Combining the Real-Time Wavelet Denoising and Long-Short-Term-Memory Neural Network for Predicting Stock Indexes

Zhixi Li and Vincent Tam
Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong, China
lizx@eee.hku.hk

***Abstract*— A stock market index can be a valuable indicator to describe the performance of a stock market in a particular region. Nevertheless, it is very difficult to forecast its future values or trends since the index data often demonstrate a high degree of fluctuations. Intrinsically, the wavelet denoising is a useful method to separate the signals from noise in many practical multi-media applications while the long-short-term-memory neural network (LSTM) is a powerful recurrent neutral network (RNN) architecture of learning and prediction models used in the field of computational intelligence. Nevertheless, few research studies have ever considered their combination for the prediction of stock or index movement. More importantly, some existing proposals trying to combine wavelet denoising with other artificial neural network architectures suffer from two major drawbacks. First, when applying the conventional one-time wavelet transform for denoising the stock data, this approach has made a serious logical flaw to include future stock data in its training phase, thus leading to impressive results in the back-testing yet actually impractical in real-world applications. In addition, the wavelet functions and decomposition levels are typically fixed in those studies for which they will not be able to produce optimal results in terms of the prediction accuracies. Hence, we propose in this paper a novel model to combine real-time wavelet denoising functions with the LSTM to predict the East Asian stock indexes in which the wavelet denoising adopts a sliding window mechanism to exclude the future data while its system configuration is flexibly optimized based on some predefined criteria. The empirical results reveal that the performance of our proposed prediction model shows significant improvements when compared to those of the original LSTM model without utilizing the wavelet denoising function. Furthermore, there are many interesting and possible directions including the integration with other deep learning networks for the future investigation of this work.**

Keywords—stock index forecasting; real-time wavelet denoising; wavelet transform; recurrent neural network; long short-term memory neural network.

I. Introduction

A stock market index is an important measurement to represent a stock market. The index is often used by the investors to manage their investment portfolios. Hence, the prediction of the stock market index movement has become important for the institutional investors as well as the general public. In addition to the north American and European markets, East Asian stock markets have been very active in these years. Thus, six East Asian stock market indexes are investigated to evaluate our proposed model in this paper, namely Hong Kong Hang Seng Index (HSI), Shanghai Stock Exchange Composite Index (SSE), Shenzhen Stock Exchange Composite Index (SZSE), Taiwan Capitalization Weighted Stock Index (TAIEX), Tokyo Nikkei Index (NIKKEI) and Korea Composite Stock Price Index (KOSPI).

As stock index data is very noisy, data denoising is one of effective ways to improve the prediction performance. The simplest way for denoising data is to use a moving average method [1]. However, the moving average method usually lags behind the real data trends. Another way is to adopt Fourier transform [2]. Nevertheless, Fourier transform cannot handle time information properly, and is not suitable to be applied to the non-stationary signals [3]. Unfortunately, stock index data is exactly time-varying and non-stationary. While wavelet transform has overcome the disadvantages of Fourier transform for expressing functions containing discontinuities and sharp peaks, and for deconstructing and reconstructing finite and non-stationary signals [4, 5]. Hence, the wavelet transform is used as a denoising technique to smooth the stock index data in this paper.

Long-short-term-memory neural network (LSTM), one of powerful architectures of Recurrent Neural Network (RNN), has been shown successful applications in many engineering domains [6-8]. Therefore, the LSTM is adopted as a prediction model in this paper.

Admittedly, there have been some research efforts on the combination of the wavelet transform and the LSTM. However, most are not aiming at financial market prediction problems, such as [9] for tourist arrivals prediction and [10] for audio onset detection. Furthermore, a few studies have attempted to apply wavelet denoising combined with other artificial neural network architectures to the stock market prediction according to [5, 11, 12].

However, there exists two major drawbacks. For one thing, in their approach, the stock data was denoised by the

978-1-5386-2726-6/17/$31.00 ©2017 IEEE

conventional one-time wavelet transform firstly and then split into two datasets: training dataset and testing dataset. But they seemed to fail to consider that the conventional one-time wavelet transform is a window-based transformation whose filter has a length. The transformation will involve the future data in a filter window, e.g. the future data next to current data point will be considered when denoising this point using this transformation. Since all data were historical data and obtained on the hand already in most studies, they were misunderstanding that it was right to do the wavelet denoising one time. In contrast, we are not able to acquire the future data in practice at all. This logical issue would lead to outstanding performance in the back-testing but uselessness in the real-world trading. For another, the wavelet function and the decomposition level are important in the wavelet denoising. However, the previous studies usually fix the function and its system configuration relying on the experience. This may not produce optimized results.

