Expert Systems With Applications 143 (2020) 113042

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Portfolio formation with preselection using deep learning from long-term financial data

Wuyu Wang[a], Weizi Li[b], Ning Zhang[a], Kecheng Liu[b,*]

[a] School of Information, Central University of Finance and Economics, 39 South College Road, Haidian District, Beijing 100081, China
[b] Informatics Research Centre, Henley Business School, University of Reading, Reading, Berkshire RG6 6UD, United Kingdom

ARTICLE INFO

Article history:
Received 20 June 2019
Revised 11 October 2019
Accepted 16 October 2019
Available online 17 October 2019

Keywords:
Asset preselection
Long-term financial data
Financial forecasting
Portfolio optimisation

ABSTRACT

Portfolio theory is an important foundation for portfolio management which is a well-studied subject yet not fully conquered territory. This paper proposes a mixed method consisting of long short-term memory networks and mean-variance model for optimal portfolio formation in conjunction with the asset preselection, in which long-term dependences of financial time-series data can be captured. The experiment uses a large volume of sample data from the UK Stock Exchange 100 Index between March 1994 and March 2019. In the first stage, long short-term memory networks are used to forecast the return of assets and select assets with higher potential returns. After comparing the outcomes of the long short-term memory networks against support vector machine, random forest, deep neural networks, and autoregressive integrated moving average model, we discover that long short-term memory networks are appropriate for financial time-series forecasting, to beat the other benchmark models by a very clear margin. In the second stage, based on selected assets with higher returns, the mean-variance model is applied for portfolio optimisation. The validation of this methodology is carried out by comparing the proposed model with the other five baseline strategies, to which the proposed model clearly outperforms others in terms of the cumulative return per year, Sharpe ratio per triennium as well as average return to the risk per month of each triennium. i.e. potential returns and risks.

1. Introduction

Portfolio management is a decision-making process in which an amount of fund is allocated to multiple financial assets, and the allocation weight is constantly changed in order to maximize the return and restrain the risk (Markowitz, 1952). Portfolio theory proposed by Markowitz in 1952, is an important foundation for portfolio management which is a well-studied subject yet not fully conquered territory. There are two issues with portfolio formation. The first one is to select assets with higher revenue, and another one is to determine the value composition of assets in the portfolio to achieve the goal of maximal potential returns with minimal risk. The quantitative approach to portfolio formation has often been adopted in investment decisions. Based on Markowitz's mean-variance (MV) model, numerous researches have discovered many model extensions and supplemented plentiful reasonable insights about the portfolio formation (Bodnar, Mazur, & Okhrin, 2017; Brown & Smith, 2011; Grauer & Hakansson, 1993; Li, Zhang, & Xu, 2013; Li, Zhang, & Xu, 2015; Liu & Loewenstein, 2002; Merton, 1969; Sharpe, 1963; Tobin, 1958, Tu & Zhou, 2010).

In the portfolio optimisation process, the expected return on an asset is a crucial factor, which means that a preliminary selection of assets is critical for portfolio management (Guerard, Markowitz, & Xu, 2015). But few researches pay attention to the preselection of assets before forming a portfolio. Asset selection has been a meaningful, but a difficult issue in the financial investment area. This line of research depends on the long-term volatility of financial time-series data in the past as well as a reliable performance forecasting of assets in the future (Huang, 2012). Traditional statistical methods are not effective in dealing with complex, multi-dimensional and noisy time-series data (Baek & Kim, 2018; Längkvist, Karlsson, & Loutfi, 2014), while early machine learning methods, such as support vector machine (SVM), principal component analysis (PCA), and artificial neural network (ANN), are not most suited for learning and storing financial time-series data over a long period (Bao, Yue, & Rao, 2017; LeCun, Bengio, & Hinton, 2015). This situation leads to the difficulties of financial assets preselection. In fact, during the investment decision-making

* Corresponding author.
E-mail address: k.liu@reading.ac.uk (K. Liu).

https://doi.org/10.1016/j.eswa.2019.113042
0957-4174/

process, it would be unsustainable to only apply complex portfolio optimisation methods without high-quality asset input (Deng & Min, 2013).

In the financial market, individual investors usually would like to know the changes in the returns of their investment assets today, the possible trends in the returns tomorrow and which measures should be adopted to help them possess the best portfolio (Zhang, Li, & Guo, 2018). Therefore, incorporating forecasting theory into the portfolio formation will be promising in financial investment (Kolm, Tütüncü, & Fabozzi, 2014). Forecasting financial time-series is always regarded as one of the most challenging tasks because of the dynamic, nonlinear, unstable and complex nature with long-term fluctuations of the financial market (Chen & Hao, 2018; Paiva, Cardoso, Hanaoka, & Duarte, 2019). But a reliable investment decision should rely on long-term observations and patterns of behaviour of asset data rather than short-term (Chong, Han, & Park, 2017; Chourmouziadis & Chatzoglou, 2016). In this case, it is necessary to observe the change and volatility of financial data over a long time in the past so as to make a good preparation for future trends forecasting and investment decisions. And numerous widely accepted empirical researches suggest that financial time-series have a memory of a period in the past, thus to some extent, financial markets are predictable. The behaviour of the asset over a long period will significantly influence the risks and returns of a portfolio, and then further affect the investment decisions (Liu & Loewenstein, 2002). However, this important point is always ignored by current researches. For instance, some apply early machine learning methods, GA (Huang, 2012), SVM (Huang, 2012; Paiva, Cardoso, Hanaoka, & Duarte, 2019), to predict and select good assets, but fail to capture long-term dependencies of financial time-series data. To overcome this limitation, we present a novel method for portfolio formation in conjunction with the asset preselection, in which long-term dependences of financial time-series data are duly considered.

The primary purpose of this paper is to construct an investment decision-making model for individual investors that combines the deep learning LSTM method which concentrates on capturing the long-term dependencies of the returns on assets and the Markowitz's MV method to form optimal portfolios. In this respect, our study has two primary contributions which fill the gaps in existing literature. Firstly, this paper develops a novel method consisting of long short-term memory networks and mean-variance model (LSTM+MV) for optimal portfolio formation. This method considers the long-term dependences on the fluctuations of financial market and captures long-time change patterns of company stocks from the time-series data. To show the benefit of the proposed method in terms of the prediction, early machine learning, and statistical models are used in our experiments as baselines to compare with the LSTM networks. Secondly, our proposed model explores in-depth the preselection process of assets before optimal portfolio formation, which guarantees high-quality inputs to the optimal portfolio. Unlike the majority of the methods which aim to improve the existing portfolio management models, this paper focuses on the preliminary phase of portfolio construction, i.e. the preselection of assets. Meanwhile, our work provides practical guidance for investors in making better investment decisions. Specifically, the systematic approach present in current paper is able to help decide which assets should be part of the portfolio and the value composition of assets in the portfolio.

The remainder of this paper is organised as follows. In Section 2, we review the development of modern portfolio theory and summarise empirical work that has used deep learning to solve issues corresponding to financial time-series data. In Section 3, we describe our methodology in detail, i.e. data source, input variable selection, the proposed model architecture. In Section 4, we present the results of the experiments and explain the results appropriately. In Section 5, we discuss our key findings, implications for theory and practice, also future work.

2. Theoretical background

2.1. Modern Portfolio theory

Markowitz (1952) proposes mean-variance (MV) methodology to solve the portfolio selection issue, which initiates the foundation of Modern Portfolio Theory (MPT). He quantifies investment return and risk by expected return and variance, respectively. The main idea of MV methodology is to maximize expected return keeping unchanged variance, or minimize variance keeping unchanged expected return. MPT has been widely accepted and studied by researchers. Tobin (1958) indicates that liquidity preference could determine how much wealth is to be invested in monetary assets and constructs an effective portfolio combined with risk-free assets as well as a special type of risky assets. Sharpe (1963) puts forward the diagonal model assuming that there is no interrelationship among securities so as to simplify the calculation, which significantly facilitates the development of portfolio theory. Some researchers notice that multi-period portfolio selection should be considered to deal with the complex financial markets. For instance, Merton (1969) extends modern portfolio theory by introducing a continuous-time model in order to achieve the goal of maximal expected utility within a constant planning region. Grauer and Hakansson (1993) apply a discrete-time dynamic investment model to compare the MV and the quadratic approximations computing the optimal portfolios. Some researches put several realistic constraints into the Markowitz's MV model. For instance, Liu and Loewenstein (2002) incorporate transaction cost into the stock trading strategy to help maximize the investors' wealth utility. Brown and Smith (2011) consider risk aversion, transaction cost, portfolio constraints into MV model and find that it would be difficult to solve portfolio optimisation issues when three more assets are involved. Moreover, some studies use robust optimisation techniques in portfolio management. Tu and Zhou (2010) involve the financial objectives into Bayesian priors to estimate uncertain parameters and they prove that Bayesian method under the objective-based priors performs better than those under alternative priors in portfolio selection. Under a Bayesian estimation framework, Bodnar, Mazur and Okhrin (2017) analyse the global minimum variance portfolio and consider investors' prior beliefs into the portfolio decisions. On the basis of random matrix theory, Bodnar, Parolya, and Schmid (2018) evaluate the global minimum variance portfolio with high-dimensional data to minimize the out-of-sample variance.

