IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 7, 2020, accepted June 4, 2020, date of publication June 17, 2020, date of current version June 26, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3002174*

# Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques

**SONDO KIM[1], SEUNGMO KU[1], WOOJIN CHANG[1,2,3], AND JAE WOOK SONG[4]**

[1]Department of Industrial Engineering, Seoul National University, Seoul 08826, South Korea
[2]Institute for Industrial Systems Innovation, Seoul National University, Seoul 08826, South Korea
[3]SNU Institute for Research in Finance and Economics, Seoul National University, Seoul 08826, South Korea
[4]Department of Industrial Engineering, Hanyang University, Seoul 04763, South Korea

Corresponding author: Jae Wook Song (jwsong@hanyang.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Ministry of Science and ICT under Grant 2018R1C1B5043835.

**ABSTRACT** This study aims to predict the direction of US stock prices by integrating time-varying effective transfer entropy (ETE) and various machine learning algorithms. At first, we explore that the ETE based on 3 and 6 months moving windows can be regarded as the market explanatory variable by analyzing the association between the financial crises and Granger-causal relationships among the stocks. Then, we discover that the prediction performance on the stock price direction can be improved when the ETE driven variable is integrated as a new feature in the logistic regression, multilayer perceptron, random forest, XGBoost, and long short-term memory network. Meanwhile, we suggest utilizing the adjusted accuracy derived from the risk-adjusted return in finance as a prediction performance measure. Lastly, we confirm that the multilayer perceptron and long short-term memory network are more suitable for stock price prediction. This study is the first attempt to predict the stock price direction using ETE, which can be conveniently applied to the practical field.

**INDEX TERMS** Econophysics, effective transfer entropy, feature engineering, information entropy, machine learning, prediction algorithms, stock markets, time series analysis.

## I. INTRODUCTION

Stock markets have been studied extensively as one of the crucial fields of economy [1]. In particular, research has been actively conducted to analyze and predict the stock market based on relationships among the dynamics of stock prices and returns. Since the stocks exhibit diverse interactions, many theoretical or empirical studies of such relationships have provided meaningful implications to investors and policy-makers developing appropriate actions regarding the market condition. Specifically, the prediction on stock price and the overall market is one of the essential tasks for investors to establish an optimal investment strategy.

Many previous studies have utilized concepts in statistical physics such as complex systems and information theory to quantify the correlations among the entities in an economic or financial system [2]–[6]. Notably, the correlation analysis is a simple and good indicator for measuring the degree of similarity between two variables. Many studies of correlation-based time series analysis have revealed the characteristics of the system using random matrix theory and network analysis [7]–[9]. Since then, the studies have discovered that a linear model such as the Pearson correlation is not sufficient enough to quantify the relationships among the stocks. More importantly, the causal relationship is not directly linked to the presence of correlation. In this context, the Granger-causality [10] has been introduced to define the causal relationship between time series. However, its function is limited to express the existence of information flow based on a linear relationship rather than measuring the amount of information flow.

To overcome such limitations of a simple linear model of a Granger-causal relationship, the concept of transfer entropy (TE), proposed by Schreiber [11], has been suggested

The associate editor coordinating the review of this manuscript and approving it for publication was K. C. Santosh.

instead to measure the amount of information flow. TE is a non-parametric measure of the amount of information transfer from a variable to a variable based on the Shannon entropy [12]. TE has been widely used in researches such as social networks, neuroscience, and financial market analysis due to its ability to capture the asymmetrical interactions within the system and to distinguish the driving and responding elements efficiently [13]–[15]. Despite its advantages, TE could involve a noise due to its requirement on a large amount of data. Thereupon, the effective transfer entropy (ETE) [16] has been proposed to obtain a more robust quantification of information flow. Since then, many studies have utilized TE and ETE to identify information flows in different financial markets [17]–[25].

Based on findings in Granger-causal relationships within the financial system, the prediction on the price or volatility of the stock market has been widely studied [26]. Especially, the prediction on stock price is a critical issue since an improved prediction performance can guarantee a higher expected return to investors. In this regard, many previous studies have attempted the prediction based on various models [27], [28] including an artificial neural network [29]–[32], support vector machine [33], [34], and random forest [35], [36]. Commonly, two different approaches are considered in machine learning-based stock price prediction. The first is to improve the existing machine learning models theoretically [37], [38], and the second is to integrate created variables such as network indicators, Google trends, and public announcements in addition to a simple set of stock price and return series [39]–[46].

In this study, we focus on the integration of entropy-driven indicators in machine learning algorithms to improve the performance of prediction in the direction of a stock price. Although many studies have devoted to integrating the network-driven indicators, an attempt to utilize the concept of entropy in stock price prediction has not been widely studied. Instead, the previous studies on TE have focused on revealing the Granger-causal relationship between various financial variables. Furthermore, the previous studies related to TE and ETE analysis have focused on the inter-market and static-interval analysis, which occurs a limited finding in intra-sector and dynamic-interval.

Therefore, we focus on the US financial market, the largest single market in the global financial system, and study 55 companies from 11 different industry sectors based on moving window methods. Indeed, the moving window method allows dynamic and time-varying observations for different crisis and non-crisis periods based on interval analysis. Once we explore that the evolution of ETE can express the dynamics of stock prices, we utilize it as an extra variable for various machine learning algorithms to predict its direction. By comparing the performances of sets with and without the ETE variable, we confirm the usability of ETE in stock price prediction.

Note that we utilize the logistic regression (LR) [47] as a linear predictive model, multilayer perceptron (MLP) as a back-propagated neural network, random forest (RF) [48] as a bagging-type ensemble method, XGBoost (XGB) [49] as a boosting-type ensemble method, and long short-term memory (LSTM) network [50] as a deep learning algorithm.

Our paper is organized as follows. Section II reviews previous studies of Granger-causality detection using TE and ETE in financial markets, as well as studies of stock price prediction using machine learning algorithms. Section III describes the ETE and machine learning algorithms used in this study. Section IV demonstrates the data and the analysis framework, including different set-ups for prediction. Section V analyzes the ETE of the US financial market, the prediction performance of stock price direction. Section VI concludes.

## II. LITERATURE REVIEW

### A. TRANSFER ENTROPY

Since the concept of Granger-causality [10] was discovered, many studies have utilized it to detect the Granger-causal relationship in various financial time-series. For instance, Mok [51] analyzed the relationship among the daily stock prices, exchange rates, and interest rates of Hong Kong using the autoregressive integrated moving average and Granger-causality test, and confirmed the sporadic unidirectional Granger-causality from stock price to interest rate as well as a weak bidirectional Granger-causality between stock price and exchange rate.

Swanson *et al.* [52] conducted a multivariate time series analysis on the data revision process for industrial production (IP) and the composite leading indicator (CLI) in the United States. They confirmed a kind of causal feedback of revision processes for IP and CLI by showing that previously available IP revisions are useful for describing CLI revisions and that IP revisions are predictable from past CLI revisions.

Soytas and Sari [53] conducted a Granger-causality test between energy consumption and GDP in nine emerging market groups and seven advanced countries. However, the Granger-causality is a model-specific approach that assumes a specific underlying dynamics such as the Vector Autoregressive (VAR) and the Vector Error Correction Model (VECM), which focuses on the existence of a Granger-causal relationship rather than the degree of a Granger-causal relationship.

In contrast, the TE [11] is a model-free measure coping with the limitation of Granger-causality in quantifying the degree of information transfer without any constraints such as linear dynamics. The previous studies on Granger-causality detection through information flow analysis using TE and ETE in financial markets are as follows. Marschinski and Kantz [16] analyzed the relationship between the Dow Jones Industrial Average (DJIA) and the German DAX Xetra Stock Index (DAX) using ETE, and discovered asymmetry information flow where more information is transferred from the DJIA to the DAX than the vice versa. Kwon and Yang [18] performed a TE analysis using daily data of 25 global financial market indices and

discovered that the US stock market has the most significant impact on the global stock market as well as confirmed that the Asia and Pacific market received information the most. Sensoy *et al.* [17] considered the direction and intensity of information flow for the exchange rate and stock price of nine emerging countries through ETE. Specifically, a low level of interaction before the 2008 financial crisis was revealed with the dominance of exchange rate over the stock prices in the crisis, whereas a robust bidirectional interaction in and after the crisis with a dominance of stock prices over the exchange rate. However, the studies of Kwon and Yang [18] and Sensoy *et al.* [17] are limited in showing an overall level of relationships without considering the relationships at industry or individual stock level.

In this context, Kwon and Yang [19] conducted a TE analysis of the Dow Jones index, S&P 500 index, and 125 stocks in the US market, and confirmed that individual stock prices are affected by the market index considering the direction of information transfer. Also, Kwon and Oh [20] performed TE analysis between the market index and stock price for nine emerging or mature stock markets, and confirmed the market index as the primary driving force for determining individual stock prices and higher asymmetric information flow of developed countries than that of emerging countries. Dimpfl and Peter [21] analyzed the information flow between the credit default swap market and the corporate bond market using the TE of 27 iTraxx companies before and after the financial crisis, and discovered the dominance of CDS market over the corporate bond market, the increment of information transmission between markets over time, and the highest importance of the CDS market in the crisis period. Sandoval [22] applied ETE analysis to 197 global financial companies selected based on the amount of market capitalization. Through assessing influence among companies as well as their network structure, he revealed the most significant impact and importance of banks and insurance companies of European and the US to the global financial markets. Chunxia *et al.* [23] analyzed the correlation between the information flow and trading volume through TE among ten sectors of the US market. As the effect of the financial crisis intensified, the information flow between sectors increased, and the financial sector always showed a massive outflow of the TE at all intervals. In particular, it is confirmed that the main information flow from the financial sector changed before and after the financial crisis. Lim *et al.* [24] analyzed the information flow of the CDS and stock market in the US through the TE into the inter and intra structure aspects, and confirmed a substantial change in the information transfer during the financial crisis as well as the precedence of the sudden change of transfer entropy in the CDS market than that of the stock market. Yue *et al.* [25] analyzed the information flows among the sectors in the Chinese stock market using the transfer entropy. They used the maximum spanning arborescence to extract information flow and the hierarchical structure of the networks. They identified the information source and sink sectors and observed that the root node sector acts as an information sink of the incoming information flow networks.

Table 1 summarizes the methodology of previous studies. In general, the previous studies tend to compress information by choosing three to five bins in the binning process to discretize the stock returns. Furthermore, the static analysis of certain time intervals has studied in the financial network analysis to identify non-linear Granger-causal relationships. Then, the studies have devoted to demonstrating the empirical findings rather than its practical application. Therefore, in this study, we set 22 bins [22] to reflect as much information in stock returns as possible. Then, we utilize the moving window method to analyze the time-varying dynamics of ETE and apply its values in predicting the direction of stock prices.

### B. STOCK PRICE PREDICTION BASED ON MACHINE LEARNING

In recent years, as the use of machine learning algorithms has attracted attention in academia, many studies have attempted and discovered the utilization of various machine learning algorithms and their adequacies in price prediction in financial markets. Note that the problem of predicting the direction

**TABLE 1. Outline of previous literature on TE and ETE.**

| Reference | Method | Dataset | moving window | binning method | Prediction |
|---|---|---|---|---|---|
| Marschinski and Kantz (2002) [16] | ETE | SMI | no | symbolic (2 to 5) | no |
| Kwon and Yang (2008) [18] | TE | SMI | no | symbolic (3) | no |
| Kwon and Yang (2008) [19] | TE | SMI / SP | no | symbolic (3) | no |
| Kwon and Oh (2012) [20] | TE | SMI / SP | no | symbolic (2) | no |
| Dimpfl and Peter (2013) [21] | ETE | CDS / COB | cross-section | symbolic(3) | no |
| Sensoy et al. (2014) [17] | ETE | SMI / ER | cross-section | symbolic(3) | no |
| Sandoval (2014) [22] | ETE | SP | no | binning (24) | no |
| Chunxia et al. (2016) [23] | TE | SI | 580 days | symbolic (3) | no |
| Lim et al. (2017) [24] | TE | CDS / SP | 550 days | symbolic (4) | no |
| Yue et al. (2020) [25] | TE | SI | no | binning (15) | no |
| The proposed approach | ETE | SP | 20 / 60 / 120 / 240 days | binning (22) | yes |

*Note: TE and ETE in the method are the abbreviations of the transfer entropy, effective transfer entropy, respectively. SMI, SP, COB, ER, and SI in the dataset are the abbreviations of the stock market index, stock price, corporate bond, exchange rate, and sector index, respectively. The cross-section in the moving window includes the analyses in the pre-crisis, crisis, and post-crisis.

of cumulative returns of stock in the future can be considered as a classification problem. For instance, Ballings *et al.* [36] compared the predictive performance of European stock prices by AUC among the RF, AdaBoost, and Kernel Factory, Neural Networks, Logistic Regression, Support Vector Machines, and K-Nearest Neighbor where the RF showed the highest performance. Patel *et al.* [35] predicted the stock price direction of the Indian stock market using the neural network, support vector machine, RF, and naive-Bayes classifier. At first, 10 technical parameters from the stock transaction data is applied as input parameters. Then, these parameters are converted into trend deterministic data. As a result, the RF showed the best prediction performance. Fischer and Krauss [54] applied the LSTM networks to predict the direction of the stock price for the constituent stocks of the S&P 500 from 1992 until 2015, where the LSTM networks outperformed the memory-free classification methods such as an RF, deep neural net, and logistic regression classifier. Bao *et al.* [55] proposed an ensemble approach consisting of the wavelet transforms, stacked autoencoders (SAEs), and LSTM to predict six stock price indices, namely CSI 300, Nifty 50, Hang Seng index, Nikkei 225, S&P 500, and DJIA. Specifically, the wavelet transforms is utilized to decompose the time series of stock prices to remove the noise; SAEs are used to generate deep high-level features for forecasting; LSTM is applied to predict the next day's stock price based on generated features. Briefly, the primary purposes of previous researches on stock price prediction using machine learning algorithms are twofold: advance in the algorithm and utilization of various variables to improve the prediction performance.

In the perspective of advancing the machine learning algorithms in price prediction, Guo *et al.* [37] proposed the advanced neural network model by incorporating the principal component analysis and radial basis function. In contrast, Qiu and Song [38] proposed a genetic algorithm-based neural network in price prediction. Tsaih *et al.* [29] proposed a rule-based system trading strategies to predict the direction of the S&P 500 stock index futures using the reasoning neural network. The results confirmed that the proposed model outperformed the backpropagation network and perceptron neural network. Tsai and Hsiao [56] proposed a multiple feature selection model combining principal component analysis, genetic algorithm, and decision trees. Then, a back-propagation neural network is applied as a prediction model, which yields the best performance in predicting the price direction of the electronic corporations in Taiwan stock exchange. Kao *et al.* [57] proposed a stock price prediction model by incorporating the wavelet transform, multivariate adaptive regression splines, and support vector regression for various emerging and mature markets. The model solved the wavelet sub-series selection problem and confirmed that the prediction accuracy outperforms the other five competing approaches. In addition, a sub-series selected by the model can identify which points in the past stock price data have had a significant impact on the predictive model configuration. Shin *et al.* [58] has applied the semi-supervised learning, the model has been used for non-time-series type data, to the time series to predict the direction of crude oil prices. In order to apply the existing model to time series prediction, the semi-supervised learning is modified by considering the similarity between different sets of time series data, labels in the stock direction, technical indicator transformation, and feature selection. Ola *et al.* [27] confirmed that daily returns of Tehran Stock Exchange stocks are a chaotic process and that prediction can be improved by applying time-series tests of the local polynomial approximation model. Zahedi and Mahdi [32] applied artificial neural network (ANN) and principal component analysis (PCA) to the 20 accounting variables of stocks listed on the Tehran Stock Exchange to confirm the superiority of stock price forecasting through ANN and the effective factor through the PCA method. Moradi *et al.* [28] analyzed and predicted stock returns by applying the lags coevolving with radial basis function networks (L-Co-R) algorithm for the Tehran Stock Exchange and London Stock Exchange, and uses the Box-Jenkins methods to analyze the Fractal market hypothesis (FMH) between the two exchanges. As a result, they discovered the fact that L-Co-R algorithm is more applicable for long-term time series, that the Box-Jenkins method performs better in short-term time series, and that the FMH is accepted in Tehran Stock Exchange but rejected for London Stock Exchange.

In the perspective of utilizing new input variable, many variables have been proposed including the price-driven measure [46], [59], volatility-driven measure [39], [46], and Google trends [40], which extended to the text mining-based variables from the financial news [41]–[45]. Specifically, Khansa and Liginlal [60] analyzed the relationship between the stock returns of information security firms and the intensity of malicious attacks using artificial neural networks and vector autoregression analysis. The results confirmed that the malicious intensity has a one-month-lagged positive effect on the stock price return of information security firms, and the time-delayed artificial neural network acts as a complementary approach to the existing VAR analysis. In this milieu, the time-delayed artificial neural network showed 95% accuracy while the regression counterpart only showed 85%. Nam and Seong [61] performed the prediction on stock price direction using various machine learning algorithms by incorporating the Granger-causality of the financial news data in the Korean market. Note that the TE is applied in Granger-causality analysis and multiple kernel learning to combine features of target firms and Granger-causal firms. The results confirmed that the proposed model outperformed the benchmarks and verified that the direction of stock price could be predicted based on the news of the Granger-causal firms even when the target firms had no news.

In summary, the RF has been recognized as a decent classifier that has been heavily used in stock price prediction. Also, the XGB is also a popular boosting-type ensemble classifier used for classification problems [62], [63].

The LSTM, a model that improves the exploding and vanishing gradient problem in the recurrent neural network, has shown decent performances in sequence learning and time series prediction, which eventually leads to studied in financial time series prediction [64], [65]. Thus, in this study, we focus on discovering a useful input variable in predicting the stock price direction by utilizing the ETE-driven network indicator based on five representative machine learning algorithms: LR as a traditional predictive model, MLP as a back-propagated neural network, RF as a bagging-type ensemble method, XGB as a boosting-type ensemble method, and LSTM as a single classifier model.

## III. METHODS

### A. EFFECTIVE TRANSFER ENTROPY

#### 1) TRANSFER ENTROPY

TE [11] is a non-parametric indicator to measure the Granger-causal relationship between two processes. Note that the TE can detect the non-linear Granger-causal relationship. It is widely used in fields such as social networks, neuroscience, and financial market analysis due to its efficient detection of the asymmetric interaction in the system. When two variables interact with each other, a time series of one variable $Y$ can affect the future time point of another time series of variable $X$. Let a time series $X$ is a Markov process of degrees $k$; then it refers that the state $x_{n+1}$ of $X$ is affected by $k$ previous states of the same variable, which can be expressed as,

$$p(x_{n+1}|x_n, x_{n-1}, \ldots, x_0) = p(x_{n+1}|x_n, x_{n-1}, \ldots, x_{n-k+1}), \quad x_i \in X \tag{1}$$

where $p(A|B)$ represents a conditional probability of $A$ given $B$, $p(A|B) = p(A, B)/p(B)$. Furthermore, let the state $x_{n+1}$ of $X$ is dependent on the $l$ previous states of $Y$, then the TE from a variable $Y$ to a variable $X$ can be defined as the average information included in $Y$ excluding the information reflected by the past state of $X$ for the next state information of $X$. Therefore, if $X$ and $Y$ denote the amount of information measured by the Shannon entropy($= -\sum_i p_i \log_2 p_i$), and the variable $x_{n+1}$ of $X$ is affected by $k$ previous states of $X$ and $l$ previous states of $Y$, the TE from the variable $Y$ to the variable $X$ can be defined as follows.

$$\begin{aligned} TE_{Y\to X}^{(k,l)} &= \sum_i p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 p(x_{n+1}|x_n^{(k)}, y_n^{(l)}) \\ &\quad - \sum_i p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 p(x_{n+1}|x_n^{(k)}) \\ &= \sum_i p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{p(x_{n+1}|x_n^{(k)}, y_n^{(l)})}{p(x_{n+1}|x_n^{(k)})} \end{aligned} \tag{2}$$

where $i = \{x_{n+1}, x_n^{(k)}, y_n^{(l)}\}$, $x_n^{(k)} = (x_n, x_{n-1}, \ldots, x_{n-k+1})$, $y_n^{(l)} = (y_n, y_{n-1}, \ldots, y_{n-l+1})$, and $p(A, B)$ is the joint probability of A and B.

The definition of TE assumes that events at some point are affected by events of $k$ and $l$ previous states. Based on the previous research [66] showing the low memory in the log-returns of stock prices, we computed the TE under the conditions $k = l = 1$, which expresses the weak form of efficient market hypothesis stating that the current price reflects all past information. Hence, the (2) can be re-defined as follows.

$$\begin{aligned} TE_{Y\to X} &= \sum_i p(x_{n+1}, x_n, y_n) \log_2 \frac{p(x_{n+1}|x_n, y_n)}{p(x_{n+1}|x_n)} \\ &= \sum_i p(x_{n+1}, x_n, y_n) \log_2 \frac{p(x_{n+1}, x_n, y_n) p(x_n)}{p(x_{n+1}, x_n) p(x_n, y_n)} \end{aligned} \tag{3}$$

where $i = \{x_{n+1}, x_n, y_n\}$.