To overcome those existing issues, a novel hybrid of the real-time wavelet denoising and the LSTM is proposed in this paper. The real-time wavelet transform is employed to denoise the stock data by a sliding window mechanism. Besides, different wavelet functions and system configurations are tried to acquire optimal results to improve the performance. Finally, the LSTM model is trained to predict one-day-ahead values and trends for the indexes based on the denoised data.

The rest of the paper is organized as follows. Section II clearly explains the methodology in details. Section III presents the relevant empirical results and Section IV analyses the comparison results. Lastly, Section V gives some conclusions on this work.

II. METHODOLOGY

A. *Real-time Wavelet Denoising*

Wavelet transform is a time-frequency decomposition that decomposes the time-series data in both time and frequency domains even if the data is non-stationary. Wavelet transform has achieved many successful applications in fields of engineering, such as signal processing, scientific calculation, image processing, etc. [13-15].

The simplest idea of classical wavelet denoising model is presented as (1):

$$f(t) = x(t) + \varepsilon(t) \tag{1}$$

where $f(t)$ is the observed signal, $x(t)$ is the real signal and $\varepsilon(t)$ is the centered Gaussian white noise.

The purpose of wavelet denoising is to filter out $\varepsilon(t)$ as far as possible. Similarly, we assume that the stock data consists of a major trend like the real signal and fluctuations like noise, as shown in (2). Moreover, we suppose that the major trend is predictable and the fluctuations are unpredictable. Therefore, the basic idea is to capture the major trend and filter out the noise. We focus on the prediction of the major trend only.

$$f(n) = x(n) + \varepsilon(n) \tag{2}$$

where $f(n)$ is the observed stock data, $x(n)$ is the predictable major smooth trend and $\varepsilon(n)$ is the unpredictable fluctuations.

Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) are two types of the wavelet transform. In the wavelet transform, mother wavelet is the basic function.

1) Continuous Wavelet Transform

The process of CWT is to summarize the signal over all time, multiplied by scaled and shifted versions of the basic wavelet function.

In CWT, suppose $\varphi(t)$ is the basic wavelet function and $\varphi_{a,b}(t) = \frac{1}{\sqrt{a}}\varphi(\frac{t-b}{a})$ is a continuous wavelet function. Then the CWT is as follows for any signal $f(t) \in L^2(R)$.

$$CWT_f(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\varphi(\frac{t-b}{a})dt \tag{3}$$

where a is the sacle coefficient while b is the shift coefficient.

The inverse wavelet transform is as shown in (4).

$$f(t) = \frac{1}{c_\varphi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{a^2}CWT_f(a,b)\varphi_{a,b}(t)dadb \tag{4}$$

2) Discrete Wavelet Transform

The CWT will be changed to DWT if let the scale coefficient a be 2^{-j} and shift coefficient b be $k2^{-j}$ as (5) shows.

$$DWT_f(j,k) = 2^{\frac{j}{2}}\int_{-\infty}^{+\infty} f(t)\varphi(2^j t - k)dt \tag{5}$$

3) Multiresolution Decomposition using Wavelet Transform

As Fig. 1 depicts, in the DWT decomposition, the original signal is decomposed into approximations and details by convolving the signal with a low-pass filter and a high-pass filter, respectively [16]. The signal passed by the low-pass filter is the input in the next iteration step and so on. The approximations contain the general pattern/trend of the signal while the details may contain some noise so we can reduce the details by some thresholding approaches [17] or even filter them out directly.

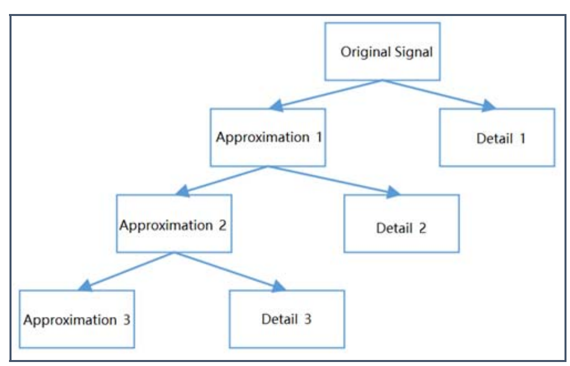

Fig. 1. Wavelet Decomposition at Level 3

4) Real-time Wavelet Denoising

To avoid involving the future data, a sliding window mechanism is employed in the proposed real-time wavelet denoising as shown in Fig.2. The denoising is conducted in a window based on the multiresolution decomposition. The window moves forward one unit each time until all data points have been covered. This mechanism instead of the one-time wavelet denoising ensures that the denoising process is real-time excluding the future data, which is practical in the real investment scenario.

Fig. 2. Real-time Wavelet Denoising with a Sliding Window

5) Optimization of Wavelet Denoising

The basic wavelet function determines the characteristics of the produced wavelet transform. Thus, it should be careful to pick up a suitable mother wavelet type. There are numerous types of wavelet functions available for different types of time-series being analyzed.

