Forecasting Apple Stock Closed Prices by LR and LSTM with Discrete Wavelet Transformation

Yuxin Yang[1,*]

[1] *Woodsworth College, University of Toronto, 27 King's College Circle, Toronto, ON, Canada, M5S1A1*
[*]*Email:aurora.yang@mail.utoronto.ca*

ABSTRACT

Stock prediction has long had a high profile among investors under the incentives of profit maximization. However, as a result of the instability and chaos of the financial stock market, predicting stock prices is challenging. To address this problem, the discrete wavelet transformation (DWT) is applied to denoise stock prices when data preprocessing. Long short-term memory (LSTM) and linear regression model (LR) are chosen to train the model. The performances of LR, LSTM, the combination of DWT and LR and the combination of DWT and LSTM are demonstrated and compared when predicting the Apple stock closed prices by using its rescaled closed price five days ago. The prediction results proved the effectiveness of DWT and illustrated LR still acts well although it is much simpler compared with LSTM in terms of RMSE, MAE, MAPE. These model- based analytic strategies and pre-programmed stock price prediction are likely to give precious guidance to investors in the pursuit of maximum benefits.

Keywords: *LR, DWT, LSTM, Apple, Forecast.*

1. INTRODUCTION

Stock price is a reflection of the performance of each company[1]. Pursuing maximized profits, enterprises and investors would like to detect the underlying future trend of stock prices so that investors can react ahead of time, keep pace with the market opportunity, gain immense financial revenues with relatively low risk in an effective way. Therefore, accurately predicting stock prices is crucial and it attracts attention from a broad range of skilled personnel to explore techniques to figure out the potential direction of stock market[2]. However, it is a tough task since tons of trading happen each second which affected by politics, unexpected event, global economic situation, full of volatility and uncertainty, arising automated, pre-programmed analytic strategies to execute orders. Based on the Efficient Market Theory by Fama[3], predicting financial market is impractical since price has already represented all available information. He also thought stock prices follow stochastic process, which are random and unpredictable[4].

Nevertheless, lots of studies have been done for predicting stock prices, arranging from simple ways to complex models, which indicates that the stock market can be predicted to some extent[5]. Mondal, Shit and Goswami(2014) stated that Linear Regression models are conventional statistical models, including Autoregression (AR), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA)[6]. Particularly, they stated that ARIMA model holds great flexibility when dealing with time series data, estimating parameter and forecasting. Additionally, Cakra and Trisedya[7] combined sentimental analysis with Linear Regression, giving rise to a surprisingly high accuracy of prediction on Indonisea stock prices. Karim and Alam[8] compared Linear Regression with Decision Tree on 1-year, six-month and three-month datasets, revealing that Linear Regression performs better for both small and big datasets. Gupta and Nagalakshmi[9] utilized Simple Linear Regression and Support Vector Regression to foresee the stock cost of the organization, finding that their anticipated movement of stock market is kind of agrees with the real stock growth. Recently, as Human-Computer Intellectualization flourishes, deep learning has become one of the most widely used approaches in forecasting stock indexes due to its ability to deal with non linear data and self-selection for features[10]. Back propagation neural network(BPNN) and Recurrent Neural Network(RNN) were widely applied in previous research[2, 11]. Also, T. Fischer and C. KraussLong noted that short-term memory (LSTM) is an another powerful models in deep learning when addressing data sequence like speaking recognition and time series[12].

© The Author(s) 2022
Y. Jiang et al. (Eds.): ICEDBC 2022, AEBMR 225, pp. 935–943, 2022.
https://doi.org/10.2991/978-94-6463-036-7_138

Sirignano and Cont[13] conducted optimization by stochastic gradient descent (SGD) and fitted three layers LSTM units and a feed-forward layer with rectified linear units (ReLUs) to predict indexes of NASDAQ stocks. McNally et al.[14] leveraged RNN and LSTM on Bitcoin dataset with Boruta algorithm and Bayesian optimization as selection of features and select LSTM parameters.

However, there is a significant problem for the models above that financial time series in models above contain enormous noise. If the noise is included in the features, fitted models tend to be volatile since we do not intend to consider those noise[15], leading to overfitting or underfitting problems[16]. Kim and Han[17] combined artificial neural networks (ANN) and genetic algorithms (GA) with optimization of feature discretization and found that selecting and initializing features are indispensable. To deal with it, S. Kumar Chandar et al.[2] suggested that Wavelet transform would filter the noise and extract the hidden information the original time series data. Li et al.[18] decomposed the Down Jones Industrial Average (DJIA) index time series by the discrete wavelet transform, concluding that the wavelet analysis improved the accuracy of predicting the trend for noise data under the genetic programming algorithm. Moghimihanjani and Vaferi[19] proposed the Recurrent Neural Network (RNN) with the wavelet decomposition and found it to be highly promising.