In summary, $TE_{Y\to X}$ is the difference between the information regarding the future value of $X_i$ obtained from $X_i$ and $Y_i$ and the information regarding the future value of $X_i$ obtained only from $X_i$. So, the positive $TE_{Y\to X}$ indicates that the variable $Y$ affects the future value of the variable $X$, which can be interpreted as the intensity of the Granger-causality to the variable $X$ of the variable $Y$. In the same context, $TE_{Y\to X}$ means the degree to which the dynamics of $Y$ affects the transition probability of $X$, which can be seen as the amount of information flow from $Y$ to $X$. Therefore, a large TE value refers to more significant information flow. Also, the TE can measure the amount of in-flow information coming from $Y$ to $X$ through $TE_{Y\to X}$ and the amount of out-flow information going from $X$ to $Y$ through $TE_{X\to Y}$, respectively, based on its asymmetric property.

#### 2) EFFECTIVE TRANSFER ENTROPY

TE is a proper measure to estimate a statistical dependency regardless of the data type. However, a relatively large amount of data is required to derive the transfer entropy. Also, a critical disadvantage of TE is its inclusion of noise due to the finite sample effects and non-stationarity of data. In this context, ETE [16] is proposed to solve the disadvantage of TE. At first, we randomly shuffle the elements of each time series to break the Granger-causal relationship between variables, yet keeping the individual probability distributions of each time series. Then, we obtain transfer entropy from this time series, called the randomized TE (RTE). Finally, we obtain ETE by subtracting the RTE from the original TE to eliminate the noise.

$$ETE_{Y\to X} = TE_{Y\to X} - RTE_{Y\to X} \tag{4}$$

In this study, we choose to use an average of 25 RTE simulations. Also, we set 22 bins to construct the discrete probability distribution for the log-returns of the stock prices. In order to reduce the influence of outliers of log-returns, we integrated the log-returns less than $-6\%$ and greater than $+6\%$ into bin #1 and #22, respectively. Then, the interval between $-6\%$ and $+6\%$ is divided by 0.6% (from bin #2 to #21).

### B. MACHINE LEARNING ALGORITHMS

#### 1) LOGISTIC REGRESSION

LR [47] is a multivariate analysis model used to predict the likelihood of an event using a linear combination of independent variables. In general, the LR is useful when the dependent variable is a binary or a multinomial categorical variable and is mainly used in the financial sector to predict the direction of stock prices or analyze the classification among companies. In this study, the LR was applied as a model representing a linear model, and the stock price direction of the cumulative log-returns was classified based on a cut-off value of 0.5.

#### 2) MULTILAYER PERCEPTRON

MLP is a kind of feed-forward artificial neural network and consists of an input layer, a hidden layer, and an output layer. Each node, except the input nodes, is a neuron that uses a nonlinear activation function and is the most common structure of a neural network learning through back-propagation. In this study, we set the number of hidden layers and their neurons as learning parameters of the MLP. Note that the number of hidden layers is composed of 4, 8, and 16. Then, the number of neurons per layer was defined as follows. In case of 4-hidden layers, the number of neurons for each layer is (256, 128, 64, 32), (128, 64, 32, 16), (64, 32, 16, 8), (32, 16, 8, 4), (16, 8, 4, 2), which yields five sets of hidden layers. In 8-hidden layers, the number of neurons for each layer of the 4-hidden layers is repeated twice (e.g., (256, 256, 128, 128, 64, 64, 32, 32)), In 16-hidden layers, the number of neurons was repeated four times, which yields a total of 15 (3 hidden layers $\times$ 5 neurons) MLPs. Finally, the ReLU was applied to the activation function.

#### 3) RANDOM FOREST

RF [48], a kind of ensemble learning method used for classification and regression analysis, is a method to obtain a classifier with high accuracy and stability by generating a multitude of decision trees and combining each predictive value. Since the RF generates multiple decision trees through bagging and learns by selecting variables randomly for each tree, it is robust against noise and outliers as well as overcomes the disadvantage that the existing single decision tree tends to incur over-fitting according to the training set. Although the training and test time increases as the number of trees increases, a large forest exhibits a relatively continuous result and better generalization ability than a small forest, which increases the stability of the model. In this study, we set the number of trees and variables as learning parameters of a random forest where the numbers of trees are 100, 150, and 200, and the numbers of randomly selected variables are from 8 to 16 by two intervals based on the square root of the lags applied when constructing the data set. In summary, a total of 15 parameter sets (3 trees $\times$ 5 variables) are used.

#### 4) EXTREME GRADIENT BOOSTING

XGB [49] is a gradient boosting algorithm that emphasizes parallel processing and optimization. The gradient boosting machine (GBM) method can be defined as follows. After constructing weak learners, the training set and consistency are evaluated to construct new weak learners using the gradient descent method as the explanatory variable. By repeating such a process, several predictive models are ensembled to construct a strong learner. The GBMs have excellent prediction performance, but there might be an overfitting problem if the proper number of splits is not obtained due to no fitting limit of the number of splits in the decision tree. The XGB is an algorithm that modified the structure of GBM to prevent the overfitting problem through regularization and enabling of parallel computation. In this study, we set the number of trees and depths as learning parameters of the XGB. The number of trees is set to 100, 150, and 200, whereas the depth is set to 2 to 10 by two intervals, similar to RF. Therefore, a total of 15 parameter sets (3 trees $\times$ 5 depths) are utilized. The learning rate was set to 0.1.

#### 5) LONG SHORT-TERM MEMORY NETWORK

LSTM [50], a kind of Recurrent Neural Network (RNN) [67], is a suitable method for classification or prediction based on time series data with unknown gaps between essential events. A typical LSTM unit consists of a cell, input gate, output gate, and forget gate. The cell maintains dependencies between the elements of the input sequence and selectively adjusts the information flow through the gates. The back-propagated error values at the output layer of the unit remain in the LSTM unit's cell, and the error values are continuously supplied to the LSTM unit's gate to learn the cut-off value. Through this process, LSTM networks can deal with vanishing gradient problems that can occur when learning through traditional RNNs. The LSTM network architecture used in this paper is defined as follows. The input size and hidden units are the numbers of features of the input data. The output is binary(up/down) data based on the cumulative log-return of stock during the prediction period. The layers consist of a fully connected layer of size 2, followed by a softmax layer and a classification layer. Specifically, we select epoch and batch size as the learning parameters of LSTM networks. Through trial and error, the epoch is set to 10, 20, 30 times, and the batch size is set to 100 intervals from 100 to 500, which yields a total of 15 parameter sets (3 epoch $\times$ 5 batch size). The parameter pairs used in machine learnings are summarized in Table 2.

#### 6) ADJUSTED ACCURACY

In finance, the performance of a portfolio is evaluated based on the risk-adjusted return, commonly known as the Sharpe ratio [68]. Therefore, we also measure the prediction performance of each algorithm based on a concept similar to the Sharpe ratio. The Sharpe ratio measures the excess return per

**TABLE 2. Parameter set-ups for machine learning algorithms.**

| Model | Parameters | Levels |
|---|---|---|
| MLP | Hidden layers<br>Neuron of first layer | 4, 8, 16<br>256, 128, 64, 32,16 |
| RF | Tree<br>Variable | 100, 150, 200<br>8, 10, 12, 14, 16 |
| XGB | Tree<br>Depth | 100, 150, 200<br>2, 4, 6, 8, 10 |
| LSTM | Epoch<br>BatchSize | 10, 20, 30<br>100, 200, 300, 400, 500 |

unit of deviation, usually referred to as risk, and is defined as,

$$S_a = \frac{E[R_a - R_f]}{\sigma_a} = \frac{E[R_a - R_f]}{\sqrt{var[R_a - R_f]}} \tag{5}$$

where $R_a$ is the asset return, $R_f$ is the risk-free return, and $\sigma_a$ is the standard deviation of the asset excess return.

In this study, we propose an adjusted accuracy as a prediction performance measure on the direction of the stock price. Note that the previous studies mainly compare the absolute value of accuracy in evaluating the results. In case of adjusted accuracy, we define $R_a$, $R_f$ and $\sqrt{var[R_a - R_f]}$ as the prediction accuracy of a single stock, the benchmark accuracy 0.5 from the expectation of binary prediction, and the standard deviation of $R_a - R_f$, respectively.

## IV. DATA AND EXPERIMENT SET-UPS

### A. DATA

In this study, the classification of industry and its constituent stocks are based on *The MSCI(Morgan Stanley Capital International) USA IMI(Investable Market Index) Sector Indexes* as of December 31, 2018. This index covers approximately 99% of the 2,400 large-, mid-, and small-cap stocks in the US stock market and divides stocks into 11 sectors based on the Global Industry Classification Standard (GICS®).

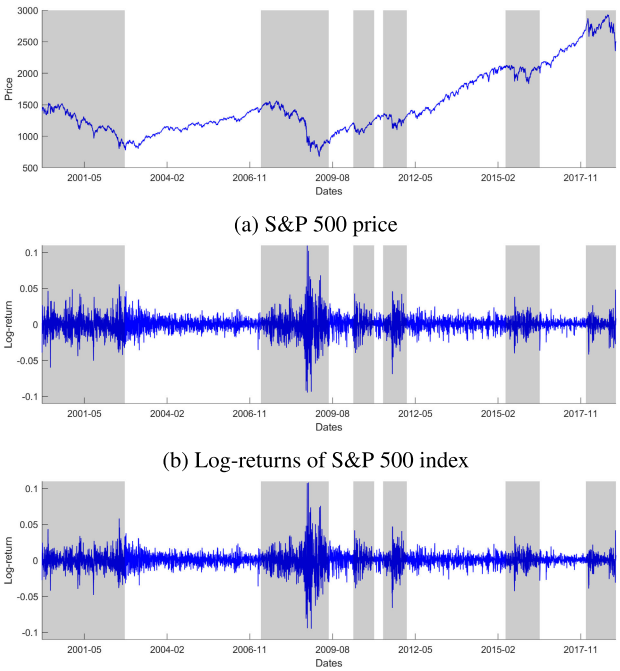
Based on the stocks in which the company continuously exists within the period of the experiment, we utilize a total of 55 stocks, the top five stocks by market capitalization in each sector. Note that all the stocks are listed on the NASDAQ and NYSE. The data period is from January 3, 2000 to December 31, 2018, which yield 4,779 daily adjusted stock prices of 55 stocks. The data is extracted from Thomson Reuters Datastream. The descriptive statistics of the log-returns of entire data are summarized in Table 3.

### B. EXPERIMENT SET-UPS

#### 1) CRISIS AND NON-CRISIS PERIODS

At first, we select the six major events to observe changes of ETE during the financial crisis a follows: Dot-com bubble, Subprime mortgage crisis, European crisis, US debt-ceiling crisis, Brexit, and US-China trade war. Minutely, the Dot-com bubble occurred in the US market when the Internet began to spread as a new business model. Therefore, the crisis period is set from the beginning of 1995 to Sept 30, 2002, when most dot-com companies went bankrupt, and Nasdaq reached its lowest point. The subprime mortgage crisis is set from Apr 1, 2007, when New Century, the US's second-largest subprime mortgage lender, requested filing for bankruptcy protection, to the first half of 2009 (Jun 30, 2009), when US Congress announced the American Recovery and Reinvestment Act of 2009. The European crisis is set from Apr 23, 2010, when the Greek government requested financial assistance to the EU and the IMF to the end of 2010 (Dec 31, 2010). The US debt-ceiling crisis is set from Apr 18, 2011, when the S&P, a renowned credit rating company, announced its first negative view in history on the US AAA sovereign-debt rating to Jan 31, 2012, when the US set an extremely low-interest rate plan to deal with the fall of its rating. The Brexit is set from May 7, 2015, the day of the 2015 UK election, where the conservative party which insisted a hold on the referendum on Brexit won the election, to Jun 23, 2016, the day of the Brexit referendum. Lastly, the US-China trade war is set from January 2018, when the US began to impose sanctions on Chinese companies, to the end of the experimental period.

From the perspective of stocks and events, we check whether the 55 stocks employed in this study could represent the US market. Then, we choose the S&P 500 as a representative indicator of the entire US market in order to examine the actual relationship between the selected financial crisis and the US market. The S&P 500 price, log-returns of S&P 500, and the average log-returns of 55 stocks are shown in Figure 1.




(a) S&P 500 price

(b) Log-returns of S&P 500 index

(c) The average log-returns of 55 stocks

**FIGURE 1. Price and log-returns of S&P 500, and the average log-returns of 55 stocks.**

**TABLE 3.** Descriptive statistics for representative stocks in S&P 500.

| SECTOR | COMPANY | TICKER | Mean | Std | Skew | Kurt | J-B | ADF |
|---|---|---|---|---|---|---|---|---|
| Telecommunication Services (TELE) | Verizon Communications | VZ | 0.0002 | 0.0156 | 0.16 | 9.88 | 9445*** | -71.1*** |
| | Disney (Walt) | DIS | 0.0003 | 0.0189 | -0.09 | 12.85 | 19330*** | -71.5*** |
| | At&T | T | 0.0001 | 0.0162 | 0.03 | 10.47 | 11101*** | -69.4*** |
| | Comcast Corp A | CMCSA | 0.0002 | 0.0209 | 0.00 | 11.35 | 13883*** | -74.5*** |
| | 21St Century Fox A | FOXA | 0.0003 | 0.0214 | 0.25 | 11.79 | 15435*** | -70*** |
| Consumer Discretionary (COND) | Amazon.Com | AMZN | 0.0006 | 0.0331 | 0.45 | 15.15 | 29545*** | -68.9*** |
| | Home Depot | HD | 0.0003 | 0.0196 | -0.98 | 26.37 | 109504*** | -68.6*** |
| | Mcdonald'S Corp | MCD | 0.0004 | 0.0145 | -0.14 | 9.71 | 8987*** | -70.7*** |
| | Nike B | NKE | 0.0008 | 0.0190 | -0.27 | 14.82 | 27867*** | -70*** |
| | Starbucks Corp | SBUX | 0.0007 | 0.0213 | 0.15 | 10.36 | 10798*** | -74.5*** |
| Energy (ENRG) | Exxon Mobil Corp | XOM | 0.0002 | 0.0152 | 0.02 | 13.50 | 21943*** | -76.1*** |
| | Chevron Corp | CVX | 0.0004 | 0.0160 | 0.05 | 13.66 | 22628*** | -73.7*** |
| | Conocophillips | COP | 0.0005 | 0.0189 | -0.31 | 8.48 | 6055*** | -71.6*** |
| | Eog Resources | EOG | 0.0007 | 0.0239 | -0.04 | 7.70 | 4390*** | -70.6*** |
| | Schlumberger | SLB | 0.0001 | 0.0223 | -0.34 | 8.93 | 7085*** | -71.4*** |
| Health Care (HLCA) | Johnson & Johnson | JNJ | 0.0003 | 0.0121 | -0.62 | 19.15 | 52207*** | -69.6*** |
| | Pfizer | PFE | 0.0002 | 0.0157 | -0.24 | 8.56 | 6201*** | -70.7*** |
| | Unitedhealth Group | UNH | 0.0008 | 0.0197 | 0.26 | 23.56 | 84202*** | -70*** |
| | Merck & Co | MRK | 0.0002 | 0.0173 | -1.43 | 32.35 | 173114*** | -68.9*** |
| | Abbott Laboratories | ABT | 0.0005 | 0.0149 | -0.48 | 12.01 | 16361*** | -68.8*** |
| Industrials (INDS) | Boeing Co | BA | 0.0005 | 0.0189 | -0.28 | 8.82 | 6803*** | -69.4*** |
| | 3M Co | MMM | 0.0004 | 0.0145 | -0.02 | 8.44 | 5884*** | -71.9*** |
| | Union Pacific Corp | UNP | 0.0007 | 0.0178 | -0.25 | 6.96 | 3164*** | -71.2*** |
| | Honeywell International | HON | 0.0003 | 0.0196 | -0.24 | 17.02 | 39193*** | -69.1*** |
| | United Technologies Corp | UTX | 0.0004 | 0.0168 | -1.23 | 29.76 | 143791*** | -73.7*** |
| Materials (MTRS) | Dowdupont | DWDP | 0.0002 | 0.0220 | -0.21 | 9.83 | 9335*** | -71.8*** |
| | Ecolab | ECL | 0.0005 | 0.0147 | -0.08 | 9.69 | 8911*** | -75.4*** |
| | Air Products & Chemicals | APD | 0.0005 | 0.0173 | -0.16 | 8.72 | 6543*** | -70.3*** |
| | Sherwin-Williams Co | SHW | 0.0007 | 0.0177 | -0.53 | 17.93 | 44571*** | -75*** |
| | Newmont Mining Corp | NEM | 0.0001 | 0.0255 | 0.19 | 7.69 | 4405*** | -74.1*** |
| Financials (FINC) | Jpmorgan Chase & Co | JPM | 0.0003 | 0.0246 | 0.26 | 17.05 | 39350*** | -74.7*** |
| | Berkshire Hathaway B | BRK-B | 0.0004 | 0.0140 | 0.74 | 15.44 | 31264*** | -69*** |
| | Bank Of America Corp | BAC | 0.0001 | 0.0292 | -0.35 | 30.17 | 147036*** | -69.6*** |
| | Wells Fargo & Co | WFC | 0.0003 | 0.0238 | 0.87 | 31.06 | 157326*** | -76.1*** |
| | Citigroup | C | -0.0004 | 0.0309 | -0.54 | 45.06 | 352428*** | -65.1*** |
| Consumer Staples (CONS) | Procter & Gamble Co | PG | 0.0002 | 0.0135 | -4.10 | 113.51 | 2444856*** | -71.9*** |
| | Coca Cola (The) | KO | 0.0003 | 0.0130 | 0.03 | 12.06 | 16343*** | -69.8*** |
| | Pepsico | PEP | 0.0003 | 0.0123 | 0.03 | 14.64 | 26955*** | -72.6*** |
| | Walmart | WMT | 0.0001 | 0.0150 | 0.10 | 10.07 | 9962*** | -71.4*** |
| | Altria Group | MO | 0.0007 | 0.0153 | -0.04 | 16.26 | 35012*** | -70.5*** |
| Information Technology (INFT) | Apple | AAPL | 0.0009 | 0.0269 | -4.36 | 121.20 | 2796451*** | -71.1*** |
| | Microsoft Corp | MSFT | 0.0002 | 0.0193 | -0.14 | 12.66 | 18580*** | -72.4*** |
| | Intel Corp | INTC | 0.0001 | 0.0235 | -0.47 | 12.08 | 16576*** | -72.1*** |
| | Cisco Systems | CSCO | 0.0000 | 0.0245 | 0.14 | 12.15 | 16669*** | -73.3*** |
| | Oracle Corp | ORCL | 0.0001 | 0.0248 | -0.02 | 11.52 | 14453*** | -74.6*** |
| Utilities (UTIL) | Nextera Energy | NEE | 0.0006 | 0.0140 | 0.16 | 12.70 | 18740*** | -71.9*** |
| | Duke Energy Corp | DUK | 0.0003 | 0.0153 | -0.24 | 14.98 | 28599*** | -72.1*** |
| | Dominion Energy | D | 0.0005 | 0.0133 | -0.61 | 13.01 | 20251*** | -70.2*** |
| | Southern Company (The) | SO | 0.0005 | 0.0119 | 0.23 | 9.55 | 8584*** | -73.6*** |
| | Exelon Corp | EXC | 0.0004 | 0.0161 | -0.05 | 11.14 | 13203*** | -72.6*** |
| Real Estate (REES) | American Tower Corp | AMT | 0.0004 | 0.0302 | -0.51 | 27.77 | 122317*** | -64.1*** |
| | Simon Property Group | SPG | 0.0006 | 0.0212 | 0.24 | 22.49 | 75696*** | -83*** |
| | Crown Castle Intl Corp | CCI | 0.0003 | 0.0299 | -0.30 | 28.10 | 125513*** | -65.1*** |
| | Prologis | PLD | 0.0004 | 0.0250 | -1.00 | 35.39 | 209679*** | -80.1*** |
| | Public Storage | PSA | 0.0006 | 0.0187 | 0.12 | 19.07 | 51452*** | -82.1*** |

*Note: Std, Skew, Kurt, J-B, ADF are the abbreviations of the standard deviation, skewness, kurtosis, Jarque-Bera tests, augmented Dickey-Fuller tests, respectively. Also, the star superscripts *, **, *** refer to 5%, 1%, and 0.1% statistical significances, respectively.

The result shows that the average log-returns are similar to that of the S&P 500. Furthermore, we set the selected events as gray boxes. In the gray boxes, the volatility of log-returns in the US market is relatively high, indicating that the selected events represent the financial crises in the US market.

