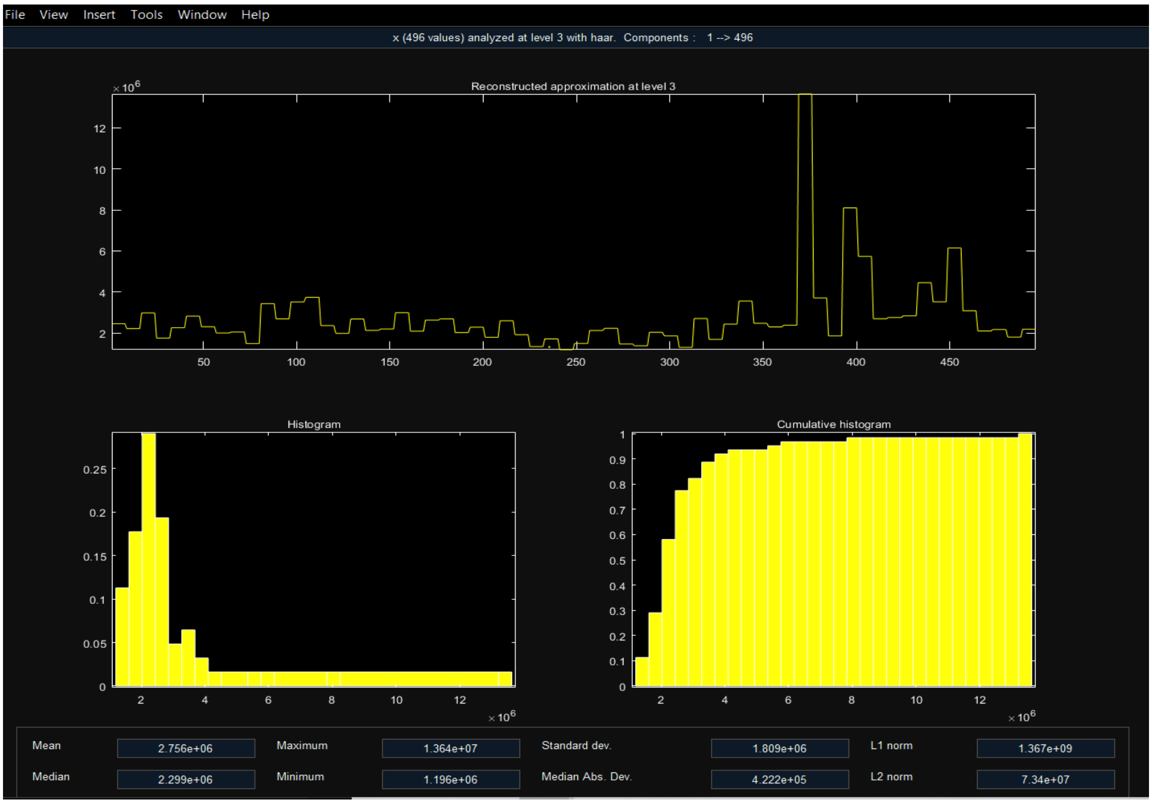

Fig.5 Histogram Analysis of approximate co-efficient values at level 3

Figure 5 represents the histogram analysis of the approximate co-efficient values at level 3 ofthe wavelet decomposition. The histogram analysis renders insights into the statistical properties of the filtered data.

Fig.6 Histogram Analysis of detailed co-efficient values at level 3

Figure 6 represents the histogram and statistical analysis of the decomposition of the data into the detailed co-efficient values. A clear difference can be seen in the normal and cumulative histograms of the detailed co-efficient values as compared to the original data or the approximate co-efficient values. A statistical analysis of the decomposition is tableted in tables

1. The parameters computed to evaluate the noise removal from the raw data Table 1 Statistical Analysis of Data

S.No.	**Parameter**	**Values**	**Class**
1.	Minimum	8.628×10^4	OriginalData
2.	Maximum	8.807×10^7	
3.	Mean	2.756×10^6	
4.	Median	2.068×10^6	
5.	Standard Deviation	4.308×10^6	
6	Mean Absolute Deviation	6.861×10^6	
7.	Minimum	3.383×10^6	Approximate Co-efficient values
8.	Maximum	3.858×10^7	
9.	Mean	7.796×10^6	
10.	Median	6.501×10^6	
11.	Standard Deviation	5.152×10^6	
12.	Mean Absolute Deviation	1.194×10^6	
13.	Minimum	-3.304×10^7	Detailed Co-efficient values
14.	Maximum	7.22×10^6	
15.	Mean	-8.081×10^5	
16.	Median	-2.9×10^5	
17.	Standard Deviation	4.732×10^6	
18.	Mean Absolute Deviation	9.876×10^5	