Although the researches on combination of wavelet analysis and deep learning exist indeed, more analysis are needed. Therefore, in this paper, hybrid models are proposed that combines the discrete wavelet transformation (DWT) in data preprocessing with machine learning models. The empirical process is summarized as follows. Firstly, the most traditional model linear regression (LR) was chosen as one of our analyzing methods. People tend to pay more attention to those complex and ignore the effect of simple one. Whether this simplest model can also perform well in predicting stock prices is what this paper wonders. After applying DWT with LR on the tested dataset of Apple stock prices from 2014 to 2016, the Rooted Mean Square Error(RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error(MAPE) are 2.86, 2.15 and 1.9 respectively. Secondly, LSTM is fitted on the training set after denoising and tested its performance on the tested dataset because of its advanced memory system. The corresponding RMSE, MAE and MAPE are 3.63, 2.96 and 2.61 respectively. Hence the combination of LR and DWT performs better than hybrid model LSTM with DWT, showing the superiority of LR on stock price prediction. Thirdly, for comparision, data with noise were also kept. LR and LSTM are also applied on the original noisy data. Compared with the hybrid models for LR and LSTM which were denoised, the corresponding standard models for LR and LSTM with the original noisy data perform worse on prediction of Apple stock prices, indicating the significance of DWT when addressing chaotic data.

The rest of the paper is organized as shown below: The section 1 presents the related information of chosen dataset of Apple stock prices and the methods that will be applied. In section 3, the explanation and discussion for the results got after applying several models are presented. Part 4 makes conclusions of this whole research and presents what can be improved or done further in the future study.

2. DATA&METHOD

2.1 Data

The original dataset all stocks 5yr, containing 500 publicly traded domestic companies in United State and sp500 indexes, comes from the website Kaggle(https://www.kaggle.com/), which contains numerous datasets after preliminarily cleaning. It seems that no one will argue that Apple company is one of the most influential, vibrant and popular technology enterprises in the market. Therefore the potential future stock trend of this impactive and innovative company attracts much attention and deserves to be figured out. Data whose company name is AAPL and date is from January 1, 2014 to December 31, 2016 are selected and the date is set to be the index. This new AAPL dataset attributes are open, high, low, close and volume. After checking whether data have a null value, the closed price for each observation are selected to be the feature and then the predictor closed prices are rescaled. Also, the corresponding closed price in five days is the label and the impact for the rest of weekend in stock market are ignored. In other words, the Apple stock closed prices are predicted by using the corresponding closed price five days ago.

Table 1 shows several numerical characteristics for the closed prices of Apple stock from January 1, 2014 to December 31, 2016. The closed Apple stock prices peaked at the date February 23, 2015, with 133.00 USD. On the January 30, 2014, the closed price(71.40 USD) was the lowest from January 1, 2014 to December 31, 2016. The 25% quantile and 75% quantile of the closed prices are 96.10 USD and 115.82 USD respectively. The mean of the closed prices is 105.64 USD, which is closed to the median closed price 107.94 USD. The standard deviation of the selected closed prices is 15.08.

Table 1. Numerical Summary for the Selected Feature

	Value
Min	71.40
Max	133.00
Mean	105.64
Median	107.94

Q1	96.10
Q3	115.82
Standard Deviation	15.08

Then this feature is rescaled for convenience of executing the following-up models. Discrete Wavelet Transform is also applied to denoising the feature, generating the data without noise and data with noise.

2.2. Method

2.2.1 Min-max Scaling

Min-max Scaling is one of the most famous normalization methods. By applying this method, variables will be mapped to the range [0, 1], which means the minimum and maximum value of a variable will be 0 and 1 respectively. The aim is to make variables measured under the same scales so that each variable has the same portion of contributions for model fitting. Cao, Stojkovic and Stojkovic [20] stated that models trained on the scaled data tend to have better performance than those trained on the unscaled data and ANN requires the normality for the input to faster and stabilize the learning process. Although in this analysis, only the Apple stock closed prices five days ago are chose as the unique feature, compared with unscaled data, fitted models are likely to be more stable and run faster with scaling, especially for deep learning neural networks. The mathematical formula is,

$$x_{scale} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x_{scale} is the value after scaling, x is the original value and x_{max} and x_{min} are the maximum and minimum value of each feature, respectively.