Furthermore, numerous scholars start to analyse portfolio issues using fuzzy set theory. Li and Xu (2013) indicate that there are often fuzzy uncertainty and random uncertainty existing in the financial market, hence, they incorporate investors' sentiments and experts' insights into the process of portfolio construction. Assuming that expected rate of returns obeys normal distribution, Li, Zhang and Xu (2013) integrate two constraints, value at risk (VaR) and risk-free assets, into a fuzzy portfolio selection model so as to find a more suitable portfolio. Li, Zhang, and Xu (2015) put forward another fuzzy portfolio selection model with background risk to obtain the effective frontier of a portfolio. Recently, with the development of big data and artificial intelligence technology, it is possible to use computers and a large number of calculations to achieve optimal portfolio management. Huang, 2012) focuses on the high-return stock selection using support using genetic algorithms (GAs) as well as vector regression (SVR), but he ignores risk factors. Based on support vector machine (SVM), Paiva, Cardoso, Hanaoka and Duarte (2019) classify the assets to achieve a certain return and determine the components of the investment portfolio. Almahdi and Yang (2017) set three optimisation objectives, annualised Sharpe ratio, Sterling ratio, and Calmar ratio, respectively, then choose the best performing algorithm to select optimal portfolio. Yunusoglu and Selim (2013) develop an expert system (ES) to support portfolio managers for investment decisions. The expert system contains three stages, the first stage is the elimination of unacceptable stock. The second stage is to evaluate stock through a comprehensive literature survey and interviews with a domain expert. The last stage is to construct a portfolio based on a mixed-integer linear programming model. Their results demonstrate that under the different risk preferences, ES performance is not particularly big difference, moreover, ES is more suitable for 6 months, 9 months and 12 months of investment period.

It is obvious that various extensions of Markowitz's MV model help enrich the modern portfolio theory and provide researchers with more research perspectives. And these extensions further confirm that MV model plays an extremely significant role in portfolio management. However, most of the related researches ignore the selection of high-quality assets, the stage before the optimal portfolio formation. Instead, they focus more on how to improve the MV model. Actually, high-quality asset input is a reliable guarantee for optimal portfolio formation during the investment process. In this regard, this paper will continue to adopt the classical MV model, moreover, we will study deeply the preliminary selection of assets in order to provide MV model with better asset inputs. At the same time, different transaction costs will be considered for simulation to visualize the performance of different models.

2.2. Return prediction with deep learning

In recent years, with the development of big data and artificial intelligence (AI) technology, more and more scholars start to use AI as support for their research solutions and prove that AI methods deal with the problem of nonlinear, nonstationary characteristics better than traditional statistical models. For example, a number of researches based on SVM (Paiva, Cardoso, Hanaoka, & Duarte, 2019), PCA (Chen & Hao, 2018, Zbikowski, 2015), GA or random forest ((Li & Xu, 2013); Mousavi, 2014), ANN (Chong, Han, & Park, 2017, Patel, Shah, Thakkar, & Kotecha, 2015) to classify, predict and optimise complex financial assets. Among these technologies, the deep learning is thought to be an appropriate method for the financial time-series forecasting solution, since it is good at processing complex, high-dimensional data as well as extracting abstract characteristics from mass data without depending on any assumptions.

The deep learning method proposed by Hinton and Salakhutdinov (2006), has become a leading application in the financial area, especially in predicting financial market movement and processing text information. Deep learning architectures mainly include deep neural networks (DNNs), deep belief networks (DBNs), recurrent neural networks (RNNs) and convolutional neural networks (CNNs) (LeCun, Bengio, & Hinton, 2015). Amongst them, DNNs are feedforward networks in which data flows from the input layer to the output layer by their single directional forward links without going backwards (Arévalo, Niño, Hernández, & Sandoval, 2016). Chong, Han, & Park (2017) testify that with regard to future returns prediction, DNN is obviously superior to a linear autoregressive model based on data from the Korean stock market. Identifying the correlation between different stocks, Lachiheba and Gouiderb (2018) come up with a DNN model with a special structure to predict the trend of stock returns over the next five minutes and the results manifest that the accuracy is improved to 71% considerably. DBNs are composed of multiple layers of latent variables, with connections between the layers but not between units within each layer (Hinton, 2009). Shen et al. 2015 construct a DBN using continuous restricted Boltzmann machines to predict the exchange rate and their results show that their method performs better than traditional methods. Unlike feedforward neural networks, RNNs can use their internal states (memory) to process sequences of inputs. For instance, Rather et al. (Rather, Agarwal, & Sastry, 2015) construct a novel hybrid model constituting autoregressive moving average model, exponential smoothing model and RNN to obtain more accurate returns prediction. Similarly, Long et al. (Long, Lu, & Cui, 2019) integrate CNN and RNN into their proposed model entitled "multi-filters neural network" aiming to see the trend of the stock price over time, finally, they verify the prediction accuracy of the model through simulation. Long short-term memory (LSTM) networks are one of the classes of recurrent neural networks (RNNs), but it has the advantage of retaining information over a long time-span compared with RNNs (Fischer & Krauss, 2018, LeCun, Bengio, & Hinton, 2015). Kraus and Feuerriegel (Kraus & Feuerriegel, 2017) analyse the text data using the long short-term memory (LSTM) networks, finally, they prove that their method increases the accuracy of the stock price prediction. Fischer and Krauss (Fischer & Krauss, 2018) take advantage of the LSTM networks to forecast stocks' directional movement and their results show that LSTM outperforms some classical machine learning models in this prediction task. Besides, Ding et al. (Ding, Zhang, Liu, & Duan, 2015) apply CNNs to predict the short-term and long-term influences of events on stock price movements and they prove that the accuracy of the model outperforms other baseline methods.

It is clear that the deep learning method is able to find complex structures in high-dimensional financial data and acquire features through simple and non-linear modules, and then transform features from lower level to higher level and more abstract features (LeCun, Bengio, & Hinton, 2015). Based on above literature review, it is easy to discover that the majority of the existing studies on predicting assets returns based on deep learning pay more attention to improve the prediction accuracy, however, few of them apply their prediction results to actual financial markets, such as portfolio management, assets selection, or trading strategy, to give investors more practical guidance. Actually, the high accuracy of prediction does not represent the optimal investment strategy. The advantages of deep learning methods in predicting can be very helpful for decision making in financial investments (Aggarwal & Aggarwal, 2017). Therefore, how to combine the prediction of deep learning to help choose the optimal investment strategy is a meaningful and promising research direction (Zhang, Li, & Guo, 2018).

3. Methodology

3.1. Data

The biggest challenge of prediction is to recognise a relation in financial time-series data between the past and the future (Paiva, Cardoso, Hanaoka, & Duarte, 2019). Since the continuity of financial stock data, the longer the sample data is involved, the more likely it is to capture history information memory (Fischer & Krauss, 2018, Long, Lu, & Cui, 2019). Hence, a large amount of long-term data is required in the empirical experiment (Chourmouziadis & Chatzoglou, 2016). In this research, we collect daily stock data from the UK Stock Exchange 100 Index (FTSE 100) from March 1994 until March 2019, covering 25 years. Since the majority of related studies have been conducted over a period of 10 years or less (Chen & Hao, 2018, Kara, Boyacioglu, & Baykan, 2011, Patel, Shah, Thakkar, & Kotecha, 2015), 15 years (Almahdi & Yang, 2017, Paiva, Cardoso, Hanaoka, & Duarte, 2019), or 25 years (Fischer & Krauss, 2018), our samples spanning 25 years can be considered to provide a sufficiently large

Table 1
Descriptive statistics for sample data.