To analyze the time-varying property of ETE, we use the moving window method and set four different sizes: 1 month (1M, 20 days), 3 months (3M, 60 days), 6 months (6M, 120 days), and 1 year (1Y, 240 days). In this regard, we can obtain 55 × 55 ETE matrix for 55 stocks for each

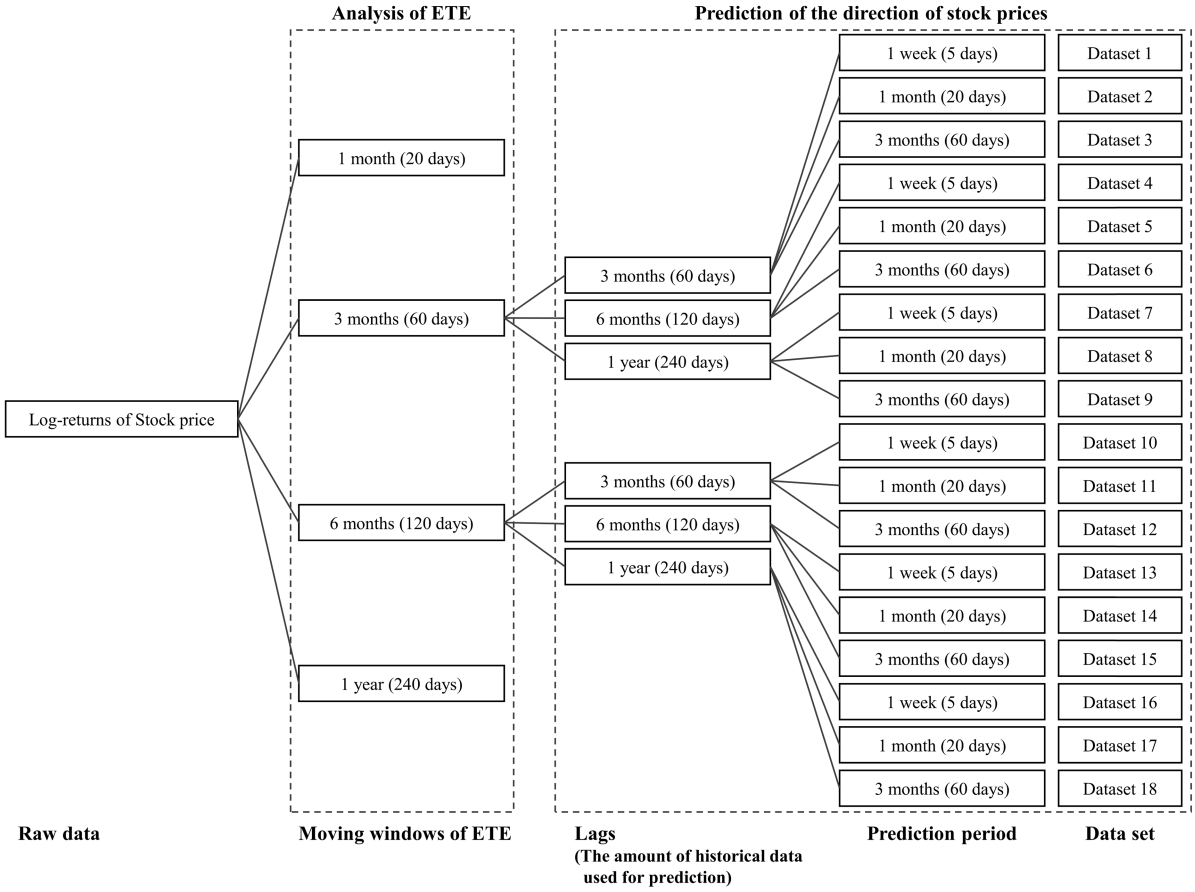



**FIGURE 2.** Dataset structure used to compute the ETE and to perform the prediction.

time point and derive a time series of the total mean value of ETE at each time point. Then, we examine the evolutions of ETE with the major US financial crises to explore appropriate sizes of moving windows size.

For the selected appropriate moving window-based ETE, we derive the average of inflow ($ETE_{Y\to X}$) and outflow ($ETE_{X\to Y}$) values at each time point and define these two indicators as the ETE network indicator. Furthermore, we perform time series analysis and interval analysis of information flow by sector through the time series of ETE network indicators.

### 2) MACHINE LEARNING FRAMEWORK

We construct the training dataset based on a various mixture of input and target variables. The input variable($x$) at each time point for each stock consist of three cases in two variables: ETE network indicators for inflow and outflow and lags for 3 months (3M, 60 days), 6 months (6M, 120 days), and 1 year (1Y, 240 days) of time series of each stock's past log-returns. Then, the target variable($y$) is the cumulative log-returns in different prediction periods for 1 week (1W, 5 days), 1 month (1M, 20 days), and 3 months (3M, 60 days).

Note that the categorized (positive & negative) cumulative log-returns are used for algorithms. This process is then repeated for the different ETEs obtained from different sizes of moving windows. Figure 2 summarizes the dataset structure used in this study.

For each dataset, we compare the prediction performances of the mixture of log-returns and ETE network indicator against that of plain log-returns. In this regard, we aim to check the following statements in terms of prediction performance: (1) the validity of ETE network indicators and (2) the most applicable machine learning algorithm. Specifically, the learning is performed as follows. At first, the entire data is divided into 50% of training and 50% of test sets. Then, we normalize the stock log-returns and inflow and outflow ETEs based on their means and standard deviations of each training set. The associated periods of training set for 3M and 6M moving windows are from 2000-01-03 to 2009-08-17 and from 2000-01-03 to 2009-09-29, respectively. Note that we only train 70% randomly sampled set from the training set to achieve the generalized training results. Then, we compare the prediction performance of test set for 3M and 6M, whose periods from 2009-08-18 to 2018-12-31 and from 2009-09-30 to 2018-12-31, respectively. Figure 3 shows an example of the proposed machine learning framework.

We define a prediction accuracy to reflect a proportion of correctly classified direction of the cumulative log-returns in the test set.

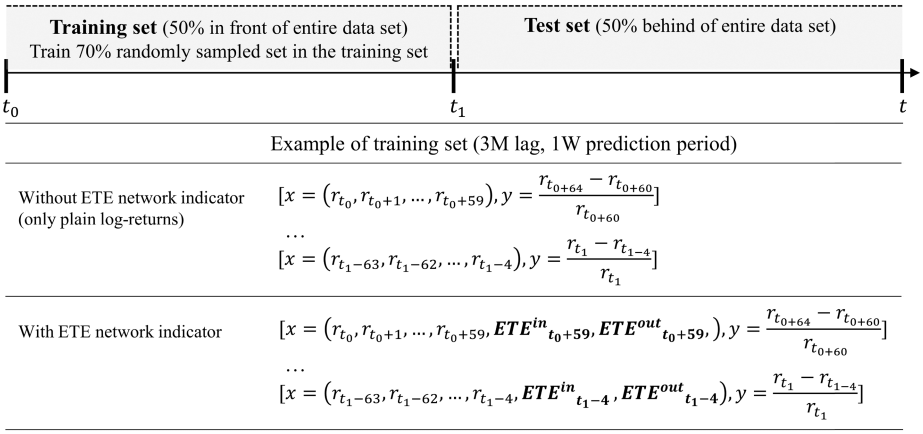



**FIGURE 3.** A scenario of the proposed machine learning framework.

Based on the prediction accuracy, we can define the prediction performance of each machine learning algorithm. For LR, the performance of the algorithm is the same as the prediction accuracy since the LR does not possess any model parameters. In contrast, for MLP, RF, XGB, and LSTM, we set 15 parameter sets, which yield 15 different results of prediction accuracy.

Based on the previous studies [35], [39], the prediction accuracy of LR for a single stock is the same as the accuracy of LR, whereas those of other four models are defined as the averages of top five results in prediction accuracy among the 15 parameter sets. In contrast, a model-specific prediction accuracy for each sector is defined as the average of the prediction accuracy of stocks associated in each sector. The same approach is used to define the overall performance of five models. We calculate the averages of the prediction accuracy of all stocks. In case of the improvement of prediction accuracy incurred by ETE network indicator, the LR is defined as the simple difference of each stock, and the MLP, RF, XGB and LSTM are defined as the average of the top five among the accuracy differences between parameter sets. Furthermore, the same approach is applied to the adjusted accuracy, which can be obtained by (5). Finally, the overall framework can be summarized as a step-by-step procedure described in Figure 4.

## V. RESULTS AND DISCUSSIONS

### A. EFFECTIVE TRANSFER ENTROPY

Figure 5 shows the effective transfer entropy of 55 major stocks in the US market for different sizes of moving windows. For each moving window, we plot the mean of TE, RTE, and ETE matrix that consists of the entire 55 stocks. In all the moving windows, we confirm that the ETE, a noise-reduced TE through RTE, tends to be more stable than a simple TE. Specifically, the ETE of the 1M moving window shows a noise-shaped oscillation, which fails to detect any particular changes in financial time-series. In contrast, the ETE of the 1Y moving window is too smooth to provide the information for the entire period, where roughly 50% is filled with zero. Therefore, we focus on the analysis of ETEs of 3M and 6M moving windows.

Figure 6 shows the average ETEs of 3M and 6M moving windows by presenting the six financial crisis as gray regions. The average ETE is relatively high in the gray regions for both cases, which suggests the implication of financial crisis based on the increased ETE. Note that the average ETE of 3M moving windows is usually smaller than that of 6M moving window, whereas the average ETE of 3M shows more volatile movement with more detailed information than that of 6M moving windows ETE. Since the ETEs of 3M and 6M seem to possess meaningful information regarding the financial markets, we further investigate the characteristics of ETE in terms of the detection of its evolution pattern and cross-sectional analysis on the financial crisis in the level of financial sectors.

Furthermore, we plot the average ETE of a sector for the total, crisis, and non-crisis periods in Figure 7 as a heatmap to investigate the information flow between individual sectors. Note that the sector ETE is defined as the average of ETEs of stocks associated in the sector. Comparing the heatmaps of the total, crisis, and non-crisis periods in the top of Figure 7a and Figure 7b, the information flow between sectors in the crisis period is significantly higher than that of non-crisis period. Note that the brighter the color, the stronger the information transfer.

Besides, we further investigate the evolutions of inflow and outflow ETE of different sectors for 3M and 6M moving windows in Figure 8. Note that Total indicates the average of all sectors, which equates the evolutions of inflow and outflow. In both 3M and 6M moving windows ETE, the inflow and outflow ETE of the financial crisis period were observed to be larger in all sectors than the non-financial crisis period. Although ETEs of two moving windows generally show the same trend, the difference between inflow and outflow ETE is prominent in 3M moving windows. However, a more stable rise and fall of ETE is observed in 6M than 3M moving windows.

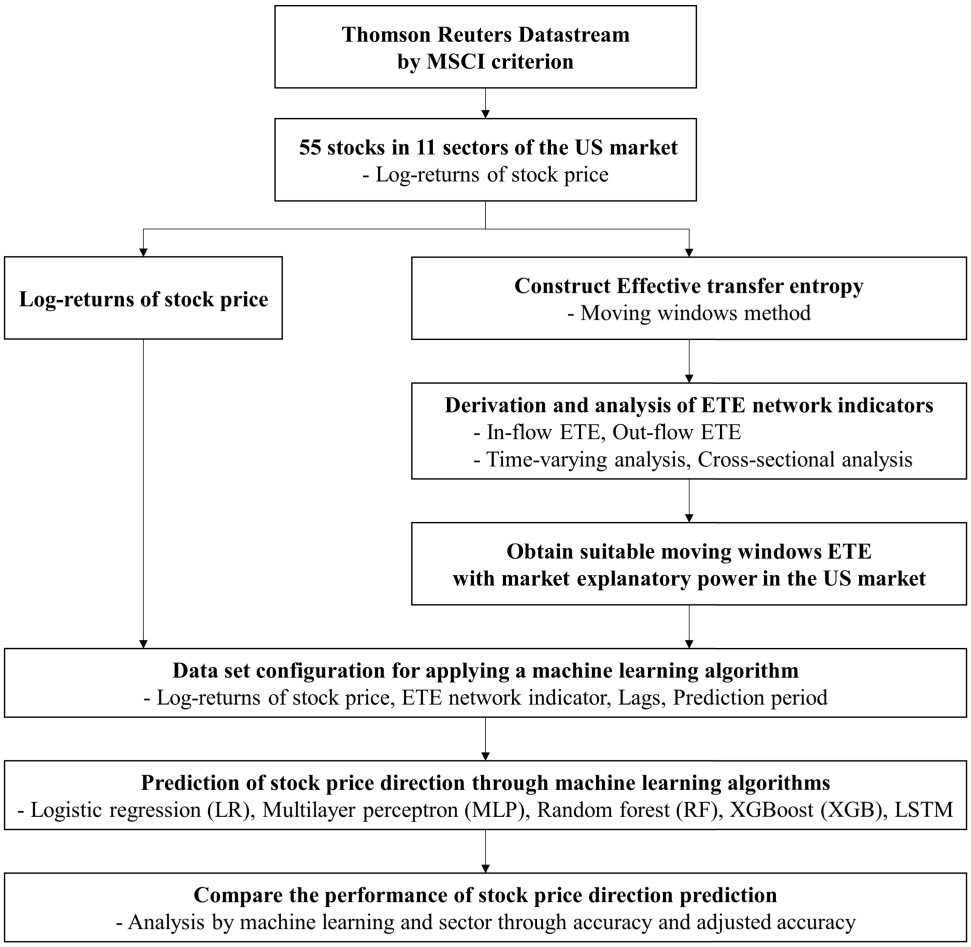



**FIGURE 4.** Procedure for the stock price prediction using the ETE.

Also, in order to check the strength of the information flow of each sector, we plot the values of outflow minus inflow in Figure 9. Note that a positive value indicates a more substantial outgoing information transfer from a specific sector to other sectors that incoming information transfer and vice versa. At first, the difference between inflow and outflow for total is zero as shown in Figure 8. In the case of 3M moving windows in Figure 9a, TELE, COND, and INFT show the most positive values in the dot-com bubble located in the first gray region from the left. In particular, INFT shows the most persistent and large positive values. On the contrary, in the same dot-com bubble period, ENRG, HLCA, and UTIL exhibit negative values most time.

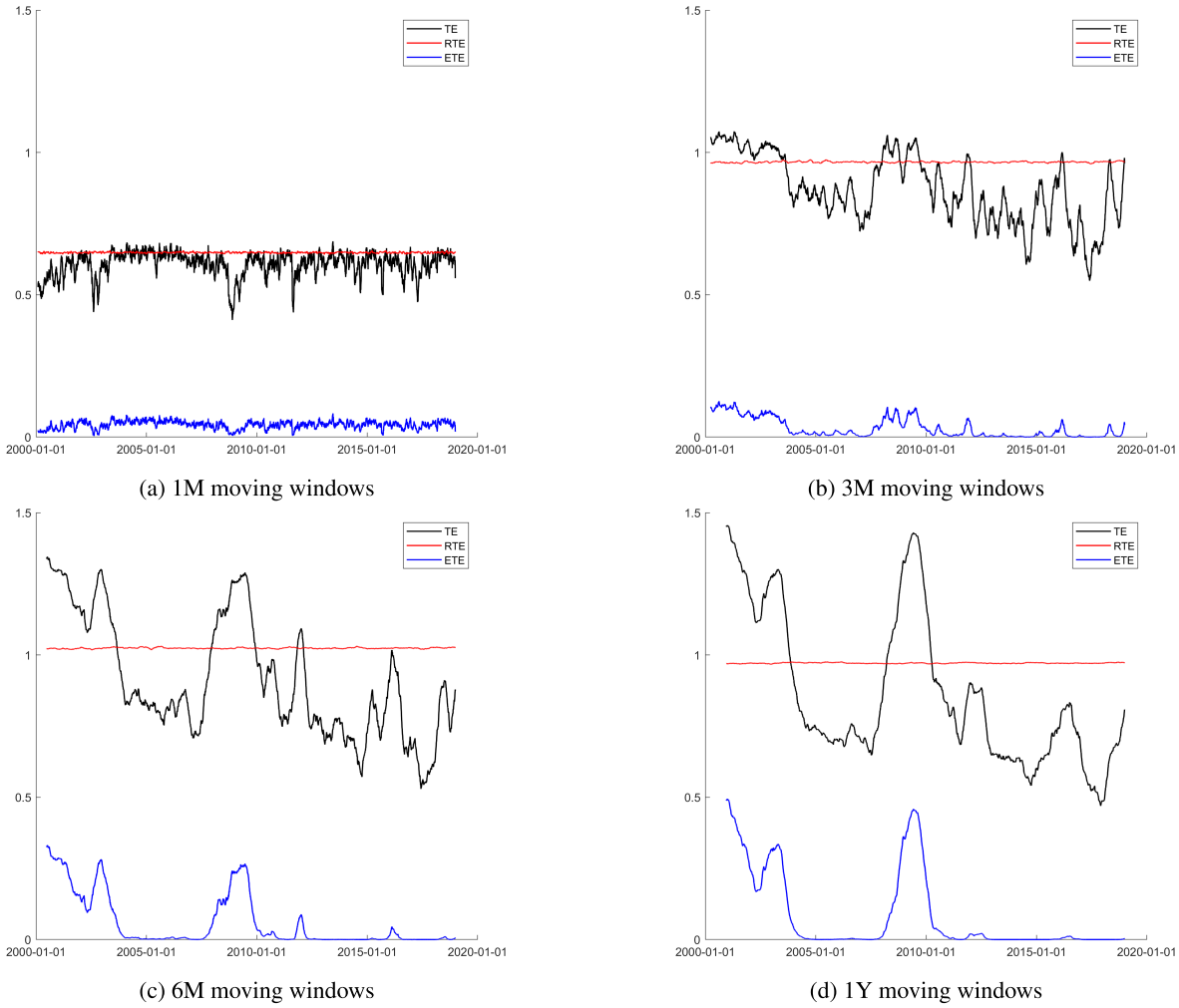
Given that the period is a dot-com bubble, the strong outgoing information transfer from the INFT is because the period is closely related to the bubble economy of information technology companies. In the Subprime mortgage and European crises, FINC and REES show the most positive values, whereas CONS, INFT, and UTIL show the most negative values. Again, FINC and REES are directly related to the subprime mortgage crisis, and the FINC is the sector most affected by the European crisis.

Similar to 3M moving windows ETE, 6M moving windows ETE in Figure 9b shows different directions and intensity of information transfer for each sector. In conclusion, it is confirmed that outflow and inflow ETEs show the characteristics of each sector according to a different time, and such information could be useful in predicting the direction of future stock prices.

### B. PREDICTION PERFORMANCE IN DIFFERENT MODELS AND PARAMETER SETS

We intend to utilize the evolution of ETE, which has a market explanatory power, as an input variable in predicting the direction of future stock price based on the LR, MLP, RF, XGB and LSTM. Especially, the prediction performance of the model using the inflow and outflow ETE values, which can be defined as ETE network indicators, as input variables are compared with that of the model without the ETE values. Table 4 summarizes the average prediction accuracy of all models for different moving windows, lags, and prediction periods.

The results show that there is no significant difference in prediction accuracy among different moving windows. As the




(a) 1M moving windows

(b) 3M moving windows

(c) 6M moving windows

(d) 1Y moving windows

**FIGURE 5.** Evolution of TE, RTE, and ETE for different moving windows.

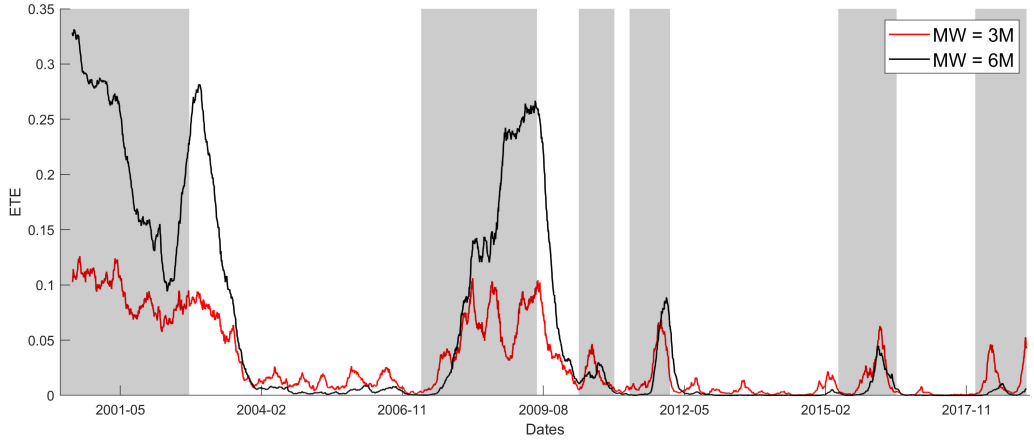
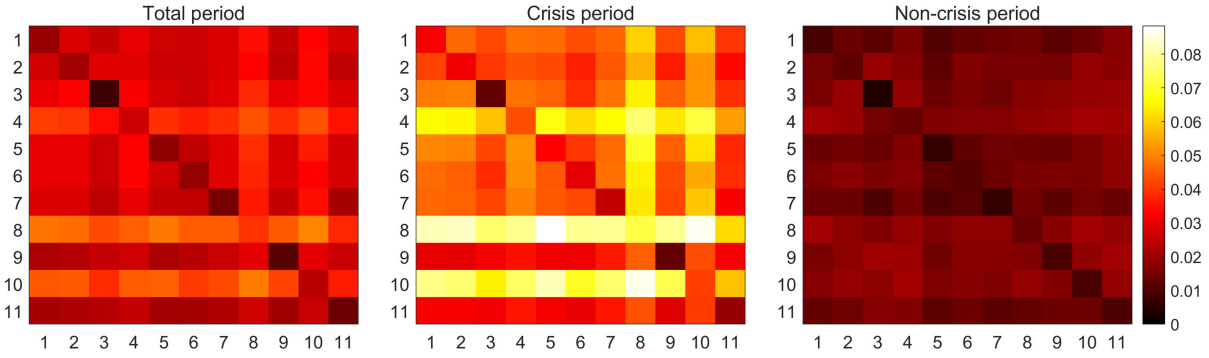


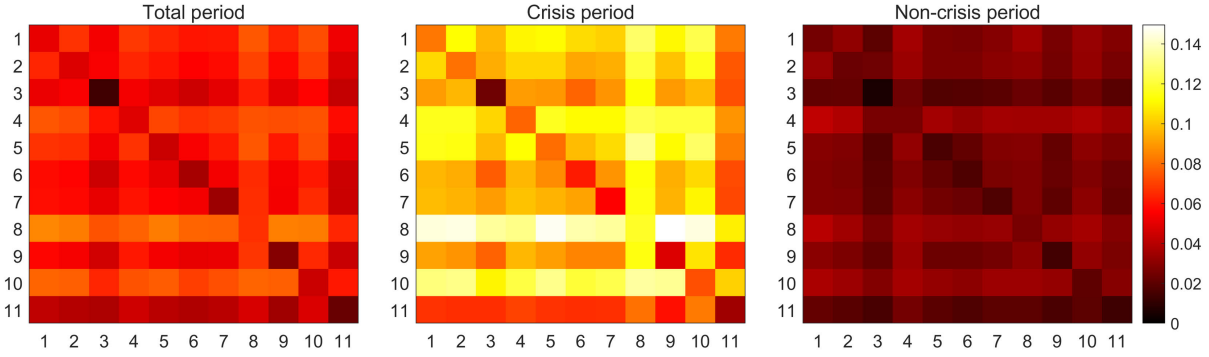

**FIGURE 6.** Evolution of average ETE in 3M and 6M moving windows.

lag increases, the prediction accuracy, in general, tends to decrease slightly, whereas the prediction accuracy tends to increase in the long term rather than the short term. More importantly, the improvement of prediction performance by utilizing the ETE network indicator is detected in all parameter sets. Also, the improvement tends to increase as the prediction period increases.