2.2.2 Wavelet Transformation

Wavelet transform is an approach giving signals the opportunity to adjust the time-frequency resolution. Abbasi, Aghaei and Fard[21] claimed that wavelet transform has superior performance in handling highly irregular data and variables that operate on distinct time scales simultaneously. Mc-Coy et al.[22] purposed that wavelet approach makes decomposition according to times and frequency and significantly reduces the processing time. Under it, original signals o(t) can be decomposed into a series of high frequency and low frequency data, self selecting the local characteristics of signals. Traditional wavelet transform consists of continuous wavelet transform (CWT) and discrete wavelet transform (DWT) and DWT is used to do following analysis. DWT decomposes a given signal into a number of sets describing the time evolution of the signal in the corresponding frequency band. DWT constructs a scaling function vector space and a wavelet function vector space at different scales and time periods[23, 24, 25]. DWT holds the scaling function (2) and the wavelet function (3) [23]. Figure 1 illustrates the decomposition of DWT. Three-layer decomposition was used in the subsequent experiments in this paper. The original signal passes through the high-frequency filter and low-frequency filter, being decomposed into detail component ("cD" component in Figure 1) and approximate component ("cA" component in Figure 1) respectively.

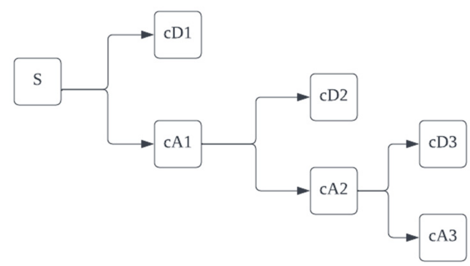

Figure 1: Decomposition of DWT

$$\phi_{jk}(t) = 2^{\frac{-j}{2}}\phi(2^{-j}t - k), j, k\epsilon Z \quad (2)$$

$$\Psi_{jk}(t) = 2^{\frac{-j}{2}}\Psi(2^{-j}t - k), j, k\epsilon Z \quad (3)$$

For financial time series, there are various kinds of wavelets, including the Daubechies, Haar, Morlet and MexicanHat. In this paper, Haar wavelet was applied to do decomposition. Other wavelet algorithms may perform better on resolution for smooth change of original data but cost too much to compute[26]. Haar wavelets are not only simple but also able to capture fluctuations between adjacent observations[27]. Here are some properties of Haar wavelets:

(1) Haar wavelets hold the tight support, having the significantly sharp drop-off performance[10].

(2) The support length is short, reducing the computation and data processing time[10].

(3) Haar wavelets are symmetric, lessening the distortion rate under signal decomposition and reconstruction[10].

As for the selection of threshold functions, soft threshold-denoising method and hard threshold-denoising method are two main streams of the threshold functions. Soft thresholding resets coefficients whose absolute values are lower than the threshold λ to 0 [28] and subtracts the absolute value of the nonzero coefficient from the threshold[10]. The expression is:

$$\omega_\lambda = \begin{cases} sign(\omega)(|\omega| - \lambda), |\omega| \geq \lambda \\ 0 \quad\quad\quad , |\omega| < \lambda \end{cases} \quad (4)$$

Hard thresholding resets coefficients whose absolute values are lower than the threshold λ to 0 and retains the nonzero coefficient[23, 28]. The soft threshold-denoising method is applied in this paper. The expression is expressed as followed,

$$\omega_\lambda = \begin{cases} \omega, |\omega| \geq \lambda \\ 0, |\omega| < \lambda \end{cases} \tag{5}$$

2.2.3 Simple Linear Regression

Chen [29] forecast the daily stock return by linear regression on the SP 500 Index ETF (SPY) in the second day. Assume that the total number of observations is n and they are collected in pairs (x1,y1), (x2, y2),..., (xn,yn). The relationship between X and Y is developed in a linear way with the response Y and the single predictor. The statistical relationship is,

$$Y = X\beta + \varepsilon \tag{6}$$

where $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$.

Here Y is a random column vector, X is a constant matrix of the predictor information, β is a constant column vector of parameters describing the intercept and slope and ϵ is a random column vector of deviation. Similarly, using the sample data, the linear relationship is,

$$Y = Xb + \hat{e} \tag{7}$$

where Y and X is the observed responses and predictors, b is the vector of coefficients representing possible estimated intercept and slope and ê is the residuals.

