Knowledge-Based Systems 235 (2022) 107608

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Forecasting price movements of global financial indexes using complex quantitative financial networks

Nohyoon Seong [a], Kihwan Nam [b,*]

[a] Management Engineering Department, College of Business, Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea
[b] Information Systems, Business School, Dongguk University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 1 April 2021
Received in revised form 19 August 2021
Accepted 14 October 2021
Available online 19 October 2021

Keywords:
Financial network
Historical price movement prediction
Technical analysis
Transfer entropy
Random matrix theory

ABSTRACT

As predicting trends in the financial market becomes more important, and artificial intelligence technology advances, there is active research on predicting stock movements by analyzing historical prices. However, few attempts have been made to utilize complex financial networks to predict price movements. Most studies focus only on the target financial index, and only a few studies examine both the target financial index and influential financial indexes. To fill the research gap, we propose a novel deep learning algorithm based on quantitative complex financial networks to forecast the price movement of global financial indexes using technical analysis. In other words, we propose a method that analyses the causal relationship between financial indexes in a cleaned correlation network, which considers the causal impact in this quantitatively constructed network. We use the random matrix theory to construct the financial network first, and we use transfer entropy to find directional impact within the network. Based on a daily historical dataset for global indexes and out-of-sample tests, the results show that the proposed method outperforms past state-of-the-art algorithms. Our findings reveal that identifying and using proper financial networks are important in predicting problems. Our study suggests that it is important to develop deep learning algorithms and to consider the financial network based on complex system theory when solving prediction problems in the financial market.

1. Introduction

Analyzing and predicting stock market trends is becoming increasingly important for both business and academics [1]. As a deeper understanding of data analysis and artificial intelligence is being established in the financial market and the number of data increases, the number of attempts to forecast in the financial market increases. Algorithms that predict financial markets not only help traders, but themselves can create portfolios that outperform market returns [2,3]. That is, financial market prediction and directional change prediction is important in the financial industry.

The concept that the market is constantly processing all of the available data and embeds it into asset prices, resulting in the immediate incorporation of any new data at a moment in time creates the basis for the efficient markets hypothesis (EMH) of Fama [4]. Fama's theory defines the market of efficiency in three levels, that is, strong, semi-strong, and weak. The weak level presents the claim of current market prices as a reflection of all data available publicly throughout history. The semi-strong level makes the assumption that all historical and currently public data is already integrated and absorbed into the prices of the traded stocks. The strong level assumes the immediate inclusion of insider and latent data in a market price. The basics of the EMH states that an asset's present price encapsulates all previous, general, and unpublicized data and that systematic outperformance of the market is impossible. The Random Walk Theory adheres to an identical distribution and states that the variation in stock prices is inter-independent. As a result, a little bearing is put on an asset's price in relation to its future movements, and predictions using the price cannot be made. The most likely way to shift the asset price is through a random walk but reliable predictions are unlikely according to this theory.

The topic of market efficiency in terms of its scope and applicability to various markets is still an active and continuous area of study, with conflicting results. In an attempt to reconcile the EMH with behavioral finance [5], researchers suggested the Adaptive Market Hypothesis (AMH) as a counter-theory. Instead of looking at market price as a subsidiary product of its cost,

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Correspondence to: Information Systems, Business School, Dongguk University, 30, Pildong-ro, 1-gil, Jung-gu, Seoul, Republic of Korea.
E-mail address: namkh@dongguk.edu (K. Nam).

https://doi.org/10.1016/j.knosys.2021.107608
0950-7051/

behavioral finance looks at it entirely as a perceived benefit. Overreaction, overconfidence, knowledge bias, and representative bias are all cognitive biases that market agents possess, suggesting that mistakes in information processing and reasoning caused by human error are predictable [6]. Urquhart and Hudson [7] conduct a systematic analytical report on the AMH, examining three of the world's most established capital markets: the UK, US, and Japanese stock markets. To differentiate stock return behaviors, the authors utilized long-run data and developed five-yearly subsamples which they subjected to linear and non-linear trials. Through linear test results, proof of the different stock markets being adaptive markets where returns oscillate between dependent and independent periods is evident. In every subsample for each sector, non-linear tests display a heavy dependence for each sector, however, the extent of the dependence varies greatly. The AMH tends to better characterize the conduct of stock returns than the EMH. According to AMH [8], financial markets do not behave randomly, and market movements can be predicted.

Therefore, many financial market forecasting studies are based on technical analysis (TA) [8,9]. Technical analysis is a method of predicting stock prices using historical data including open price, high price, low price, close price, and volume size. For technical analysis, most of the studies compute technical indicators (TIs), which are mathematical calculations based on historical prices and volume to predict the financial market [8,9]. For example, Teixeira and Oliveira [10] suggested a method to combine technical analysis and the nearest neighbor classification to improve the performance of the technical analysis. Long et al. [11] suggested a multi-filters network for feature extraction with technical analysis. Some studies investigated the importance of prediction horizon and window size on prediction performance [12,13].

Most of the studies used each company's technical indicators and predicted the movement of the corresponding company's stock price movement. However, recent research suggests that considering the network effect may give better results than simply using each company's information in predicting stock prices [1,3]. Financial networks have several structures such as shareholder networks, production networks, and the Global Industry Classification Standard (GICS). However, most financial networks have a limitation in that they are already established through external information, which is constructed only in a limited field. For example, shareholder networks and production networks use external data sources, and GICS is a pre-defined economic-based industry classification system.

Therefore, to fill the void in the literature, we present a method of constructing a financial network quantitatively and propose a predictive methodology using this network effect which results in higher performance than those established in existing studies.

The primary purpose of the research is to predict financial market trends. Accordingly, we propose a powerful prediction model, controlling for the two key points. First, the current article utilizes the meaningful data extraction method. Just as the output cannot be meaningful if the input is not meaningful, we do not simply put all the data we can reach for prediction but classify and extract the meaningful data and utilize it for market prediction. This method is extremely important particularly when data is dotted with unavoidable noise like financial market analytics. Second, we focus on the relationship among financial indexes for market trends prediction. As financial indexes are closely related to each other, the direction of influence can be either one-sided or mutual. However, despite the importance of considering causation, the most predictive analysis assumes correlation among financial indexes. Therefore, current research proposes a novel framework that only considers the direction of influence that is significant to another. These two methodologies are very useful for financial market trends analysis because (1) the financial market itself is jam-packed with noise, often called nearly random, which makes prediction exceedingly challenging, and (2) all financial indexes have significant influence either one-sided or mutual. Therefore, to deal with the complicated nature of the financial market, we propose a novel framework for predictive analysis on the financial market.

We combine two methodologies to build a financial network: the Random Matrix Theory (RMT) and Transfer Entropy (TE). RMT is a theory that calculates the true correlation between each time series in a large time-series dataset [14,15]. The empirical correlation of financial markets is inaccurate due to white noise and market-wide effects [16]. Therefore, in order to implement an accurate financial network, a true correlation matrix must be obtained using RMT, which is able to prevent spurious correlation. However, since RMT is a correlation-based method, it is disadvantaged in that it cannot consider a directional relationship. Furthermore, transfer entropy is a way to find causality in time series. It is important to consider causality because a specific financial index affects other financial indexes [17], but an inverse relationship does not exist in many cases. Transfer entropy is suitable for financial networks because it can confirm a nonlinear relationship, unlike Granger causality [1]. However, because transfer entropy detects causality only when used with time series, it can have spurious causality [18]. Therefore, this paper complements the shortcomings of each methodology and maximizes the advantages by combining the two financial networks. Each of the two methodologies is widely used. When finding the direction of the causal relationship considered in the Transfer Entropy (TE) method, detailed noise control is difficult because the network changes depending on what information is included and calculates the causal relationship on the entire network. In other words, removing the noise first and then creating the network will yield different results than when creating the network and removing the noise. Therefore, this study proposes how to optimize the model by applying Random Matrix Theory (RMT) to the existing limit of Transfer Entropy (TE) to derive the maximized effect.

For the time series classification, we presented a deep learning algorithm that combines state-of-the-art algorithms [19,20]. We used the dual-stage attention structure to consider the dynamics and temporal dynamics of the input features. The input features were considered for interaction between features via CNN, and feature recalibration was performed for squeeze-and-excitation [19]. After that, the multi-layer LSTM was used for temporal dynamics, and the importance of variables was evaluated through the attention layer. In this way, the proposed method adaptively selects the most relevant input features as well as captures the long-term temporal dependencies of a time series in an appropriate manner.

To justify the effectiveness of the proposed method, we obtained 52 daily financial index datasets from January 1971 to December 2019. Among these, we predicted the 13 most important financial indexes in the Korean financial market based on the recommendation of a renowned financial company. The reason for using global financial indexes is to show that the proposed method works well even where financial networks are not clearly established. The results are as follows. Considering the only RMT network effect showed better accuracy and a higher F1-score than simply using the company's TA, and considering the only TE network effect also showed better accuracy and a higher F1-score as compared to using only TA. The proposed method, which combines the effects of RMT and TE, showed statistically superior performance than the method considering network effects and the method only considering TA. Besides accuracy and the F1-score, this method also showed higher returns and a Sharpe ratio in portfolio management through back-testing.

The current article contributes to academia with the following. First, in predicting the financial market, we combined a complex system theory with a deep learning algorithm and showed the effectiveness, extending the work of Nam and Seong [1]. We are one of the first studies to integrate complex system theory with deep learning. Second, most of the previous studies predicted price movements at individual levels, and only a few studies searched for relevant assets based on pre-defined financial networks. To the best of our knowledge, we are the first to predict price movements by considering the causality and correlation network based on a quantitative financial network. Finally, we combined the state-of-the-art methodology with time series classification, taking into account both the interaction between features and temporal dynamics.

We organize the remainder of this paper as follows. Section 2 provides a literature review on technical analysis, time series classification algorithms, and complex networks. Section 3 describes historical price datasets, transfer entropy, random matrix theory, time series classification algorithm, and performance metrics. Section 4 describes the experimental results. Section 5 explains the limitations and future works.