Stock	Mean	Std.	Maximum	Minimum
TES	255.66	104.65	492.0	67.33
AST	2923.12	1142.28	6317.0	658.41
BAR	318.90	141.05	710.69	47.0
BP	468.58	109.99	712.0	174.5
BAT	1808.8	1468.73	5643.0	217.59
HAL	357.44	349.40	1648.0	81.5
HH	611.21	165.72	951.6	171.09
JM	1572.69	967.46	3823.0	263.85
LG	141.78	65.36	23.0	294.4
MSG	396.84	111.54	749.0	170.75
PEA	887.68	316.77	2301.79	429.5
RELX	714.0	352.18	1782.0	348.82
RB	2076.30	1877.32	6026.35	103.0
RDSB	1753.62	451.63	761.02	2841.0
SG	978.12	283.06	424.34	1801
SJ	343.75	77.27	594.0	214.6
SCH	1416.98	875.33	3773.0	346.01
ST	1257.59	536.65	2553.0	487.17
SG	978.12	283.06	1801.0	424.34
SSE	984.88	433.62	1696.0	272.5
VG	158.88	65.81	408.57	32.29

volume data to generate statistically significant results. Our sample data involves the historical series of adjusted open prices, close prices, the highest prices, the lowest prices, and the trading volume of assets. Numerous scholars agree on that holding tens of thousands of different stocks as a portfolio is not realistic for individual investors (Almahdi & Yang, 2017, Kocuk & Cornuéjols, 2018, Ranguelova, 2001, Tanaka, Guoa, & Turksen, 2000). For instance, Tanaka et al. (Tanaka, Guoa, & Turksen, 2000) select 9 securities as the sample to form the optimal portfolio. Almahdi and Yang (Almahdi & Yang, 2017) construct a five-asset portfolio. Hence, this paper randomly chooses twenty-one stocks from FTSE 100 as sample data, which is sufficiently large for the asset preselection before forming the portfolio for individual investors. The names of these sample stocks are "BP" (BP), "Barclays" (BAR), "Tesco" (TES), "Vodafone Group" (VG), "Halma" (HAL), "Johnson Matthey" (JM), "HSBC Holdings" (HH), "Sainsbury J" (SJ), "Marks & Spencer Group" (MSG), "Astrazeneca" (AST), "British American Tobacco" (BAT), "PEARSON" (PEA), "Relx" (RELX), "SSE" (SSE), "Legal & General" (LG), "Royal Bank" (RB), "Royal Dutch Shell B" (RDSB), "Sage Group" (SG), "Schroders" (SCH), "Seven Trent" (ST) and "Smiths Group" (SG). Their abbreviations are used for convenience, respectively. Table 1 shows the descriptive statistics of close prices for the 21stocks selected from FTSE 100. As can be seen, stock AST has the highest daily mean prices: 2923.12, stock LG has the lowest standard deviation: 65.36, stock VG follows, with 65.81.

3.2. LSTM networks

LSTM networks were introduced by Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997) as an alternative method to learn sequential patterns. LSTM networks are one of the classes of recurrent neural networks (RNNs), but it has the advantage to retaining information over a long time-span compared with RNNs (Fischer & Krauss, 2018, LeCun, Bengio, & Hinton, 2015). Graves and Schmidhuber (Graves & Schmidhuber, 2005) demonstrate that LSTM networks are able to overcome the previously inherent problems and memorize temporal patterns over a long period time.

LSTM networks are comprised of an input layer, several hidden layers, and an output layer. The most important characteristic of LSTMs is memory cells contained in the hidden layers. Fig. 1 illustrates the structure of an LSTM memory cell. As we can see, for each memory cell, x_t and h_t correspond to the input and hidden state respectively, at time t, and i_t, o_t and f_t, are the gates which are called input, output and forget gates, respectively, s_t is adjusting its cell state. It is worth noting that the input gate decides which data can be added into the memory cell, the output gate decides which data from the memory cell can be used as output, and the forget gate decides which data should be deleted from the memory cell. The calculations for each state and gate are performed as the following formulas.

$$f_t = sigmoid\left(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f\right) \tag{1}$$

$$\tilde{s}_t = sigmoid\left(W_{\tilde{s},x}x_t + W_{\tilde{s},h}h_{t-1} + b_{\tilde{s}}\right) \tag{2}$$

$$i_t = sigmoid\left(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i\right) \tag{3}$$

$$s_t = f_t * s_{t-1} + i_t * \tilde{s}_t \tag{4}$$

$$o_t = sigmoid\left(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o\right) \tag{5}$$

$$h_t = o_t * \tanh(s_t) \tag{6}$$

Where $W_{f,x}, W_{f,h}, W_{\tilde{s},x}, W_{\tilde{s},h}, W_{i,x}, W_{i,h}, W_{o,x}$ and $W_{o,h}$ are weight matrices, $b_f, b_{\tilde{s}}, b_i$, and b_o are bias vectors of the respective gates. Those bias vectors are added to increase the flexibility of the model to fit the data. Bias vectors $b_{\tilde{s}}, b_i$, and b_o are initialized to zero, but the bias b_f for the forget gate in LSTM is initialized to 1.0 (Jozefowicz, Zaremba, & Sutskever, 2015). The symbol of * indicates element-wise multiplication. Because of this selective process of information, LSTM is able to deal with longer temporal patterns.

3.3. Mean-variance model

Mean-variance (MV) model proposed by Markowitz (Markowitz, 1952) in order to solve the optimal portfolio selection issue, which initiates the foundation of Modern Portfolio Theory (MPT). In this model, the investment return and risk are quantified by expected return and variance, respectively. Santos and Tessari (Santos & Tessari, 2012) hold the view that the core of the portfolio selection for investors is to decide which portfolio is the best on the basis of risk and expected returns. Hereby, rational investors always prefer the lower risk portfolios with constant expected returns or the higher expected return portfolios with the constant risk level. To solve this issue, a set of optimal solutions is generated, named an efficient investment frontier. The model can be described by the following formulas:

$$\underset{w_i, \ldots, w_n}{Min} \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \delta_{ij} \tag{7}$$

$$\underset{w_i, \ldots, w_n}{Max} \sum_{i=1}^{n} w_i \mu_i \tag{8}$$

$$\text{Subject to}: \begin{cases} \sum_{i=1}^{n} w_i = 1 \\ 0 \le w_i \le 1, \forall\ i = 1, \ldots, n \end{cases} \tag{9}$$

Where w_i *and* w_j represent the initial value invested in the portfolio or asset i and asset j. δ_{ij} specifies covariance between assets i and asset j. μ_i is expected return on asset i. Following Paiva et al. (Paiva, Cardoso, Hanaoka, & Duarte, 2019), a variable λ called risk aversion coefficient is integrated into the model to depict investors' behavior corresponding to the risk investment choices. A mono-objective formulation is as following:

$$\underset{w_i, \ldots, w_n}{Min} \lambda\left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \delta_{ij}\right] - (1-\lambda)\left[\sum_{i=1}^{n} w_i \mu_i\right] \tag{10}$$

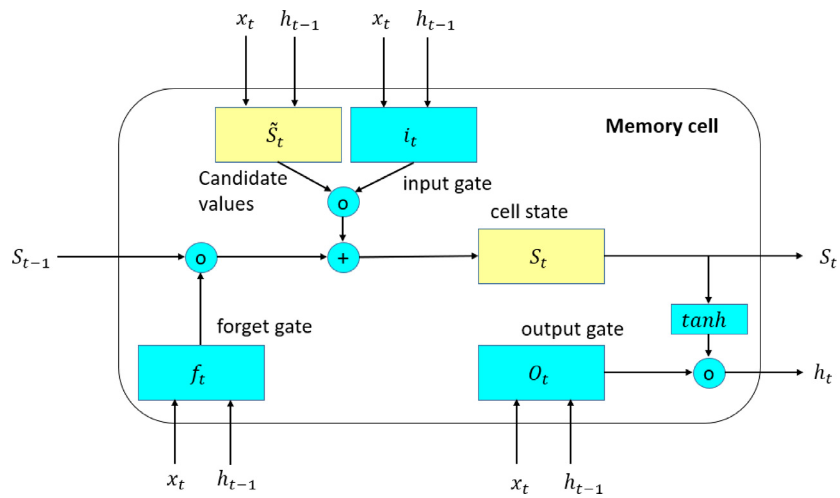

Fig.1. Structure of LSTM memory cell following Fischer and Krauss (Fischer & Krauss, 2018).

$$\text{Subject to}: \begin{cases} \sum_{i=1}^{n} w_i = 1 \\ 0 \le w_i \le 1, \forall\ i = 1, \ldots, n \end{cases} \tag{11}$$

As a result, a group of optimal portfolios constitutes an effective frontier can be derived and introduced to the investor. So, the investor could select the point among these possible solutions according to his or her risk preference.