To estimate unknown β, the line of the best fit that minimizes the sum of the residuals over the whole data is supposed to be found. To get such estimators $\widehat{\beta_0}$ and $\widehat{\beta_1}$, the residual sum of squares(RSS) is minimized and the expression is

$$RSS = \sum_{i=1}^{n} \hat{e}^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 \tag{8}$$

To find the optimal $\widehat{\beta_0}$ and $\widehat{\beta_1}$, RSS is differentiated with respect to b0 and b1 and the following equations are generated.

$$\frac{\partial RSS}{\partial b_0} = -2\sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0 \tag{9}$$

$$\frac{\partial RSS}{\partial b_1} = -2\sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) x_i = 0 \tag{10}$$

After simplification, the least square estimators for simple linear model with only one predictor appear as followed.

$$\widehat{\beta_0} = \overline{y} - \widehat{\beta_1}\overline{x} \tag{11}$$

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} x_i y_i - n(\overline{x})(\overline{y})}{\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2} \tag{12}$$

In this paper, the only predictor used is the Apple stock closed prices five days ago. Therefore, simple linear regression is used.

2.2.4 LSTM

Long Short-Term Memory(LSTM) is a variation of RNN. What makes LSTM outstanding is its special memory system containing multiple gates and a memory cell, solving the long-term dependency problem[1]. Moghar and Hamiche[30] also figured out that compared with RNN who can not memorize for a long time, LSTM tends to be outstanding in predicting long time data. They stated that the the memorization of earlier stages and information transformation are controlled by gates and memory line are incorporated simultaneously. Figure 3[10] illustrates the structure of LSTM.

The forget gate regulates the amount of information of the previous cell state that can pass through [1]. What to be thrown away can be determined by a sigmoid layer "forget gate layer". It looks at the hidden layer $h_{t-1}$and the current layer x_t and creates a number in the range [0, 1] for each number in the cell state C_{t-1}[31]. 0 represents completely forgetting it and 1 represents completely remembering it [31]. This function f_t can be written as

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \tag{13}$$

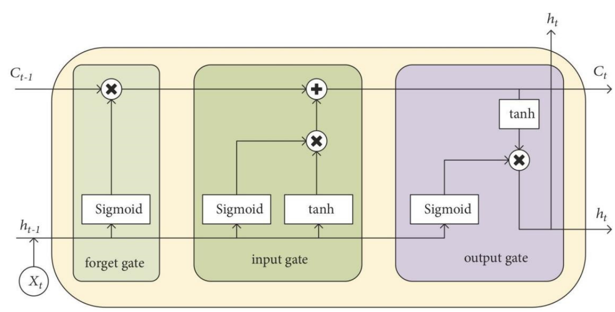

Figure 2. Structure of LSTM

The input gate controls the information of input signal $[h_{t-1}, x_t]$ that can be stored in the cell state. The sigmoid layer here is the input gate layer, deciding what to be updated, which expresses as (14). The tanh layer here creates a new candidate vector $\widetilde{C_t}$ which can be added to the cell state[31], illustrating as (15). Replacing $C_{t-1}$with C_t can be realized by forgetting what have been decided to throw away before and putting the scaled candidate vector into the new cell state[31]. The equation is (16)

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{14}$$

$$\widetilde{C_t} = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{15}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C_t} \tag{16}$$

The output gate decides what to be outputted. Which part of the cell stated will be the output id determined by the sigmoid layer[31], expressing as (17). The tanh activation function pushes the values in the range [-1, 1][18]. Finally, the multiplication produces the desirable output, expression as (18).

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (17)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (18)$$

There also exists some variations for LSTM we mentioned above with different activation functions. The task of activation functions is to transform the input to get the output of the node. "The sigmoid and hyperbolic tangent activation functions cannot be used in networks with many layers due to the vanishing gradient problem. The rectified linear activation function just overcomes the vanishing gradient problem, allowing models to learn faster and perform better." [32] Therefore in this paper, LSTM with ReLU activation function, 10 memory days, 3 LSTM layers and 1 dense layer is applied.