Therefore, we suggest that the ETE network indicators can be used to improve the performance of the stock price forecast in the US market.




(a) 3M moving windows ETE




(b) 6M moving windows ETE

**FIGURE 7.** Heatmaps of the average ETE in different periods. **Note: Numbers from 1 to 11 sequentially represent the TELE, COND, ENRG, HLCA, INDS, MTRS, FINC, CONS, INFT, UTIL, REES, respectively.

**TABLE 4.** Overall accuracies in different dataset structure.

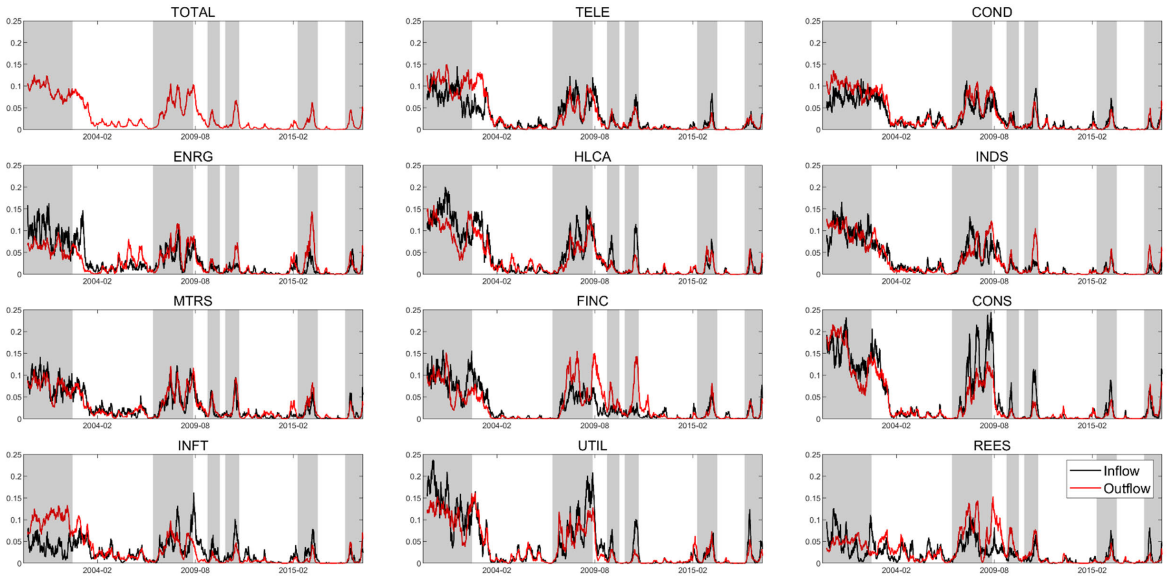
| MW | Lag | With ETE (%) | | | Without ETE (%) | | | Improvement (%p) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M |
| 3M | 3M | 53.53 | 56.68 | 60.90 | 53.30 | 56.50 | 60.87 | 1.19 | 1.41 | 1.85 |
| | 6M | 53.25 | 56.46 | 60.66 | 53.05 | 56.18 | 60.22 | 1.14 | 1.44 | 2.01 |
| | 1Y | 52.98 | 55.77 | 59.88 | 52.75 | 55.54 | 59.13 | 1.18 | 1.35 | 2.31 |
| 6M | 3M | 53.59 | 56.89 | 61.04 | 53.29 | 56.95 | 61.08 | 1.31 | 1.22 | 1.67 |
| | 6M | 53.38 | 56.86 | 60.67 | 53.08 | 56.40 | 60.45 | 1.23 | 1.65 | 1.89 |
| | 1Y | 53.00 | 55.86 | 60.10 | 52.75 | 55.45 | 59.44 | 1.22 | 1.64 | 2.34 |

*Note: MW, with ETE, without ETE, and Improvement are the abbreviations of the moving window, the performance when the stock price direction is predicted only by its plain log-returns, the performance when predicted by the mixture of log-returns and the ETE network indicator, and improvement of prediction accuracy incurred by ETE network indicator, respectively.

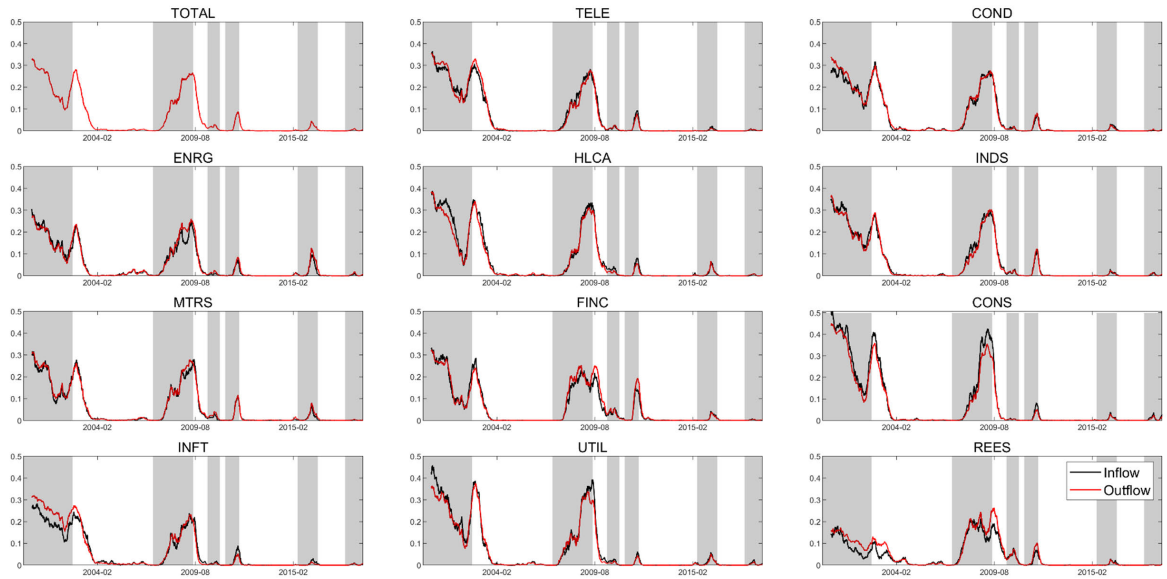
Besides, the detailed results on the prediction accuracy in different models can be investigated by evaluating the prediction accuracy and adjusted accuracy in (5). Table 5 summarizes prediction performances for different machine learning algorithms, moving windows, lags, prediction periods, and integration of ETE indicator. Also, we mark the top five accuracies and the adjusted accuracy of a total of 30 sets of models and parameters.

From the perspective of accuracy, the improvements by integrating ETE network indicators are detected in all models. The longer the prediction period, the higher the accuracy, and there is little difference in accuracy between the moving windows. In LR and LSTM, the smaller the lag, the higher the accuracy. However, the MLP, RF, XGB did not show any change in accuracy for different lags. Since the top five of the accuracy is all present in the RF and most of the top five of the improvement present in the MLP, the absolute prediction performance is highest in the RF and the improvement of the accuracy through the ETE network indicator is the largest in the MLP.

From the perspective of adjusted accuracy, all models except the LR show lower adjusted accuracy over the longer term. Since most of the top five of the adjusted accuracy present in MLP, it implies that the prediction performance of MLP is high and consistent. Specifically, in most of the top five of the improvement, there are two LSTMs in the 1W prediction period, and the rest are all in the MLP. The RF and XGB show the adjusted accuracy improvement observed in




(a) 3M moving windows ETE




(b) 6M moving windows ETE

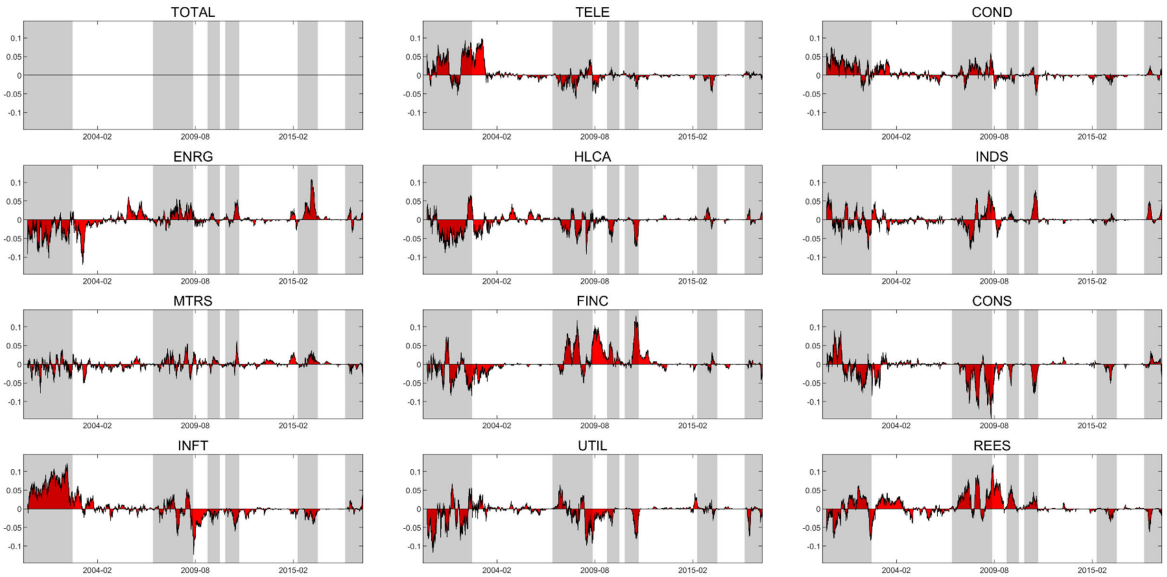
**FIGURE 8.** **Evolution of inflow and outflow ETE in different sectors.**

the 1W prediction period with significantly smaller improvement in the 1M and 3M prediction period. Although the prediction performance of the LR is similar to those of other models in terms of accuracy, the adjusted accuracy shows that the prediction performance of the LR is not stable as those of other models.

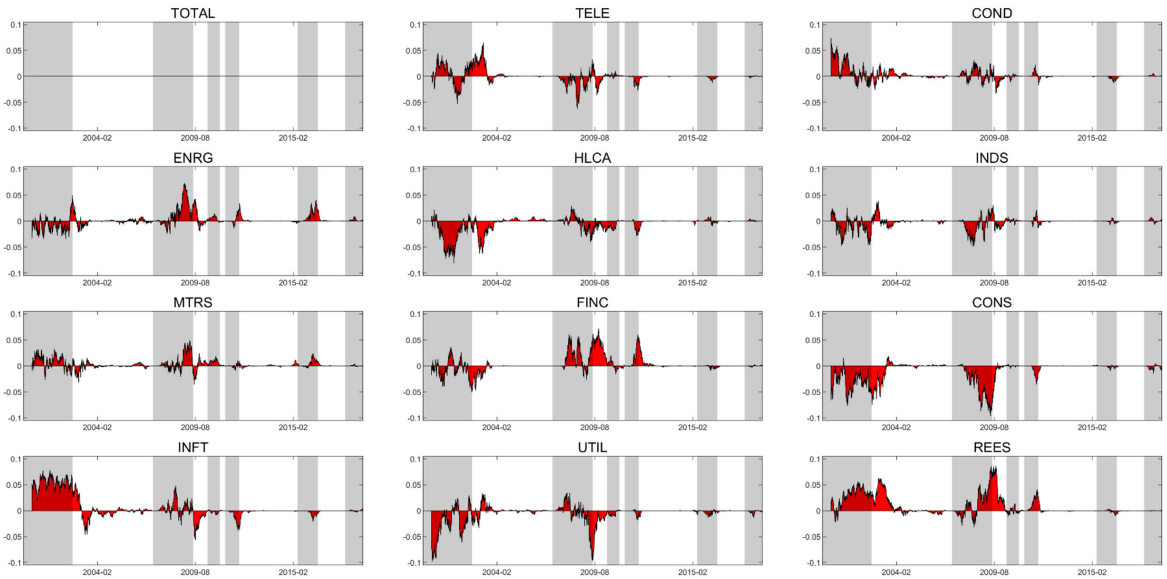
Overall, we discover that all five machine learning algorithms have improved the accuracy through the ETE network indicator and suggest that the MLP and LSTM are the most suitable models for predicting future stock price direction predictions when considered the accuracy and adjusted accuracy simultaneously. Note that the RF has the advantage of high accuracy, whereas the MLP and LSTM have the characteristic of consistent improvement in prediction performance.

### C. PREDICTION PERFORMANCE IN DIFFERENT SECTORS

Since the adequacy of models established, the detailed analysis of the prediction performances in the different sectors is evaluated based on the accuracy and adjusted accuracy of the MLP, RF, XGB, and LSTM. Note that we exclude the LR due to its poor performance in the adjusted accuracy. Table 6 and 7 summarize the accuracy and adjusted accuracy of each sector, respectively, for different moving windows, lags, prediction periods, and integration of ETE indicator.




(a) 3M moving windows ETE




(b) 6M moving windows ETE

**FIGURE 9.** Evolution of 'outflow - inflow' ETE in different sectors.

Table 6 demonstrates that accuracy tends to increase in all four algorithms in a more extended prediction period except for the TELE sector of RF. The improvement of accuracy is observed for all conditions of MLP and almost all sectors in LSTM. For the 1W prediction period of RF and XGB, the improvement of accuracy is generally observed in all sectors regardless of lag. However, the decrease in accuracy is observed in some sectors in 1M or 3M prediction periods.

Table 7 shows that the adjusted accuracy of four algorithms tends to be higher for short-term predictions than long-term. The improvement by using the ETE network indicators tend to vary by sector and condition where TELE, FINC, and INFT show the increase of adjusted accuracy in most conditions.

Furthermore, we summarize the accuracy and adjusted accuracy with their improvements in each sector for different moving windows based on the average of performance for all lags and prediction periods in Table 8. Although the performance of machine learning algorithms differs for each sector, we depict sectors with good performance regardless of the algorithm.

In order to further analyze the prediction performance on each sector, we draw the scatter plots for accuracy and adjusted accuracy in Figure 10. Note that the sector-by-sector

**TABLE 5.** Accuracy and adjusted accuracy for different machine learning algorithms.

| Model | MW | Lag | | With ETE | | | Without ETE | | | Improvement | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M |
| LR | 3M | 3M | Acc | 52.89 | 56.40 | 61.12 | 52.29 | 55.84 | 60.81 | 0.61 | 0.56 | 0.32 |
| | | | Acc (adj) | 1.11 | 1.15 | 1.11 | 0.93 | 0.96 | 0.95 | 0.45 | 0.20 | 0.07 |
| | | 6M | Acc | 52.54 | 55.57 | 60.20 | 51.92 | 54.79 | 59.30 | 0.62 | 0.78 | 0.90 |
| | | | Acc (adj) | 1.20 | 1.17 | 1.12 | 0.95 | 0.99 | 0.96 | 0.64 | 0.31 | 0.26 |
| | | 1Y | Acc | 51.86 | 53.94 | 57.96 | 51.42 | 53.38 | 56.73 | 0.44 | 0.56 | 1.23 |
| | | | Acc (adj) | 1.13 | 1.00 | 0.95 | 0.94 | 0.95 | 0.82 | 0.58 | 0.31 | 0.45 |
| | 6M | 3M | Acc | 53.22 | 56.98 | 61.19 | 52.44 | 56.67 | 61.16 | 0.78 | 0.30 | 0.03 |
| | | | Acc (adj) | 1.35 | 1.20 | 1.09 | 1.01 | 1.21 | 1.03 | 0.85 | 0.17 | 0.01 |
| | | 6M | Acc | 52.51 | 56.26 | 60.14 | 52.00 | 55.38 | 59.81 | 0.51 | 0.88 | 0.33 |
| | | | Acc (adj) | 1.20 | 1.38 | 1.08 | 1.13 | 1.28 | 1.06 | 0.54 | 0.54 | 0.11 |
| | | 1Y | Acc | 51.79 | 53.99 | 58.15 | 51.26 | 53.24 | 56.98 | 0.53 | 0.75 | 1.16 |
| | | | Acc (adj) | 1.01 | 0.93 | 0.94 | 0.93 | 0.89 | 0.86 | 0.74 | 0.47 | 0.42 |
| MLP | 3M | 3M | Acc | 53.26 | 56.50 | 60.58 | 53.36 | 56.09 | 60.44 | 2.46† | 3.90† | 5.60† |
| | | | Acc (adj) | 2.23† | 1.79† | 1.44† | 2.23† | 1.65 | 1.36† | 1.97 | 1.81† | 1.47† |
| | | 6M | Acc | 53.29 | 56.47 | 60.35 | 53.24 | 56.01 | 59.72 | 2.36 | 3.67† | 4.95 |
| | | | Acc (adj) | 2.14† | 1.83† | 1.43 | 2.11† | 1.75† | 1.32 | 2.11† | 1.59† | 1.46† |
| | | 1Y | Acc | 53.23 | 56.22 | 60.77 | 53.01 | 56.09 | 59.49 | 2.60† | 3.36 | 6.01† |
| | | | Acc (adj) | 1.99 | 1.86† | 1.44† | 1.96† | 1.91† | 1.39† | 1.89 | 1.89† | 1.41 |
| | 6M | 3M | Acc | 53.25 | 56.16 | 60.81 | 53.16 | 56.42 | 60.01 | 2.55† | 3.47† | 5.62† |
| | | | Acc (adj) | 2.03† | 1.72 | 1.48† | 2.11† | 1.69† | 1.39† | 1.86 | 1.54 | 1.57† |
| | | 6M | Acc | 53.33 | 56.57 | 60.52 | 53.08 | 56.03 | 60.14 | 2.58† | 4.13† | 4.99† |
| | | | Acc (adj) | 2.11† | 1.73† | 1.45† | 2.14† | 1.86† | 1.57† | 2.29† | 1.74† | 1.48† |
| | | 1Y | Acc | 53.35 | 56.23 | 60.97 | 53.09 | 55.98 | 60.19 | 2.72† | 4.05† | 5.81† |
| | | | Acc (adj) | 2.11† | 1.79† | 1.58† | 1.95 | 2.12† | 1.47† | 2.00† | 1.72† | 1.59† |
| RF | 3M | 3M | Acc | 54.65† | 57.88 | 61.94 | 54.61† | 58.18 | 61.99 | 0.66 | 0.09 | 0.29 |
| | | | Acc (adj) | 1.59 | 1.26 | 1.07 | 1.58 | 1.27 | 0.97 | 0.87 | 0.03 | 0.07 |
| | | 6M | Acc | 54.53 | 58.66† | 62.66† | 54.61† | 58.76† | 62.55† | 0.50 | 0.27 | 0.39 |
| | | | Acc (adj) | 1.40 | 1.40 | 1.09 | 1.49 | 1.37 | 1.03 | 0.85 | 0.19 | 0.17 |
| | | 1Y | Acc | 54.76† | 58.78† | 62.85† | 54.79† | 58.80† | 62.67† | 0.54 | 0.31 | 0.52 |
| | | | Acc (adj) | 1.51 | 1.46 | 1.12 | 1.52 | 1.43 | 1.08 | 0.91 | 0.35 | 0.31 |
| | 6M | 3M | Acc | 54.59† | 58.13† | 62.41† | 54.51 | 58.58† | 62.31† | 0.69 | -0.06 | 0.44 |
| | | | Acc (adj) | 1.51 | 1.39 | 1.22 | 1.50 | 1.46 | 1.06 | 0.96 | -0.04 | 0.08 |
| | | 6M | Acc | 54.80† | 58.74† | 62.77† | 54.71† | 58.94† | 62.55† | 0.70 | 0.21 | 0.56 |
| | | | Acc (adj) | 1.81 | 1.47 | 1.18 | 1.71 | 1.53 | 1.07 | 1.13 | 0.18 | 0.15 |
| | | 1Y | Acc | 54.65† | 58.61† | 63.14† | 54.60† | 58.60† | 62.89† | 0.55 | 0.33 | 0.58 |
| | | | Acc (adj) | 1.50 | 1.36 | 1.17 | 1.45 | 1.33 | 1.12 | 1.10 | 0.39 | 0.27 |
| XGB | 3M | 3M | Acc | 52.74 | 55.35 | 59.17 | 52.69 | 55.79 | 60.06 | 0.86 | 0.56 | 0.21 |
| | | | Acc (adj) | 1.40 | 1.17 | 1.07 | 1.44 | 1.17 | 1.02 | 0.93 | 0.23 | 0.05 |
| | | 6M | Acc | 52.70 | 55.53 | 59.77 | 52.83 | 56.03 | 60.41 | 0.81 | 0.52 | 0.31 |
| | | | Acc (adj) | 1.42 | 1.25 | 1.22 | 1.50 | 1.33 | 1.11 | 0.96 | 0.20 | 0.09 |
| | | 1Y | Acc | 52.53 | 55.28 | 59.26 | 52.53 | 55.41 | 59.70 | 0.88 | 0.67 | 0.43 |
| | | | Acc (adj) | 1.26 | 1.17 | 1.12 | 1.31 | 1.24 | 1.04 | 1.21 | 0.34 | 0.12 |
| | 6M | 3M | Acc | 52.92 | 55.67 | 59.44 | 52.78 | 55.96 | 60.52 | 0.99 | 0.66 | -0.11 |
| | | | Acc (adj) | 1.56 | 1.27 | 1.11 | 1.46 | 1.28 | 1.09 | 1.28 | 0.30 | -0.02 |
| | | 6M | Acc | 53.03 | 56.16 | 59.58 | 52.96 | 56.21 | 60.43 | 0.89 | 0.75 | 0.17 |
| | | | Acc (adj) | 1.56 | 1.48 | 1.12 | 1.54 | 1.42 | 1.12 | 1.46 | 0.39 | 0.03 |
| | | 1Y | Acc | 52.65 | 55.67 | 59.62 | 52.65 | 55.55 | 60.01 | 0.94 | 0.90 | 0.58 |
| | | | Acc (adj) | 1.33 | 1.29 | 1.09 | 1.43 | 1.27 | 1.06 | 1.71 | 0.45 | 0.09 |
| LSTM | 3M | 3M | Acc | 54.10 | 57.29 | 61.71 | 53.54 | 56.61 | 61.04 | 1.36 | 1.95 | 2.82 |
| | | | Acc (adj) | 1.76 | 1.49 | 1.25 | 1.62 | 1.27 | 1.11 | 1.30 | 0.80 | 0.69 |
| | | 6M | Acc | 53.16 | 56.08 | 60.34 | 52.65 | 55.33 | 59.13 | 1.43 | 1.97 | 3.50 |
| | | | Acc (adj) | 1.70 | 1.43 | 1.22 | 1.66 | 1.42 | 1.11 | 1.48 | 0.99 | 1.17 |
| | | 1Y | Acc | 52.52 | 54.62 | 58.54 | 52.02 | 54.01 | 57.09 | 1.42 | 1.87 | 3.36 |
| | | | Acc (adj) | 1.73 | 1.40 | 1.17 | 1.65 | 1.61 | 1.08 | 2.45† | 1.18 | 1.39 |
| | 6M | 3M | Acc | 53.98 | 57.49 | 61.37 | 53.53 | 57.14 | 61.40 | 1.53 | 1.71 | 2.33 |
| | | | Acc (adj) | 1.70 | 1.43 | 1.18 | 1.78 | 1.48 | 1.20 | 1.62 | 0.95 | 0.70 |
| | | 6M | Acc | 53.26 | 56.55 | 60.35 | 52.63 | 55.44 | 59.31 | 1.46 | 2.27 | 3.42 |
| | | | Acc (adj) | 1.59 | 1.51 | 1.15 | 1.55 | 1.51 | 1.16 | 1.98 | 1.45 | 1.26 |
| | | 1Y | Acc | 52.58 | 54.81 | 58.62 | 52.15 | 53.86 | 57.12 | 1.34 | 2.15 | 3.57 |
| | | | Acc (adj) | 1.76 | 1.38 | 1.09 | 1.83 | 1.43 | 1.06 | 2.80† | 1.49 | 1.45 |