3.4. Proposed model: LSTM+MV

Many researches always ignore the fact that the purpose of forecasting financial market is not to show off the accuracy of a model but to apply these good results into the real market so as to give investors more practical and meaningful guidance. During the investment decision-making process, high-quality asset input would be very helpful for optimal portfolio formation. Given the important role that MV method plays in portfolio management area, we will continue to adopt this classical model, moreover, we will study deeply the preliminary selection of assets in order to provide MV model with better asset inputs. In this regard, this study puts forward a mixed method named LSTM+MV combining the advantages of deep learning LSTM method in time-series forecasting with the effectiveness of MV model in portfolio optimisation, aiming to improve the investment decision-making process.

There are two stages in our proposed model. In the first stage, LSTM method is applied to predict the return of the sample stocks in the next period. All the predicted results will be sorted in descending order and the top stocks will enter into the next phase. In the second stage, the Markowitz's MV model will be used to obtain the capital allocation proportion for each stock that has been entered.

3.4.1. Input variable selection

The selection of input variables is extremely necessary for time-series prediction tasks. In light of previous literature, technical indicators are effective features to describe and reflect the real market situation. For instance, Chen and Hao (Chen & Hao, 2018) suggest that Exponential Moving Average (EMA), Relative Strength Index (RSI) and Momentum Index (MoM) are correlated with changes in the stock market. Kara et al. (Kara, Boyacioglu, & Baykan, 2011) select ten technical indicators as input feature subsets. Also, financial time-series forecasting is always explained by the lagged observations. For example, Fisher and Krauss (Fischer & Krauss, 2018) use a return time sequence length of 240 for training. Paiva et al. (Paiva, Cardoso, Hanaoka, & Duarte, 2019) use several lagged variables of return as inputs to predict the future return of stocks. Hereby, after referring to the views of domain papers, we make feature selection by recursive feature elimination (RFE). To be specific, RFE works by recursively removing features and building a model on those features that remain. It uses the model accuracy to identify which features contribute the most to predicting the target feature (return in $t+1$ period). We use RFE with the logistic regression algorithm to select the features with a ratio greater than 0.3. Fig. 2 shows the results of feature selection using RFE. We finally choose twenty important indicators as input variables, including five technical indicators and fifteen lagged observations about return. The values of all technical indicators are standardized in the range of $(-1, +1)$, in order to avoid the errors caused by different indicators of different numerical ranges. Table 2 summarises the selected input variables. Among the variables are return measures based on open, close, high, low prices, and volume. A brief explanation of each indicator is as following.

(1) Simple return

Set P_t^i as the price process of stock i at time t, with $i\varepsilon\{1, 2\ldots, n\}$ and $R_t^{m,i}$ as the simple return for stock i over t periods, i.e., $R_t^{m,i} = \frac{P_t^i}{P_{t-m}^i}$.

(2) Relative Strength Index (RSI)

RSI, a momentum indicator, is able to measure the magnitude of the rise and fall in prices recently. It is very effective in assessing the overbought/oversold condition of an asset. According to the parameters of this indicator in existing researches (Chen & Hao, 2018, Paiva, Cardoso, Hanaoka, & Duarte, 2019, Patel, Shah, Thakkar, & Kotecha, 2015), this paper set the period as 14.

(3) Momentum Index (MoM)

MoM is an extremely popular indicator measuring a security's rate-of-change, which refers to the force or speed of movement. Following existing researches (Chen & Hao, 2018, Paiva, Cardoso, Hanaoka, & Duarte, 2019, Patel, Shah, Thakkar, & Kotecha, 2015), in this paper, the period is set to 10.

(4) True range (TR)

TR is the maximum change in the price of the day compared to the previous day.

(5) Average true range (ATR)

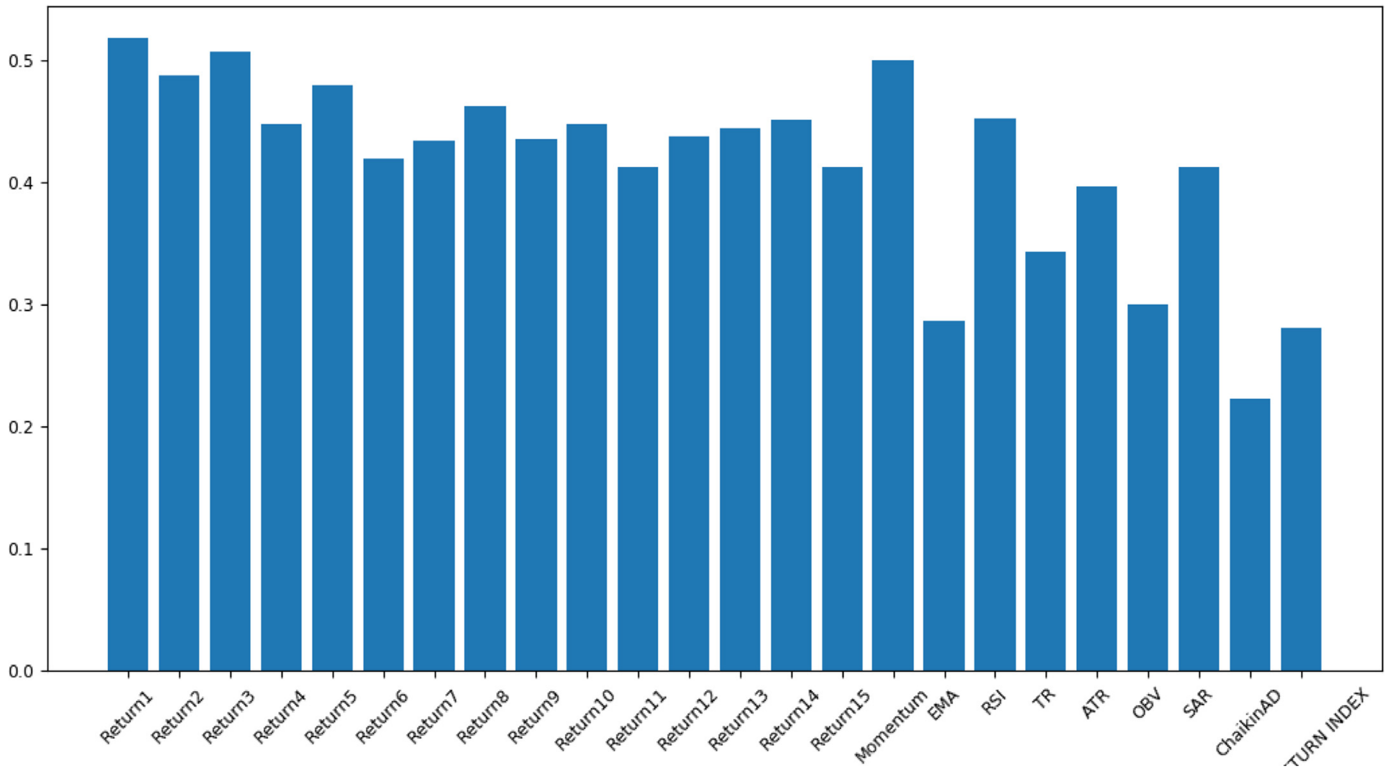

Fig. 2. Feature selection results.

Table 2
Input features summary.

Attribute	Details	Attribute	Details
1	$r_1 = \ln(\frac{close\ price_i}{close\ price_{i-1}})$	11	$r_{11} = \ln(\frac{high\ price_{i-3}}{open\ price_{i-3}})$
2	$r_2 = \ln(\frac{close\ price_{i-1}}{close\ price_{i-2}})$	12	$r_{12} = \ln(\frac{low\ price_i}{open\ price_i})$
3	$r_3 = \ln(\frac{close\ price_{i-2}}{close\ price_{i-3}})$	13	$r_{13} = \ln(\frac{low\ price_{i-1}}{open\ price_{i-1}})$
4	$r_4 = \ln(\frac{close\ price_{i-3}}{close\ price_{i-4}})$	14	$r_{14} = \ln(\frac{low\ price_{i-2}}{open\ price_{i-2}})$
5	$r_5 = \ln(\frac{high\ price_i}{open\ price_i})$	15	$r_{15} = \ln(\frac{low\ price_{i-3}}{open\ price_{i-3}})$
6	$r_6 = \ln(\frac{high\ price_i}{open\ price_{i-1}})$	16	Relative Strength Index (close price, period =14)
7	$r_7 = \ln(\frac{high\ price_i}{open\ price_{i-2}})$	17	Momentum Index (close price, period =10)
8	$r_8 = \ln(\frac{high\ price_i}{open\ price_{i-3}})$	18	True range (high, low, and close price)
9	$r_9 = \ln(\frac{high\ price_{i-1}}{open\ price_{i-1}})$	19	Average true range (high, low and close price, period = 14))
10	$r_{10} = \ln(\frac{high\ price_{i-2}}{open\ price_{i-2}})$	20	Parabolic SAR (high and low price, acceleration = 0.02, maximum = 0)

ATR is a technical analysis indicator that reflects market volatility through decomposing the entire range of an asset price for a period.