Table 1 clearly indicates that there is a distinct visible similarity between the histogram of the original data and the approximate co-efficient values of the data, whereas the detailed co- efficient values bear a clear difference in both magnitude, distribution and the polarity of the values. This clearly indicates that the noise and disturbance in the noise floor affects the detailed co -efficient values much more compared to the approximate co-efficient values. A similaranalysis has been done for the Infosys and Reliance datasets

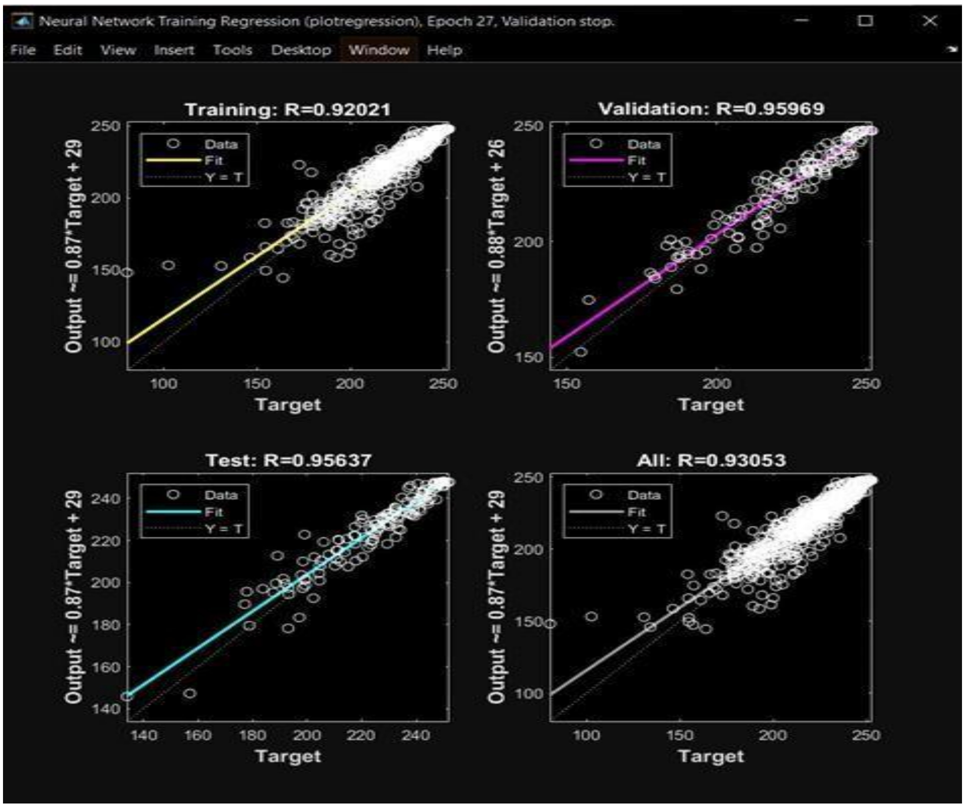

Fig.7 Regression of the Proposed Model

The R (regression) values have been depicted in figure 8 for the training, testing, validation and average overall cases. It can be observed that the proposed system attains an average regressionof 93%. A high value of regression indicates the closeness in the forecasted and actual values.