*Note: MW, with ETE, without ETE, Improvement, Acc, and Acc (adj) are the abbreviations of the moving window, the performance when the stock price direction is predicted only by its plain log-returns, the performance when predicted by the mixture of log-returns and the ETE network indicator, improvement of prediction accuracy/adjusted accuracy incurred by ETE network indicator, accuracy, and adjusted accuracy, respectively. Also, † is the top five accuracy and the adjusted accuracy of a total of 30 sets of models and parameters.

mean of the prediction performance is on the $X$-axis, whereas the mean of improvement achieved by integrating ETE network indicator is on the $Y$-axis. We set the means of $X-$ and $Y-$ axes as red dashed lines and the $\pm 1\sigma$, an indifferent region, as a gray box. Specifically, the first quadrant of the scatter plot is the best-case scenario for the proposed

**TABLE 6.** Accuracy and its improvement by ETE network indicators.

| Model | Sector | MW = 3M | | | | | | | | | MW = 6M | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Lag = 3M | | | Lag = 6M | | | Lag = 1Y | | | Lag = 3M | | | Lag = 6M | | | Lag = 1Y | | |
| | | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M |
| MLP | TELE | 52.22 | 54.78 | 55.16 | 52.54 | 55.97 | 55.52 | 51.61 | 54.25 | 54.22 | 51.51 | 54.28 | 57.06 | 52.07 | 54.29 | 56.48 | 52.13 | 56.13 | 55.71 |
| | | (2.43) | (4.81) | (5.15) | (2.19) | (5.88) | (5.37) | (2.51) | (3.48) | (4.52) | (2.10) | (5.65) | (7.37) | (2.73) | (2.27) | (4.95) | (2.42) | (5.52) | (5.78) |
| | COND | 53.59 | 59.27 | 61.91 | 53.38 | 58.49 | 62.07 | 52.64 | 56.83 | 59.97 | 54.44 | 57.50 | 63.63 | 53.49 | 57.97 | 59.97 | 54.46 | 57.41 | 62.67 |
| | | (3.65) | (6.61) | (7.88) | (3.52) | (5.49) | (6.61) | (3.01) | (3.01) | (6.93) | (3.61) | (3.71) | (9.87) | (3.70) | (5.11) | (7.71) | (4.27) | (7.13) | (8.28) |
| | ENRG | 52.51 | 53.21 | 55.05 | 52.96 | 53.22 | 53.91 | 53.02 | 53.11 | 55.16 | 52.64 | 52.86 | 54.86 | 53.28 | 53.02 | 54.81 | 52.76 | 53.02 | 55.79 |
| | | (1.86) | (2.44) | (2.70) | (2.21) | (1.88) | (2.11) | (2.00) | (1.75) | (2.49) | (1.71) | (1.58) | (3.75) | (2.23) | (2.08) | (2.62) | (1.74) | (1.96) | (4.42) |
| | HLCA | 52.89 | 56.52 | 58.15 | 52.41 | 55.82 | 59.28 | 52.80 | 56.16 | 59.47 | 52.56 | 55.75 | 57.11 | 53.31 | 55.67 | 58.43 | 52.98 | 55.92 | 58.66 |
| | | (2.48) | (3.85) | (5.36) | (2.72) | (4.14) | (7.72) | (2.37) | (3.93) | (7.48) | (2.78) | (3.65) | (5.22) | (3.17) | (5.02) | (5.77) | (2.47) | (3.05) | (5.28) |
| | INDS | 54.33 | 59.27 | 67.02 | 53.73 | 58.69 | 64.44 | 54.26 | 58.45 | 67.89 | 53.90 | 58.08 | 64.70 | 53.47 | 58.80 | 65.00 | 54.08 | 56.78 | 65.15 |
| | | (2.49) | (3.47) | (9.02) | (2.33) | (3.45) | (5.49) | (3.52) | (4.50) | (10.74) | (2.92) | (3.70) | (6.90) | (2.18) | (4.66) | (5.59) | (3.51) | (4.10) | (4.25) |
| | MTRS | 52.90 | 55.55 | 62.22 | 53.36 | 57.03 | 61.50 | 53.13 | 57.01 | 61.56 | 53.57 | 55.35 | 62.18 | 53.59 | 56.05 | 62.35 | 52.88 | 57.22 | 63.13 |
| | | (2.60) | (2.57) | (6.94) | (2.05) | (4.53) | (4.69) | (2.24) | (3.16) | (6.39) | (2.24) | (4.54) | (5.19) | (2.28) | (5.41) | (5.05) | (1.72) | (4.66) | (5.62) |
| | FINC | 52.43 | 54.57 | 57.15 | 52.27 | 54.40 | 57.57 | 52.44 | 54.40 | 57.25 | 52.22 | 54.34 | 56.86 | 52.05 | 55.19 | 57.38 | 52.28 | 54.51 | 57.88 |
| | | (2.05) | (3.99) | (3.81) | (1.58) | (2.15) | (4.21) | (2.78) | (2.72) | (3.25) | (2.00) | (3.32) | (3.08) | (1.53) | (2.94) | (3.89) | (2.01) | (2.63) | (5.23) |
| | CONS | 53.69 | 56.29 | 62.50 | 53.34 | 56.87 | 62.30 | 53.84 | 56.97 | 62.78 | 53.65 | 57.33 | 63.24 | 54.23 | 57.86 | 62.30 | 53.57 | 56.00 | 61.82 |
| | | (2.50) | (2.68) | (3.56) | (1.72) | (3.08) | (5.88) | (2.62) | (4.09) | (4.85) | (3.03) | (1.70) | (4.57) | (2.21) | (4.62) | (4.86) | (2.56) | (2.73) | (5.32) |
| | INFT | 53.06 | 55.31 | 57.74 | 52.97 | 54.54 | 57.58 | 53.19 | 55.54 | 58.12 | 53.33 | 55.22 | 57.86 | 53.00 | 55.30 | 58.16 | 53.34 | 54.75 | 57.72 |
| | | (2.66) | (5.45) | (6.89) | (3.58) | (3.56) | (5.23) | (3.23) | (4.29) | (6.72) | (2.95) | (5.21) | (6.48) | (2.85) | (4.62) | (5.43) | (3.37) | (4.07) | (6.82) |
| | UTIL | 54.17 | 58.63 | 65.02 | 54.69 | 58.74 | 64.50 | 53.99 | 57.96 | 64.83 | 54.04 | 58.68 | 65.59 | 54.05 | 59.08 | 64.45 | 53.61 | 58.15 | 64.83 |
| | | (2.07) | (3.02) | (4.02) | (2.27) | (2.95) | (3.67) | (2.07) | (3.32) | (4.39) | (1.92) | (2.27) | (3.64) | (3.06) | (4.75) | (4.63) | (3.32) | (4.04) | (4.25) |
| | REES | 54.10 | 58.05 | 64.42 | 54.52 | 57.42 | 65.19 | 54.56 | 57.78 | 67.24 | 53.88 | 58.39 | 65.82 | 54.04 | 59.01 | 66.35 | 54.81 | 58.60 | 67.34 |
| | | (2.32) | (4.03) | (6.28) | (1.82) | (3.30) | (3.48) | (2.25) | (2.69) | (8.41) | (2.84) | (2.84) | (5.77) | (2.50) | (3.97) | (4.39) | (2.57) | (4.73) | (8.69) |
| RF | TELE | 51.64 | 55.22 | 52.59 | 51.82 | 57.41 | 53.67 | 52.51 | 55.59 | 55.09 | 52.41 | 55.86 | 56.64 | 51.69 | 57.28 | 54.25 | 52.04 | 56.43 | 54.98 |
| | | (1.16) | (1.52) | (0.93) | (0.96) | (1.29) | (-0.08) | (0.75) | (0.89) | (0.45) | (1.50) | (0.18) | (4.42) | (1.08) | (0.81) | (0.45) | (1.18) | (1.23) | (-0.17) |
| | COND | 55.20 | 57.27 | 62.60 | 55.07 | 57.66 | 62.43 | 54.60 | 57.63 | 61.46 | 55.09 | 60.20 | 63.06 | 55.88 | 60.20 | 62.79 | 55.14 | 58.66 | 60.89 |
| | | (0.52) | (-1.82) | (-0.89) | (0.17) | (-1.19) | (0.27) | (0.53) | (-0.16) | (1.70) | (0.62) | (0.43) | (-2.98) | (1.02) | (0.02) | (-0.08) | (0.57) | (1.16) | (0.68) |
| | ENRG | 53.67 | 53.45 | 56.12 | 53.76 | 53.88 | 56.73 | 53.74 | 54.33 | 57.31 | 53.76 | 53.00 | 53.51 | 53.72 | 53.82 | 56.20 | 53.76 | 53.97 | 57.26 |
| | | (0.47) | (-0.37) | (-0.83) | (0.36) | (-0.08) | (-0.37) | (0.16) | (0.03) | (-0.04) | (0.43) | (-0.55) | (-3.19) | (0.20) | (-0.25) | (-1.04) | (0.16) | (-0.16) | (-0.02) |
| | HLCA | 53.09 | 57.11 | 61.14 | 53.82 | 58.00 | 60.32 | 53.51 | 57.65 | 60.47 | 52.71 | 56.18 | 61.89 | 54.08 | 57.54 | 61.84 | 53.28 | 57.19 | 61.67 |
| | | (0.84) | (0.60) | (3.58) | (0.85) | (0.28) | (2.12) | (1.02) | (0.24) | (1.21) | (0.72) | (-0.42) | (3.85) | (1.10) | (0.09) | (3.97) | (0.91) | (0.12) | (2.34) |
| | INDS | 56.99 | 63.37 | 66.41 | 56.23 | 63.79 | 70.67 | 56.89 | 64.16 | 68.93 | 56.15 | 62.81 | 65.51 | 56.08 | 63.14 | 68.80 | 56.80 | 63.78 | 68.93 |
| | | (0.57) | (0.16) | (-2.34) | (0.46) | (-0.10) | (-1.07) | (0.40) | (0.14) | (0.02) | (0.46) | (-0.11) | (-2.64) | (0.60) | (-0.18) | (-2.32) | (0.45) | (0.11) | (-0.25) |
| | MTRS | 54.85 | 54.53 | 65.31 | 54.63 | 59.52 | 65.85 | 54.77 | 59.65 | 65.38 | 54.60 | 58.19 | 65.63 | 54.92 | 59.24 | 65.65 | 54.53 | 59.96 | 64.65 |
| | | (0.45) | (-3.37) | (-0.90) | (0.25) | (0.07) | (0.15) | (0.49) | (0.20) | (0.76) | (0.86) | (-0.49) | (0.28) | (0.77) | (0.12) | (0.53) | (0.71) | (0.33) | (0.91) |
| | FINC | 53.66 | 56.45 | 58.00 | 53.39 | 55.73 | 57.88 | 53.39 | 56.85 | 57.73 | 53.39 | 56.50 | 58.82 | 53.39 | 56.08 | 59.30 | 53.23 | 56.00 | 59.36 |
| | | (0.98) | (0.55) | (1.31) | (0.77) | (0.90) | (0.37) | (0.94) | (0.54) | (-0.05) | (0.78) | (0.42) | (1.23) | (0.77) | (1.04) | (2.01) | (0.76) | (0.68) | (0.49) |
| | CONS | 55.70 | 59.12 | 63.72 | 56.02 | 59.64 | 64.54 | 55.57 | 59.59 | 65.35 | 55.59 | 58.90 | 64.70 | 56.18 | 58.98 | 65.40 | 55.72 | 59.82 | 66.34 |
| | | (0.40) | (0.11) | (-0.34) | (0.42) | (0.18) | (0.31) | (0.33) | (0.17) | (-0.05) | (0.64) | (0.01) | (1.14) | (0.45) | (-0.35) | (0.83) | (0.30) | (-0.09) | (0.23) |
| | INFT | 53.72 | 58.02 | 58.30 | 52.49 | 57.53 | 58.31 | 54.48 | 59.14 | 60.71 | 54.04 | 56.51 | 60.63 | 54.22 | 57.76 | 58.71 | 53.82 | 56.83 | 61.66 |
| | | (1.20) | (3.82) | (4.91) | (0.86) | (1.80) | (3.08) | (1.07) | (1.36) | (2.05) | (0.98) | (1.00) | (5.62) | (1.30) | (1.29) | (3.18) | (0.83) | (0.61) | (2.41) |
| | UTIL | 56.18 | 60.63 | 67.71 | 56.11 | 60.64 | 67.49 | 56.35 | 60.63 | 67.61 | 56.36 | 60.95 | 67.12 | 56.37 | 60.86 | 67.24 | 56.36 | 60.85 | 67.65 |
| | | (0.43) | (0.01) | (-0.14) | (0.25) | (0.01) | (-0.25) | (0.13) | (0.00) | (-0.15) | (0.22) | (0.10) | (-0.60) | (0.23) | (0.01) | (-0.47) | (0.09) | (0.00) | (-0.09) |
| | REES | 56.49 | 61.46 | 69.46 | 56.47 | 61.45 | 71.36 | 56.52 | 61.35 | 71.35 | 56.38 | 60.29 | 69.02 | 56.22 | 61.25 | 70.30 | 56.43 | 61.17 | 71.14 |
| | | (0.25) | (-0.22) | (-2.14) | (0.14) | (-0.14) | (-0.23) | (0.16) | (-0.03) | (-0.16) | (0.35) | (-1.19) | (-2.25) | (0.18) | (-0.30) | (-0.97) | (0.12) | (-0.32) | (-0.13) |
| XGB | TELE | 52.05 | 53.74 | 55.72 | 51.57 | 54.11 | 55.35 | 51.30 | 53.20 | 54.62 | 51.29 | 54.22 | 56.98 | 52.14 | 55.37 | 57.04 | 51.37 | 54.63 | 58.00 |
| | | (1.82) | (2.29) | (4.81) | (1.34) | (1.39) | (3.33) | (0.93) | (1.39) | (2.65) | (1.31) | (2.68) | (4.94) | (1.33) | (1.93) | (5.03) | (1.18) | (2.08) | (5.05) |
| | COND | 53.21 | 54.54 | 60.55 | 53.04 | 55.59 | 60.62 | 52.37 | 52.80 | 57.15 | 53.21 | 57.36 | 60.83 | 53.51 | 56.87 | 58.35 | 53.11 | 54.96 | 57.89 |
| | | (0.33) | (-1.56) | (0.03) | (0.30) | (-1.20) | (0.12) | (1.15) | (-0.72) | (1.87) | (1.38) | (0.77) | (-0.66) | (0.90) | (0.89) | (-0.76) | (0.98) | (1.02) | (2.69) |
| | ENRG | 52.60 | 52.63 | 53.64 | 52.28 | 52.70 | 53.73 | 52.56 | 53.02 | 54.06 | 52.80 | 52.02 | 53.21 | 52.65 | 52.12 | 53.33 | 52.25 | 52.51 | 53.37 |
| | | (0.91) | (1.12) | (0.98) | (0.78) | (1.33) | (0.58) | (1.05) | (1.42) | (0.36) | (1.24) | (0.63) | (0.97) | (1.00) | (0.57) | (-0.05) | (0.99) | (0.18) | (-0.34) |
| | HLCA | 51.24 | 54.10 | 52.76 | 51.17 | 52.88 | 56.31 | 51.39 | 53.30 | 54.50 | 52.53 | 55.21 | 58.26 | 52.01 | 56.34 | 59.38 | 52.13 | 55.20 | 57.90 |
| | | (0.70) | (0.09) | (-2.61) | (0.37) | (-0.97) | (0.41) | (0.38) | (0.73) | (-0.70) | (0.60) | (1.66) | (1.14) | (0.81) | (1.04) | (2.85) | (0.89) | (3.16) | (2.23) |
| | INDS | 53.25 | 57.90 | 64.95 | 52.94 | 58.91 | 64.17 | 53.02 | 58.89 | 64.01 | 53.11 | 57.95 | 63.86 | 53.09 | 58.58 | 64.64 | 52.32 | 57.70 | 64.29 |
| | | (0.92) | (0.83) | (1.08) | (1.07) | (2.18) | (0.35) | (1.20) | (1.88) | (0.84) | (1.15) | (0.98) | (-0.19) | (0.92) | (1.03) | (0.67) | (1.10) | (0.78) | (1.00) |
| | MTRS | 51.36 | 52.88 | 59.64 | 51.73 | 54.13 | 60.59 | 51.85 | 53.48 | 60.05 | 52.49 | 55.11 | 60.27 | 52.79 | 55.22 | 61.14 | 52.45 | 55.22 | 60.72 |
| | | (0.25) | (-1.41) | (-1.15) | (0.65) | (-1.14) | (-0.28) | (0.24) | (-0.89) | (-0.52) | (1.27) | (-0.24) | (-0.64) | (0.97) | (0.35) | (0.19) | (0.95) | (1.03) | (0.47) |
| | FINC | 51.76 | 53.31 | 56.02 | 52.03 | 53.68 | 56.89 | 52.01 | 54.45 | 57.15 | 52.15 | 54.53 | 57.49 | 51.46 | 54.65 | 58.50 | 51.98 | 54.17 | 58.73 |
| | | (1.02) | (0.29) | (0.73) | (1.24) | (0.85) | (0.58) | (1.17) | (1.53) | (1.42) | (1.06) | (1.07) | (0.85) | (0.34) | (1.82) | (2.35) | (1.16) | (1.42) | (2.72) |
| | CONS | 52.58 | 56.55 | 62.19 | 52.61 | 56.47 | 61.53 | 52.05 | 56.20 | 62.25 | 53.84 | 56.63 | 60.25 | 53.51 | 57.16 | 60.51 | 53.33 | 56.46 | 59.96 |
| | | (0.52) | (0.32) | (0.24) | (0.40) | (0.37) | (0.37) | (0.56) | (0.63) | (-0.26) | (0.69) | (0.55) | (-1.84) | (1.02) | (0.81) | (-1.00) | (0.71) | (0.79) | (-2.17) |
| | INFT | 53.06 | 55.41 | 54.78 | 53.46 | 55.21 | 57.44 | 52.44 | 55.50 | 57.11 | 52.56 | 54.25 | 58.31 | 53.14 | 55.19 | 58.79 | 52.52 | 54.63 | 59.23 |
| | | (1.03) | (3.01) | (1.11) | (1.39) | (2.11) | (1.38) | (0.72) | (1.43) | (1.33) | (1.05) | (1.41) | (3.69) | (1.03) | (1.35) | (3.07) | (1.04) | (0.42) | (3.40) |
| | UTIL | 54.45 | 59.50 | 66.43 | 54.25 | 59.54 | 66.35 | 54.55 | 59.80 | 66.82 | 53.92 | 57.77 | 61.04 | 53.87 | 58.47 | 61.17 | 53.45 | 59.24 | 62.30 |
| | | (0.89) | (1.02) | (-0.47) | (0.69) | (1.04) | (-0.14) | (1.09) | (0.54) | (0.14) | (0.18) | (-1.30) | (-5.26) | (0.47) | (-0.64) | (-5.40) | (0.50) | (-0.23) | (-4.68) |
| | REES | 54.55 | 58.28 | 64.17 | 54.66 | 57.63 | 64.44 | 54.32 | 57.42 | 64.18 | 54.21 | 57.37 | 63.34 | 55.17 | 57.76 | 62.56 | 54.27 | 57.69 | 63.47 |
| | | (1.14) | (0.16) | (-2.39) | (0.72) | (-0.27) | (-3.23) | (1.24) | (-0.59) | (-2.39) | (0.92) | (-0.97) | (-4.18) | (1.04) | (-0.90) | (-5.11) | (0.87) | (-0.70) | (-3.94) |
| LSTM | TELE | 51.24 | 55.44 | 56.43 | 52.43 | 55.09 | 55.66 | 51.54 | 52.99 | 54.31 | 52.19 | 55.68 | 55.92 | 51.76 | 54.99 | 55.80 | 51.75 | 53.11 | 54.15 |
| | | (1.38) | (4.28) | (5.80) | (1.86) | (2.72) | (4.27) | (1.27) | (1.79) | (2.65) | (2.10) | (3.01) | (4.41) | (1.48) | (2.21) | (3.96) | (1.22) | (1.66) | (2.81) |
| | COND | 54.69 | 58.22 | 64.50 | 52.93 | 56.19 | 62.57 | 52.75 | 54.58 | 58.19 | 53.97 | 56.94 | 59.44 | 53.56 | 56.81 | 58.36 | 52.57 | 54.15 | 55.43 |
| | | (2.01) | (1.76) | (4.79) | (1.04) | (1.60) | (5.43) | (1.48) | (1.73) | (3.88) | (1.64) | (0.85) | (-0.36) | (1.49) | (1.49) | (2.35) | (1.29) | (2.30) | (2.78) |
| | ENRG | 53.44 | 52.93 | 54.31 | 53.32 | 52.27 | 53.81 | 52.53 | 52.07 | 53.08 | 53.44 | 52.90 | 54.68 | 52.94 | 52.69 | 54.39 | 52.63 | 52.41 | 53.73 |
| | | (1.40) | (1.01) | (0.66) | (1.52) | (0.98) | (0.62) | (1.25) | (0.84) | (1.51) | (1.49) | (1.20) | (1.91) | (1.12) | (1.54) | (1.96) | (1.22) | (1.01) | (2.36) |
| | HLCA | 53.55 | 56.81 | 58.95 | 52.47 | 55.58 | 57.81 | 52.36 | 54.52 | 56.52 | 52.77 | 56.59 | 58.31 | 52.25 | 55.74 | 57.68 | 52.25 | 54.81 | 56.14 |
| | | (1.56) | (2.16) | (4.01) | (1.91) | (1.86) | (3.47) | (1.63) | (1.81) | (3.49) | (1.26) | (1.71) | (3.72) | (1.23) | (2.08) | (3.20) | (1.06) | (1.64) | (3.06) |
| | INDS | 54.66 | 59.37 | 65.81 | 52.74 | 57.74 | 64.08 | 52.16 | 54.97 | 61.41 | 54.33 | 60.17 | 66.73 | 52.75 | 58.45 | 65.61 | 52.05 | 55.15 | 61.52 |
| | | (1.23) | (2.84) | (0.89) | (1.61) | (3.39) | (4.16) | (1.34) | (2.60) | (4.70) | (1.30) | (2.22) | (1.93) | (1.68) | (3.41) | (5.04) | (1.41) | (3.26) | (5.04) |
| | MTRS | 53.96 | 55.72 | 63.43 | 52.78 | 54.92 | 62.10 | 51.91 | 54.32 | 58.63 | 53.26 | 57.26 | 63.57 | 53.15 | 56.61 | 62.06 | 52.06 | 55.73 | 59.24 |
| | | (0.76) | (0.85) | (2.16) | (0.75) | (0.49) | (3.29) | (1.22) | (0.94) | (3.02) | (0.91) | (1.74) | (2.39) | (1.42) | (1.67) | (3.57) | (1.15) | (2.02) | (3.51) |
| | FINC | 53.55 | 56.06 | 57.66 | 52.83 | 55.14 | 57.88 | 52.15 | 54.11 | 56.36 | 53.73 | 56.39 | 59.07 | 52.48 | 55.17 | 58.02 | 52.29 | 53.83 | 56.97 |
| | | (1.25) | (2.02) | (1.74) | (1.77) | (2.05) | (1.92) | (1.40) | (1.45) | (2.05) | (1.62) | (2.08) | (1.98) | (1.28) | (1.96) | (2.72) | (1.40) | (1.62) | (2.73) |
| | CONS | 54.84 | 58.61 | 63.95 | 53.40 | 56.77 | 62.68 | 52.78 | 54.51 | 61.32 | 54.91 | 58.98 | 63.72 | 54.03 | 56.99 | 62.86 | 52.66 | 55.12 | 61.87 |
| | | (1.52) | (1.67) | (2.61) | (0.89) | (2.62) | (4.09) | (1.53) | (2.71) | (3.96) | (1.35) | (1.66) | (2.26) | (1.73) | (2.81) | (3.83) | (1.48) | (2.57) | (3.91) |
| | INFT | 53.78 | 56.66 | 59.47 | 52.36 | 54.88 | 56.31 | 52.45 | 53.07 | 56.05 | 53.76 | 56.97 | 58.59 | 52.63 | 55.60 | 55.39 | 52.65 | 53.30 | 55.67 |
| | | (2.06) | (3.40) | (7.62) | (1.30) | (2.10) | (6.18) | (1.30) | (1.59) | (3.81) | (2.53) | (2.33) | (5.82) | (1.84) | (2.53) | (4.03) | (1.42) | (1.86) | (3.30) |
| | UTIL | 55.63 | 60.10 | 66.90 | 54.53 | 59.28 | 64.83 | 53.36 | 58.19 | 63.60 | 56.03 | 60.61 | 67.21 | 55.11 | 59.63 | 66.04 | 53.93 | 57.84 | 64.37 |
| | | (1.02) | (1.24) | (1.49) | (1.58) | (2.80) | (3.64) | (1.74) | (3.31) | (4.70) | (1.48) | (1.83) | (2.15) | (1.29) | (3.18) | (4.20) | (1.72) | (3.27) | (5.26) |
| | REES | 55.81 | 60.32 | 67.34 | 55.02 | 59.02 | 66.03 | 53.73 | 57.45 | 64.45 | 55.42 | 59.89 | 67.81 | 55.16 | 59.39 | 67.59 | 53.48 | 57.54 | 65.75 |
| | | (0.73) | (0.22) | (-0.78) | (1.46) | (1.06) | (1.47) | (1.46) | (1.75) | (3.18) | (1.15) | (0.15) | (-0.54) | (1.55) | (2.07) | (2.74) | (1.34) | (2.44) | (4.52) |