(6) Parabolic SAR

The parabolic SAR is used to determine the direction in which asset prices rise or fall, besides, it will remind us when the direction of the price changes, in other words, it will adjust as prices change so as to attract investors' attention.

3.4.2. Generation of training and testing sets

Since the continuity of time-series data, we consider each training-testing set as a "study period", involving a training period of 750 days and a testing period of 250 days (Fischer & Krauss, 2018). We divide our sample data from March 1994 until March 2019 into twenty-two study periods with overlapping training-testing sets. In each study period, the data in the first 750 days is used for training with rolling windows, the rest data fully out-of-sample in the last 250 days is performed for testing based on the trained parameters. Then, the entire network will roll forward 250 days, leading to twenty-two non-overlapping testing sets. Details can be seen in Fig. 3. The blue area represents the whole span of our sample, from March 1994 until March 2019. The yellow area indicates the training set, 750 days. The red area is the testing set, 250 days. The red and yellow areas together form our "study period", 1000 days.

3.4.3. Process of optimal portfolio formation

The proposed model LSTM+MV in this paper is on the basis of technical analysis as well as historical asset prices identification. On this account, we follow the assumption of Fama (1965) who holds the view that history behaviour trend of the price change in individual assets is inclined to repeat in the future. The primary objective of the LSTM method here is to forecast the relative return rate of each stock in $t+1$ trading day on the basis of the information before time t. In LSTM networks of our proposed model, some sequences of input features are required for training, that is, the values of input features at points in consecutive times. With regard to the training of the LSTM networks, three advanced methods are applied through Keras. First, Adam (Kingma & Ba, 2014) is used as the optimiser to improve the neural network. This selection is inspired from some existing researches (Kingma

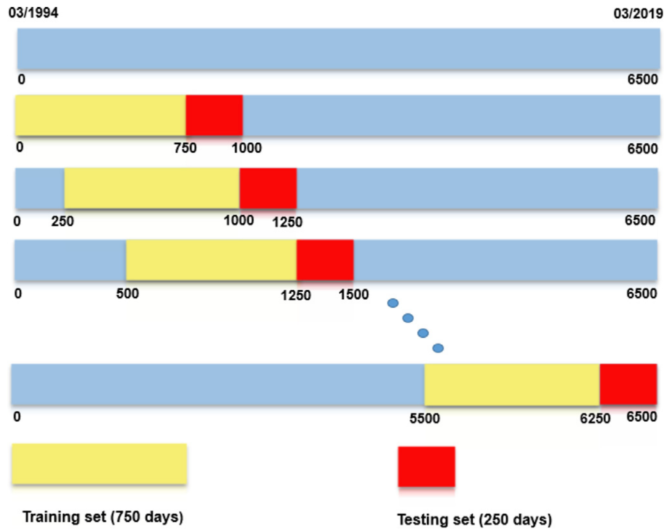

Fig. 3. overlapping training-testing sets.

& Ba, 2014, Kraus & Feuerriegel, 2017, Reimers & Gurevych, 2017), as they testify that Adam is appropriate for deep LSTM networks and has a better performance in optimising the neural network. Second, referring to Srivastava et al. (Srivastava et al., 2014), we make use of the dropout regularization technique on the hidden layer. In this case, randomly selected neurons are dropped during training times, along with corresponding input and output connections, which is able to reduce overfitting efficiently (Fischer & Krauss, 2018, Srivastava et al., 2014). In the case of Adam optimiser, we also carry an initial experiment using a part of the sample, the result shows that the model performance decreases as the dropout rate increases, hence, we set the dropout rate relatively low as 0.1. Third, we perform random search method to dynamically find a good combination of hyperparameters based on the above settings. Plenty of empirical evidence has shown the effectiveness of random search in optimising the parameters (Bergstra & Bengio, 2012, Greff et al., 2017). The random search samples the following hyperparameters: (1) the sequence length, ranging from 30 to 250; (2) the number of epochs, ranging from 10 to 100, (3) neuron activation function; (4) the number of neurons per hidden layer, ranging from 2 to 200. Finally, the specified topology of the LSTM network is confirmed. We set 20 features and 72 timesteps in input layer. And in LSTM layer, we set 60 hidden neurons and 0.1 for dropout rate. In dense layer, we apply 16 neurons and relu activation function. Also, we set one neuron and sigmoid activation function in output layer, which is a standard configuration (Fischer & Krauss, 2018). Since the optimal sequence length is 72, approximately covering the data of three testing months. Thus, overlapping sequences of 72 consecutives are generated. In total, 22 study periods contain about 429,000 of those sequences, in which approximately 321,750 are utilized for in-sample training, and 107,250 are utilized for out-of-sample predictions. For each study period, there are about 19,500 of those sequences. Suppose that we would like to find whether an asset has the potential to reach higher return in $t+1$. Then, we will collect all data of that asset before the trading session at t_0 in order to achieve this goal. According to LSTM principles, the data series from previous days would be put into the model to implement the experiment.

Once all the assets are predicted, one by one, we will rank all stocks for each period $t+1$ in descending order of this predicted return. Only the top k of the ranking with the higher return assets that are considered to qualify to enter into the next phase. The purpose of the second stage is to obtain the capital allocation proportion for each asset. And the Markowitz's MV model will be used to carry on this stage. It is worth clarifying that the proposed model does not take into account investors' risk preferences and risk-free assets, thus, the portfolios exclusively compose of risky assets. According to the way of Malkiel (Malkiel, 2007) letting a blindfolded monkey throw darts at a newspaper's financial pages, we also create a function in python to randomly generate 50,000 portfolios. From a statistical perspective, 50,000 random portfolios basically cover most possible portfolios with different weights and can be considered representative enough (Fischer & Krauss, 2018). Furthermore, all these 50,000 portfolios will be screened in accordance with MV optimisation rules so that better portfolio can be found. In the end, the available resources will be allocated to the portfolio with the lowest variance. As such, when the assets and the respective investment proportions are confirmed, the next step is to allocate capital at the opening of the next trading day. We will go long the top k assets during the investment day. The detailed process of the proposed method is shown in Fig. 4.

3.4.4. Benchmark models for prediction: SVM, RAF, DNN, and ARIMA

In order to benchmark the LSTM, three representative machine learning models, support vector machine, random forest, deep neural network, as well as a traditional statistical model named Autoregressive Integrated Moving Average that is often applied for time-series prediction. We will introduce the principles of each model in the following paragraphs.

Support Vector Machine: This technique aims to solve issues related to classification, regression estimation, pattern recognition and time series (Paiva, Cardoso, Hanaoka, & Duarte, 2019). Support vector regression (SVR), proposed by Drucker et al. (Drucker et al., 1997), is a version of support vector machine (SVM) for regression. SVR is able to deal with continuous values and find the best regression hyperplanes in order to estimate the dependent variable value (Loureiro, Miguéis, & Silva, 2018).

Random Forest: The algorism derives from the decision trees and is developed to improve the accuracy of decision trees and overcome the high sensitivity to small changes in data. It is generally accepted that it is an advanced machine learning model that usually gets good results and seldom needs tuning (Fischer & Krauss, 2018).

Deep neural network: DNN consists of multiple hidden layers, one input and one output layer (Loureiro, Miguéis, & Silva, 2018). To be specific, this paper applies a feedforward neural network with 20 input neurons and the activation with relu (Li & Yuan, 2017), 30 neurons in the first hidden layer, 3 neurons in the second hidden layer (Fischer & Krauss, 2018), and one neuron in the output layer. Dropout is set to 0.2.