2. Literature review

2.1. Key related research

Technical indicators are mathematical calculations to forecast market direction with historical prices and volume. As reported by Atsalakis and Valavanis [8], approximately 20% of the financial market prediction studies use technical analysis. In other words, technical analysis is a very important tool and is widely used in financial market forecasting. For example, Teixeira and Oliveira [10] suggested a method to combine technical analysis and the nearest neighbor classification to improve the performance of prediction. Kara et al. [12] investigated how the performance of prediction depends on a forecast horizon and a window size. Long et al. [11] suggested a feature engineering algorithm with technical indicators based on the multi-filter structure of a convolutional neural network and recurrent neural network. In addition, there is some research on the characteristics of technical indicators. Shynkevich et al. [13] suggested how to make technical indicators by changing the input window length for each prediction horizon. Pan et al. [21] emphasized the importance of utilizing mixed frequency sampling.

Most of the studies used each company's technical indicators and predicted the movement of the corresponding company's stock price movement. However, recent research suggests that considering the network effect rather than simply using each company's information shows better performance in predicting stock prices. Many studies have constructed a network based on GICS [1,3]. Seong and Nam [3] clustered companies within GICS and showed that homogeneity was different for each sector, thus the network configuration should be different for each sector. Nam and Seong [1] first used transfer entropy to find causal relationships in companies based on GICS, and when predicting the stock price of a target company, they showed higher performance after considering causal relationships in companies. They suggested that it is important to implement a network structure well rather than simply consider GICS as a network in prediction. Kim et al. [22] also used transfer entropy to find causal relationships and predict the stock price. They established causal relationships in the cross-sector and found that transfer entropy was helpful in predicting the stock price. However, these studies were limited in that their stock price prediction methodology was based on GICS [1,22]. This may cause spurious causality [18], and this has a disadvantage of poor extensibility in predicting financial indexes where a pre-defined GICS does not exist. Also, Chen et al. [23] suggested a graph convolutional network for stock market prediction. The authors constructed the graph based on Spearman's correlation with stock prices. However, the paper has a limitation that simple empirical correlation does not fully represent stock market relationships because of the white noise and market co-movement [15,16]. To overcome these limitations, we propose a novel financial network that can achieve stock market relationships considering white noise and market co-movement, and alleviate spurious causality which will be used to predict the stock price.

2.2. Transfer entropy

Transfer entropy is a non-parametric measure for a directed transfer of information between two time series [24]. Transfer entropy is used to find causality in nonlinear times series [1]. The Wiener–Granger Causality is widely used to check for causality in time series [25,26]. However, Wiener–Granger Causality is limited in that it cannot detect causality in cases of stock herding behavior [27,28] and in complex systems with critical points [29]; therefore it is inappropriate in establishing nonlinear causal relationships. As a nonlinear causality needs to be used in the stock market, we use transfer entropy [30].

Many studies used transfer entropy for causality detection in the financial market [1,17,22]. First, Marschinski and Kantz [17] analyzed the causality with transfer entropy between two financial time series. They found that the magnitude of causality differs from each other. However, the study has a limitation in that the authors did not find the statistical significance of the relationship. Therefore, to overcome the limitation, we compute transfer entropy by considering statistical significance. Nam and Seong [1] found causal relationships using transfer entropy between individual stock prices in the Korean market. They found that the causal relationship helps to predict stock movements. However, the study has a limitation in that the authors measured transfer entropy in the GICS sector. This has a limitation that it is difficult to use in financial markets without pre-defined industry classification, and since GICS is not a classification system created for prediction, it is difficult to use it for prediction. Therefore, in order to overcome these limitations, we suggest a financial network that is able to apply to markets that do not exist in the pre-defined industry classification system and is suitable for prediction without having issues such as spurious causality, and we use transfer entropy within that financial network.

2.3. Random matrix theory

A correlation has always been an important issue in finance. As the Markowitz portfolio used correlations, the correlation became the basis of the portfolio [31]. In addition, correlations are also essential in finding related indexes or companies. Christodoulakis [32] suggested an evolving correlation matrix and found stock co-movement based on a correlation coefficient. Jung et al. [33] analyzed the financial market structure by performing agglomerative hierarchical clustering through correlation analyses.

However, a true correlation is difficult to achieve based on empirical correlation in the financial market due to white noise and market-wide movement [16]. Therefore, in econophysics, the Random Matrix Theory (RMT) was proposed, which cleans the correlation matrix in financial markets and finds true correlations using empirical correlations [15].

Laloux et al. [34] conducted one of the early studies that applied the RMT to the financial market. They found that the financial correlation matrix has a random nature, which implies that the use of RMT is necessary to establish true correlations in the financial market. Utsugi et al. [35] found the random eigenvalues in financial correlations. They also suggested a method to cluster based on correlation with RMT. Kim et al. [16] also used RMT to establish true correlation and suggested a method to a cluster using a cleansed correlation matrix. The results of clustering suggested that the RMT reflects the GICS and the characteristics of the cross-sector are also confirmed. Garcia [36] suggested that there is a true correlation between financial indexes by analyzing the global financial indexes and Twitter sentiment.

It is important to apply RMT in the financial market, however, it is also important to apply this properly [37]. In the literature, five RMT methodologies were suggested: basic linear shrinkage [38], advanced linear shrinkage [39], eigenvalue clipping [14], eigenvalue substitution [40], and rotational invariant estimator [41]. However, many estimators are less efficient than the basic linear shrinkage [37]. Among these, the rotational invariant estimator is considered state-of-the-art. Therefore, we used a rotational invariant estimator to establish a true correlation [15].

2.4. Time series classification algorithm

Several studies have been conducted using the methodology of predicting time series. Many studies attempt to predict price movements using machine algorithms such as Support Vector Machine [42], Artificial Neural Network [42], genetic fuzzy systems [43], k Nearest Neighbors [44], and Elman Neural Network [45] by analyzing variables through technical analysis. Some attempts have been made to combine ARIMA and machine learning techniques [46]. Recently, using the Hidden Markov model has made a breakthrough in time series prediction [47].

However, with the development of deep learning algorithms, time series classification has shown much higher performance than existing methods. In previous studies, it was common to make models using CNN or LSTM. Zheng et al. [48] created a time series classification model that put each feature of the time series into a 1-D CNN, performed feature extraction, and synthesized the features with a fully connected layer. Bai et al. [49] developed 1-D CNN architecture by using causal 1-D CNN which assumed that there was no information “leakage” from the future to the past. They showed that causal 1-D CNN has low memory requirements and captures local information, making it more efficient than RNN networks. Wang et al. [50] showed that ResNet, which is widely used in image classification, also works well for time series classification. Malhotra et al. [51] proposed a multilayered recurrent neural network that is pre-trained with a sequence auto-encoder and found that a pre-trained model yielded a significantly improved performance.

There are several recent studies in which the attention structure was used to increase performance and combine the advantages of using both CNN and RNN series. Long et al. [11] proposed a multi-filter structure based on the integration of convolutional neural networks and recurrent neural network for feature extraction. Qin et al. [20] proposed a recurrent neural network-based model with attention to capture long-term temporal dependencies and recalibrate features. Karim et al. [19] proposed a network that combines the advantages of CNN and LSTM. The model showed high performance by applying squeeze-and-excitation [52] to CNN for feature recalibration and applying an attention network to LSTM to fully utilize the features. Following the state-of-the-art papers, we used an attention-based algorithm and a methodology that combines the advantages of CNN and LSTM in this study.

3. Proposed method

In this chapter, we specifically described how to construct a model that predicts price movements by finding influential indexes with causality analysis and correlated networks through technical analysis. First of all, we describe the historical data. Afterward, we described transfer entropy and the random matrix theory, both of which are utilized mainly in econophysics and to examine the causal relationship between indexes, and to examine the true correlation network between indexes by removing noises. Finally, the time series classification algorithm and what evaluation indicators were utilized were explained. This series of processes is illustrated in Fig. 1.

3.1. Data and labeling

In order to predict global financial market trends, we have collected daily historical data of a total of 52 financial indexes. The list of indexes is available in Table 1. We used all the indexes for the RMT and TE network and for the relevant features. We used 13 indexes, which are bolded in Table 1, for the prediction and algorithm comparison. The 13 indexes were selected because these were the most important indexes in the Korean financial market and they were recommended by a renowned financial company.

The dataset was downloaded from Bloomberg. The dataset contains daily open, close, high, and low prices, and the trading volume for the corresponding business day. The dataset included those from every business day from January 1, 1971, to December 31, 2019. The dataset is divided into three sets: a training set, a validation set, and a test set. The training set included data from January 1, 1971, to December 31, 2014, the validation set included data from January 1, 2015, to December 31, 2016, and the test set included data from January 1, 2017, to December 31, 2019.

In the following experiments, the binary directions of future price movements are predicted. We predict stock movements based on three prediction horizons: short-term, medium-term, and long-term to analyze the results in various aspects. For the short-term, the prediction horizon was one day, for the medium-term, the prediction horizon was two weeks, and for the long-term, the prediction horizon was one month. The label assignment is created to reflect the motion behind the prediction horizon of the close price, as in Eq. (1) where h refers to the prediction horizon.

$$Label\,(t,h) = \begin{cases} 1 & if \ \frac{P_{t+h}-P_t}{P_t} \geq 0 \\ 0 & if \ \frac{P_{t+h}-P_t}{P_t} < 0 \end{cases} \tag{1}$$

The list of all indexes that were used in the experiment and the up-label distributions are shown in Table 2.