In the previous studies, the wavelet function and decomposition level are typically fixed when reducing the noise for all stocks or indexes. This practice seems to ignore that different stock or index may be sensitive to different wavelet function and configuration. Hence the selection of the wavelet function and configuration for each stock index should be optimized. One curve will be selected as the optimal one according to its volatility, and similar directional accuracy between the real data and the denoised data in the training period.

To exclude the future data, the optimization of wavelet functions, sliding windows and denoising levels is conducted in the training period strictly and then be applied to the testing dataset. Each wavelet function, as listed in Table I, is tried with various combinations together, as (6) shows.

$$window\ size \in [2,4,8,16,32,64,128,256,512]$$
$$decomposition\ level \in [1,2,3,4,5] \qquad (6)$$

Then the volatility as expressed in (7) and the similar directional accuracy as expressed in (8) of each curve are calculated.

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2} \qquad (7)$$

where r_i is the daily change of the i^{th} data point, n is the length of the data and $\bar{r}$ is the mean of daily changes.

$$similar\ directional\ accuracy = \left(1 - \frac{\sum_{i=1}^{n-1}|sgn(Pd_{i+1}-Pr_i)-sgn(Pr_{i+1}-Pr_i)|}{n-1}\right) * 100\% \qquad (8)$$

where Pr_i is the real value of the i^{th} data point, Pd_i is the denoised value of the i^{th} data point, and n is the length of the data.

After a lot of trials, we found the smoothest curve, i.e. the smallest volatility, with 60%-70% of the similar directional accuracy, could be selected as an optimized denoised curve.

TABLE I. WAVELET FAMILIES

Wavelet Family	Functions
Daubechies	db1, db2, db3, db4, db5, db6, db7, db8, db9, db10, db11, db12, db13, db14, db15, db16, db17, db18, db19, db20, db21, db22, db23, db24, db25, db26, db27, db28, db29, db30, db31, db32, db33, db34, db35, db36, db37, db38
Symlets	sym2, sym3, sym4, sym5, sym6, sym7, sym8, sym9, sym10, sym11, sym12, sym13, sym14, sym15, sym16, sym17, sym18, sym19, sym20
Coiflets	coif1, coif2, coif3, coif4, coif5, coif6, coif7, coif8, coif9, coif10, coif11, coif12, coif13, coif14, coif15, coif16, coif17
Biorthogonal wavelets	bior1.1, bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8
Reverse biorthogonal wavelets	rbio1.1, rbio1.3, rbio1.5, rbio2.2, rbio2.4, rbio2.6, rbio2.8, rbio3.1, rbio3.3, rbio3.5, rbio3.7, rbio3.9, rbio4.4, rbio5.5, rbio6.8

B. Long Short-Term Memory Neural Network

LSTM introduces memory cells into the neural network. Besides, the LSTM has input, output and forget gates, which is able to associate the memory cells using gates with the time-series input in time. Therefore, the LSTM is believed to grasp the dynamic changes of the stock data.

As Fig. 3 presents, there are two LSTM layers between the input and output layers in our proposed LSTM model. The number of memory cells on the input layer is consistent with of the number of the input features, and the numbers of memory cells of two LSTM layers are 1/2 and 1/3 of that of the input layer's, respectively. The final one is a dense output layer with a sigmoid activation function, which outputs the regression results.

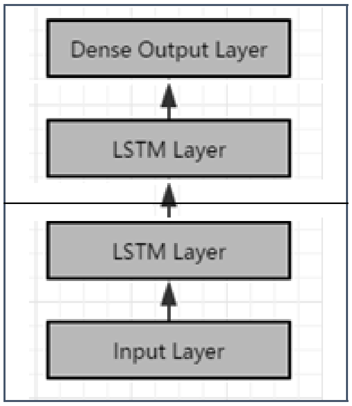

Fig. 3. Topology of the proposed LSTM model

III. EXPERIMENTAL RESULTS

A. Data Preparation

Six East Asian stock indexes were evaluated in the experiment. All data were obtained from the Bloomberg terminal. In addition to the prices and volumes, some common technical indicators were calculated as the extra input features, including EMA, MA, RSI, MACD, ROC, Willr, CCI, SAR and MFI.

The ranges of the training dataset and prediction dataset are specified as follows:

Training Dataset: *2010-01-01* to *2015-12-31* *(6 years)*

Testing Dataset: *2016-01-01* to *2016-12-31* *(1 year)*

B. The Proposed System

Fig. 4 depicts the flowchart of the proposed system to predict indexes through combining the wavelet denoising technique and the long short-term memory neural network (LSTM). First, the original data was fed as the data input. Second, the optimization of the real-time wavelet denoising was conducted where the wavelet function, the size of the sliding window and the decomposition level were optimized based on the criteria defined in the Section II.