Autoregressive Integrated Moving Average model: ARIMA is a classical econometric model, fitted to predict time-series data in future, and ARIMAX extends ARIMA model by including exogenous variables (Pektas & Cigizoglu, 2013). This paper uses ARIMAX model as one of the baseline models.

3.4.5. Baseline strategies for portfolio formation

In reality, except for MV model, equal-weighted portfolio and Black-Litterman (BL) model are also popular. It is worth noting that we originally used the BL model as one of the baselines, but in the end, we found that we could not get a prominent and consistent result to explain. Maybe the parameters of different models need to be adjusted or due to some other reasons, we have not figured out. Therefore, we decide not to discuss BL model in this paper. These following baseline strategies are based on the LSTM+MV model proposed in the prior section and used to compare with this model's changes and performance.

(1) **Alternative model: Machine learning + MV**

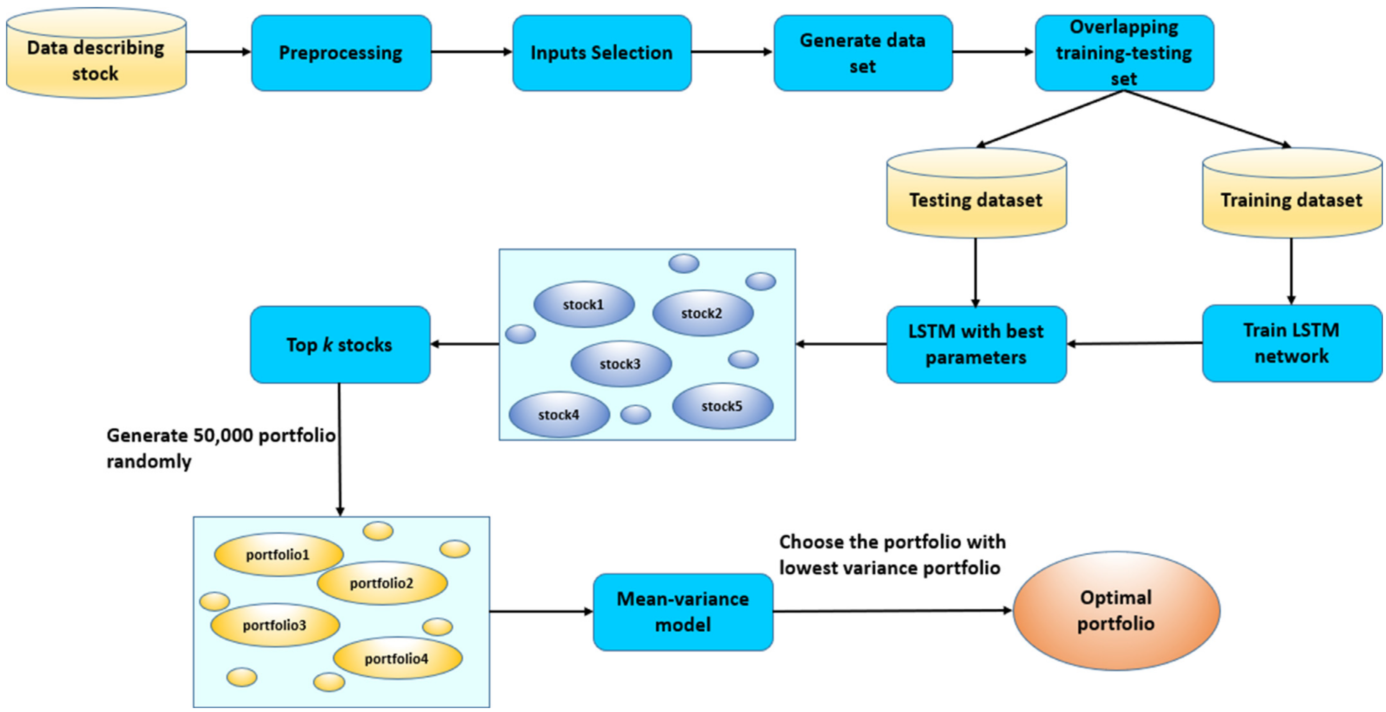

Fig. 4. The scheme of proposed model.

This kind of model's design is similar to the logical structure of the LSTM+MV model. The main objective is to find out whether different prediction results of asset return will have an impact on the formation of the final optimal portfolio. To be specific, assets returns in $t+1$ will be predicted by one machine learning method with better forecasting performance in the first stage, and assets with higher return in the future will be chosen into the second stage. Notice that the number of assets selected must be as same as the number defined in the LSTM+MV model. The second stage, portfolio optimisation, applying the Markowitz's MV method is maintained.

(2) **Alternative model: Machine learning + 1/N**

The objective of this baseline strategy is to examine the portfolio optimisation effect between MV and 1/N (equal-weighted), in the case of the same initial selection of assets. Specifically, one machine learning method with better forecasting performance in the first stage will be used to predict assets returns in $t+1$, and then rank these assets according to the predicted results. Finally, the top k assets will enter into the second stage and receive the same proportion of investment. Notice that k should be consistent with the number defined in the LSTM+MV model.

(3) **Alternative model: Random+ MV or 1/N**

This kind of baseline strategy differs from the previous baselines in terms of the asset preselection phase. The asset preselection is randomly undertaken without relying on any predictions, but the number of assets should be the same as the number defined by the other models. To be specific, we will randomly select a certain number of assets from all our samples and then apply Markowitz's MV method or 1/N optimisation separately to optimise the portfolio. The objective of this kind of baseline strategy is to examine the necessity of asset preselection using machine learning.

4. Experiments and results

4.1. Results analysis in the first stage: prediction

In this section, we use five criterions to evaluate predictive accuracy, mean square error (MSE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE) and coefficient of determination (R^2). Tables 3 to 5 summarise the best results achieved for each model applied according to the different evaluation metrics employed. As can be seen from three tables, the majority of indicators corresponding to LSTM model perform better than the index value of other models, but several exceptions also exist. For example, the MAE and MAPE indicator of stock SSE where the prediction result of LSTM is larger than that of SVM and DNN respectively. Another example is that the R^2 of 3 stocks (BP, JM, SG) predicting by SVM are higher than that of LSTM, and the R^2 of 2 stocks (BAT, PEA) using RAF are higher than that of LSTM too.

Mean square error (MSE): It is an indicator measuring the average squared difference between the observed values and predicted values. From Table 3, Table 4 and Table 5, we find the following average MSE result: 0.0033 for LSTM model, 0.0047 for SVM model, 0.008 for DNN model, 0.64 for RAF model and 25.49 for ARIMA model.

Root-mean-square error (RMSE): It is another effective indicator measuring differences between the observed values and predicted values. As can be seen from Tables 3 to 5, LSTM exhibits favourable mean RMSE 0.0543, followed by SVM (0.0649), then the indicator for DNN and RAF equals to 0.0852 and 0.0723, but 3.817 for ARIMA model.

Mean absolute percentage error (MAPE): It measures the prediction deviation proportion in terms of the true value. After comparing different models in terms of MAPE, we can get the average results: 91.44 for LSTM model, 106.93 for SVM model, 161.58 for DNN model, 137.53 for RAF model and 2560 for ARIMA model.

Mean absolute error (MAE): It is a measure of the accuracy of a forecasting method. We see that the LSTM has the lowest mean

Table 3
Comparison of prediction performance.