3.2. Technical analysis

Stock market prices reflect expectations and form trends and patterns. Technical analysis is a tool to predict future prices using market sentiment through price data analysis. The technical indicator is a snapshot of the market sentiment through mathematical analysis of past market movements. To determine technical indicators, we needed to choose which indicators to use and up to what time in the past these indicators could consider that would aid in making better predictions of future price movements. First, technical indicators have different inputs for each indicator. Some indicators need all Open, High, Low, Close, and Volume types, some indicators need Open, High, Low, Close types, and some indicators only need the Close type. Since each index has different types of data available, we have created

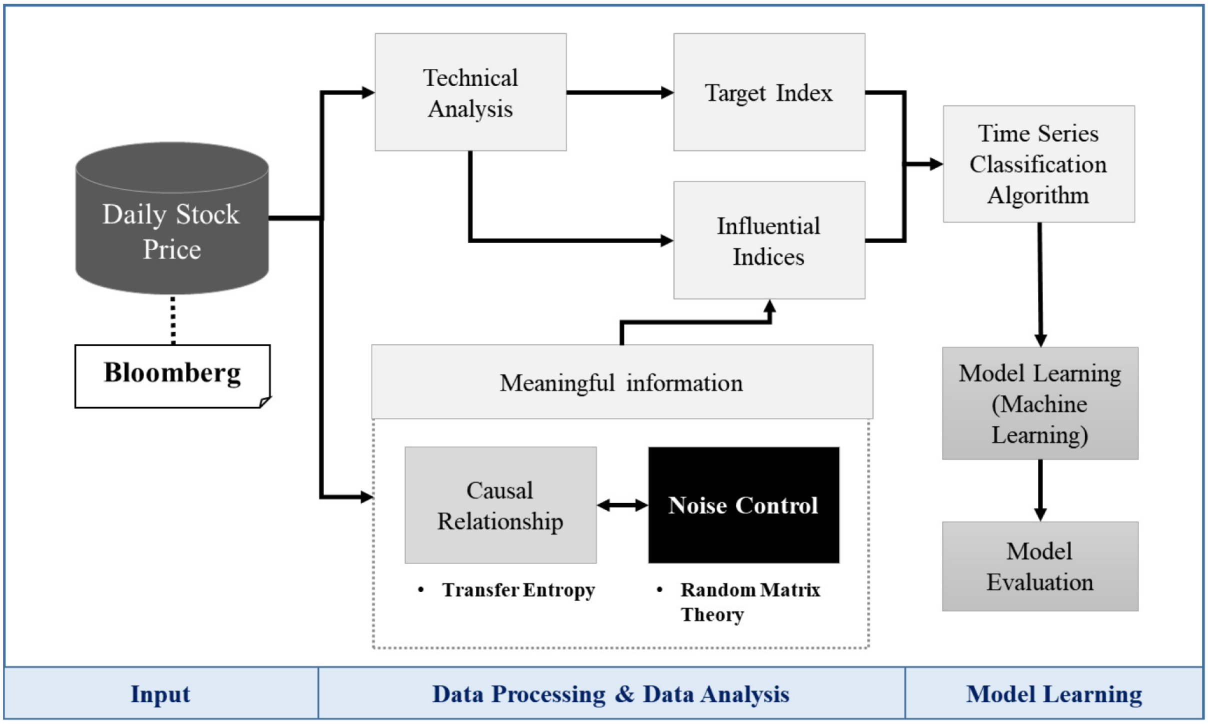

Fig. 1. Proposed approach.

Table 1
indexes used in the paper.

Index ticker	Index name	Index ticker	Index name
MXWD index	MSCI ACWI Index	IGUUDC01 index	US Corporate High Quality Corporate Bond 1Y Yield Index
MXEF index	MSCI Emerging Markets Index	IGUUDC10 index	US Corporate High Quality Corporate Bond 10Y Yield Index
MXWOU index	MSCI World ex USA Index	CSI BARC index	BarCap US Corp HY YTW - 10 Year Spread Index
SPX index	S&P 500 Index	BASPCAAA index	US Corporate AAA 10 Year Spread Index
DJI index	Dow Jones Industrial Average Index	**LUATTRUU index**	Bloomberg Barclays US Treasury TR Value Index
UKX index	FTSE 100 Index	**LUACTRUU index**	Bloomberg Barclays US Corporate TR Value Index
KOSPI index	KOSPI Index	**LBUTTRUU index**	Bloomberg Barclays US Treasury Inflation Notes TR Value Index
AS51 index	S&P/ASX 200 INDEX	**H0A0 index**	ICE BofA US High Yield Index
IBOV index	BRAZIL IBOVESPA INDEX	**JPEICORE index**	J.P. Morgan EMBI Global Core Index
SPTSX60 index	S&P/TSX 60 INDEX	**SPGSGCTR index**	S&P GSCI Gold Total Return Index
SHSZ300 index	CSI 300 INDEX	**SPGSCLTR index**	S&P GSCI Crude Oil Total Return Index
CAC index	CAC 40 INDEX	**SPGSICTR index**	S&P GSCI Copper Total Return Index
DAX index	DAX INDEX	BBREIT index	Bloomberg US REITs Index
NIFTY index	Nifty 50 Index	**DJUSRE index**	Dow Jones US Real Estate Index
JCI index	JAKARTA COMPOSITE INDEX	AUDUSD curncy	AUD-USD X-RATE
FTSEMIB index	FTSE MIB INDEX	USDBRL curncy	USD-BRL X-RATE
NKY index	NIKKEI 225 Index	CADUSD curncy	CAD-USD X-RATE
MEXBOL index	S&P/BMV IPC Index	USDCNY curncy	USD-CNY X-RATE
RTSI$ index	RUSSIAN RTS INDEX	USDINR curncy	USD-INR X-RATE
IBEX index	IBEX 35 INDEX	USDIDR curncy	USD-IDR X-RATE
XU030 index	BIST 30 Index	EURUSD curncy	EUR-USD X-RATE
USGG1M index	US Generic Govt 1 Month Yield Index	USDJPY curncy	USD-JPY X-RATE
USGG12M index	US Generic Govt 12 Month Yield Index	USDMXN curncy	USD-MXN X-RATE
USGG3YR index	US Generic Govt 3 Year Yield Index	USDRUB curncy	Russian Ruble SPOT (TOM)
USGG10YR index	US Generic Govt 10 Year Yield Index	USDTRY curncy	USD-TRY X-RATE
USGG30YR index	US Generic Govt 30 Year Yield Index	GBPUSD curncy	GBP-USD X-RATE
DXY Curncy	US Dollar Index	USDKRW curncy	USD-KRW X-RATE

technical indicators by dividing each indicator into three. This is described in Table 3.

Second, we needed to determine the input window length that we would consider for each prediction horizon. According to the literature [13], for short-term prediction, we set 3, 5, and 7 business days as the input window length, and we used the features of all the window lengths. For example, if it was a short-term prediction, and there are n technical indicators for the index, we had 3*n technical indicators. For medium-term prediction, we set 7, 10, 15, and 20 business days as the input window length. For long-term prediction, we set 10, 15, 20, and 30 business days as the input window length.

3.3. Transfer entropy

In information theory, entropy is an average level of information in the possible outcomes. Following the definition of entropy, entropy is the sum of all possible outcomes of the variable. In a mathematical way, entropy is defined as Eq. (2). Similarly, joint entropy, which is the joint of a pair of outcomes (i, j), and

Table 2
Label distribution in the test set.

Up label distribution (%)			
	Short-term	Medium-term	Long-term
DJUSRE index	53.714	57.692	58.753
H0A0 index	64.879	68.037	68.072
JPEICORE index	56.952	63.770	67.513
KOSPI index	54.0246	54.843	57.844
LBUTTRUU index	53.0585	61.436	65.241
LUACTRUU index	56.516	63.431	70.080
LUATTRUU index	53.192	58.644	60.106
MXEF index	54.151	59.898	59.259
MXWOU index	54.406	59.387	63.602
SPGSCLTR index	55.232	56.748	57.748
SPGSGCTR index	53.245	56.0265	53.642
SPGSICTR index	50.728	51.788	54.172
SPX index	56.499	70.027	71.883
Mean index	**55.122**	**60.133**	**62.147**

conditional entropy, which is the entropy of I conditioned on J, are defined as in Eqs. (3) and (4) respectively.

$$\mathrm{H}\left(\mathrm{I}\right)=-\sum p_{I}\left(i\right)\,logp_{I}(i) \tag{2}$$

$$\mathrm{H}\left(\mathrm{I,J}\right)=-\sum\sum p_{IJ}\left(i,j\right)\,logp_{IJ}(i,j) \tag{3}$$

$$\mathrm{H}\left(\mathrm{I|J}\right)=-\sum\sum p_{IJ}\left(i,j\right)\,logp_{I|J}(i|j) \tag{4}$$

The mutual entropy is a measure of the mutual dependence of two time-series, which is defined as in Eq. (4) [24,53–55].

$$\mathrm{M}\left(\mathrm{I,J}\right)=\sum p_{IJ}\left(i,j\right)\,log\frac{p(i,j)}{p_{I}\left(i\right)p_{J}(j)} \tag{5}$$

Transfer entropy is defined as a directional mutual entropy [24]. Transfer entropy is defined as in Eq. (6), assuming that we have two time-series I and J, and we use k number of data points of I and l number of data points of J, where $\mathrm{i_t}$ and j_t are data points at time t of I and J processes, and $\mathrm{i}_{\mathrm{t}}^{\mathrm{k}}$ and j_t^l are k-dimensional delay vectors that are l-dimensional delay vectors.

$$\mathrm{TE}_{J\to I}\overset{def}{=}\sum p\left(i_{t+1},i_t^k,j_t^l\right)log\frac{p(i_{t+1}|i_t^k,j_t^l)}{p(i_{t+1}|i_t^k)} \tag{6}$$

In other words, TE provides information about the direction of interaction between two systems, which is conceptually represented in Fig. 2.

To measure the transfer entropy between the global financial market indexes, we followed the steps in the literature [54,56]. First, to make the distribution of prices consistent and scaling, a log-return transformation is implemented as in Eq. (7).

$$\mathrm{y_t}\overset{def}{=}\log\frac{Close_t}{Open_t} \tag{7}$$

Second, the transfer entropy is calculated with all combinations of two different indexes. For the calculation of the transfer entropy, we used Kraskov–Stögbauer–Grassberger estimator [54]. Third, the parameter selection of k, the target history length, in Eq. (6) is important for accurate measurements. The ideal value of k is ∞ in the non-Markov process [56]. However, $\mathrm{k}\to\infty$ is an impossible choice in the real-world dataset. Since the k showed a small difference in the results when k is high compared to the number of time-series data points [57], we chose k as 256 for target history length considering that the number of time-series data points was approximately 10,000. We also chose k as 1 for the source history length as recommended by the literature [24].