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the improvement. The unit for accuracy and improvement are percent and percent point, respectively.

**TABLE 7.** Adjusted accuracy and its improvement by ETE network indicators.

| Model | Sector | MW = 3M | | | | | | | | | MW = 6M | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Lag = 3M | | | Lag = 6M | | | Lag = 1Y | | | Lag = 3M | | | Lag = 6M | | | Lag = 1Y | | |
| | | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M | 1W | 1M | 3M |
| MLP | TELE | 1.82 | 4.09 | 1.17 | 5.42 | 1.68 | 2.27 | 1.73 | 2.16 | 0.72 | 1.12 | 1.18 | 1.12 | 1.38 | 3.26 | 1.09 | 1.31 | 1.95 | 1.62 |
| | | (2.81) | (2.36) | (1.73) | (1.82) | (1.39) | (1.87) | (2.49) | (2.26) | (2.06) | (2.43) | (1.31) | (2.07) | (2.15) | (3.54) | (1.31) | (2.36) | (2.12) | (1.07) |
| | COND | 2.39 | 3.36 | 2.63 | 2.97 | 5.58 | 2.25 | 1.20 | 1.78 | 1.40 | 3.52 | 2.01 | 3.07 | 1.46 | 1.72 | 1.96 | 4.24 | 1.77 | 2.78 |
| | | (2.07) | (1.85) | (1.37) | (2.45) | (2.18) | (2.35) | (1.29) | (2.13) | (1.35) | (1.50) | (2.50) | (1.60) | (3.60) | (3.32) | (1.13) | (1.76) | (3.00) | (1.75) |
| | ENRG | 1.82 | 1.30 | 1.10 | 1.82 | 1.45 | 0.90 | 2.48 | 1.42 | 1.10 | 2.13 | 1.14 | 0.93 | 2.18 | 1.40 | 0.93 | 1.56 | 1.43 | 1.29 |
| | | (3.35) | (1.65) | (3.03) | (3.06) | (10.30) | (2.61) | (3.68) | (1.17) | (1.04) | (2.68) | (3.87) | (2.62) | (3.75) | (4.95) | (2.27) | (2.06) | (1.67) | (1.35) |
| | HLCA | 2.24 | 0.97 | 0.67 | 1.88 | 0.89 | 0.68 | 1.94 | 0.99 | 0.76 | 2.88 | 0.91 | 0.59 | 2.24 | 0.85 | 0.65 | 1.92 | 0.95 | 0.77 |
| | | (1.68) | (2.28) | (1.94) | (3.33) | (3.04) | (1.32) | (1.87) | (2.67) | (2.29) | (2.02) | (2.11) | (1.30) | (3.56) | (2.85) | (2.05) | (3.50) | (1.62) | (5.54) |
| | INDS | 4.10 | 3.82 | 4.01 | 4.59 | 4.38 | 3.00 | 3.11 | 3.30 | 5.00 | 6.11 | 3.69 | 2.78 | 4.76 | 2.86 | 2.78 | 3.10 | 5.75 | 2.56 |
| | | (1.09) | (2.24) | (2.96) | (3.26) | (4.98) | (7.91) | (1.61) | (1.78) | (2.26) | (1.80) | (4.39) | (1.72) | (3.14) | (2.79) | (2.92) | (2.18) | (2.33) | (3.00) |
| | MTRS | 1.87 | 1.79 | 1.80 | 1.83 | 2.11 | 1.73 | 2.21 | 2.32 | 1.88 | 1.29 | 1.69 | 2.43 | 1.82 | 1.74 | 1.94 | 1.83 | 2.97 | 2.15 |
| | | (2.31) | (1.81) | (1.48) | (2.59) | (1.31) | (1.41) | (1.98) | (1.77) | (1.53) | (2.57) | (1.57) | (2.57) | (2.97) | (1.35) | (1.58) | (2.12) | (2.04) | (1.56) |
| | FINC | 2.12 | 1.54 | 1.30 | 1.43 | 1.36 | 1.52 | 1.93 | 1.64 | 1.40 | 1.75 | 2.07 | 1.21 | 2.42 | 1.50 | 1.89 | 2.13 | 1.35 | 1.77 |
| | | (2.95) | (2.60) | (1.42) | (2.43) | (2.64) | (0.95) | (1.80) | (2.17) | (1.11) | (4.75) | (2.01) | (1.23) | (3.90) | (1.82) | (1.36) | (1.53) | (1.44) | (3.04) |
| | CONS | 2.38 | 2.25 | 2.75 | 1.48 | 2.86 | 2.18 | 1.56 | 1.94 | 2.13 | 3.59 | 3.01 | 2.36 | 3.30 | 2.93 | 1.97 | 2.21 | 1.67 | 2.31 |
| | | (1.65) | (3.07) | (1.05) | (1.57) | (2.44) | (1.60) | (2.07) | (2.00) | (2.02) | (2.57) | (1.80) | (2.70) | (5.29) | (2.52) | (2.01) | (2.27) | (1.45) | (2.87) |
| | INFT | 1.54 | 1.52 | 0.84 | 1.57 | 1.05 | 0.75 | 1.69 | 3.00 | 0.94 | 1.75 | 1.45 | 0.83 | 1.57 | 1.48 | 0.92 | 1.45 | 1.21 | 0.83 |
| | | (3.69) | (3.27) | (2.43) | (2.23) | (2.28) | (1.14) | (2.51) | (1.80) | (2.41) | (1.64) | (2.12) | (2.95) | (2.13) | (1.22) | (1.68) | (3.03) | (1.46) | (1.16) |
| | UTIL | 2.52 | 2.27 | 1.86 | 6.10 | 2.42 | 2.11 | 3.65 | 2.63 | 2.19 | 2.39 | 2.45 | 2.00 | 2.54 | 2.74 | 2.06 | 2.70 | 2.49 | 2.04 |
| | | (1.88) | (1.66) | (1.45) | (3.60) | (2.04) | (1.47) | (1.78) | (2.02) | (1.37) | (1.67) | (4.66) | (2.08) | (1.66) | (1.48) | (1.21) | (2.89) | (1.96) | (1.39) |
| | REES | 5.43 | 3.22 | 2.29 | 2.92 | 12.25 | 4.76 | 5.95 | 4.04 | 12.04 | 3.24 | 4.71 | 5.15 | 3.17 | 5.31 | 3.73 | 7.60 | 8.00 | 5.46 |
| | | (2.36) | (1.66) | (1.16) | (1.81) | (1.21) | (2.39) | (2.22) | (2.05) | (1.15) | (1.73) | (1.46) | (1.32) | (1.65) | (2.04) | (1.37) | (10.96) | (2.76) | (1.82) |
| RF | TELE | 0.62 | 0.94 | 0.17 | 0.59 | 1.35 | 0.24 | 0.67 | 0.92 | 0.36 | 0.80 | 1.89 | 0.65 | 0.47 | 2.07 | 0.28 | 0.60 | 1.28 | 0.31 |
| | | (0.86) | (0.49) | (0.31) | (0.75) | (0.47) | (-0.04) | (0.91) | (0.77) | (0.29) | (1.67) | (0.11) | (0.69) | (2.05) | (0.29) | (0.19) | (2.19) | (0.74) | (-0.05) |
| | COND | 1.54 | 0.78 | 1.29 | 1.05 | 0.68 | 0.76 | 0.92 | 0.75 | 0.70 | 1.09 | 1.33 | 1.14 | 2.47 | 1.00 | 0.84 | 1.20 | 0.74 | 0.63 |
| | | (0.49) | (-0.37) | (-0.14) | (0.25) | (-0.43) | (0.17) | (0.96) | (-0.10) | (0.57) | (0.64) | (0.24) | (-0.88) | (1.11) | (0.17) | (-0.06) | (0.98) | (1.05) | (1.03) |
| | ENRG | 2.14 | 1.08 | 0.98 | 1.84 | 1.14 | 1.07 | 1.86 | 1.36 | 1.29 | 2.09 | 0.89 | 0.51 | 1.86 | 1.12 | 1.13 | 2.02 | 1.25 | 1.32 |
| | | (1.83) | (-0.50) | (-0.57) | (2.81) | (-0.29) | (-0.34) | (1.06) | (0.76) | (-0.44) | (0.95) | (-0.33) | (-1.63) | (1.10) | (-0.95) | (-1.09) | (1.38) | (-0.40) | (-0.45) |
| | HLCA | 0.85 | 0.62 | 0.56 | 0.89 | 0.71 | 0.51 | 0.81 | 0.66 | 0.52 | 0.76 | 0.50 | 0.64 | 1.51 | 0.64 | 0.63 | 0.81 | 0.61 | 0.62 |
| | | (0.97) | (1.27) | (0.45) | (1.53) | (0.46) | (0.46) | (1.04) | (0.79) | (0.48) | (0.50) | (-0.33) | (0.45) | (1.38) | (1.62) | (0.48) | (2.32) | (0.58) | (0.51) |
| | INDS | 9.90 | 5.07 | 1.47 | 3.05 | 4.68 | 4.22 | 6.62 | 4.63 | 1.82 | 3.31 | 4.58 | 1.29 | 3.71 | 3.78 | 2.67 | 8.29 | 4.42 | 1.93 |
| | | (1.49) | (0.20) | (-0.77) | (1.30) | (-0.26) | (-0.96) | (1.50) | (0.54) | (0.07) | (3.94) | (-0.26) | (-0.82) | (1.81) | (-0.35) | (-1.03) | (1.63) | (0.47) | (-0.73) |
| | MTRS | 1.53 | 0.55 | 1.46 | 1.51 | 1.62 | 1.64 | 1.52 | 1.67 | 1.66 | 1.39 | 1.42 | 1.84 | 1.54 | 1.62 | 1.86 | 1.46 | 1.84 | 1.71 |
| | | (0.90) | (-0.46) | (-0.64) | (0.56) | (0.22) | (0.11) | (2.37) | (0.51) | (0.37) | (1.67) | (-0.28) | (0.54) | (2.27) | (0.89) | (0.71) | (1.61) | (0.78) | (0.68) |
| | FINC | 2.01 | 1.78 | 1.04 | 1.86 | 1.67 | 1.05 | 1.73 | 2.18 | 0.82 | 1.93 | 1.81 | 1.08 | 2.29 | 1.87 | 1.43 | 1.65 | 1.29 | 1.06 |
| | | (2.06) | (0.91) | (0.92) | (2.72) | (1.59) | (0.42) | (1.30) | (1.09) | (-0.06) | (1.64) | (1.22) | (1.34) | (2.19) | (1.61) | (1.21) | (1.30) | (0.70) | (0.41) |
| | CONS | 1.35 | 1.58 | 1.50 | 1.65 | 1.97 | 1.64 | 1.17 | 1.92 | 1.87 | 1.26 | 1.52 | 2.23 | 1.84 | 1.57 | 2.62 | 1.34 | 2.31 | 2.93 |
| | | (1.16) | (0.54) | (-0.25) | (1.03) | (0.65) | (0.17) | (0.96) | (0.61) | (-0.43) | (1.76) | (0.04) | (0.14) | (1.02) | (-0.48) | (0.18) | (1.05) | (-0.25) | (0.45) |
| | INFT | 1.11 | 2.82 | 0.84 | 0.58 | 1.65 | 0.84 | 2.42 | 3.24 | 1.20 | 1.22 | 1.90 | 1.29 | 2.04 | 2.45 | 0.90 | 1.08 | 1.19 | 1.54 |
| | | (0.96) | (0.82) | (0.65) | (1.35) | (0.84) | (0.79) | (1.46) | (0.79) | (0.71) | (1.26) | (0.32) | (0.55) | (1.57) | (0.57) | (0.59) | (1.60) | (1.01) | (0.69) |
| | UTIL | 3.25 | 2.19 | 1.99 | 3.22 | 2.19 | 1.98 | 3.45 | 2.19 | 1.98 | 3.16 | 2.20 | 1.91 | 3.22 | 2.22 | 1.92 | 3.26 | 2.22 | 1.94 |
| | | (0.74) | (0.67) | (-0.42) | (0.71) | (0.45) | (-0.45) | (1.19) | (0.00) | (-0.45) | (0.97) | (0.52) | (-0.50) | (0.75) | (0.45) | (-0.44) | (1.10) | (0.45) | (-0.45) |
| | REES | 5.89 | 10.38 | 3.70 | 5.27 | 10.30 | 7.00 | 5.06 | 9.40 | 6.89 | 6.59 | 4.45 | 4.11 | 5.64 | 8.95 | 4.61 | 5.28 | 8.28 | 7.43 |
| | | (2.36) | (-0.42) | (-0.75) | (1.98) | (-0.44) | (-0.49) | (1.73) | (-0.27) | (-0.45) | (1.32) | (-0.57) | (-0.87) | (0.72) | (-0.44) | (-0.47) | (1.53) | (-0.44) | (-0.45) |
| XGB | TELE | 0.98 | 0.86 | 0.92 | 0.93 | 1.57 | 1.08 | 0.79 | 1.07 | 0.62 | 1.14 | 2.16 | 1.05 | 1.21 | 3.04 | 1.44 | 1.20 | 1.32 | 1.34 |
| | | (2.93) | (0.97) | (1.67) | (1.16) | (0.66) | (1.31) | (1.74) | (0.71) | (1.56) | (2.44) | (1.56) | (1.35) | (2.40) | (1.38) | (1.51) | (1.73) | (1.20) | (1.55) |
| | COND | 1.58 | 0.79 | 1.37 | 1.22 | 0.79 | 1.38 | 0.66 | 0.31 | 0.76 | 1.99 | 1.29 | 1.37 | 1.57 | 1.39 | 0.86 | 1.01 | 0.63 | 0.74 |
| | | (0.73) | (-0.37) | (0.01) | (0.39) | (-0.32) | (0.02) | (1.69) | (-0.17) | (0.31) | (1.40) | (0.45) | (-0.16) | (1.32) | (0.37) | (-0.12) | (3.24) | (0.34) | (0.45) |
| | ENRG | 1.85 | 1.28 | 0.58 | 1.58 | 1.64 | 0.71 | 1.29 | 2.80 | 0.75 | 1.68 | 0.89 | 0.57 | 1.40 | 0.83 | 0.60 | 1.32 | 1.39 | 0.56 |
| | | (3.00) | (0.66) | (1.17) | (1.34) | (0.66) | (0.77) | (1.79) | (0.84) | (0.24) | (1.29) | (0.25) | (0.47) | (1.46) | (0.27) | (-0.03) | (4.18) | (0.11) | (-0.13) |
| | HLCA | 0.90 | 0.56 | 0.25 | 0.81 | 0.50 | 0.55 | 1.21 | 0.64 | 0.41 | 2.63 | 0.88 | 0.60 | 1.33 | 0.88 | 0.65 | 1.70 | 0.82 | 0.50 |
| | | (0.30) | (0.05) | (-0.29) | (0.32) | (-0.27) | (0.12) | (0.45) | (1.45) | (-0.28) | (1.16) | (0.92) | (0.22) | (2.41) | (0.64) | (0.68) | (2.37) | (0.81) | (0.31) |
| | INDS | 1.50 | 2.64 | 2.68 | 1.45 | 3.33 | 2.49 | 2.00 | 3.62 | 2.76 | 1.71 | 2.74 | 1.50 | 2.21 | 3.21 | 1.74 | 1.23 | 2.74 | 1.49 |
| | | (0.89) | (0.64) | (0.19) | (2.27) | (1.00) | (0.06) | (1.69) | (1.19) | (0.15) | (1.67) | (1.24) | (-0.04) | (2.75) | (2.60) | (0.15) | (1.27) | (0.67) | (0.24) |
| | MTRS | 0.97 | 0.62 | 1.13 | 1.09 | 0.78 | 1.38 | 1.51 | 0.85 | 1.20 | 1.53 | 1.06 | 1.35 | 2.02 | 1.27 | 1.50 | 1.68 | 1.71 | 1.39 |
| | | (0.23) | (-0.57) | (-0.25) | (0.62) | (-0.31) | (-0.06) | (0.31) | (-0.32) | (-0.10) | (1.10) | (-0.23) | (-0.28) | (1.32) | (0.50) | (0.04) | (2.69) | (0.84) | (0.10) |
| | FINC | 1.25 | 1.36 | 1.12 | 1.84 | 1.51 | 1.36 | 1.70 | 3.25 | 1.43 | 1.91 | 1.68 | 1.27 | 1.14 | 1.98 | 1.96 | 1.52 | 1.67 | 2.22 |