Next, the denoised data was produced using the above optimized wavelet denoising and split into two datasets: training dataset and testing dataset. The LSTM was trained on the training dataset to build a prediction model. The model produced the prediction results when the testing dataset was fed.

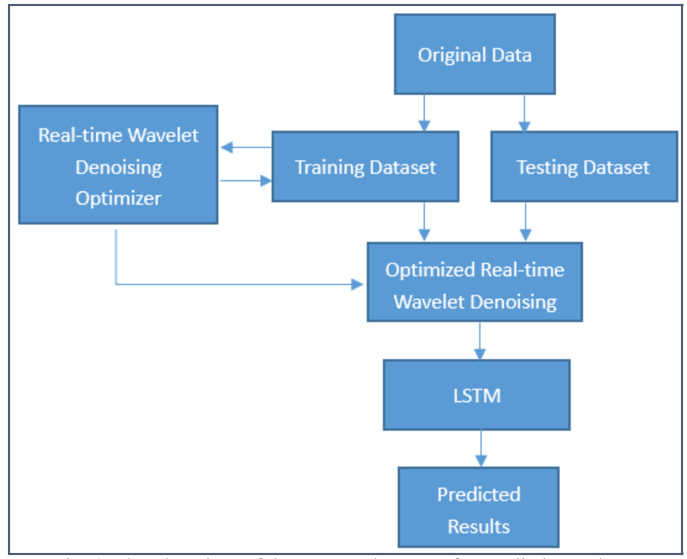

Fig. 4. The Flowchart of the Proposed System for Predicting Indexes

C. Experimental Results

Table II, IV, VI, VIII, X & XII list the optimal wavelet denoising function and system configuration for each stock index, respectively. For instance, the optimal wavelet function, decomposition level and sliding window size for HSI are *db27*, *3* and *512*, respectively.

Fig. 5, 7, 9, 11, 13 & 15 show the prediction performance for each index accordingly based on the original LSTM model without utilizing the wavelet denoising function. While Fig. 6, 8, 10, 12 & 16 are the results based on the denoised data produced by our proposed wavelet denoising approach where the blue curves are the predicted values by the LSTM and the black ones are the denoised values.

Table III, V, VII, IX, XI & XIII list the comparative LSTM prediction results for different data inputs. The mean absolute percentage error (MAPE) is a metric to evaluate the regression error while the directional accuracy is for the classification accuracy of one-day ahead trends prediction.

1) Hong Kong Hang Seng Index (HSI)

TABLE II. OPTIMIZED WAVELET DENOISING FOR HSI

Wavelet Function	**Decomposition Level**	**Sliding Window Size**
db14	3	256

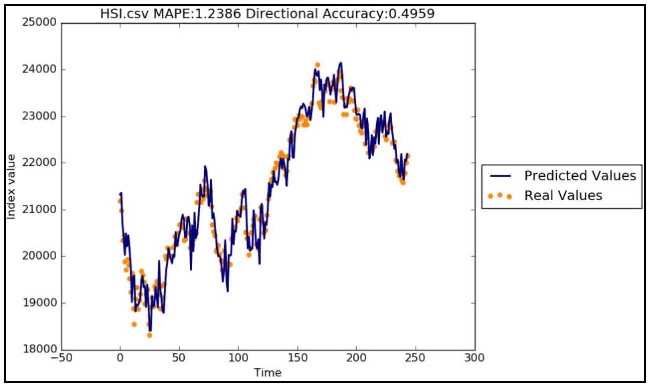

Fig. 5. Prediction Performance for the Original HSI

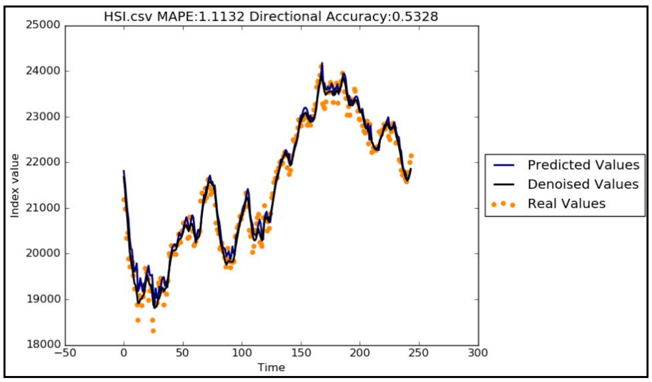

Fig. 6. Prediction Performance for the Denoised HSI

TABLE III. COMPARATIVE RESULTS FOR HSI

Data Input	Mean Absolute Percentage Error	Directional Accuracy
Original Data	1.2386%	49.59%
Wavelet Denoised Data	1.1132%	53.28%
Back-testing	0.8623%	66.39%

2) Shanghai Stock Exchange Composite Index (SSE)