Finally, to measure statistical significance, we used permutation testing [1]. The null hypothesis of permutation testing is that there is no directed relationship. In order to find the distribution of transfer entropy, we identified a surrogate process that satisfies $\mathrm{p^s}\left(\mathrm{i_t}|\mathrm{i_{t-1}^k}\right)=\mathrm{p}(\mathrm{i_t}|\mathrm{i_{t-1}^k},\mathrm{j_{t-1}^l})$ by subsampling. In other words, a surrogate process J^s has the same statistical properties as J but should not have a direct relationship to I [30]. Transfer entropy was asymptotically distributed as in Eq. (8), where dJ and dI are the dimensionalities of processes J and I [30]. With the distribution, we examine the statistical significance of the output. We used JIDT [58] to measure the transfer entropy and get the statistical significance.

$$\mathrm{TE}_{J^s\to I}\sim\frac{\aleph^2}{2N}(in\ nats),\ degree\ of\ freedom=ldJdI \tag{8}$$

Finally, we need to select a threshold of the p-value and establish a meaningful causal relationship. However, since the p-value threshold affects prediction performance in predicting price movement, we consider the p-value threshold as a parameter to find the optimal value. [1]. Therefore, we find the best p-value thresholds among 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5, which is the suggested range in literature [1] by using Bayesian optimization [59].

3.4. Random matrix theory

We considered N differential financial assets for daily historical data. We defined the vector of returns $\vec{\mathrm{r}}_{\mathrm{t}}=(r_{1t},r_{2t},\ldots,r_{Nt})$ for each day, $\mathrm{t}=\{1,\ldots,\tau\}$. Following the procedure [37], we standardized the return by removing the sample mean of each asset and normalizing each return by an estimated standard deviation of its daily volatility. For standard deviation, we chose the cross-sectional daily volatility to reduce a non-stationarity. The final return matrix is $\mathrm{X}\in\mathrm{R}^{\mathrm{NX}\tau}$. As a result, the sample correlation matrix is defined in Eq. (8). The T refers to the transpose operator.

$$E\overset{def}{=}\frac{1}{\tau}XX^T \tag{9}$$

The sample correlation has the following eigenvalues λ_k and the corresponding eigenvectors u_k.

$$E=\sum_{k=1}^{N}\lambda_k u_k u_k^*,\ where\ \lambda_1\geq\lambda_2\geq\cdots\geq\lambda_N\geq 0 \tag{10}$$

To get a true correlation from the sample correlation, we used a debiased rotational invariant estimator (hereafter RIE), which is the state-of-the-art random matrix theory approach [41]. RIE is based on the fact that one can compute the overlap between true and sample eigenvectors [60]. Therefore, our goal is to find $\Xi^{rie}=\sum_{k=1}^{N}\xi_k^{rie}u_k u_k^T$, and we need to find ξ_k^{rie}. The procedure of RIE to find ξ_k^{rie} are as follows.

First, we construct an optimal estimator of E based on the optimal shrinkage function [41]. Given the eigenvalues λ_k of the sample correlation matrix and $\mathrm{q}=\mathrm{N}/\tau$, we define the complex variable $\mathrm{z_k}=\lambda_k-i/\sqrt{N}$ for $\mathrm{k}\in(1,\mathrm{N})$.

Then, the ξ_k without debias is as follows.

$$\xi_k=\frac{\lambda_k}{\left|1-q+qz_k s_k\left(z_k\right)\right|^2},\ where\ s_k\left(z_k\right)=\frac{1}{N}\sum_{j=1,j\neq k}^{N}\frac{1}{z_k-\lambda_j} \tag{11}$$

To debias the ξ_k, we need to find the condition Γ_{k}. It may be calculated using Eq. (12), where $\mathrm{g_{mp}}\left(z\right)$ is the Stieltjes transform of the Marchenko–Pastur density and λ_N is the smallest eigenvalue.

$$\Gamma_{\mathrm{k}}=\sigma^2\frac{\left|1-q+qz_k g_{mp}\left(z_k\right)\right|^2}{\lambda_k} \tag{12}$$

$$\mathrm{g_{mp}}\left(z\right)=\frac{z+\sigma^2\left(q-1\right)-\sqrt{z-\lambda_N}\sqrt{z-\lambda_+}}{2qz\sigma^2},$$

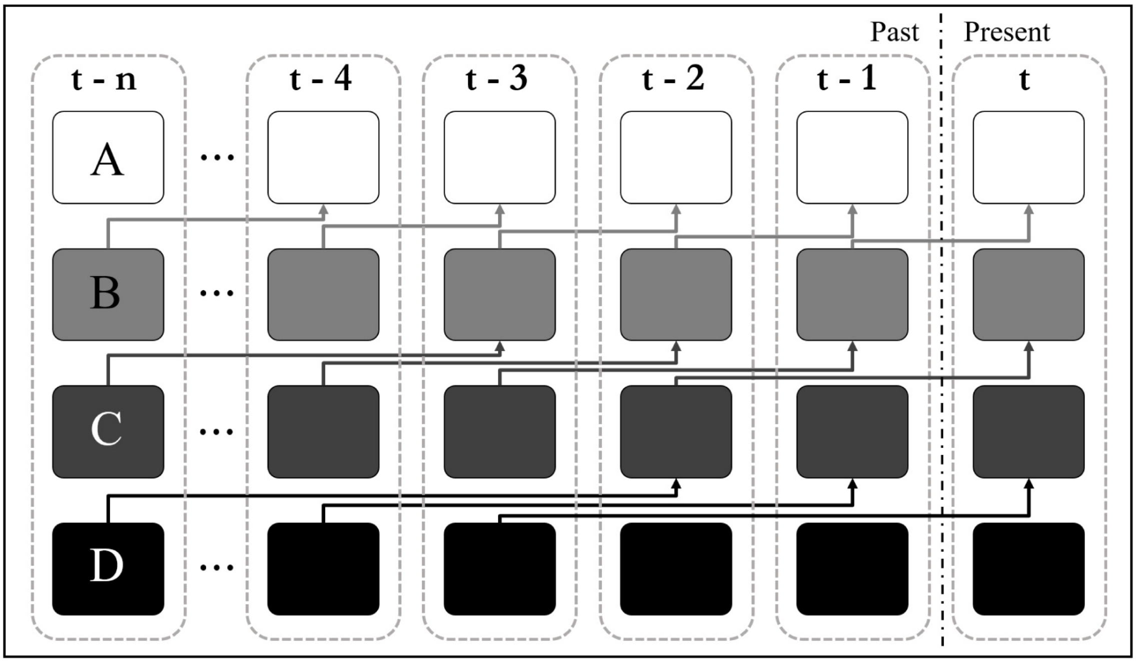

Fig. 2. Conceptual representation of transfer entropy.

Table 3
Technical indicators by input type.

Input type	Technical indicators	Index ticker
Close	Moving Average, Exponential Moving Average, Rate of Change, Standard Deviation, Bollinger Bands, Triple Exponential Average, Fourier Transform	All indexes except those below
Open, High, Low, Close	All technical indicators in Input Type Close, Average True Range, Average Directional Movement Index, Commodity Channel Index, Relative Strength Index, Williams %R, Slow Stochastic Oscillator, Vortex Indicator, Coppock Curve, Keltner Channel	AUDUSD, CAC, CADUSD, EURUSD, GBPUSD, USDCNY, USDINR, USDJPY, USDKRW, USDMXN, USDTRY, USDIDR, USDBRL, USDRUB, AS51, DAX, DJI
Open, High, Low, Close, Volume	All technical indicators above, Accumulation/Distribution Indicator, Money Flow Index, On-Balance-Volume, Force, Ease of Movement, Rate of Change (Volume)	KOSPI, SPX

$$\text{where } \lambda_+ = \lambda_N \left(\frac{1+\sqrt{q}}{1-\sqrt{q}}\right), \sigma^2 = \frac{\lambda_N}{\left(1-\sqrt{q}\right)^2} \tag{13}$$

Then, the debiased RIE is as follows.

$$\xi_k^{rie} = \begin{cases} \Gamma_k \xi_k^{rie} \text{ if } \Gamma_k > 1, \\ \xi_k^{rie} \text{ otherwise} \end{cases} \tag{14}$$

With Eq. (14), we can determine the true correlation based on the empirical correlation. Using the true correlation, we find relevant top-n indexes with the highest correlation. We find the optimal n with Bayesian optimization in the range of 1, 3, 5, 7, 10, 15, 20 [59].

Fig. 3 is a visualization of the results from the random matrix theory. In the left figure, the original is the heatmap of the empirical correlation matrix, and the right one is the heatmap of the correlation matrix refined from the random matrix theory. It can be seen that the correlation matrix from RMT has a lower overall correlation than empirical correlation. This is because the market-wide movement has been removed from random matrix theory [16]. The figure on the right is the distribution of the eigenvalues of the correlation matrix with noise removed from the random matrix theory. The red line represents the Marchenko–Pastur distribution. In theory, the noise of the eigenvalue of the financial correlation matrix follows the Marchenko–Pastur distribution, and it can be seen that the effect of noise is very small for the correlation matrix from random matrix theory.

3.5. Complex quantitative financial network

After calculating the TE network and RMT network, the final influential indexes are computed as the overlap from indexes selected from TE and indexes selected from RMT for each financial index. Fig. 4 shows the proposed complex quantitative financial network. The red dots show the 13 target financial indexes and the black dots show the other financial indexes. TE that has a p-value of <0.1 and correlation from RMT has a value of >0.1 is selected in Fig. 4.