2.2.5 Hybrid method

In this research, the hybrid methods for stock prediction are proposed. On the one hand, the raw data are used, followed by performing the basic data preprocessing without denoising, created labels, breaking them into training datasets and testing datasets and applying LR and LSTM respectively. On the other hand, DWT is applied on the data with labels, broken into training datasets and testing datasets whose size were the same as their correspondingly training and testing datasets without DWT, and then LR and LSTM are applied respectively. Finally the performances of LR, LSTM and two hybrid models which were DWT + LR and DWT + LSTM are evaluated and compared. Figure 3 shows the general procedure of this research.

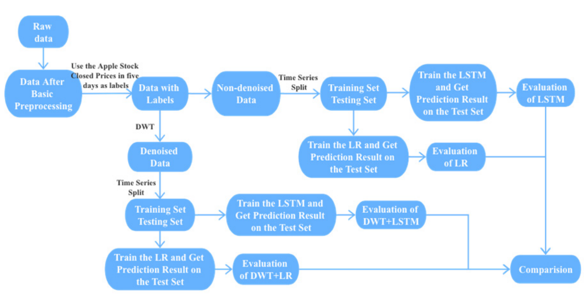

Figure 3: Flowchart of analysis

3. RESULTS

In this section, firstly the differences of the feature before and after removing the noise are visualized. Figure 4 shows that the blue curve is much more fluctuating than the yellow one, indicating that the curve of Apple closed prices becomes significantly smoother after DWT-denoising method.

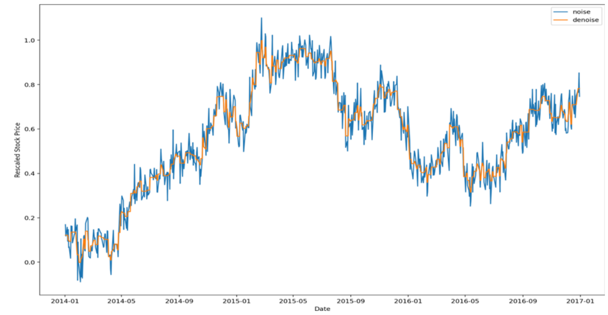

Figure 4: Plots of Apple Closed Price Before and After Denoising

Then four models are conducted to predict the Apple stock closed prices by using its closed price five days ago, which were LR, DWT + LR, LSTM and DWT + LSTM. To assess the performance of these developed predicting models, three metrics will be used to evaluate these models. These criteria measure the distance between the true values in the testing set and the values predicted using the developed models in the testing set, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). For three metrics used to measure accuracy for models, the larger the better. Table 2 shows detailed formulas and points that should be minded.

Table 2: Evaluation Metrics for Models

	Formula	Attention Points
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i}$	It brings the unit back to the actual unit, measures the average magnitude of the error and is easy to interpret our model accuracy.
MAE	$\frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert$	It is not sensitive to outliers because of the less weight of outliers.
MAPE	$\frac{100}{n}\sum_{i=1}^{n} \frac{y_i - \hat{y}_i}{y_i}$	It allows error to be compared across data with different scales but has problems when the true observation is 0.

The evaluation for each model in terms of RMSE, MAE and MAPE are illustrated in the Table 3. LR with DWT holds the smallest RMSE, MAE and MAPE among 4 proposed models, which are 2.86, 2.15 and 1.90 respectively. The RMSE, MAE and MAPE for LSTM without denoising are the highest compared with the other three models, which are 5.05, 4.47 and 3.93 respectively. For both LR and LSTM, after DWT, all three metrics will be lowered compared with each corresponding model without DWT. Moreover, the proposed two LSTM models tend to have larger RMSE, MAE and MAPE than the other two LR models.

Table 3: Evaluation values for Models

	LR	DWT + LR	LSTM	DWT + LSTM
RMSE	3.24	2.86	5.05	3.61
MAE	2.40	2.15	4.47	2.96
MAPE	2.12	1.90	3.93	2.61

Additionally, the experimental results of the performances for the linear regression, linear regression with DWT, LSTM and LSTM with DWT of Apple closed stock price prediction in the testing sets are visualized from Figure 5 to Figure 8.

Figure 5 is the plot of predicting stock prices by linear regression, showing that the trend of the predicted prices is almost the same as the the trend of true prices but has the problem of lag in response. Without removing noise, the linear regression almost just uses the corresponding price five days ago as the prediction prices.

Figure 6 shows the predicting stock prices by linear regression with DWT. Lag in response seems to be improved and the overall predicted prices are close to the corresponding true prices.

Figure 7 shows the performance of LSTM without DWT on predicting stock prices. There exists apparent distinctions between predicted stock prices and true stock prices from approximately October 25, 2016, to nearly November 18, 2016.