TABLE IV. OPTIMIZED WAVELET DENOISING FOR SSE

Wavelet Function	Decomposition Level	Sliding Window Size
bior4.4	3	128

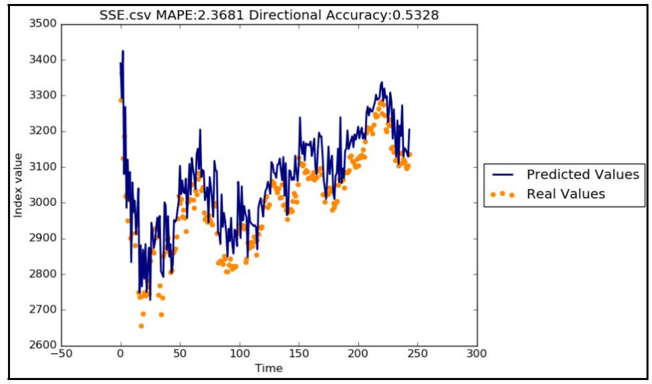

Fig. 7. Prediction Performance for the Original SSE

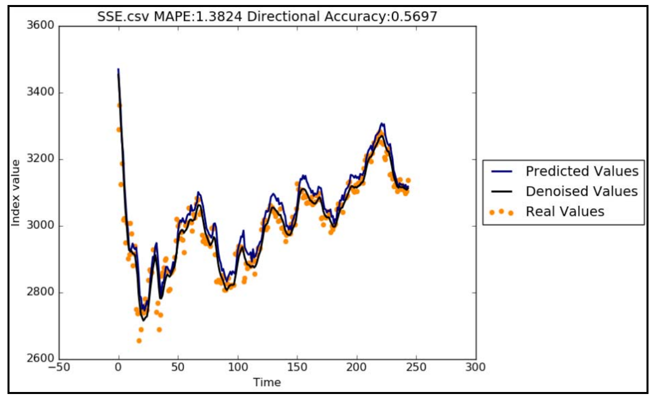

Fig. 8. Prediction Performance for the Denoised SSE

TABLE V. COMPARATIVE RESULTS FOR SSE

Data Input	Mean Absolute Percentage Error	Directional Accuracy
Original Data	2.3681%	53.28%
Wavelet Denoised Data	1.3824%	56.97%
Back-testing	1.4953%	60.66%

3) Shenzhen Stock Exchange Composite Index (SZSE)

TABLE VI. OPTIMIZED WAVELET DENOISING FOR SZSE

Wavelet Function	Decomposition Level	Sliding Window Size
db3	4	128

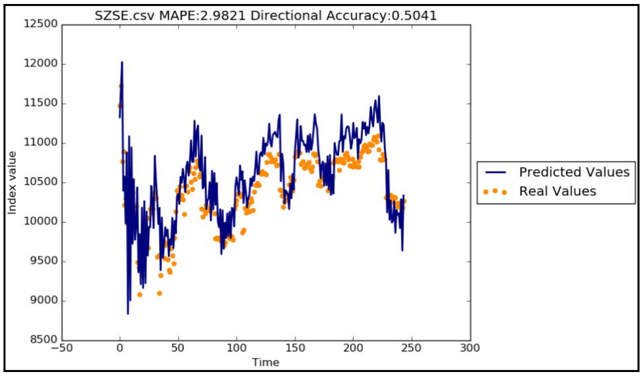

Fig. 9. Prediction Performance for the Original SZSE

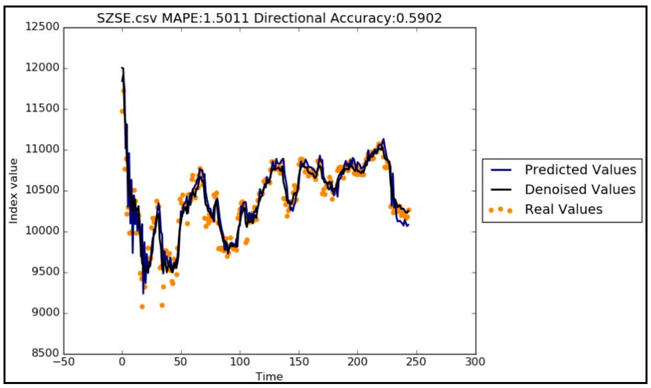

Fig. 10. Prediction Performance for the Denoised SZSE

TABLE VII. COMPARATIVE RESULTS FOR SZSE

Data Input	Mean Absolute Percentage Error	Directional Accuracy
Original Data	2.9821%	50.41%
Wavelet Denoised Data	1.5011%	59.02%
Back-testing	1.8605%	59.84%

4) Taiwan Capitalization Weighted Stock Index (TAIEX)