3.6. The proposed method

We used a combination of two state-of-the-art algorithms for time series classification [19,20]. Fig. 5 shows the overall process. Assuming that there are n influential indexes to the target index, TI_{0t} denotes the technical indicators of the target index at time t, and TI_{kt} denotes those of the influential index k at time t. Technical indicators enter the convolutional network in consideration of cross-sectional interaction, and the features that emerged enter the attention LSTM network in consideration of temporal dynamics, predicting the price movement after h days.

First, we used a one-dimensional convolutional neural network to consider the interaction between features. This extracts interactions from the cross-sectional information at each time

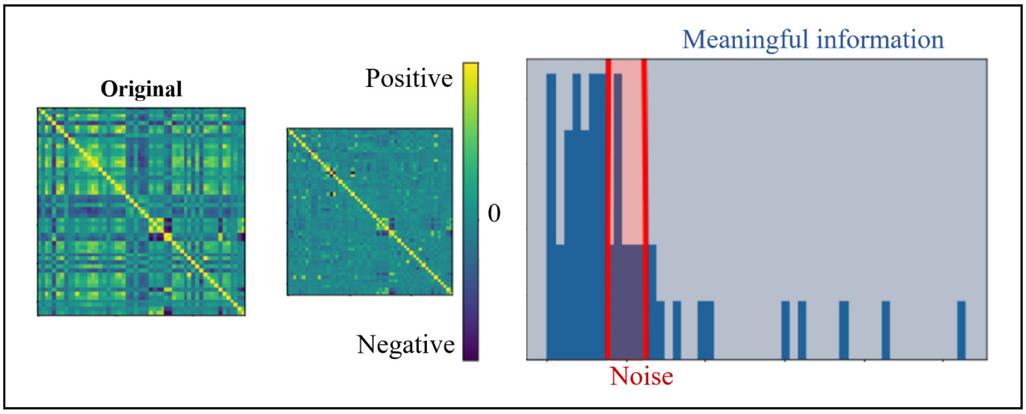

Fig. 3. Results from Random Matrix Theory.

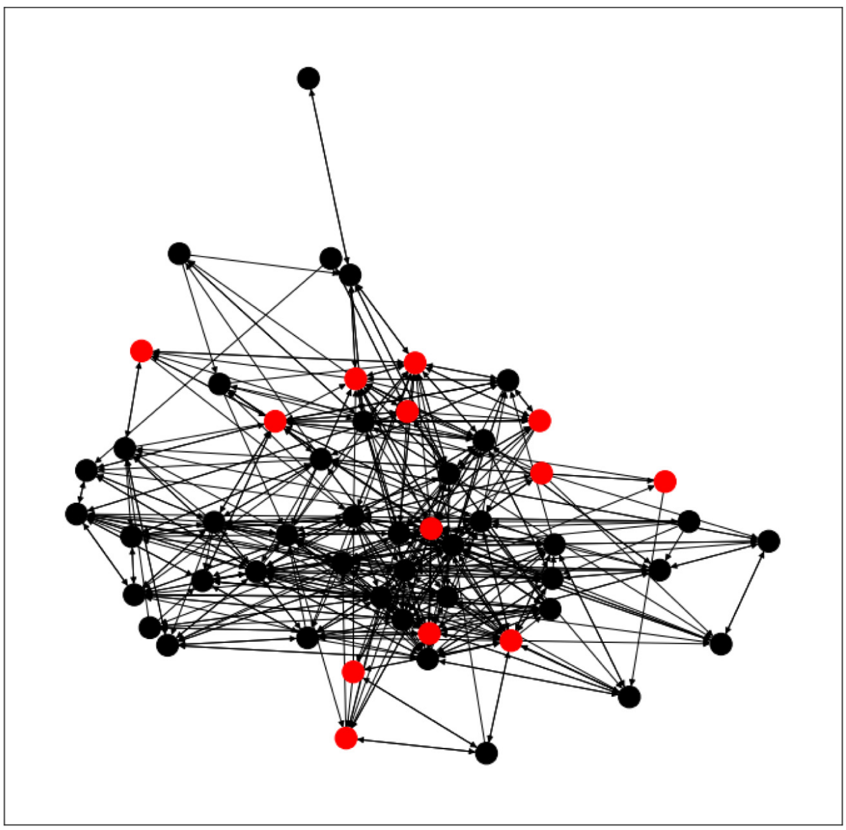

Fig. 4. Complex quantitative financial networks.

step t. Afterward, features are represented through feature recalibration in squeeze-and-excitation.

Squeeze-and-Excitation (SE) recalibrates the feature map after the CNN layer [52]. The Squeeze-and-Excitation consists of two parts: squeeze and excitation. Squeeze refers to an operation that extracts only the important information on each channel. Assuming that X is an input feature, c is the channel index, and H and W are the sizes of the channel, squeeze operation is expressed by Eq. (15) where * is a convolution. In detail, the squeeze operation makes feature maps of size (H, W, C) into vectors of size $(1, 1, C)$, and compresses information.

$$z_c = \frac{1}{H \cdot W} \sum \sum u_c(i, j), \text{ where } u_c = \sum v_c * X \tag{15}$$

Excitation operation calculates channel-wise dependencies. Assuming that W_1 and W_2 are fully connected layers and σ is a sigmoid function, the excitation operator is as follows in Eq. (16). In detail, the excitation operation is a process of restoring information compressed in Eq. (15).

$$s_c = \sigma(W_2 ReLU(W_1 z_c)) \tag{16}$$

After the squeeze and excitation operations, the final outputs of the operators are multiplied by the original values as follows in Eq. (17). Since all s_c values from the excitation operator are between 0 and 1, they are scaled according to the importance of the channels.

$$\tilde{x} = s_c \cdot u_c \tag{17}$$

After repeating the process and undergoing 1-D CNN and global average pooling, the feature F_t that takes into account the cross-sectional interaction of technical analysis is calculated.

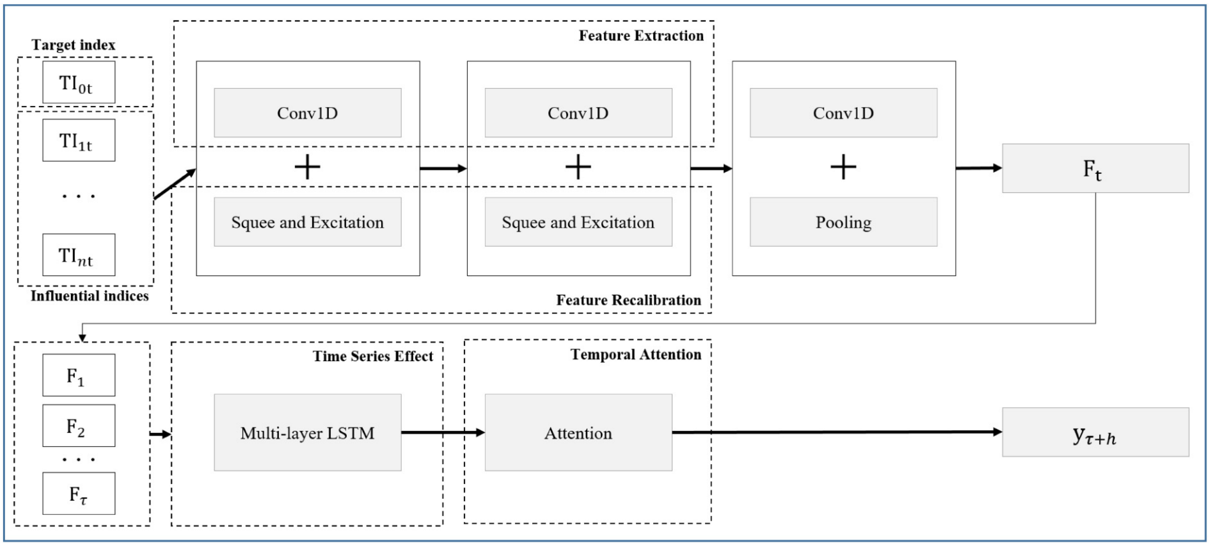

Fig. 5. The proposed time series classification algorithm.

Following the CNN with SE, a temporal attention LSTM is used to capture temporal dynamics. First, the features undergo a multi-layer LSTM then undergo the attention layer. The attention mechanism is as follows in Eqs. (18) and (19), where l_t^i determines how similar the hidden state and the output of the CNN module, and h_i is a hidden state of the LSTM. Here, β_t^i stands for attention score for each feature, and c_t is a context vector reflecting the importance of each feature.

$$\beta_t^i = \frac{\exp(l_t^i)}{\sum \exp\left(l_t^j\right)} \tag{18}$$

$$c_t = \sum \beta_t^i h_i \tag{19}$$

After this process, the final features enter the fully connected layer and predict the movement of the price of the financial index h days later. All parameters in the model such as the number of units, channels, LSTM layers, and dropout are optimized using Bayesian optimization during the validation period [59].

3.7. Evaluation

Three years of historical data from January 2017 to December 2019 were used as the test set in this study. The training period included all historical data from January 1971 to December 2014 and the validation period was from January 2015 to December 2017. The Bayesian parameter search is performed during the validation period and the following reported results are during the test period for out-of-sample-test [59].

We use accuracy and F1-score for balanced performance measures according to literature [61]. With the confusion matrix in Table 4, Accuracy is defined as Accuracy $\stackrel{def}{=} \frac{TP+TN}{TP+FP+FN+TN}$ and F1-score is defined as F1 − score $\stackrel{def}{=} 2*\frac{precision*recall}{precision+recall}$, where precision $\stackrel{def}{=} \frac{TP}{TP+FP}$ and recall $\stackrel{def}{=} \frac{TP}{TP+FN}$.

Furthermore, another evaluation metric, the Portfolio value, gives practical implications. We calculated the final value of the portfolio by constructing a portfolio similar to that in a real situation through back-testing during the test period, and we also calculated the Sharpe ratio to confirm robustness of the proposed method [62]. A higher sharpe ratio means stable, high returns. The sharpe ratio is in Eq. (20), where R_p is the return of portfolio, R_f is the risk-free rate (Here, 3% in a year), and σ_p is the standard deviation of the portfolio.