In Figure 8, after applying DWT, gaps between predicted stock prices and true stock prices from October 25, 2016 to November 18, 2016 have been significantly reduced while stock price prediction from approximately December 8, 2016 to nearly December 23, 2016 is lack of accuracy.

It is clear that the prediction of two proposed linear regression models get closer to the real stock prices than two proposed LSTM models. Also, after using the wavelet preprocessing, Linear Regression and LSTM tends to display slightly better than those without DWT due to the slightly smaller distances between predicted stock prices and real stock prices.

Therefore, DWT is indeed helpful to predict stock prices. After denoising, models are fitted better and the accuracy of prediction is proved in terms of RMSE, MAE and MAPE. Furthermore, the simple linear regression is likely to more suitable in this analysis than LSTM. And the hybrid model, the combination of DWT and LR tends to exhibit the best performance among the four proposed models.

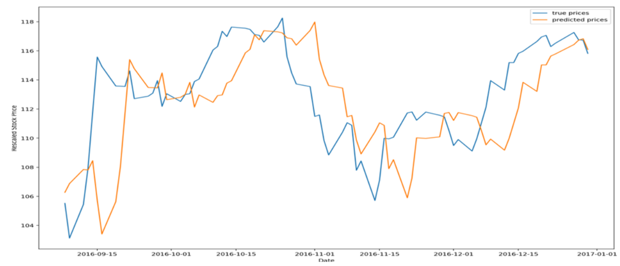

Figure 5: Plots of predicting stock prices for Linear Regression

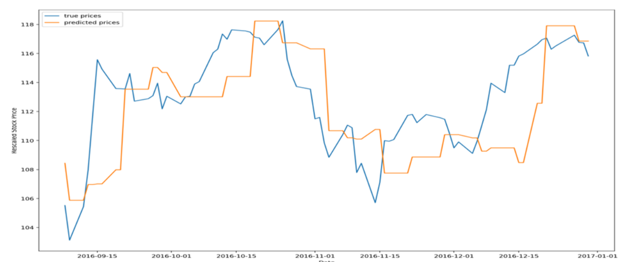

Figure 6: Plots of predicting stock prices for Linear Regression with DWT

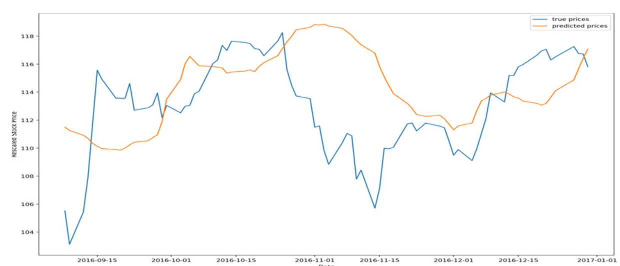

Figure 7: Plots of predicting stock prices for LSTM

Figure 8: Plots of predicting stock prices for LSTM with DWT

4. CONCLUSION

To assist investors to predict the future trend of stock prices, this study proposed hybrid forecasting models intermingling the advantage of denoising method and machine learning models for the prediction of Apple stock closed prices by using the closed prices five days ago. The selected feature is Apple stock closed prices. After min-max rescaling, all the value of the feature is between 0 and 1 and LSTM can execute faster. For hybrid models, labeled data were decomposed by DWT and the hybrid LR and LSTM models were trained based on the denoised data with the obtained features. The hybrid LR holds lower RMSE, MAE, MAPE than hybrid LSTM on their testing sets. Then the original LR and LSTM are fitted without denoising, having the same comparison result as two hybrid models. Finally, the hybrid LR and LSTM were compared with the original LR and LSTM on their testing sets in terms of RMSE, MAE, MAPE. The result of comparision indicates the superiority of applying the wavelet transform. DWT improves the prediction ability of LR and LSTM models. Also, due to the high dependency of the time series stock prices, LR seems to be an advisable choice to predict the particularly Apple stock prices. Therefore, when enterprises hesitate to choose whose stocks to invest and yearn to figure out the trend of stock prices of a particular company, linear regression with DWT perhaps an alternative model to predict stock prices, providing a direction for investors.