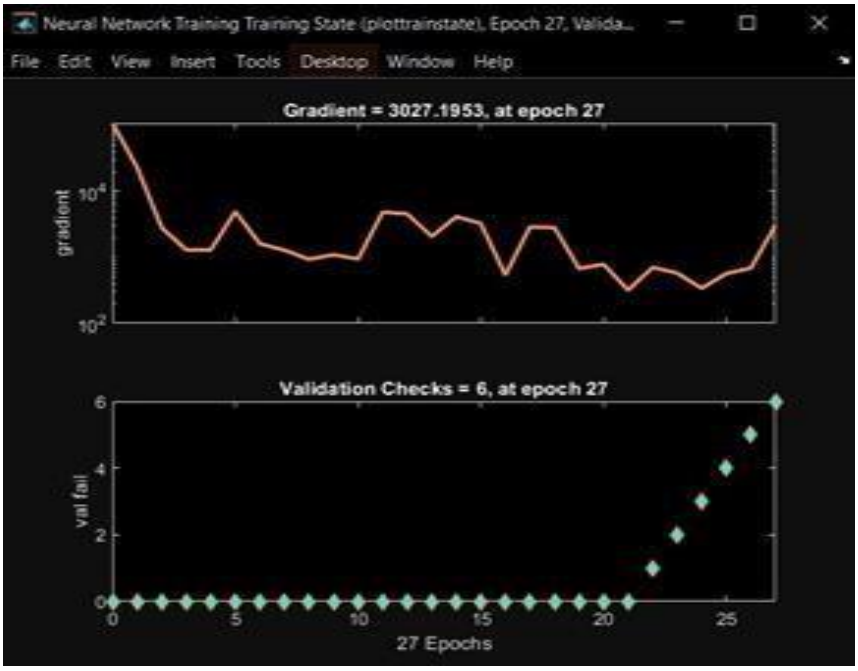

Fig.8 Training States of the Network

The proposed network employs a modified scaled version of the gradient descent algorithm and it can be seen that even though the algorithm used is backpropagation, the training convergedat 27 iterations with a near monotonic decrease in the gradient, The validation checks ensure that the objective function stabilizes prior to training.