Table 4
Confusion matrix.

		Prediction	
		Up	Down
Actual	Up	TP	FN
	Down	FP	TN

Table 5
Accuracy.

Accuracy	Short-term	Medium-term	Long-term
Baseline	0.551	0.601	0.621
Lin et al. [63]	0.568	0.618	0.629
Lai et al. [64]	0.582	0.639	0.660
TA model [12,13]	0.586	0.643	0.660
RMT model	0.591	0.647	0.672
TE model [22]	0.589	0.643	0.673
The proposed method	**0.602**	**0.660**	**0.688**

$$Sharpe\ ratio = \frac{R_p - R_f}{\sigma_p} \tag{20}$$

4. Result

4.1. Experiment result

We compared the accuracy of seven approaches: the baseline, two state of the art algorithms [63,64], the proposed method, technical analysis with TE causal indexes (TE model), technical analysis with RMT relevant indexes (RMT model), and technical analysis only (TA model). Lin et al. [63] extracted the feature through the first order difference of technical analysis features and predicted the stock price using the deep stacked LSTM model. Lai et al. [64] used the 2-D CNN to consider spatial patterns (local dependency among technical analysis features) and the Recurrent Skip Network to consider temporal patterns.

The baseline accuracy results from a method that predicts only the dominant label. These results were only taken from the test period, and are described in Table 5. In Table 5, the algorithm with the highest accuracy for each prediction horizon is in bolded text.

Table 6
T-test results.

P-value	Short-term	Medium-term	Long-term
PM-TA	0.006***	0.000***	0.000***
PM-RMT	0.083*	0.021**	0.016**
PM-TE	0.001***	0.007***	0.004***
RMT-TA	0.526	0.427	0.082*
TE-TA	0.637	0.875	0.842
TE-RMT	0.232	0.503	0.139

Note:
*Denote significance at the 10% levels.
**Denote significance at the 5% levels.
***Denote significance at the 1% levels.

Table 7
F1-score.

F1-score	Short-term	Medium-term	Long-term
Lin et al. [63]	0.571	0.638	0.649
Lai et al. [64]	0.583	0.642	0.662
TA model [12,13]	0.586	0.643	0.660
RMT model	0.602	0.660	0.688
TE model [22]	0.594	0.646	0.671
The proposed method	**0.667**	**0.720**	**0.738**

Lin et al. [63] shows higher accuracy than the baseline, which means that it can better predict price movement with the features of the time series. Lai et al. [64] shows higher accuracy than Lin et al. [63], which suggests that it is important to consider interactions among technical analysis variables. TA model shows slightly higher accuracy than Lai et al. [64], validating the performance of the proposed deep learning architecture.

Both the TE model and RMT model show higher accuracy than the TA model. This shows that it is important to consider influential indexes rather than simply using target indexes when predicting price movement. In all prediction horizons, the proposed method shows high accuracy in comparison with all other groups. It suggests that since simply using the TE model network may cause issues such as spurious causality, and using only the RMT model network does not confirm the directional relationship, it is important to consider two networks at the same time.

In order to show that the difference in this result is statistically significant, a t-test of prediction accuracy was performed. For the excess accuracy, which is calculated as accuracy from the model minus the baseline accuracy, a t-test was performed for all combinations of the four methods. The results are shown in Table 6. In order to check the network effect, we only compared the TA model, the RMT model, the TE model, and the proposed method.

Table 6 displays the p-value of the t-test, and the PM refers to the proposed method. The proposed model showed statistically significantly higher performance in all prediction horizons than all other methods However, the RMT model and TE model showed statistically insignificant differences except in terms of long-term prediction of the RMT model. The results show that the proposed method shows statistically better performance than the other methods. In addition, the results also suggest that using the financial network has a statistically significant effect on model performance, and how the network is configured has a statistically significant effect on model performance so that it is important to build the financial network according to the appropriate method, such as the proposed method.

To give more robust results, we also provided the F1-score. The results are described in Table 7. The F1-scores show the consistent results. Moreover, the difference between the F1-scores is greater than that of accuracy. This means that the proposed method makes a more stable prediction than other models.

Table 8
Accuracy by p-value threshold of TE.

p-value threshold from TE	Short-term	Medium-term	Long-term
0.05	**0.584**	0.639	0.664
0.1	0.581	**0.640**	**0.667**
0.2	0.578	0.638	0.665
0.3	0.573	0.636	0.661
0.4	0.567	0.632	0.659
0.5	0.562	0.629	0.654
Baseline	0.551	0.601	0.621

Table 9
Accuracy by top n of RMT.

Top n from RMT	Short-term	Medium-term	Long-term
1	0.580	0.629	0.649
3	0.587	0.632	0.654
5	**0.589**	0.635	0.661
7	0.586	0.639	0.665
10	0.583	**0.643**	**0.670**
15	0.578	0.641	0.663
20	0.572	0.638	0.660
Baseline	0.551	0.601	0.621

4.2. Additional result

We also checked the results by changing the p-value and top n, which are the parameters of TE and RMT. In this experiment, the RMT network is not used when constructing a network using TE, and the TE is not used when constructing a network using RMT.

Table 8 shows the mean accuracy for financial indexes movement to the p-value thresholds. The short-term prediction horizon shows the highest accuracy when the p-value equals 0.05, while the medium-term prediction and the long-term prediction horizon expenses sector shows the highest accuracy when the p-value equals 0.1. Table 9 shows the mean accuracy for financial indexes movement to the top n nearest indexes from RMT. The short-term prediction horizon shows the highest accuracy when the top n equals 5, while the medium-term prediction horizon and long-term prediction horizon show the highest accuracy when the top n equals 10. Here, the reason that the accuracy is lower than the TE/RMT model in Table 5 is that the TE/RMT model finds the optimal p-value for each financial index.

In Tables 8 and 9, we can see that the prediction accuracy significantly changes by the p-value threshold and top n. This means that two parameters should be optimized according to the prediction horizon and financial index. In addition, performance gradually decreases at a p-value threshold/top n that exceeds the best p-value threshold/top n in all prediction horizons. It suggests that if a high p-value threshold/top n is set, information on financial indexes with meaningless relationships is also reflected on the model, which results in lower performance. Also, accuracy decreases at a p-value threshold/top n below the best parameters. It suggests that failure to reflect all meaningful relationships adversely affects performance.

To compare the performance of the model from different angles, we compared the complexity. Complexity was defined as the number of parameters according to the literature [65]. Table 10 shows the number of parameters for each model. Complexity is the average of each financial index prediction model. Also, Fig. 6 shows the complexity versus accuracy for each model. The accuracy in Fig. 6 is the short-term accuracy.

Lin et al. [63] has a middle level of complexity, but does not show high accuracy. The TA model and Lai et al. [64] shows

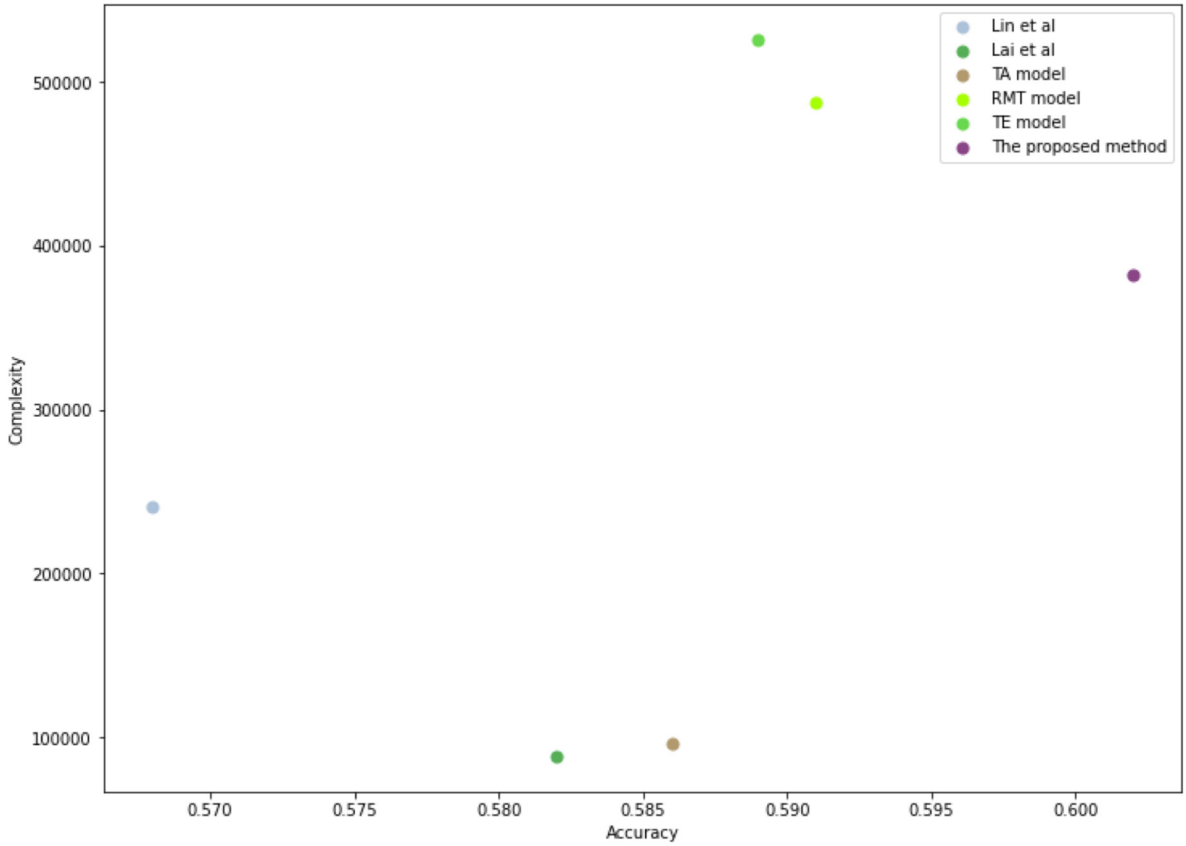

Fig. 6. Complexity comparison.