TABLE VIII. OPTIMIZED WAVELET DENOISING FOR TAIEX

Wavelet Function	Decomposition Level	Sliding Window Size
db5	3	128

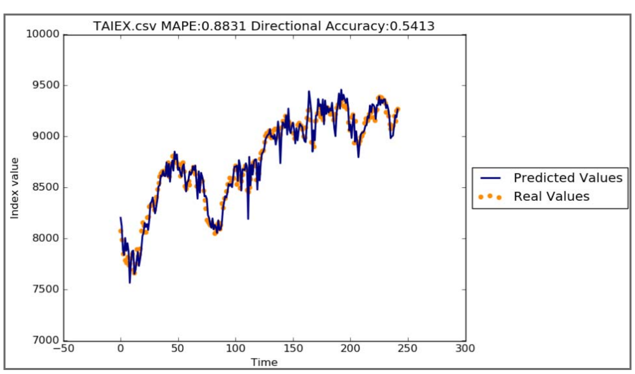

Fig. 11. Prediction Performance for the Original TAIEX

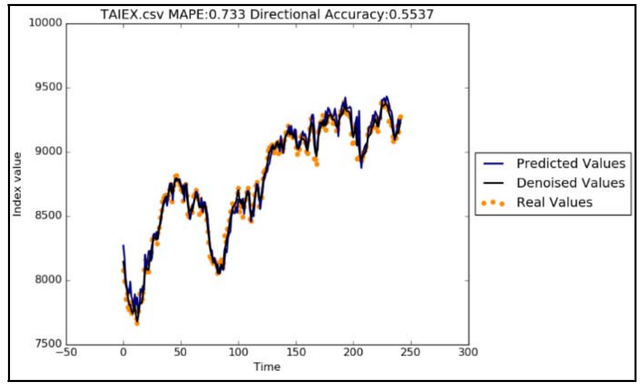

Fig. 12. Prediction Performance for the Denoised TAIEX

TABLE IX. COMPARATIVE RESULTS FOR TAIEX

Data Input	Mean Absolute Percentage Error	Directional Accuracy
Original Data	0.8831%	54.13%
Wavelet Denoised Data	0.7330%	55.37%
Back-testing	0.6531%	65.29%

5) Tokyo Nikkei Index (NIKKEI)

TABLE X. OPTIMIZED WAVELET DENOISING FOR NIKKEI

Wavelet Function	Decomposition Level	Sliding Window Size
db2	3	32

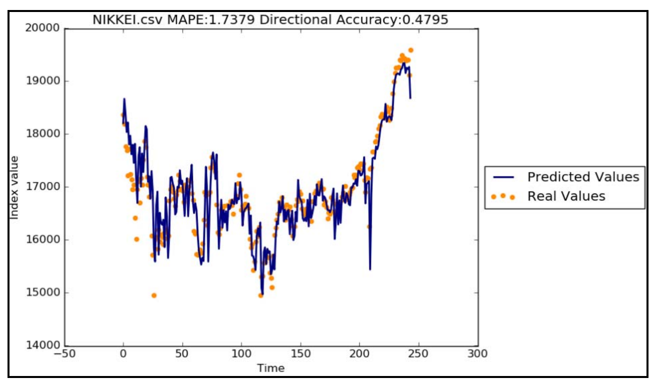

Fig. 13. Prediction Performance for the Original NIKKEI

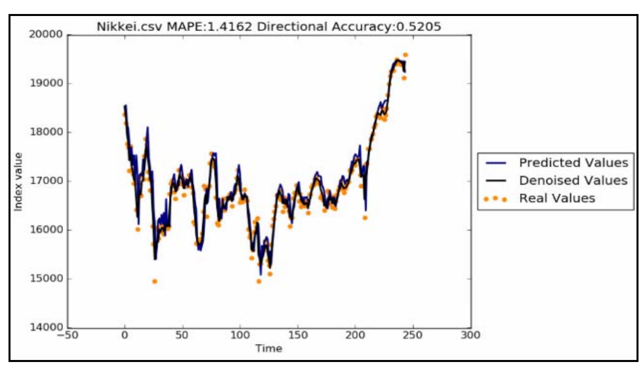

Fig. 14. Prediction Performance for the Denoised NIKKEI

TABLE XI. COMPARATIVE RESULTS FOR NIKKEI

Data Input	Mean Absolute Percentage Error	Directional Accuracy
Original Data	1.7379%	47.95%
Wavelet Denoised Data	1.4162%	52.05%
Back-testing	1.6428%	61.07%

6) Korea Composite Stock Price Index (KOSPI)