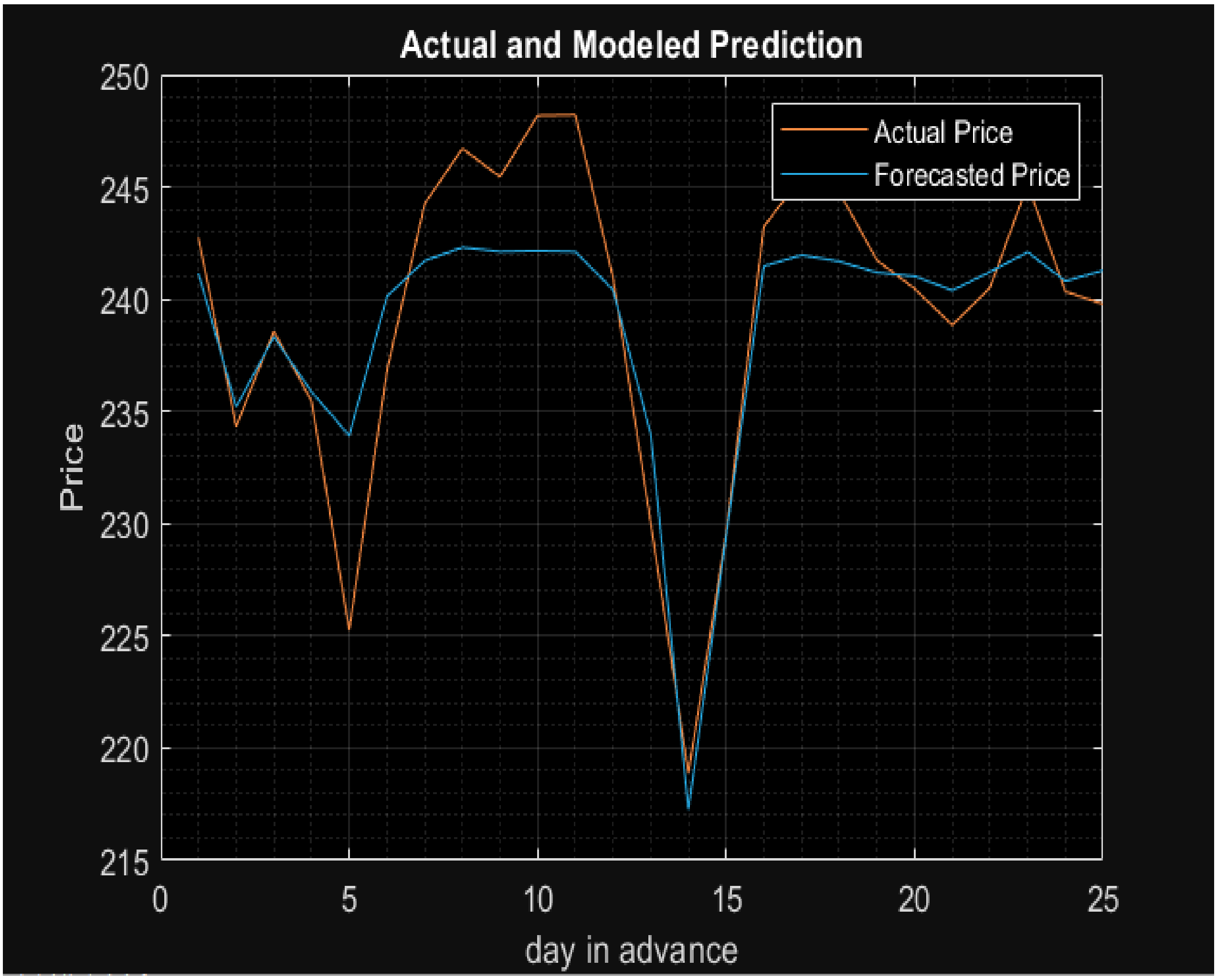

Fig.9 MAPE of the system for SBI dataset

The PCA and DWT decomposition has been performed for all the three datasets used in the study. The final result has been obtained in terms of the actual and predicted share prices. The comparative analysis of the actual and predicted values indicates the accuracy of the proposed approach. The mean absolute percentage error (MAPE) of the system is the metric used in this study to computed the prediction accuracy. The regression analysis also shows the similarity between the predicted and actual values of the stock prices. The training parameters for the SBI dataset have been tabulated in Table 2, while Table 3 summarizes the MAPE and regression values of the proposed algorithm for all the three data sets.