Table 10
Average complexity comparison.

Algorithm	Number of parameters
Lin et al. [63]	240,652
Lai et al. [64]	88,649
TA model [12,13]	96,517
RMT model	487,035
TE model [22]	525,826
The proposed method	381,633

the lowest complexity. However, they have relatively low accuracy compared to RMT model and TE model. Also, because the RMT model and the TE model consider the network effect, they have high complexity. This shows the accuracy complexity trade off [66].

However, the proposed method shows similar complexity to Lin et al. [63], while the accuracy shows the best performance among models. Even, while the proposed method shows lower complexity than the RMT model and TE model, it shows high accuracy. This means that using the intersection of RMT network and TE network has a positive effect on both accuracy and complexity.

4.3. Back testing for portfolio management

To give more robust results, we implemented back-testing taking into consideration the four strategies: the proposed method, technical analysis with TE causal indexes (TE model), technical analysis with RMT relevant indexes (RMT model), and technical analysis only (TA model). Back-testing was only implemented in the test period only. We conducted back-testing as a long-short portfolio. Back-testing received to buy or sell signals from the models. 10% of the total holdings were traded at every prediction horizon. That is, when the prediction horizon is short-term, trades are made daily; when the prediction horizon is medium-term, trades are made every two weeks; and when the prediction horizon is long-term, trades are made every month. For ease of implementation, we set a 0.15% fee for the transaction cost of buy and sell, and not 0.3% for sell. We set the initial money as 100,000. In a mathematical way, back-testing is described in Eqs. (21) and (22), where p_{it} and r_{it} are prediction probability and return for financial index i and time t, S_{it} is a signal, m_t is money at time t. Here, α is a portion of trading money, which is 0.1 in this experiment, and β is a transaction fee, which is 0.0015 in this experiment.

$$S_{it} = \begin{cases} 1 & if\ p_{it} \geq 0.5 \\ -1 & if\ p_{it} < 0.5 \end{cases} \tag{21}$$

$$m_t = m_{t-1} * \alpha + \left(1 + \overrightarrow{r_t^T} \cdot \overrightarrow{S_{t-1}}\right) * (1-\beta) * m_{t-1} * (1-\alpha),$$
$$where\ m_0 = 100,000 \tag{22}$$

The back-testing results are shown in Figs. 7 to 9. The red line represents the proposed method, the green line represents the TE model, the magenta line represents the RMT model, and the black line represents the TA model.

The proposed method shows superior performance than other approaches in the short-term and medium-term prediction horizons. As in Table 11, this result is consistent when calculating the Sharpe ratio. The proposed method shows a more stable performance than other methods. In addition, both the TE model and RMT model show better performance than the TA model, and this shows that it is important to use the features of related indexes rather than simply predicting only one's own time series. Overall, the result is consistent with previous results.

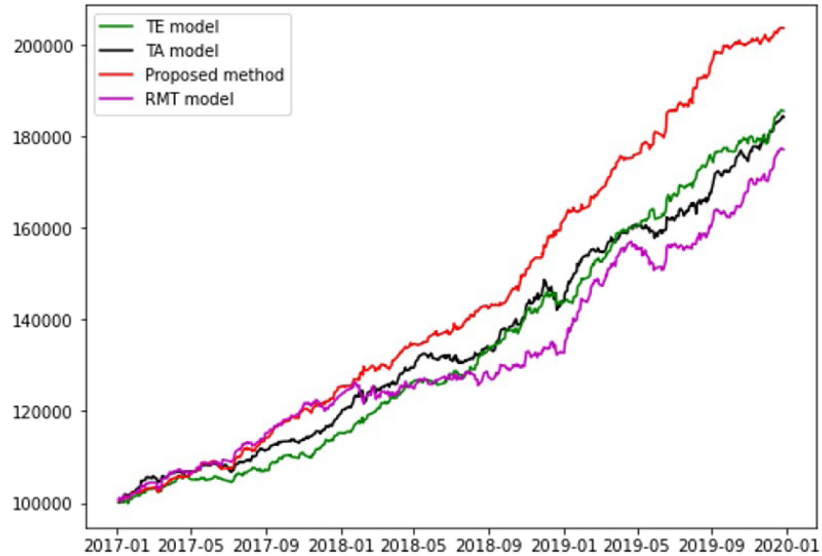

Fig. 7. Portfolio with a short-term prediction horizon.

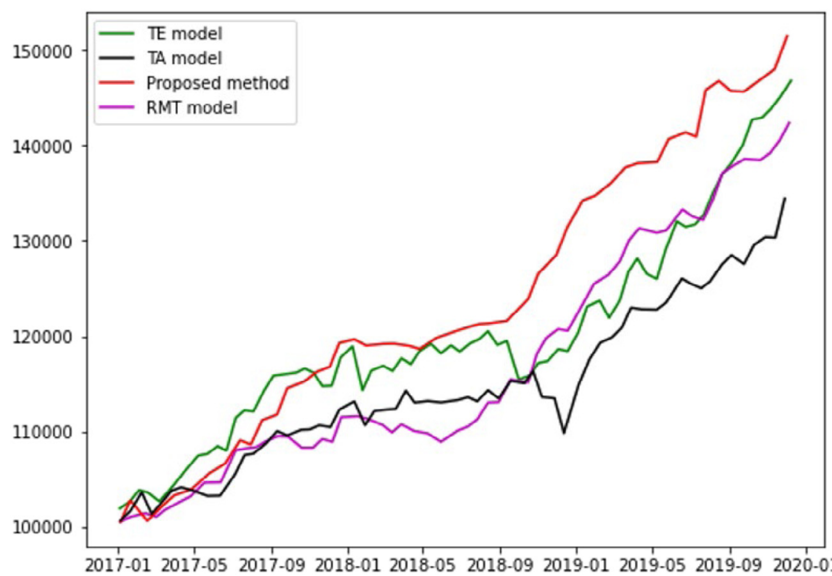

Fig. 8. Portfolio with a medium-term prediction horizon.

Table 11
Back-testing results: Sharpe ratio.

Sharpe ratio	Short-term	Medium-term	Long-term
TA model	2.843	0.935	0.810
RMT model	3.152	1.955	0.879
TE model	4.221	1.176	1.551
The proposed method	**5.490**	**2.895**	**3.239**

5. Conclusion

We proposed a novel deep learning model to predict the global financial index price movement based on technical analysis using complex financial networks. The proposed method suggests a quantitative financial network based on the complex theory in econophysics and shows the effects of the network on model performance. The proposed network is a combination of two types of financial networks: the cleaned correlation network and the causal network in global financial indexes. In our proposed method, the random matrix theory is used to construct a cleaned correlation network, and transfer entropy is used to find causality within the network. The cleaned correlation network has the advantage of preventing spurious correlation but has the disadvantage of not considering the directional impact. Also, the causal network can consider the directional impact, but there is an issue of spurious causality. In order to complement the disadvantages of the two networks and maximize the advantages, we combine the two networks. Based on a daily historical dataset for global indexes and out-of-sample tests, our experimental results reveal that the proposed method outperforms past state-of-the-art methods. In addition, the method of constructing a financial network by combining the two networks presented in this paper showed higher model performance than predicting the price movement of financial indexes using each network. Our findings reveal that identifying and using proper financial networks are important in prediction problems.

Limitations of our study include that this method cannot predict all global financial indexes, as it can only predict important financial indexes. Future work should focus on predicting the

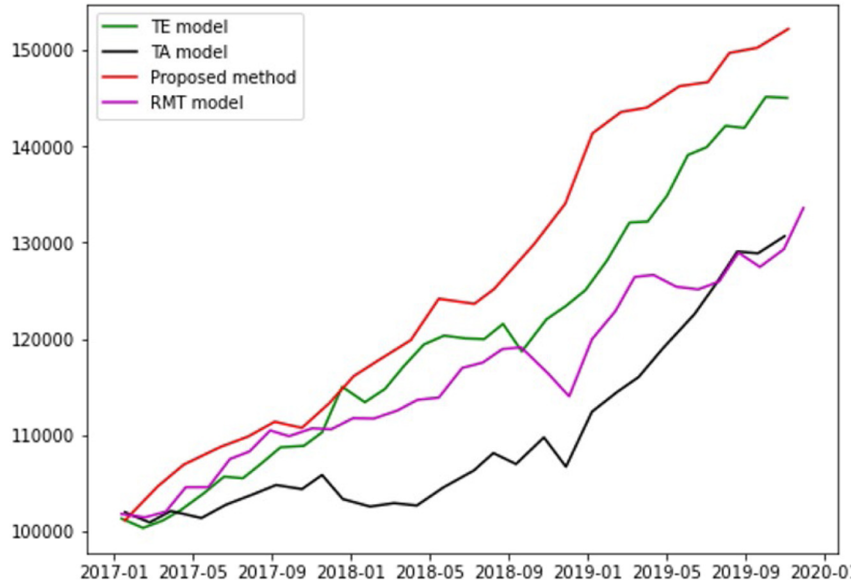

Fig. 9. Portfolio with a long-term prediction horizon.