| | | (2.02) | (0.27) | (0.87) | (2.11) | (1.67) | (0.28) | (3.13) | (2.74) | (0.82) | (2.96) | (1.51) | (0.37) | (0.69) | (5.85) | (0.79) | (2.62) | (13.72) | (1.95) |
| | CONS | 1.02 | 1.28 | 1.67 | 1.23 | 1.33 | 1.36 | 1.29 | 1.41 | 2.20 | 1.35 | 1.00 | 4.29 | 1.58 | 1.40 | 3.40 | 1.26 | 1.24 | 4.73 |
| | | (1.26) | (0.32) | (0.09) | (1.48) | (0.62) | (0.15) | (0.47) | (1.20) | (-0.08) | (1.23) | (0.47) | (-0.26) | (1.30) | (0.73) | (-0.17) | (1.67) | (0.96) | (-0.38) |
| | INFT | 1.45 | 1.32 | 0.45 | 1.85 | 1.28 | 0.78 | 0.92 | 1.49 | 0.76 | 0.93 | 0.89 | 0.76 | 1.33 | 1.34 | 0.82 | 0.92 | 1.12 | 0.94 |
| | | (1.80) | (0.73) | (0.17) | (0.96) | (0.65) | (0.40) | (1.17) | (0.76) | (0.38) | (1.82) | (0.49) | (0.53) | (1.47) | (0.42) | (0.67) | (3.99) | (0.32) | (0.70) |
| | UTIL | 2.27 | 2.32 | 2.03 | 2.50 | 2.81 | 1.96 | 2.53 | 2.60 | 2.10 | 1.56 | 1.29 | 0.90 | 1.52 | 1.63 | 0.89 | 1.34 | 1.85 | 1.01 |
| | | (1.56) | (1.62) | (-0.93) | (1.29) | (1.65) | (-0.22) | (1.21) | (1.07) | (0.27) | (0.18) | (-0.32) | (-0.51) | (0.59) | (-0.19) | (-0.48) | (0.57) | (-0.12) | (-0.48) |
| | REES | 15.53 | 3.80 | 1.68 | 3.55 | 2.90 | 1.78 | 3.16 | 3.32 | 1.55 | 3.74 | 2.86 | 1.55 | 4.06 | 4.06 | 1.27 | 3.02 | 3.71 | 1.37 |
| | | (2.71) | (0.17) | (-0.55) | (3.00) | (-0.14) | (-0.82) | (6.03) | (-0.66) | (-0.61) | (1.67) | (-0.53) | (-0.48) | (2.84) | (-0.72) | (-0.52) | (0.99) | (-0.56) | (-0.39) |
| LSTM | TELE | 0.68 | 1.82 | 0.77 | 3.05 | 2.52 | 1.00 | 1.94 | 1.28 | 0.82 | 0.90 | 1.65 | 0.63 | 1.14 | 3.08 | 0.89 | 1.30 | 1.41 | 0.62 |
| | | (1.26) | (5.35) | (2.06) | (2.45) | (2.33) | (2.47) | (1.97) | (6.89) | (2.57) | (1.29) | (0.99) | (1.08) | (2.19) | (2.06) | (2.16) | (2.46) | (2.70) | (6.00) |
| | COND | 1.86 | 1.25 | 1.80 | 1.36 | 1.12 | 1.51 | 1.47 | 1.22 | 1.09 | 1.08 | 0.83 | 0.85 | 1.60 | 1.28 | 0.78 | 1.48 | 0.76 | 0.51 |
| | | (1.52) | (0.41) | (1.23) | (0.73) | (0.45) | (1.92) | (1.51) | (0.59) | (1.68) | (1.66) | (0.38) | (-0.11) | (1.50) | (0.57) | (0.91) | (2.99) | (0.88) | (0.77) |
| | ENRG | 2.12 | 1.03 | 0.82 | 1.85 | 0.86 | 0.70 | 2.40 | 1.04 | 0.68 | 2.02 | 0.98 | 0.93 | 1.75 | 0.95 | 0.95 | 2.03 | 1.25 | 1.04 |
| | | (1.79) | (2.35) | (0.59) | (1.85) | (1.07) | (0.26) | (2.43) | (1.58) | (0.66) | (2.64) | (2.04) | (2.90) | (1.32) | (2.07) | (1.07) | (2.62) | (0.83) | (2.89) |
| | HLCA | 1.60 | 0.72 | 0.50 | 1.43 | 0.73 | 0.50 | 2.70 | 0.82 | 0.53 | 1.15 | 0.65 | 0.48 | 2.04 | 0.64 | 0.50 | 2.28 | 0.80 | 0.50 |
| | | (1.72) | (1.61) | (0.70) | (3.81) | (1.62) | (1.23) | (3.67) | (2.33) | (1.72) | (2.40) | (0.85) | (0.69) | (5.42) | (1.13) | (1.01) | (5.68) | (1.47) | (1.58) |
| | INDS | 2.69 | 2.47 | 1.59 | 1.86 | 2.50 | 1.93 | 1.54 | 3.43 | 1.76 | 1.98 | 3.11 | 1.67 | 2.25 | 3.11 | 1.94 | 1.73 | 2.91 | 1.91 |
| | | (0.70) | (0.67) | (0.18) | (0.91) | (0.94) | (1.32) | (1.56) | (1.22) | (1.62) | (1.64) | (0.95) | (0.49) | (1.57) | (2.26) | (1.58) | (2.66) | (3.29) | (1.52) |
| | MTRS | 1.69 | 1.24 | 1.71 | 1.79 | 1.35 | 1.68 | 1.35 | 1.43 | 1.74 | 1.35 | 1.52 | 1.84 | 1.57 | 2.16 | 1.59 | 2.36 | 2.83 | 1.54 |
| | | (1.83) | (0.37) | (0.74) | (1.07) | (0.37) | (1.59) | (2.61) | (0.65) | (1.77) | (1.25) | (1.41) | (1.02) | (2.46) | (1.79) | (1.46) | (2.51) | (2.15) | (1.55) |
| | FINC | 2.35 | 2.08 | 0.94 | 2.08 | 1.85 | 1.50 | 1.81 | 2.61 | 2.01 | 2.14 | 2.19 | 1.35 | 1.65 | 1.80 | 1.76 | 1.67 | 1.62 | 1.86 |
| | | (1.67) | (1.43) | (0.56) | (1.84) | (3.00) | (0.92) | (5.86) | (3.07) | (1.14) | (2.20) | (2.34) | (1.06) | (1.88) | (2.09) | (1.52) | (3.39) | (2.87) | (1.84) |
| | CONS | 1.24 | 1.74 | 1.99 | 1.68 | 1.82 | 1.69 | 1.13 | 1.26 | 1.60 | 2.05 | 1.88 | 1.81 | 1.35 | 1.82 | 1.92 | 1.18 | 1.28 | 1.65 |
| | | (1.68) | (1.45) | (1.61) | (1.45) | (1.23) | (1.62) | (2.18) | (1.27) | (1.43) | (3.10) | (1.52) | (1.38) | (1.69) | (1.46) | (1.69) | (2.64) | (1.30) | (1.52) |
| | INFT | 1.67 | 1.74 | 1.17 | 0.80 | 1.09 | 0.72 | 1.33 | 1.07 | 0.66 | 1.52 | 1.78 | 0.90 | 0.92 | 1.32 | 0.52 | 1.21 | 0.89 | 0.59 |
| | | (1.21) | (0.93) | (1.37) | (2.31) | (1.20) | (2.67) | (2.63) | (1.03) | (1.49) | (1.47) | (0.95) | (1.83) | (2.01) | (1.93) | (0.99) | (1.94) | (1.12) | (1.04) |
| | UTIL | 2.71 | 2.16 | 1.98 | 2.01 | 2.17 | 1.85 | 3.59 | 2.93 | 2.02 | 3.07 | 2.21 | 1.94 | 2.72 | 2.29 | 1.98 | 4.21 | 3.08 | 2.15 |
| | | (3.14) | (1.15) | (0.86) | (2.27) | (1.41) | (0.90) | (2.85) | (2.13) | (1.15) | (4.04) | (2.08) | (1.07) | (2.77) | (1.59) | (1.08) | (2.83) | (2.40) | (1.53) |
| | REES | 6.40 | 5.44 | 3.06 | 3.84 | 3.30 | 2.45 | 2.60 | 2.60 | 2.32 | 6.58 | 3.81 | 3.16 | 4.44 | 3.59 | 3.05 | 2.99 | 3.53 | 2.61 |
| | | (1.34) | (0.24) | (-0.22) | (1.20) | (1.02) | (0.39) | (3.00) | (1.79) | (1.52) | (2.13) | (0.12) | (-0.21) | (2.16) | (1.32) | (0.88) | (4.20) | (2.14) | (2.11) |

*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the improvement.

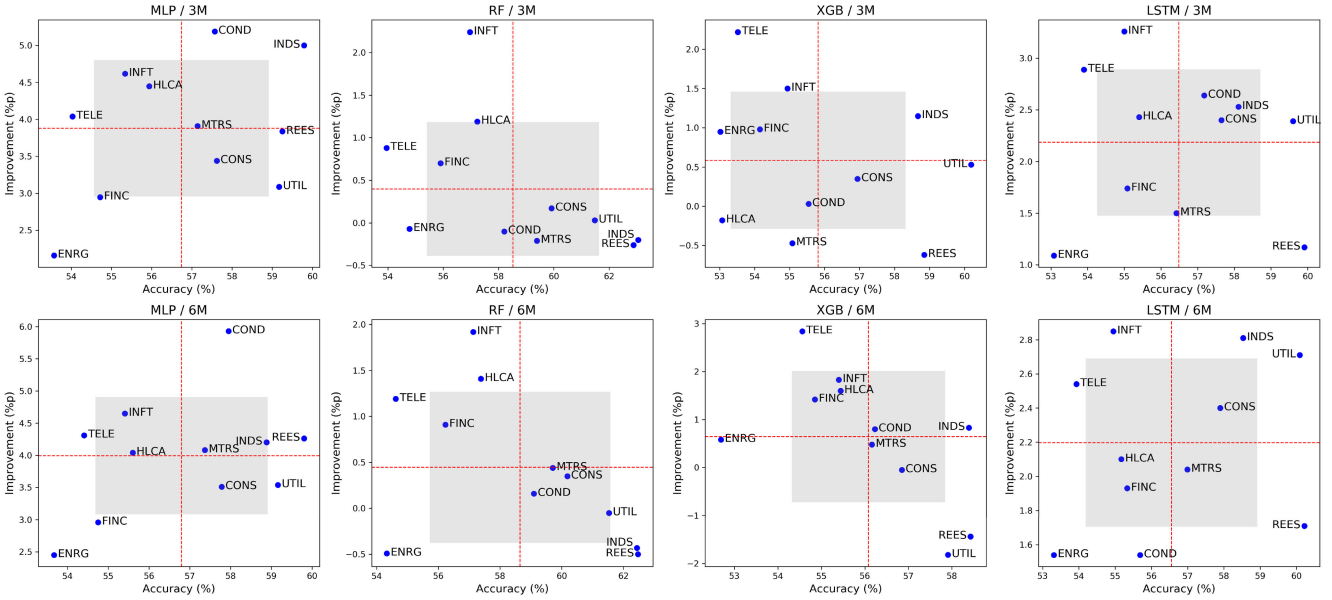
**TABLE 8.** Accuracy, adjusted accuracy, and their improvements in different sectors.

| **Accuracy** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **MLP** | | **RF** | | **XGB** | | **LSTM** | |
| MW | 3M | 6M | 3M | 6M | 3M | 6M | 3M | 6M |
| TELE | 54.03 | 54.41 | 53.95 | 54.62 | 53.52 | 54.56 | 53.90 | 53.93 |
| | (4.04) | (4.31) | (0.88) | (1.19) | (2.22) | (2.84) | (2.89) | (2.54) |
| COND | 57.57 | 57.95 | 58.21 | 59.10 | 55.54 | 56.23 | 57.18 | 55.69 |
| | (5.19) | (5.93) | (-0.10) | (0.16) | (0.03) | (0.80) | (2.64) | (1.54) |
| ENRG | 53.57 | 53.67 | 54.78 | 54.33 | 53.02 | 52.69 | 53.08 | 53.31 |
| | (2.16) | (2.45) | (-0.07) | (-0.49) | (0.95) | (0.58) | (1.09) | (1.54) |
| HLCA | 55.94 | 55.60 | 57.23 | 57.38 | 53.07 | 55.44 | 55.40 | 55.17 |
| | (4.45) | (4.04) | (1.19) | (1.41) | (-0.18) | (1.60) | (2.43) | (2.10) |
| INDS | 59.79 | 58.88 | 63.05 | 62.44 | 58.67 | 58.39 | 58.11 | 58.53 |
| | (5.00) | (4.20) | (-0.20) | (-0.43) | (1.15) | (0.83) | (2.53) | (2.81) |
| MTRS | 57.14 | 57.37 | 59.39 | 59.71 | 55.08 | 56.16 | 56.42 | 56.99 |
| | (3.91) | (4.08) | (-0.21) | (0.44) | (-0.47) | (0.48) | (1.50) | (2.04) |
| FINC | 54.72 | 54.75 | 55.90 | 56.23 | 54.15 | 54.85 | 55.08 | 55.33 |
| | (2.95) | (2.96) | (0.70) | (0.91) | (0.98) | (1.42) | (1.74) | (1.93) |
| CONS | 57.62 | 57.78 | 59.92 | 60.18 | 56.94 | 56.85 | 57.65 | 57.90 |
| | (3.44) | (3.51) | (0.17) | (0.35) | (0.35) | (-0.05) | (2.40) | (2.40) |
| INFT | 55.34 | 55.41 | 56.97 | 57.13 | 54.94 | 55.40 | 55.00 | 54.95 |
| | (4.62) | (4.65) | (2.24) | (1.92) | (1.50) | (1.83) | (3.26) | (2.85) |
| UTIL | 59.17 | 59.16 | 61.48 | 61.53 | 60.19 | 57.91 | 59.60 | 60.09 |
| | (3.09) | (3.54) | (0.03) | (-0.05) | (0.53) | (-1.82) | (2.39) | (2.71) |
| REES | 59.25 | 59.80 | 62.88 | 62.47 | 58.85 | 58.43 | 59.91 | 60.22 |
| | (3.84) | (4.26) | (-0.26) | (-0.50) | (-0.62) | (-1.44) | (1.17) | (1.71) |
| **Adjusted Accuracy** | | | | | | | | |
| | **MLP** | | **RF** | | **XGB** | | **LSTM** | |
| MW | 3M | 6M | 3M | 6M | 3M | 6M | 3M | 6M |
| TELE | 2.34 | 1.56 | 0.65 | 0.93 | 0.98 | 1.55 | 1.54 | 1.29 |
| | (2.09) | (2.04) | (0.53) | (0.88) | (1.41) | (1.68) | (3.04) | (2.33) |
| COND | 2.62 | 2.50 | 0.94 | 1.16 | 0.98 | 1.21 | 1.41 | 1.02 |
| | (1.89) | (2.24) | (0.16) | (0.47) | (0.25) | (0.81) | (1.12) | (1.06) |
| ENRG | 1.49 | 1.44 | 1.42 | 1.35 | 1.39 | 1.03 | 1.28 | 1.32 |
| | (3.32) | (2.80) | (0.48) | (-0.16) | (1.16) | (0.87) | (1.40) | (2.04) |
| HLCA | 1.23 | 1.31 | 0.68 | 0.75 | 0.65 | 1.11 | 1.06 | 1.00 |
| | (2.27) | (2.73) | (0.83) | (0.83) | (0.20) | (1.06) | (2.05) | (2.25) |
| INDS | 3.92 | 3.82 | 4.61 | 3.78 | 2.50 | 2.06 | 2.20 | 2.29 |
| | (3.12) | (2.70) | (0.35) | (0.52) | (0.90) | (1.17) | (1.01) | (1.77) |
| MTRS | 1.95 | 1.98 | 1.46 | 1.63 | 1.06 | 1.50 | 1.55 | 1.86 |
| | (1.80) | (2.04) | (0.44) | (0.99) | (-0.05) | (0.67) | (1.22) | (1.73) |
| FINC | 1.58 | 1.79 | 1.57 | 1.60 | 1.65 | 1.70 | 1.91 | 1.78 |
| | (2.01) | (2.34) | (1.22) | (1.29) | (1.55) | (3.38) | (2.17) | (2.13) |
| CONS | 2.17 | 2.59 | 1.63 | 1.96 | 1.42 | 2.25 | 1.57 | 1.66 |
| | (1.94) | (2.61) | (0.49) | (0.43) | (0.61) | (0.62) | (1.55) | (1.81) |
| INFT | 1.44 | 1.28 | 1.63 | 1.51 | 1.14 | 1.01 | 1.14 | 1.07 |
| | (2.42) | (1.93) | (0.93) | (0.91) | (0.78) | (1.16) | (1.65) | (1.48) |
| UTIL | 2.86 | 2.38 | 2.49 | 2.45 | 2.35 | 1.33 | 2.38 | 2.63 |
| | (1.92) | (2.11) | (0.31) | (0.32) | (0.84) | (-0.08) | (1.76) | (2.15) |
| REES | 5.88 | 5.15 | 7.10 | 6.15 | 4.14 | 2.85 | 3.56 | 3.75 |
| | (1.78) | (2.79) | (0.36) | (0.04) | (1.01) | (0.25) | (1.14) | (1.65) |

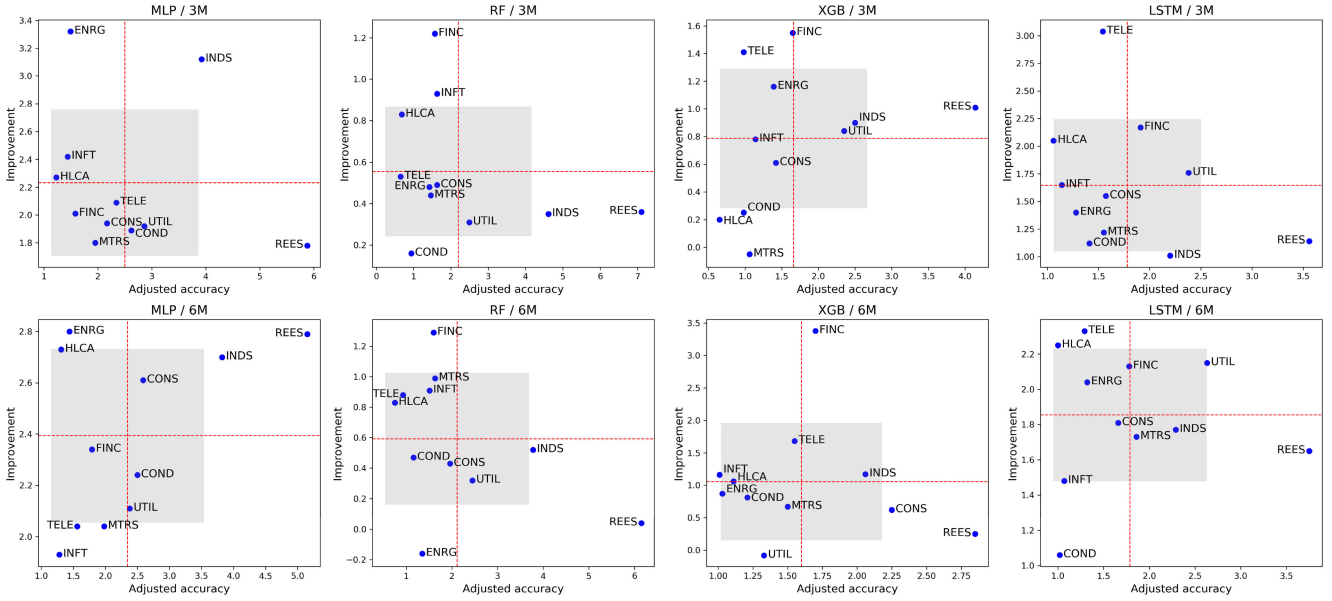
*Note: MW is the abbreviations of the moving window. The numbers in parenthesis indicate the improvement. The unit for accuracy and its improvement are percent and percent point, respectively.

prediction framework containing the sectors with high prediction accuracy and improvement. The second (fourth) quadrant encompasses sectors that have poor (decent) prediction performance but have decent (poor) performance improvement from ETE indicator. The third quadrant is the worst-case scenario containing sectors with poor performance in both prediction performance and performance improvement from ETE indicator.