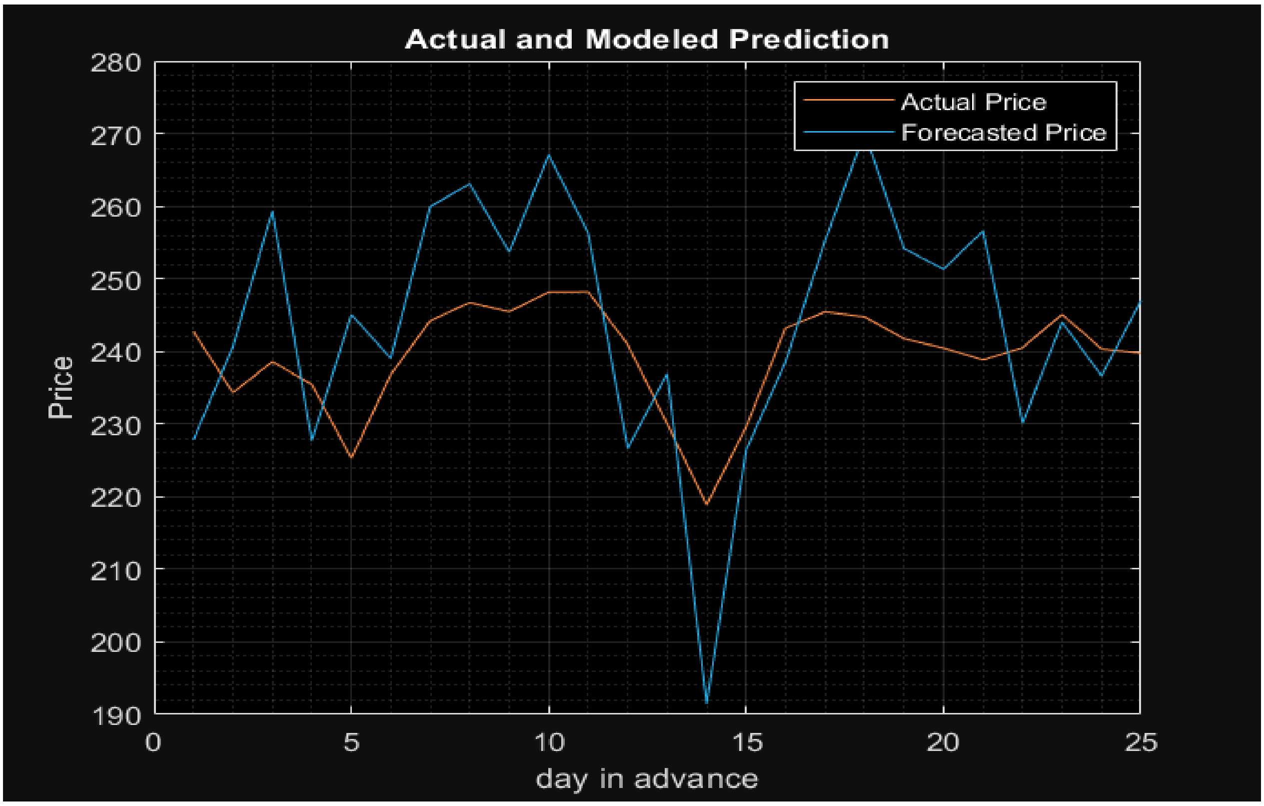

Fig.10 MAPE of the system for Infosys dataset

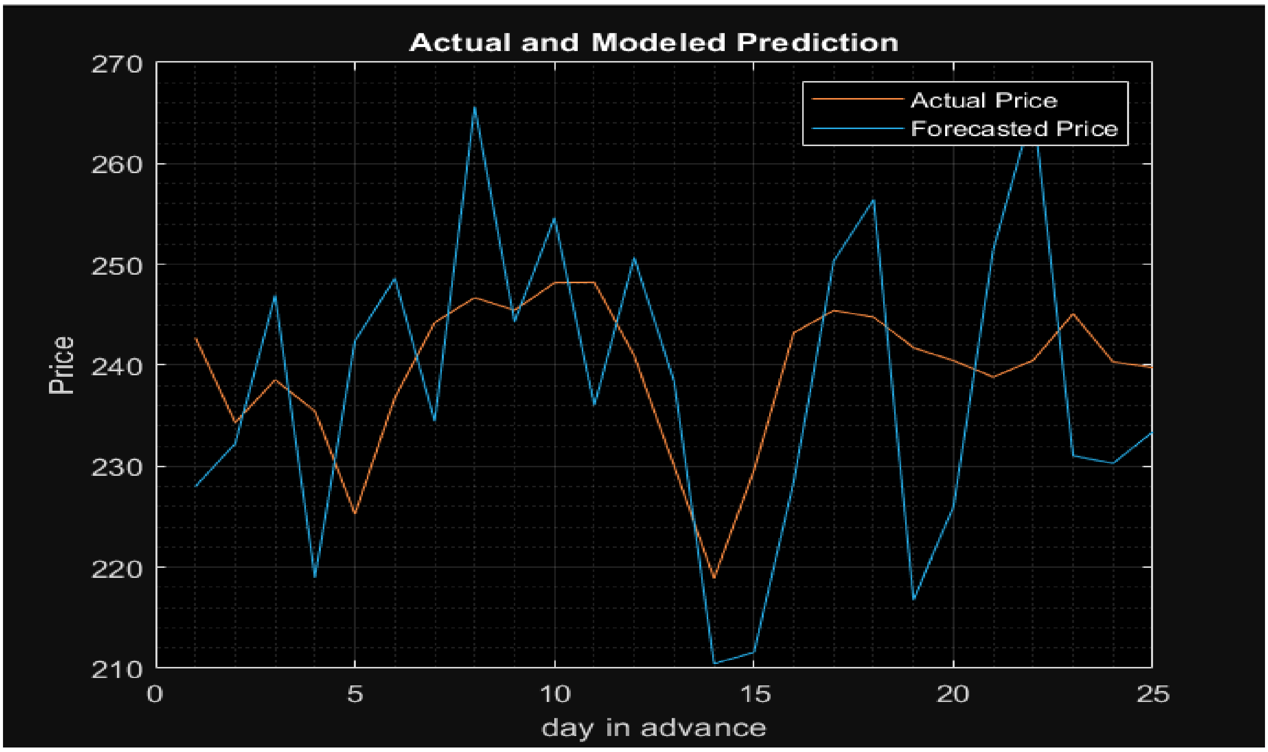

Fig.11 MAPE of the system for Reliance dataset

Three different datasets have been used in the proposed work which are that of SBI, Infosys and Reliance share prices, obtained from [28]. The actual and modelled forecasting values are depicted in figure 9. The mean absolute percentage error of the system is found to be 8.72% for the SBI dataset. This yields and accuracy of 91.28% which is relatively high compared to the existing literature [1]-[3]. A similar analysis has been adopted for the Infosys and Reliancedatasets. The MAPE and Regression Values for the Infosys and TCS datasets are 11.22, 0.915, and 13.15, .0.91 respectively. The value of the MAPE and accuracy are suggestive of the fact that the proposed system is capable to filter out the noisy values from the original noise floor and forecast the stock prices with relatively high accuracy. This statistical analysis of the approximate and detailed co-efficient values also indicates the same as the histogram of the detailed coefficient values are significantly w.r.t. to the original data while the detailed co-efficient values are convergent with the actual data. This indicates that the noise effects have been removed by iterative filtering employing the DWT. The scaled version of the neural network used has 10 hidden layers with a log - sigmoid activation function trained with the scaled version of the gradient descent. This allows faster convergence compared to the conventional gradient de scent.

Table 2 Summary of Computed Parameter Values for SBI dataset

S. No.	**Parameter**	**Value**
1.	Iterations	27
2.	Training Regression	0.9202 1
3.	Testing Regression	0.9563 7
4.	Validation Regression	0.9596 9
5.	Overall Regression	0.9305 3
6.	MAPE	8.72%
7.	Accuracy	91.28%

Table 3 Summary of MAPE and Regression Values

S. No.	**Dataset**	**MAPE**	**Regression (Overall)**
1.	SBI	8.72	0.93
2.	Infosys	11.22	0.915
3.	Reliance	13.15	0.91

Table 4 Comparison with existing techniques in terms of predictionaccuracy

S. No.	Technique	Accuracy
1.	Transfer Entropy and Machine Learning [1]	57%
2.	Augmented Textual Feature Based Learning [2]	60%
3.	LSTM with Sentiment Analysis [3]	49.6%
4.	**Proposed Technique (Mean Accuracy)**	88.97%

5. Conclusion