Stock	LSTM					SVM				
	MSE	RMSE	MAPE	MAE	R²	MSE	RMSE	MAPE	MAE	R²
TES	0.0031	0.0557	67.99	0.0364	0.4209	0.0042	0.0651	167.92	0.0427	0.3108
AST	0.0032	0.0568	165.53	0.0324	0.2806	0.0089	0.0942	190.68	0.0512	0.1856
BAR	0.0007	0.0265	6.63	0.0159	0.2631	0.0050	0.0608	22.47	0.0335	0.1200
BP	0.0053	0.0727	123.00	0.0413	0.1121	0.0054	0.0732	114.88	0.0450	0.1379
BAT	0.0019	0.0439	7.76	0.0296	0.1259	0.0064	0.0731	25.50	0.0404	0.0862
HAL	0.0050	0.0709	266.49	0.0378	0.2288	0.0054	0.0735	221.27	0.0397	0.1359
HH	0.0015	0.0390	25.10	0.0214	0.2395	0.0024	0.0484	55.71	0.0290	0.1349
JM	0.0063	0.0797	221.26	0.0495	0.1327	0.0099	0.0997	226.03	0.0431	0.1718
LG	0.0009	0.0293	17.42	0.0155	0.1585	0.0010	0.0317	18.71	0.0176	0.1210
MSG	0.0029	0.0540	17.06	0.0301	0.2630	0.0040	0.0629	19.62	0.0390	0.2141
PEA	0.0018	0.0426	9.06	0.0227	0.1100	0.0026	0.0513	10.95	0.0323	0.1816
REL	0.0028	0.0532	35.08	0.0267	0.1557	0.0029	0.0539	37.78	0.0297	0.1145
RB	0.0002	0.0146	1.98	0.0094	0.4578	0.0003	0.0162	2.22	0.0104	0.3768
RDSB	0.0068	0.0825	158.40	0.0417	0.2960	0.0062	0.0785	154.90	0.0417	0.1093
SG	0.0039	0.0622	74.90	0.0302	0.2545	0.0047	0.0684	71.50	0.0367	0.3592
SJ	0.0028	0.0525	16.29	0.0258	0.6139	0.0031	0.0552	17.00	0.0279	0.1218
SCH	0.0011	0.0329	265.81	0.0189	0.1638	0.0014	0.0368	232.70	0.0224	0.1553
ST	0.0046	0.0680	43.34	0.0388	0.1441	0.0070	0.0837	77.83	0.0450	0.1454
SG	0.0052	0.0726	61.96	0.0394	0.2809	0.0077	0.0876	94.77	0.0510	0.2001
SSE	0.0041	0.0642	196.59	0.0346	0.4572	0.0042	0.0645	245.40	0.0335	0.2977
VG	0.0045	0.0672	138.70	0.0385	0.3448	0.0071	0.0844	237.70	0.0469	0.2810

Table 4
Comparison of prediction performance.

Stock	DNN					RAF					
	MSE	RMSE	MAPE	MAE	R²	MSE	RMSE	MAPE	MAE	MSE	R²
TES	0.0130	0.1160	317.26	0.0590	0.1300	0.0058	0.0758	189.56	0.0526	0.0058	0.3593
AST	0.0100	0.1010	194.48	0.0480	0.1410	0.0046	0.0676	162.97	0.0409	0.0046	0.2119
BAR	0.0020	0.0450	10.04	0.0220	0.0800	0.0056	0.0643	23.25	0.0370	0.0056	0.1563
BP	0.0150	0.1210	484.08	0.0590	0.1020	0.0213	0.1402	653.1	0.0817	0.0213	0.1261
BAT	0.0040	0.0630	10.44	0.0340	0.1770	0.0025	0.0495	8.87	0.0285	0.0025	0.2855
HAL	0.0080	0.0900	273.28	0.0460	0.1120	0.0062	0.0787	234.59	0.0413	0.0062	0.2100
HH	0.0060	0.0750	89.25	0.0380	0.1340	0.0069	0.0832	49.11	0.0472	0.0069	0.2047
JM	0.0150	0.1220	495.23	0.0660	0.1210	0.0256	0.1503	296.76	0.0847	0.0256	0.1026
LG	0.0030	0.0540	34.60	0.0230	0.6890	0.0023	0.0475	24.14	0.0282	0.0023	0.3095
MSG	0.0060	0.0760	35.77	0.0390	0.1630	0.0045	0.0674	31.37	0.0362	0.0045	0.2057
PEA	0.0060	0.0780	17.88	0.0340	0.0360	0.0020	0.0442	9.69	0.0265	0.0020	0.2403
REL	0.0047	0.0684	53.22	0.0407	0.1130	0.0046	0.0678	43.69	0.0404	0.0046	0.1153
RB	0.0010	0.0300	3.71	0.0150	0.1530	0.0002	0.0140	1.91	0.0085	0.0002	0.3735
RDSB	0.0170	0.1310	221.60	0.0670	0.1030	0.0095	0.0972	155.13	0.0666	0.0095	0.1051
SG	0.0060	0.0790	95.74	0.0380	0.1293	0.0045	0.0672	81.20	0.0470	0.0045	0.1064
SJ	0.0090	0.0950	38.24	0.0450	0.1407	0.0037	0.0607	19.67	0.0363	0.0037	0.2391
SCH	0.0030	0.0550	270.87	0.0260	0.0840	0.0014	0.0368	291.32	0.0244	0.0014	0.1528
ST	0.0090	0.0930	122.99	0.0500	0.0826	0.0064	0.0797	48.82	0.0474	0.0064	0.1515
SG	0.0120	0.1080	196.68	0.0600	0.1973	0.0062	0.0784	64.88	0.0460	0.0062	0.1032
SSE	0.0090	0.0940	191.07	0.0460	0.2891	0.0050	0.0703	312.05	0.0407	0.0050	0.1439
VG	0.0090	0.0950	236.85	0.0490	0.0100	0.0060	0.0771	186.13	0.0450	0.0060	0.2093

MAE of 0.0303, followed by SVM (0.0361), 0.0431 for DNN model, 0.0432 for RAF model and 3.1913 for ARIMA model.

Coefficient of determination (R^2): This is a measure of how well the model can be explained. The R^2 of RAF, SVM, and DNN is a little higher than that of LSTM in terms of several stocks, but on average, the LSTM model has the highest R^2 of 0.2621, followed by RAF (0.1958) and SVM (0.1886), 0.1518 for DNN model and 0.0699 for ARIMA model. We can see that R^2 of LSTM ranges from 0.1100 to 0.6139, similar to several existing financial researches (Fischer & Krauss, 2018, Gatev, Goetzmann, & Rouwenhorst, 2006). To be specific, the main purpose of the preselection phase is to forecast the return of assets and select assets with higher potential returns. Unlike researches on explanatory modelling aiming to explain causal relationships and the importance of each indicator, predictive modelling is primarily concerned with accuracy and error in order to predict future observations (Gandhmal & Kumar, 2019, Shmueli, 2010). In this case, the effectiveness of this kind of model is primarily determined by accuracy measures, such as RMSE and MSE, rather than the value of R^2 (Alexander, Tropsha, & Winkler, 2015, Gandhmal & Kumar, 2019).

Concerning stock market prediction, MSE, RMSE, MAPE, and MAE are generally regarded as popular performance metrics since they can clearly present the average model prediction error (Gandhmal & Kumar, 2019, Kao, Chiu, Lu, & Yang, 2013, Weng et al., 2018). For several other works, it is difficult to evaluate these metrics through direct comparison due to the difference in datasets. But we can compare the results with widely used methods in related researches. From Tables 3 to 5, the average values of MSE, RMSE, MAE for LSTM model are 0.0033, 0.0543 and 0.0303 respectively, which have showed superior performance in forecasting stock returns against existing works (Gandhmal & Kumar, 2019, Sadaei, Enayatifar, Lee, & Mahmud, 2016, Ticknor, 2013, Weng et al., 2018).

In conclusion, the LSTM model predictions are superior to other baseline methods in both accuracy and direction. And the predicted performance of SVM and RAF is second only to LSTM, but

Table 5
Comparison of prediction performance.

Stock	ARIMA				
	MSE	RMSE	MAPE	MAE	R^2
TES	32.22	5.68	8618.9	5.89	0.2427
AST	171.13	13.08	1222.3	10.37	0.1717
BAR	7.90	2.81	389.45	2.56	0.1399
BP	14.83	3.85	7780.1	3.94	0.1038
BAT	1.67	1.29	200.90	0.80	0.0444
HAL	50.50	7.11	19,703	5.79	0.0805
HH	3.89	1.97	888.75	1.36	0.0772
JM	2.63	1.62	2701.7	1.05	0.0176
LG	3.08	1.75	518.96	1.53	0.1103
MSG	8.75	2.96	639.18	1.72	0.1648
PEA	7.25	2.69	416.42	2.43	0.1239
REL	67.52	2.60	1065.3	2.14	0.1534
RB	2.25	1.50	214.61	1.56	0.2945
RDSB	11.52	3.39	2297.2	1.99	0.0678
SG	3.50	1.87	884.67	1.66	0.0570
SJ	11.70	3.42	795.99	3.54	0.1286
SCH	39.27	6.27	34,652	5.74	0.1363
ST	40.70	6.38	3414.5	6.49	0.3774
SG	51.26	7.16	5028.2	4.07	0.2443
SSE	2.66	1.63	6030.1	1.53	0.0299
VG	1.26	1.12	3880.9	0.86	0.0137

far better than DNN and ARIMA model. Besides, traditional statistics model ARIMA performs worst. For example, for stock TES, the MSE in ARIMA equals to 32.21, which is 5000 times bigger than MSE (0.0031) in LSTM.