Despite the hybrid LR is outstanding among four proposed models in this research, inevitably, the main drawback is still that it is dataset specific. In other words, this model that yields outstanding performance results for Apple stocks may not perform well on an another stock. It is hard for the hybrid LR to be applied in an entire stock market. Also, there are various denoising wavelets in wavelet transform and different activation functions in LSTM. This paper does not thoroughly investigate their influence in forecasting stock trend, which can be continued to fulfilled. Furthermore, the memory days, numbers of layers and units do not especially accurately chosen. Thus, the LSTM built may not be the optimal LSTM. Construct more extra functions to automatically optimize the LSTM may be worth trying in the future research.

REFERENCES

[1] D. Song, A.M.C. Baek, N. Kim, Forecasting stock market indices using padding-based fourier transform denoising and time series deep learning models, IEEE Access, vol. 9, 2021, pp. 83786-83796. Doi: 10.1109/ACCESS.2021.3086537

[2] [2] S.Kumar Chandar, M. Sumathi, S. N. Sivanandam, Prediction of Stock Market Price using Hybrid of Wavelet Transform and Artificial Neural Network, Indian Journal of Science and Technology 9(8) (2016). DOI: 10.17485/ijst/2016/v9i8/87905

[3] E. F. Fama, Efficient capital markets: II, J.Finance 46(5) (1991) DOI: https://doi.org/10.1111/j.1540-6261.1991.tb04636.x

[4] D. M. Q. Nelson, A. C. M. Pereira and R. A. de Oliveira, Stock market's price movement prediction with LSTM neural networks, International Joint Conference on Neural Networks (IJCNN), 2017, pp. 1419-1426. Doi: 10.1109/IJCNN.2017.7966019

[5] N. Tripathy, S. Jaipuria, Forecasting Stock Market using Discrete Wavelet Transforms and Artificial Neural Networks Model, The Empirical Economics Letters 19(11) (2020) 1263-1277. This paper can be achieved from: https://www.researchgate.net/publication/346266024_forecasting_stock_market_using_Discrete_Wavelet_Transforms_and_Artificial_Neural_Networks_Model

[6] P. Mondal, L. Shit, S. Goswami, Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices, International Journal of Computer Science Engineering and Applications 4(2) (2014) 13-29. DOI: 10.5121/ijcsea.2014.4202

[7] Y. E. Cakra and B. Distiawan Trisedya, Stock price prediction using linear regression based on sentiment analysis, International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2015, pp. 147-154. Doi: 10.1109/ICACSIS.2015.7415179

[8] R. Karim, M. K. Alam, M. R. Hossain, Stock Market Analysis Using Linear Regression and Deci- sion Tree Regression, International Conference on Emerging Smart Technologies and Applications, 2021. DOI: 10.1109/eSmarTA52612.2021.9515762

[9] A. Gupta, T.J.Nagalakshmi, Stock Price Prediction using Linear Regression in Ma- chine Learning, International Journal of Innovative Technology and Exploring Engineering, 8(12) (2019) 2278-3075. This paper can be achieved from: https://www.ijitee.org/wp-content/uploads/papers/v8i12/L39321081219.pdf

[10] X. Liang, Z. Ge, L. Sun, M. He, H. Chen, LSTM with Wavelet Transform Based Data Prepro-cessing for Stock Price Prediction, Mathematical Problems in Engineering, vol. 2019, 2019. Doi: https://doi.org/10.1155/2019/1340174

[11] M. Jarrah, N. Salim, A Recurrent Neural Network and a Discrete Wavelet Transform to Predict the Saudi Stock Price Trends, International Journal of Advanced Computer Science and Applications 10(4) (2019). DOI: 10.14569/IJACSA.2019.0100418

[12] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, European Journal of Operational Research 270(2) (2019) 654-669. Doi: https://doi.org/10.1016/j.ejor.2017.11.054.

[13] J. Sirignano , R. Cont, Universal features of price formation in financial markets: perspectives from

deep learning, Econometrics: Econometric Statistical Methods - General eJournal, 2018. Doi: https://doi.org/10.2139/ssrn.3141294.