This paper presents a stock market forecasting model based on the modified version of the gradient descent and back propagation for neural networks. The neural network is designed with 10 hidden layers. A modification in terms of the scaled version of the gradient to be employed while computing the gradients in each iteration is proposed with a search vector updating the gradient direction in each iteration. Rigorous data cleaning is employed using the principal component analysis and the discrete wavelet transform. A 3rd level decomposition of the data is done and subsequent statistical analysis is performed to correlate the noise floor with the actual data to be analysed. It has been successfully shown that discarding the detailed co-efficient values helps in data cleaning and retaining the approximate co-efficient values results in subsequent accurate pattern recognition in the data. The performance of the system has bene evaluated in terms of the regression, mean absolute percentage error and accuracy of the system. The system attains an average accuracy of 88.97 which is significantly higher compared to existing literature. The novelty of the approach lies in the fact that the approach uses the gradient boosting methodology for all the co-efficient values of the DWT decomposition as training values. Moreover, the moving average acts as an additional input the algorithm to find recent patterns in the data.

References

Attigeri G, Pai R (2015) Stock market prediction: A big data approach, TENCON 2015 - 2015 IEEE Region 10 Conference, 1-5.

Baek Y, Kim H (2018) ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module, Journal of Expert System and Applications, Elsevier, 113, 457-480.

Bansal G, HasijaV, Chamola V, Kumar N, Guizani M (2019) Smart Stock Exchange Market: A Secure Predictive Decentralized Model, 2019 IEEE Global Communications Conference (GLOBECOM), 1-6.

Billah M, Waheed S, Hanifa A (2016) Stock market prediction using an improved training algorithm of neural network", 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering

(ICECTE), , pp. 1-4.

Bouktif S, Fiaz S, Awad M, Mosavi A (2020) Augmented Textual Features-Based Stock Market Prediction, IEEE Access 8, 40269 – 40282.

Cakra Y, Trisedya B (2015) Stock price prediction using linear regression based on sentiment analysis, 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 147-154.

Devi K, Bhaskaran V (2015) Cuckoo optimized SVM for stock market prediction, 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp: 1-5.

Eapen J, Bein D, Verma A (2019) Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction, 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), 264-270.