price movement for various global financial indexes with additional information on global financial indexes. Additionally, we did not use a pre-defined network and we did not compare the proposed network with a pre-defined network. Future studies should predict the price movement by combining the pre-defined network and the proposed network. Finally, since the number of nodes was small, we did not use a graph neural network. Future research should use a graph neural network based on several nodes in the company-level stock market.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Nam, N. Seong, Financial news-based stock movement prediction using causality analysis of influence in the Korean stock market, Decis. Support Syst. 117 (2019) 100–112.
[2] C.-H. Chen, C.-H. Yu, A series-based group stock portfolio optimization approach using the grouping genetic algorithm with symbolic aggregate approximations, Knowl.-Based Syst. 125 (2017) 146–163.
[3] N. Seong, K. Nam, Predicting stock movements based on financial news with segmentation, Expert Syst. Appl. 164 (2021) 113988.
[4] E.F. Fama, The behavior of stock-market prices, J. Bus. 38 (1) (1965) 34–105.
[5] A.W. Lo, Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis, J. Invest. Consult. 7 (2) (2005) 21–44.
[6] G. Friesen, P.A. Weller, Quantifying cognitive biases in analyst earnings forecasts, J. Financial Mark. 9 (4) (2006) 333–365.
[7] A. Urquhart, R. Hudson, Efficient or adaptive markets? Evidence from major stock markets using very long run historic data, Int. Rev. Financ. Anal. 28 (2013) 130–142.
[8] G.S. Atsalakis, K.P. Valavanis, Surveying stock market forecasting techniques–Part II: Soft computing methods, Expert Syst. Appl. 36 (3) (2009) 5932–5941.
[9] X. Zhang, Y. Hu, K. Xie, S. Wang, E. Ngai, M. Liu, A causal feature selection algorithm for stock prediction modeling, Neurocomputing 142 (2014) 48–59.
[10] L.A. Teixeira, A.L.I. De Oliveira, A method for automatic stock trading combining technical analysis and nearest neighbor classification, Expert Syst. Appl. 37 (10) (2010) 6885–6890.
[11] W. Long, Z. Lu, L. Cui, Deep learning-based feature engineering for stock price movement prediction, Knowl.-Based Syst. 164 (2019) 163–173.
[12] Y. Kara, M.A. Boyacioglu, 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. 38 (5) (2011) 5311–5319.
[13] Y. Shynkevich, T.M. McGinnity, S.A. Coleman, A. Belatreche, Y. Li, Forecasting price movements using technical indicators: Investigating the impact of varying input window length, Neurocomputing 264 (2017) 71–88.
[14] J.-P. Bouchaud, M. Potters, Financial applications of random matrix theory: a short review, 2009, arXiv preprint arXiv:0910.1205.
[15] J. Bun, J.-P. Bouchaud, M. Potters, Cleaning large correlation matrices: tools from random matrix theory, Phys. Rep. 666 (2017) 1–109.
[16] D.-H. Kim, H. Jeong, Systematic analysis of group identification in stock markets, Phys. Rev. E 72 (4) (2005) 046133.
[17] R. Marschinski, H. Kantz, Analysing the information flow between financial time series, Eur. Phys. J. B 30 (2) (2002) 275–281.
[18] Z. He, K. Maekawa, On spurious Granger causality, Econom. Lett. 73 (3) (2001) 307–313.
[19] F. Karim, S. Majumdar, H. Darabi, S. Harford, Multivariate LSTM-FCNs for time series classification, Neural Netw. 116 (2019) 237–245.
[20] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, G. Cottrell, A dual-stage attention-based recurrent neural network for time series prediction, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017, pp. 2627–2633.
[21] Y. Pan, Z. Xiao, X. Wang, D. Yang, A multiple support vector machine approach to stock index forecasting with mixed frequency sampling, Knowl.-Based Syst. 122 (2017) 90–102.
[22] S. Kim, S. Ku, W. Chang, J.W. Song, Predicting the direction of US stock prices using effective transfer entropy and machine learning techniques, IEEE Access 8 (2020) 111660-111682.
[23] W. Chen, M. Jiang, W.-G. Zhang, Z. Chen, A novel graph convolutional feature based convolutional neural network for stock trend prediction, Inform. Sci. 556 (2021) 67–94.
[24] T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85 (2) (2000) 461.
[25] K. Hlaváčková-Schindler, M. Paluš, M. Vejmelka, J. Bhattacharya, Causality detection based on information-theoretic approaches in time series analysis, Phys. Rep. 441 (1) (2007) 1–46.
[26] F.A. Razak, H.J. Jensen, Quantifying 'causality'in complex systems: understanding transfer entropy, PLoS One 9 (6) (2014) e99462.
[27] P. Gopikrishnan, B. Rosenow, V. Plerou, H.E. Stanley, Quantifying and interpreting collective behavior in financial markets, Phys. Rev. E 64 (3) (2001) 035106.
[28] T.D.M. Peron, F.A. Rodrigues, Collective behavior in financial markets, Europhys. Lett. 96 (4) (2011) 48004.
[29] T. Lux, M. Marchesi, Scaling and criticality in a stochastic multi-agent model of a financial market, Nature 397 (6719) (1999) 498.
[30] T. Bossomaier, L. Barnett, M. Harré, J.T. Lizier, An Introduction to Transfer Entropy, Springer, 2016.
[31] H.M. Markowitz, Foundations of portfolio theory, J. Finance 46 (2) (1991) 469–477.

[32] G.A. Christodoulakis, Common volatility and correlation clustering in asset returns, European J. Oper. Res. 182 (3) (2007) 1263–1284.
[33] S.S. Jung, W. Chang, Clustering stocks using partial correlation coefficients, Physica A 462 (2016) 410–420.
[34] L. Laloux, P. Cizeau, M. Potters, J.-P. Bouchaud, Random matrix theory and financial correlations, Int. J. Theor. Appl. Finance 3 (03) (2000) 391–397.
[35] A. Utsugi, K. Ino, M. Oshikawa, Random matrix theory analysis of cross correlations in financial markets, Phys. Rev. E 70 (2) (2004) 026110.
[36] A. García, Global financial indexes and twitter sentiment: A random matrix theory approach, Physica A 461 (2016) 509–522.
[37] J. Bun, J.-P. Bouchaud, M. Potters, My beautiful laundrette: Cleaning correlation matrices for portfolio optimization, 2016, in.
[38] L. Haff, Empirical Bayes estimation of the multivariate normal covariance matrix, Ann. Statist. (1980) 586–597.
[39] O. Ledoit, M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, J. Empir. Financ. 10 (5) (2003) 603–621.
[40] N. El Karoui, Spectrum estimation for large dimensional covariance matrices using random matrix theory, Ann. Statist. 36 (6) (2008) 2757–2790.
[41] J. Bun, R. Allez, J.-P. Bouchaud, M. Potters, Rotational invariant estimator for general noisy matrices, IEEE Trans. Inform. Theory 62 (12) (2016) 7475–7490.
[42] L.-J. Cao, F.E.H. Tay, Support vector machine with adaptive parameters in financial time series forecasting, IEEE Trans. Neural Netw. 14 (6) (2003) 1506–1518.
[43] E. Hadavandi, H. Shavandi, A. Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, Knowl.-Based Syst. 23 (8) (2010) 800–808.
[44] J. Arroyo, C. Maté, Forecasting histogram time series with k-nearest neighbours methods, Int. J. Forecast. 25 (1) (2009) 192–207.
[45] Y. Wang, L. Wang, F. Yang, W. Di, Q. Chang, Advantages of direct input-to-output connections in neural networks: The Elman network for stock index forecasting, Inform. Sci. 547 (2021) 1066–1079.
[46] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing 50 (2003) 159–175.
[47] R.S. Mamon, R.J. Elliott, Hidden Markov Models in Finance, Springer, 2007.
[48] Y. Zheng, Q. Liu, E. Chen, Y. Ge, J.L. Zhao, Time series classification using multi-channels deep convolutional neural networks, in: International Conference on Web-Age Information Management, Springer, 2014, pp. 298–310.
[49] S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, 2018, arXiv preprint arXiv:1803.01271.
[50] Z. Wang, W. Yan, T. Oates, Time series classification from scratch with deep neural networks: A strong baseline, in: 2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017, pp. 1578–1585.
[51] P. Malhotra, V. TV, L. Vig, P. Agarwal, G. Shroff, TimeNet: Pre-trained deep recurrent neural network for time series classification, 2017, arXiv preprint arXiv:1706.08838.
[52] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[53] T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley & Sons, 2012.
[54] A. Kraskov, Synchronization and Interdependence Maesures and their Applications to the Electroencephalogram of Epilepsy Patients and Clustering of Data, in, Universität Wuppertal, Fakultät für Mathematik und Naturwissenschaften> Physik> Dissertationen, 2004.
[55] N. Seong, Deep spatiotemporal attention network for fine particle matter 2.5 concentration prediction with causality analysis, IEEE Access 9 (2021) 73230–73239.
[56] J.T. Lizier, M. Prokopenko, A.Y. Zomaya, Local information transfer as a spatiotemporal filter for complex systems, Phys. Rev. E 77 (2) (2008) 026110.
[57] J.T. Lizier, Measuring the dynamics of information processing on a local scale in time and space, in: Directed Information Measures in Neuroscience, Springer, 2014, pp. 161–193.
[58] J.T. Lizier, JIDT: An information-theoretic toolkit for studying the dynamics of complex systems, Front. Robot. AI 1 (2014) 11.
[59] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, in: Advances in Neural Information Processing Systems, 2012, pp. 2951–2959.
[60] O. Ledoit, S. Péché, Eigenvectors of some large sample covariance matrix ensembles, Probab. Theory Related Fields 151 (1–2) (2011) 233–264.
[61] H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller, Deep learning for time series classification: a review, Data Min. Knowl. Discov. 33 (4) (2019) 917–963.
[62] W.F. Sharpe, The sharpe ratio, J. Portf. Manag. 21 (1) (1994) 49–58.
[63] Y.-F. Lin, T.-M. Huang, W.-H. Chung, Y.-L. Ueng, Forecasting fluctuations in the financial index using a recurrent neural network based on price features, IEEE Trans. Emerg. Top. Comput. Intell. (2020).
[64] G. Lai, W.-C. Chang, Y. Yang, H. Liu, Modeling long-and short-term temporal patterns with deep neural networks, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 95–104.
[65] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, Deep residual learning for image compression, in: CVPR Workshops, 2019.
[66] M. Fazzolari, R. Alcalá, F. Herrera, A multi-objective evolutionary method for learning granularities based on fuzzy discretization to improve the accuracy-complexity trade-off of fuzzy rule-based classification systems: D-MOFARC algorithm, Appl. Soft Comput. 24 (2014) 470–481.