[14] B. Weng, L. Lu, X. Wang, F. M. Megahed, W. Martinez, Predicting short-term stock prices using ensemble methods and online data sources, Expert Systems with Applications, vol. 112, 2018, pp. 258-273. Doi: https://doi.org/10.1016/j.eswa.2018.06.016

[15] H. Hassani, A. Dionisio, M. Ghodsi, The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, Nonlinear Analysis Real World Applications 11(1) (2010) 492–502. Doi: 10.1016/j.nonrwa.2009.01.004

[16] J. Wang, J. Jiang, R. Yu, Robust back propagation algorithm as a chemometric tool to prevent the overfitting to outliers, Chemometrics and Intelligent Laboratory Systems, 34(1) (1996) 109–115.Doi: 10.1016/0169-7439(96)00005-6

[17] K. Kim, I. Han, Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index, Expert Systems with Applications, 19(2) (2000) 125–132. Doi: https://doi.org/10.1016/S0957-4174(00)00027-0

[18] J. Li, Z. Shi, X. Li, Genetic programming with wavelet-based indicators for financial fore- casting, Transactions of the Institute of Measurement and Control 28(3) (2006) 285-297. Doi: https://doi.org/10.1191/0142331206tim177oa

[19] M. Moghimihanjani, B. Vaferi, A Combined Wavelet Transform and Recurrent Neural Networks Scheme for Identification of Hydrocarbon Reservoir Systems From Well Testing Signals, J. Energy Resour. Technol 143(1) (2021). Doi: https://doi.org/10.1115/1.4047595

[20] X. H. Cao, I. Stojkovic, Z. Obradovic, A robust data scaling algorithm to improve classification ac- curacies in biomedical data, BMC bioinformatics 17(1) (2016). Doi: https://doi.org/10.1186/s12859-016-1236-x

[21] N. M. Abbasi, M. A. Aghaei, M. M. Fard, Forecasting Stock Market Using Wavelet Transforms and Neural Networks and ARIMA (Case study of price index of Tehran Stock Exchange), Inter- national Journal of Applied Operational Research 5(3) (2015) 31-40. This paper can be achieved from: https://www.semanticscholar.org/paper/Forecasting-Stock-Market-Using-Wavelet-Transforms-Abbasi-Aghaei/df61efcaabd0adec2c5dfe8960710dee49f0f7fd

[22] E. J. McCoy, A. T. Walden, Wavelet analysis and synthesis of stationary long-memory processes, Journal of computational and Graphical statistics 5(1) (1996) 26-56.

[23] D. Wu, X. Wang, S. Wu, A Hybrid Method Based on Extreme Learning Machine amd Wavelet Transform Denoising for Stock Prediction, Entropy 23(4) (2021) 440. Doi: https://doi.org/10.3390/e23040440

[24] A. Altunkaynak, M. Ozger, Comparison of discrete and continuous wavelet–multilayer percep- tron methods for daily precipitation prediction, Journal of Hydrologic Engineering 21(7) (2016). Doi: https://doi.org/10.1061/(ASCE)HE.1943-5584.0001376

[25] D. Chen, S. Wan, J. Xiang, F.S. Bao, A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG. PLoS One 12(3) (2017). Doi: 10.1371/journal.pone.0173138

[26] T. T. Khuat, Q. C. Le, B. L. Nguyen, M. H. Le, Forecasting Stock Price using Wavelet Neural Network Optimized by Directed Artificial Bee Colony Algorithm, Journal of telecommunications and information technology (2016). This paper can be achieved from: https://www.semanticscholar.org/paper/Forecasting-Stock-Price-using-Wavelet-Neural-by-Bee-Khuat-Le/fdee569e5e210f94894ddb8db60265fc6b6de761

[27] D. Jothimani, R. Shankar, S. S. Yadav, Discrete Wavelet Transform-Based Prediction of Stock Index: A Study on National Stock Exchange Fifty Index, Journal of Finanical Management and Analysis 28(2) (2015) 35-49. Doi: https://doi.org/10.48550/arXiv.1605.07278

[28] Z. German-Sallo, Nonlinear filtering in data signal denoising, RoEduNet IEEE International Conference, 2010, pp. 85-88. This paper can be achieved from: https://ieeexplore.ieee.org/document/5541593

[29] S. Chen, Forecasting Daily Stock Market Return with Multiple Linear Regres- sion, Mathematics Senior Capstone Papers, 2020. This paper can be achieved from: https://digitalcommons.latech.edu/mathematics-senior-capstone-papers/19

[30] A. Moghar, M. Hamiche, Stock Market Prediction Using LSTM Recurrent Neu- ral Network, Procedia Computer Science, vol. 170,

2020, pp. 1168-1173. Doi: https://doi.org/10.1016/j.procs.2020.03.049

[31] C. Olah, Understanding LSTM Networks, 2015. This paper can be achieved from: http://colah.github.io/posts/2015-08-Understanding-LSTMs/

[32] J. Brownlee, A Gentle Instruction to the Rectified Linear Unit (ReLU). URL: https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural- networks/

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.