Guo Y, Han S, Shen C, Li Y, Yin X, Bai Y (2018) An adaptive SVR for high-frequency stock price forecasting, IEEE Access, 6, 11397 – 11404.

Kim S, Ku S, Chang S, Song J, (2020) Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques, IEEE Access, 8, 111660 –111682.

Li X, Wu P, Wang W (2020) Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong, Information Processing & Management, Elsevier 57, 5, 1-19.

Lincy G, John C (2016) A multiple fuzzy inference systems framework for daily stock trading with application to NASDAQ stock exchange, Journal of Expert Systems with Applications, Elsevier, 44, 298-311.

Meesad P, Rasel R (2013) Predicting stock market price using support vector regression", 2013 International Conference on Informatics, Electronics and Vision (ICIEV), pp. 1-6.

Nelson D, Pereira A, Oliveira R (2017) Stock market's price movement prediction with LSTM neural networks, 2017 International Joint Conference on Neural Networks (IJCNN), 1419-1426

Patel J, Shah S, Thakkar P, Kotecha K (2015) Predicting stock market index using fusion of machine learning techniques, Journal of Expert Systems with Applications, Elsevier, 42, 4, 2162-2172.

Porshnev A, Redkin I, Shevchenko A (2013) Machine Learning in Prediction of Stock Market Indicators Based on Historical Data and Data from Twitter Sentiment Analysis, 13th International Conference on Data Mining Workshops, 440-444.

Raimundo M, Okamoto (2018) SVR-wavelet adaptive model for forecasting financial time series", 2018 International Conference on Information and Computer Technologies (ICICT), 111-114.

Sadaei H, Enayatifar R, Lee M, Mahmud M (2016) A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting, Journal of Applied Soft Computing, Elsevier, 40, 132-149.

Selvin S, Vinayakumar R, Gopalakrishnan E, Menon V, Soman P (2017) Stock price prediction using LSTM, RNN and CNN-sliding window model, 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 1643-1647.

Sim C, Gan S, Soon K, On C, Alfred R, Anthony P (2013) A review of stock market prediction with Artificial neural network (ANN), 2013 IEEE International Conference on Control System, Computing and Engineering, 477-482.

Smailović J, Grčar M, Lavrač N, Žnidaršič M, (2013) Predictive Sentiment Analysis of Tweets: A Stock Market Application, Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, Springer, 77-88.

Walden A, "Wavelet Analysis for Discrete Time Series", European Congress of Mathematics, Springer 2001, pp: 627-641.

Wen M; Li P, Zhang L, Chen Y (2019) Stock Market Trend Prediction Using High-Order Information of Time Series IEEE Access, 7, 28299 – 28308.

Xu M, Lan Y, Jiang S (2015) Unsupervised Learning PartBased Representation for Stocks Market Prediction, 2015 8th International Symposium on Computational Intelligence and Design (ISCID), 63- 66.

Yoshihara A, Fujikawa K, Seki K, Uehara K (2014), Predicting stock market trends by r ecurrent deep neural networks, Pacific Rim International Conference on Artificial Intelligence PRICAI 2014: PRICAI 2014: Trends in Artificial Intelligence, Lecture Notes in Computer Science, 8862, 759-769.

Zhao Z, Rao R, Tu S, Shi J, (2017) Time-weighted LSTM model with redefined labeling for stock trend prediction", 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), 1210-1217.
https://in.finance.yahoo.com/quote

About Authors

Amanpreet Singh received B.Tech. degree in Computer Science and Engineering from Punjab Technical University, Jalandhar, Punjab in 2010, M. Tech. degree in Computer Science and Engineering from Guru Nanak Dev University, Amritsar, Punjab in 2012. Currently, he is working as assistant professor at College of Engineering and Management Punjabi University Neighborhood Campus, Rampura-Phul, Punjab, India. His area of interest is Machine learning. He has published papers in national and international journals.

Dr. R. K. Bathla is working as a professor in Computer Science Department at Desh Bhagat University, Mandi Gobindgarh, Punjab, India. He has published over 100 research papers in SCI, Scopus, Springer and IEEE. He has 25 patents in his name at national and international level. He hasguided 8 Ph.D. scholars, 4 have been awarded degree. In over 50 international journals, he is in editorial board or working as reviewer.