Figure 10a and 10b shows the results on accuracy and adjusted accuracy, respectively. At first, in the accuracy aspect in Figure 10a, the best sector for prediction is INDS located in the first quadrant of all MLP, XGB, and LSTM cases and the fourth quadrant of RF. Sectors in the second quadrant with high accuracy improvement are INFT, TELE, HLCA. In particular, INFT and TELE are located in the second quadrants under all conditions. INFT can be found to deviate significantly from the indifferent region in terms of accuracy in RF, LSTM, confirming significant performance improvements due to the ETE network indicators. Likewise, TELE shows the decent imrpovement in XGB. Although HLCA is mostly in indifferent region, it is located in the second quadrant under all conditions except




(a) Accuracy




(b) Adjusted accuracy

**FIGURE 10.** **Prediction performance vs. prediction improvement.**

the case of XGB in 3M and LSTM in 6M. Sectors mostly found in the fourth quadrant with high prediction accuracy are UTIL and REES. UTIL is located in the fourth quadrant in all conditions except for the LSTM, whereas REES is distant from the indifferent region except for the MLP in 6M. Sectors that are not biased in a particular quadrant are FINC, CONS, MTRS, and COND. FINC is located in the second quadrant of RF and XGB and the third quadrant of MLP and LSTM, showing decent improvement in ensemble methods. CONS is located in the indifferent region under all conditions, but it was located in the first quadrant of LSTM and the fourth quadrant of the rest of the models with high accuracy performance. MTRS is mostly in the indifferent region without a specific trend. COND was located in the first quadrant of MLP and LSTM in 3M, showing higher accuracy and improvement, and mostly in the indifferent region in the remaining conditions. Lastly, the sector belonging to the third quadrant of the worst-case scenario is ENRG. Specifically, ENRG was located in the third quadrant under all conditions except XGB in 3M, indicating poor performance in both prediction accuracy and its improvement.

Then, in the adjusted accuracy aspect in Figure 10b, INDS, UTIL, and REES are all located in the first and fourth quadrant except for UTIL of XGB in 6M, showing consistent high

accuracy. In particular, the high improvement of consistent accuracy is also found in INDS for MLP, UTIL for LSTM, and REES for MLP in 6M and XGB in 3M. HCLA except for XGB in 3M and INFT except for MLP in 6M, LSTM in 6M, and XGB in 3M are located in the second quadrant, showing the high improvement in the consistent prediction performance. Also, the high improvement in consistent accuracy is observed in TELE of LSTM, FINC of ensemble methods including RF and XGB, and ENRG of MLP. Sectors belonging to the third quadrant of the worst-case scenario are CONS, MTRS, and COND. Especially, MTRS is located in the third quadrant of MLP in 6M and XGB in 3M, whereas COND is located in the third quadrant of RF in 3M, XGB in 3M, and LSTM in 6M. Note that CONS is mostly located in the indifferent region of the third quadrant.

In conclusion, the first quadrant sector consistently outperforms the other sectors in all cases, while the third quadrant sector consistently performs poorly. Thus, we claim that there is a more suitable sector for the application of ETE network indicators in predicting the direction of the stock.

## VI. CONCLUSION

Throughout the paper, we analyze the Granger-causal relationship of major US stocks according to financial events based on the time-varying ETE using the moving window method. Then, we utilize and examine the ETE network indicator as a feature to improve the prediction performance of the direction of the stock price through machine learning algorithms.

Many previous studies are describing Granger-causal relationships in the financial system using TE, which can identify asymmetry information flow between components. In this context, ETE used in this paper is an advanced method for controlling noise in measuring the information flow, which is a disadvantage of TE. Furthermore, the previous research on TE has focused on analyzing market phenomena in connection with Granger-causal relationships, whereas this paper focuses on a more practical question such as the utilization of ETE in the financial market. Thus, the novelty of this paper lies in the fact that, in our best knowledge, this is the first attempt to integrate the ETE of an individual US stock to analyze the US market and to predict the direction of the stock price.

The findings of this paper can be summarized as follows. At first, we discover that the time-varying ETE based on the 3M and 6M moving windows have market explanatory power using 55 stocks from 11 sectors and six cases of financial crises in the US financial market. Secondly, we detect the increases in the influence of sectors related to the financial crisis and the absolute size of information flow in the market. Thirdly, the utilization of ETE network indicators as new features improves the prediction on the stock price direction for all cases of the LR, MLP, RF, XGB and LSTM. In particular, the smaller the lag for prediction, the longer the prediction period, the higher the prediction accuracy. Fourthly, we identify the MLP and LSTM as more suitable machine learning algorithms in predicting the direction of stock price based on the adjusted accuracy, newly introduced performance measure based on the concept of risk-adjusted return. Lastly, we reveal the suitable sectors for the utilization of ETE network indicators.

The limitation of this research that should be addressed in future studies is the computation time to obtain the time-varying ETE. Especially, it is necessary to optimize the ETE computation algorithm to apply the mechanism to all stocks in S&P 500 broadly. Also, as a follow-up task, we are planning to propose and test the portfolio investment strategies using the prediction results and the Black-Litterman model.

## REFERENCES

[1] E. Ghysels and D. R. Osborn, *The Econometric Analysis of Seasonal Time Series*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
[2] R. N. Mantegna and H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Financ*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[3] J. D. Noh, "Model for correlations in stock markets," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 61, no. 5, p. 5981, 2000.
[4] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, "Topology of correlation-based minimal spanning trees in real and model markets," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 68, no. 4, Oct. 2003, Art. no. 046130.
[5] L. Zunino, B. M. Tabak, A. Figliola, D. G. Pérez, M. Garavaglia, and O. A. Rosso, "A multifractal approach for stock market inefficiency," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 26, pp. 6558–6566, Nov. 2008.
[6] K. T. Chi, J. Liu, and F. C. Lau, "A network perspective of the stock market," *J. Empirical Finance*, vol. 17, no. 4, pp. 659–667, 2010.
[7] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, "Random matrix approach to cross correlations in financial data," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 65, no. 6, Jun. 2002, Art. no. 066126.
[8] M. J. Kim, Y. B. Kwak, and S. Y. Kim, "Dependence structure of the korean stock market in high frequency data," *Phys. A, Stat. Mech. Appl.*, vol. 390, no. 5, pp. 891–901, Mar. 2011.
[9] S. Kumar and N. Deo, "Correlation and network analysis of global financial indices," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 86, no. 2, Aug. 2012, Art. no. 026101.
[10] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica, J. Econ. Soc.*, vol. 37, no. 3, pp. 424–438, Aug. 1969.
[11] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, p. 461, 2000.
[12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.
[13] M. Kim, D. Newth, and P. Christen, "Macro-level information transfer in social media: Reflections of crowd phenomena," *Neurocomputing*, vol. 172, pp. 84–99, Jan. 2016.
[14] L. Faes, G. Nollo, and A. Porta, "Compensated transfer entropy as a tool for reliably estimating information transfer in physiological time series," *Entropy*, vol. 15, no. 1, pp. 198–219, Jan. 2013.
[15] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—A model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 45–67, 2011.
[16] R. Marschinski and H. Kantz, "Analysing the information flow between financial time series," *Eur. Phys. J. B*, vol. 30, no. 2, pp. 275–281, Nov. 2002.
[17] A. Sensoy, C. Sobaci, S. Sensoy, and F. Alali, "Effective transfer entropy approach to information flow between exchange rates and stock markets," *Chaos, Solitons Fractals*, vol. 68, pp. 180–185, Nov. 2014.
[18] O. Kwon and J.-S. Yang, "Information flow between stock indices," *EPL (Europhys. Lett.)*, vol. 82, no. 6, Jun. 2008, Art. no. 068003.
[19] O. Kwon and J.-S. Yang, "Information flow between composite stock index and individual stocks," *Phys. A, Stat. Mech. Appl.*, vol. 387, no. 12, pp. 2851–2856, May 2008.

[20] O. Kwon and G. Oh, "Asymmetric information flow between market index and individual stocks in several stock markets," *EPL (Europhys. Lett.)*, vol. 97, no. 2, Jan. 2012, Art. no. 028007.

[21] T. Dimpfl and F. J. Peter, "Using transfer entropy to measure information flows between financial markets," *Stud. Nonlinear Dyn. Econometrics*, vol. 17, no. 1, pp. 85–102, 2013.

[22] L. Sandoval,"Structure of a global network of financial companies based on transfer entropy," *Entropy*, vol. 16, no. 8, pp. 4443–4482, Aug. 2014.

[23] Y. Chunxia, Z. Xueshuai, J. Luoluo, H. Sen, and L. He, "Study on the contagion among American industries," *Phys. A, Stat. Mech. Appl.*, vol. 444, pp. 601–612, Feb. 2016.

[24] K. Lim, S. Kim, and S. Y. Kim, "Information transfer across intra/inter-structure of CDS and stock markets," *Phys. A, Stat. Mech. Appl.*, vol. 486, pp. 118–126, Nov. 2017.

[25] P. Yue, Q. Cai, W. Yan, and W.-X. Zhou, "Information flow networks of Chinese stock market sectors," *IEEE Access*, vol. 8, pp. 13066–13077, 2020.

[26] M. P. Clements, P. H. Franses, and N. R. Swanson, "Forecasting economic and financial time-series with non-linear models," *Int. J. Forecasting*, vol. 20, no. 2, pp. 169–183, 2004.

[27] M. R. Ola, M. J. Nooghabi, and M. M. Rounaghi, "Chaos process testing (using local polynomial approximation model) in predicting stock returns in tehran stock exchange," *Asian J. Res. Banking Finance*, vol. 4, no. 11, pp. 100–109, 2014.

[28] M. Moradi, M. J. Nooghabi, and M. M. Rounaghi, "Investigation of fractal market hypothesis and forecasting time series stock returns for Tehran Stock Exchange and London Stock Exchange," *Int. J. Finance Econ.*, pp. 1–17, 2019, doi: 10.1002/ijfe.1809.

[29] R. Tsaih, Y. Hsu, and C. C. Lai, "Forecasting S&P 500 stock index futures with a hybrid AI system," *Decis. Support Syst.*, vol. 23, no. 2, pp. 161–174, Jun. 1998.

[30] E. Guresen, G. Kayakutlu, and T. U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10389–10397, Aug. 2011.

[31] A. M. Rather, A. Agarwal, and V. N. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3234–3241, Apr. 2015.

[32] J. Zahedi and M. M. Rounaghi, "Application of artificial neural network models and principal component analysis method in predicting stock prices on tehran stock exchange," *Phys. A, Stat. Mech. Appl.*, vol. 438, pp. 178–187, Nov. 2015.

[33] K.-J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, nos. 1–2, pp. 307–319, Sep. 2003.

[34] Y. Kara, M. A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5311–5319, May 2011.

[35] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, Jan. 2015.

[36] M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Syst. Appl.*, vol. 42, no. 20, pp. 7046–7056, Nov. 2015.

[37] Z. Guo, H. Wang, J. Yang, and D. J. Miller, "A stock market forecasting model combining two-directional two-dimensional principal component analysis and radial basis function neural network," *PLoS ONE*, vol. 10, no. 4, Apr. 2015, Art. no. e0122385.

[38] M. Qiu and Y. Song, "Predicting the direction of stock market index movement using an optimized artificial neural network model," *PLoS ONE*, vol. 11, no. 5, May 2016, Art. no. e0155133.

[39] T. K. Lee, J. H. Cho, D. S. Kwon, and S. Y. Sohn, "Global stock market investment strategies based on financial network indicators using machine learning techniques," *Expert Syst. Appl.*, vol. 117, pp. 228–242, Mar. 2019.

[40] H. Hu, L. Tang, S. Zhang, and H. Wang, "Predicting the direction of stock markets using optimized neural networks with Google trends," *Neurocomputing*, vol. 285, pp. 188–195, Apr. 2018.

[41] S. Feuerriegel and J. Gordon, "Long-term stock index forecasting based on text mining of regulatory disclosures," *Decis. Support Syst.*, vol. 112, pp. 88–97, Aug. 2018.

[42] M. Hagenau, M. Liebmann, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decis. Support Syst.*, vol. 55, no. 3, pp. 685–697, Jun. 2013.

[43] T. Geva and J. Zahavi, "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news," *Decis. Support Syst.*, vol. 57, pp. 212–223, Jan. 2014.

[44] S. W. K. Chan and J. Franklin, "A text-based decision support system for financial sequence prediction," *Decis. Support Syst.*, vol. 52, no. 1, pp. 189–198, Dec. 2011.

[45] Y. Shynkevich, T. M. McGinnity, S. A. Coleman, and A. Belatreche, "Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning," *Decis. Support Syst.*, vol. 85, pp. 74–83, May 2016.

[46] A. Ntakaris, G. Mirone, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Feature engineering for mid-price prediction with deep learning," *IEEE Access*, vol. 7, pp. 82390–82412, 2019.

[47] D. R. Cox, "The regression analysis of binary sequences," *J. Roy. Stat. Soc., Ser. B, Methodol.*, vol. 20, no. 2, pp. 215–232, Jul. 1958.

[48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[49] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[51] H. M. Mok, "Causality of interest rate, exchange rate and stock prices at stock market open and close in hong kong," *Asia Pacific J. Manage.*, vol. 10, no. 2, pp. 123–143, Oct. 1993.

[52] N. R. Swanson, E. Ghysels, and M. Callan, "A multivariate time series analysis of the data revision process for industrial production and the composite leading indicator," in *Cointegration, Causality and Forecasting: Festschrift in Honor of Clive WJ Granger*. London, U.K.: Oxford Univ. Press, 1999.

[53] U. Soytas and R. Sari, "Energy consumption and GDP: Causality relationship in G-7 countries and emerging markets," *Energy Econ.*, vol. 25, no. 1, pp. 33–37, Jan. 2003.

[54] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *Eur. J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018.

[55] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, Jul. 2017, Art. no. e0180944.

[56] C.-F. Tsai and Y.-C. Hsiao, "Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches," *Decis. Support Syst.*, vol. 50, no. 1, pp. 258–269, Dec. 2010.

[57] L.-J. Kao, C.-C. Chiu, C.-J. Lu, and C.-H. Chang, "A hybrid approach by integrating wavelet-based feature extraction with MARS and SVR for stock index forecasting," *Decis. Support Syst.*, vol. 54, no. 3, pp. 1228–1244, Feb. 2013.

[58] H. Shin, T. Hou, K. Park, C.-K. Park, and S. Choi, "Prediction of movement direction in crude oil prices based on semi-supervised learning," *Decis. Support Syst.*, vol. 55, no. 1, pp. 348–358, Apr. 2013.

[59] S. Pyo, J. Lee, M. Cha, and H. Jang, "Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the korean stock markets," *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0188107.

[60] L. Khansa and D. Liginlal, "Predicting stock market returns from malicious attacks: A comparative analysis of vector autoregression and time-delayed neural networks," *Decis. Support Syst.*, vol. 51, no. 4, pp. 745–759, Nov. 2011.

[61] K. Nam and N. Seong, "Financial news-based stock movement prediction using causality analysis of influence in the korean stock market," *Decis. Support Syst.*, vol. 117, pp. 100–112, Feb. 2019.

[62] J. Nobre and R. F. Neves, "Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets," *Expert Syst. Appl.*, vol. 125, pp. 181–194, Jul. 2019.

[63] M. Jiang, J. Liu, L. Zhang, and C. Liu, "An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms," *Phys. A, Stat. Mech. Appl.*, vol. 541, Mar. 2020, Art. no. 122272.

[64] K. Chen, Y. Zhou, and F. Dai, "A lstm-based method for stock returns prediction: A case study of China stock market," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Nov. 2015, pp. 2823–2824.

[65] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1419–1426.

[66] L. Sandoval, "To lag or not to lag? How to compare indices of stock markets that operate on different times," *Phys. A, Stat. Mech. Appl.*, vol. 403, pp. 227–243, Jun. 2014.
[67] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, p. 533, 1986.
[68] W. F. Sharpe, "The sharpe ratio," *J. Portfolio Manage.*, vol. 21, no. 1, pp. 49–58, Oct. 1994.

**SONDO KIM** received the B.Sc. degree in mathematics and the B.A. degree in economics from Sungkyunkwan University (SKKU), Seoul, South Korea, in 2014, the M.Sc. degree in industrial engineering from the Seoul National University, Seoul, in 2016, where he is currently pursuing the Ph.D. degree in industrial engineering. His research interests include econophysics, machine learning, data-mining, time-series analysis, financial engineering and applications of these areas in trend prediction, investment strategy, crisis management, option pricing, and other topics related to data-driven analytics.

**SEUNGMO KU** received the B.Sc. degree in industrial and management engineering, and mathematics from KAIST, Daejeon, South Korea, in 2015. He is currently pursuing the joint M.Sc. and Ph.D. degree in industrial engineering with Seoul National University, Seoul, South Korea. His research interests and specialties are focused on data-driven analytics with emphasis on the econophysics, machine learning (deep learning and natural language processing), time-series analysis, and applications of these areas in prediction problem, pattern recognition, anomaly detection, risk assessment, demand forecasting, and other topics related to business analytics.

**WOOJIN CHANG** received the B.Sc. degree in polymer science and fiber engineering from Seoul National University, Seoul, South Korea, in 1997, and the M.Sc. degree in operations research and the Ph.D. degree in industrial engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1998 and 2002, respectively. From 2002 to 2003, he was an Assistant Professor with the Department of Decision Science and Engineering System, Rensselaer Polytechnic Institute, Troy, NY, USA. He joined the Department of Industrial Engineering, Seoul National University, in 2004, where he is currently a Professor running the Financial Risk Engineering Laboratory.

**JAE WOOK SONG** received the B.Sc. degree from the Georgia Institute of Technology, in 2010, and the Ph.D. degree from Seoul National University, in 2016, in industrial engineering. Then, he had been working as a Senior Data Scientist with Samsung Electronics and an Assistant Professor of data science with Sejong University. In 2019, he joined the Department of Industrial Engineering, Hanyang University, where he is currently an Assistant Professor running the Financial Innovation and Analytics Laboratory. His research interests include inter- and multi-disciplinary approaches for data-driven innovations in financial markets, risk management, investment decisions, and applied analytics.