TABLE XII. OPTIMIZED WAVELET DENOISING FOR KOSPI

Wavelet Function	Decomposition Level	Sliding Window Size
coif10	3	512

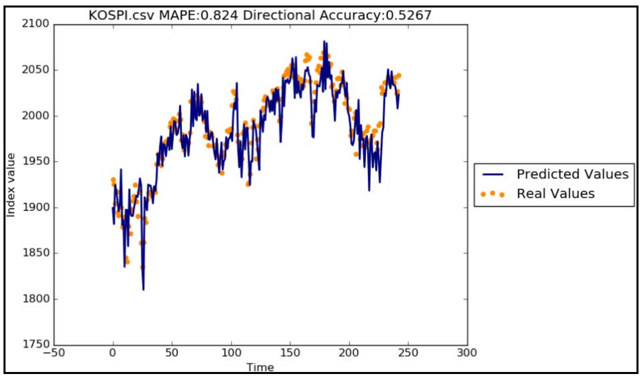

Fig. 15. Prediction Performance for the Original KOSPI

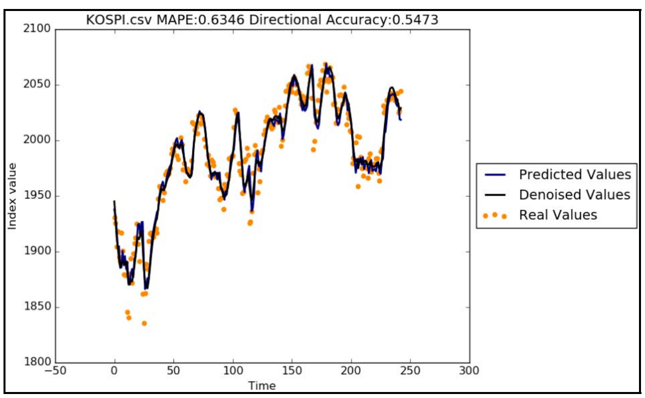

Fig. 16. Prediction Performance for the Denoised KOSPI

TABLE XIII. COMPARATIVE RESULTS FOR KOSPI

Data Input	Mean Absolute Percentage Error	Directional Accuracy
Original Data	0.8240%	52.67%
Wavelet Denoised Data	0.6346%	54.73%
Back-testing	0.5532%	67.49%

IV. ANALYSIS AND DISCUSSION

A. The Original Data Versus The Denoised Data

As shown in Fig. 5, 7, 9, 11, 13 & 15, the predicted values fluctuate significantly without adopting the wavelet denoising. On the one hand, the prediction model fails to predict the market accurately. For instance, the predicted curves of the original SSE and SZSE almost stand above the real ones all the time, as shown in Fig. 7 & 9. On the other hand, the model is so sensitive to the drastic market changes (i.e. the noise) that the model is not able to predict the major trend of the market well especially when the market is volatile, as seen in Fig. 5 (HSI), 13 (NIKKEI) & 15 (KOSPI). In the real trading, this original model may generate a lot of fake or wrong signals due to the fluctuations of the market. As a result, this may cause the investors a great loss.

By contrast, all denoised curves are smoother than the original data points because most market fluctuations have been filtered out by the wavelet denoising, as shown in Fig. 6, 8, 10, 12, 14 & 16. Thus, the model is able to grasp the major trend of the market. In practice, this could help the investors to take potential huge profit following the trend trading.

By analyzing all empirical results obtained in Table III, V, VII, IX, XI & XIII, the prediction performance of LSTM based on the proposed wavelet denoising approach acquires apparent improvements for the regression errors compared with those without the wavelet denoising. We can see that the MAPEs have been reduced by ***0.13%***, ***0.99%***, ***1.48%***, ***0.15%***, ***0.32%*** and ***0.19%*** for HSI, SSE, SZSE, TAIEX, NIKKEI and KOSPI, respectively, after employing the wavelet denosing approach.

Though our prediction problem is pre-defined to forecast one-day ahead values and trends, the directional accuracy is more valuable in the real applications. We can obverse that the directional accuracies have been improved, to some degree, for all indexes as well, i.e. increased from *49.59%* to *53.28%* (***+3.69%***), *53.28%* to *56.97%* (**+3.69%**), *50.41%* to *59.02%* (***+8.61%***), *54.13%* to *55.37%* (***+1.24%***), *47.95%* to *52.05%* (***+4.10%***) and *52.67%* to *54.73%* (*+**2.06%***) for HSI, SSE, SZSE, TAIEX, NIKKEI and KOSPI, respectively.

It is interesting that some indexes attain significant improvements, such as SZSE and NIKKEI, whereas some acquire just slight improvements, such as TAIEX and KOSPI. We think it may be due to the different volatilities of the indexes. To investigate on this, the volatilities of all indexes are calculated according to (7) and summarized with the performance improvements in Table XIIII.

TABLE XIIII. SUMMARY OF INDEXES VOLATILITIES AND PERFORMANCE IMPROVEMENTS

Index	Volatility	MAPE Decrement	Directional Accuracy Increment
SZSE	0.85%	1.48%	8.61%
NIKKEI	0.78%	0.32%	4.10%
SSE	0.65%	0.99%	3.69%

HSI	0.56%	0.13%	3.69%
TAIEX	0.41%	0.15%	1.24%
KOSPI	0.33%	0.19%	2.06%

We can observe clearly that the higher volatility original indexes tend to get the higher performance improvements after using the wavelet denoising. This suggests that the high degree of fluctuations may have been removed in the high-volatility indexes so that the model can predict the market much better.

B. *Back-testing Results*

The back-testing results are indicated in Table III, V, VII, IX, XI & XIII. The back-testing is based on the conventional one-time wavelet transform as described above. Apparently, the back-testing results are more impressive, i.e. *66.39%, 60.66%, 59.84%, 65.29%, 61.07%* and *67.49%* for HSI, SSE, SZSE, TAIEX, NIKKEI and KOSPI, respectively. As a matter of fact, this is fine in the back-testing. However, this practice involving the future data is completely out of question in the real trading scenario.

V. Concluding Remarks

In this work, we propose a new framework combining the real-time wavelet denoising with the LSTM neural network to forecast the movement of stocks or indexes. Our proposal rectified some logical flaw in some previous studies that carelessly include the future data during the training phase. In addition, we developed a real-time wavelet denoising approach with a sliding window mechanism. The system configuration of the resulting framework can be flexibly enhanced by an optimization approach. To demonstrate its feasibility and effectiveness, a prototype of our proposed framework is applied onto six East Asian stock indexes with its experimental results carefully compared and analyzed against those of the original model without utilizing the denoising function. The findings reveal that the proposed model attains a very significant improvement, i.e. a higher prediction accuracy and lower regression error than those of the original model.

In addition, the back-testing is conducted based on the conventional denoising function. Though the obtained back-testing results seem impressive, it is impractical in the real-world trading activities since the future stock or index data is required as part of the input for the prediction task. Thus, such impressive back-testing results should be analyzed more carefully before applying to any real investment activity.

Acknowledgment

The authors are grateful to Professor Edward Tsang for his fruitful discussion, and also the generous supports from the Hong Kong Applied Science and Technology Research Institute Company Limited (ASTRI) and the University of Hong Kong to conduct this research project.

References

[1] H. Hassani, A. Dionisio, and M. Ghodsi, "The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets," Nonlinear Analysis: Real World Applications, vol. 11, pp. 492-502, 2010.

[2] R. M. Alrumaih and M. A. Al-Fawzan, "Time series forecasting using wavelet denoising: an application to Saudi Stock Index," Journal of King Saud University, Engineering Sciences, vol. 2, pp. 221-234, 2002.

[3] M. Sifuzzaman, M. Islam, and M. Ali, "Application of wavelet transform and its advantages compared to Fourier transform," 2009.

[4] C. K. Chui, An introduction to wavelets: Elsevier, 2016.

[5] L. Wang and S. Gupta, "Neural networks and wavelet de-noising for stock trading and prediction," Time Series Analysis, Modeling and Applications, pp. 229-247, 2013.

[6] F. A. Gers, D. Eck, and J. Schmidhuber, "Applying LSTM to time series predictable through time-window approaches," in International Conference on Artificial Neural Networks, 2001, pp. 669-676.

[7] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in Acoustics, speech and signal processing (icassp), 2013 ieee international conference on, 2013, pp. 6645-6649.

[8] M. Sundermeyer, R. Schlüter, and H. Ney, "LSTM Neural Networks for Language Modeling," in Interspeech, 2012, pp. 194-197.

[9] P. Sugiartawan, R. Pulungan, and A. K. Sari, "Prediction by a Hybrid of Wavelet Transform and Long-Short-Term-Memory Neural Network," INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, vol. 8, pp. 326-332, 2017.

[10] E. Marchi, G. Ferroni, F. Eyben, L. Gabrielli, S. Squartini, and B. Schuller, "Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks," in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014, pp. 2164-2168.

[11] S. K. Chandar, M. Sumathi, and S. Sivanandam, "Prediction of stock market price using hybrid of wavelet transform and artificial neural network," Indian Journal of Science and Technology, vol. 9, 2016.

[12] T.-J. Hsieh, H.-F. Hsiao, and W.-C. Yeh, "Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on artificial bee colony algorithm," Applied soft computing, vol. 11, pp. 2510-2525, 2011.

[13] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," IEEE transactions on image processing, vol. 9, pp. 1532-1546, 2000.

[14] L. Satish and B. Nazneen, "Wavelet-based denoising of partial discharge signals buried in excessive noise and interference," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 10, pp. 354-367, 2003.

[15] J. S. Vitter and M. Wang, "Approximate computation of multidimensional aggregates of sparse data using wavelets," in ACM SIGMOD Record, 1999, pp. 193-204.

[16] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE transactions on pattern analysis and machine intelligence, vol. 11, pp. 674-693, 1989.

[17] P. Hedaoo and S. S. Godbole, "Wavelet thresholding approach for image denoising," International Journal of Network Security & Its Applications (IJNSA), vol. 3, pp. 16-21, 2011.