4.2. Results analysis in the second stage: optimal portfolio formation

4.2.1. Determination of the portfolio size

Firstly, we analyse the characteristics of portfolios consisting of k assets. Most of the researches corresponding to portfolio formation for individual investors focus on only fewer than 10 assets (Almahdi & Yang, 2017, Kocuk & Cornuéjols, 2018, Tanaka, Guoa, & Turksen, 2000), because holding too many different stocks is hard for an individual investor to manage. Ranguelova (Ranguelova, 2001) indicate that individual investors usually hold three or four stocks in their account on average. Paiva et al., (Paiva, Cardoso, Hanaoka, & Duarte, 2019) discover that the portfolio with seven assets performs better than others with different numbers of assets. Hereby, assuming an individual investor holding less than or equal to 10 assets is realistic. Based on the above discussion, we choose $k \in \{4, 5, 6, 7, 8, 9, 10\}$, and then compare the performance of the model LSTM+MV with the other baseline strategies according to the dimensions annualised standard deviation, annualised mean return, annualised Sharpe ratio, and Sortino ratio before transaction costs.

As can be seen from Fig. 5, there are four subgraphs. Specifically, the Y-axis of four sub-graphs represents mean return, standard deviation, Sharpe ratio, and Sortino ratio, the X-axis of four subgraphs represents the same meaning, that is, different models with different portfolio sizes. From Fig. 5, it is clear that irrespective of the portfolio size k, the LSTM+MV shows greater performance than the other strategies in three dimensions of annualised mean return, Sharpe ratio, and Sortino ratio. To be specific, annualised returns prior to transaction costs are at 0.16, compared to 0.09 for the LSTM+1/N, 0.11 for the SVM+MV, 0.07 for the SVM+1/N, 0.09 for the RAF+MV and 0.06 for RAF+1/N for $k = 8$. For other portfolio sizes, like $k = 10$, the LSTM+MV also achieves the highest mean returns per year. Concerning annualised standard deviation, a risk metric, differences among models are not obvious, the LSTM +MV is on a similar level as the other models, with slightly higher values for $k = 6, 7, 8$, thus we could not distinguish which models are good or bad on this metric easily. In this study, we set risk-free ratio as 0.0125, according to the British treasuring bill rate in recent 10 years. With respect to Sharpe ratio, return per unit of risk, is highest for the LSTM+MV. For example, when $k = 9$, Sharpe ratio before transaction cost is 0.58, compared to 0.40 for the LSTM+1/N, 0.46 for the SVM+MV, 0.36 for the SVM+1/N, 0.38 for the RAF+MV, 0.28 for RAF+1/N. Sortino ratio, measuring the risk-adjusted return of an investment portfolio. A clear advantage of the LSTM+MV can be seen for portfolios of each size. From the perspective of different portfolio sizes, it is easy to find that the four indicators perform better overall in each model when $k = 10$ than other sizes. Specifically, in model LSTM+MV, the portfolio with $k = 10$ not only has a high mean return 0.136, Sharpe ratio 0.58 and Sortino ratio 13.7, but also has a lower standard deviation of 0.21. And the same is true for the analysis of other models. From the above analysis, we focus the portfolio with $k = 10$ in our subsequent analyses, which is also consistent with the research of Fischer and Krauss (Fischer & Krauss, 2018).

4.2.2. Details on financial performance

It is worth clarifying that this paper only considers brokerage cost as transaction cost because the investor is able to control brokerage cost directly (Paiva, Cardoso, Hanaoka, & Duarte, 2019). According to Brooks et al. (Brooks, Rew, & Ritson, 2001), brokerage costs for purchasing and selling the stocks of FTSE 100 index is from 0.00 bps to 0.30 bps. Referring to the parameters of several empirical research (Almahdi and Yang, 2016; (Guerard, Markowitz, & Xu, 2015, Paiva, Cardoso, Hanaoka, & Duarte, 2019)), we decide to simulate transaction costs as 0.10 bps, 0.05 bps to present the results finally. Tables 6 to 8 provide insights into the financial performance of the LSTM+MV, compared to the baselines, without transaction cost, including transaction cost (0.1 bps, 0.05 bps) separately. Hence, Panel A, B, and C depict daily return characteristics, daily risk characteristics and annualised risk-return metrics respectively.

Return characteristics: In panel A of Table 6, we can see that the LSTM+MV exhibits favourable daily mean return 0.0005, and the SVM+MV has the lowest standard deviation as 0.0116. After including transaction cost 0.05 bps, in panel A of Table 6, we can find that all the models have almost the same daily return 0.0003, the LSTM+MV model and SVM+1/N model have the lowest standard deviation. After including transaction cost 0.1 bps, in panel A of Table 7, SVM+1/N model has a better risk level, the daily standard deviation equals to 0.0126.

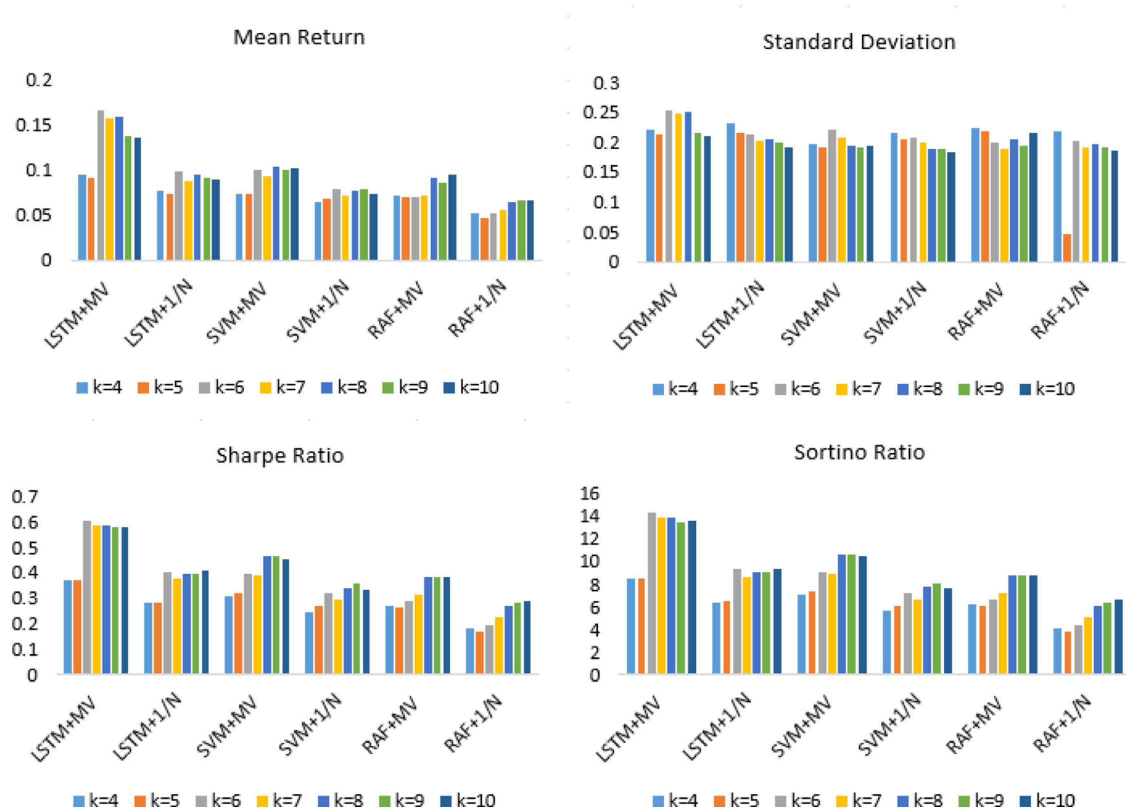

Fig. 5. Annualised performance characteristics for portfolios of different sizes.

Table 6
Performance characteristics for portfolios without transaction cost.

		LSTM+MV	LSTM+1/N	SVM+MV	SVM+1/N	RAF+MV	RAF+1/N
A	Mean return	0.0005	0.0004	0.0004	0.0003	0.0004	0.0003
	Standard deviation	0.0134	0.0121	0.0124	0.0116	0.0137	0.0118
	Maximum	0.1003	0.1100	0.0953	0.0965	0.1241	0.1052
	Minimum	−0.0748	−0.0953	−0.0744	−0.1014	−0.0749	−0.1005
B	1-percent VaR	0.0330	0.0337	0.0336	0.0327	0.0385	0.0328
	1-percent CVaR	0.0439	0.0446	0.0451	0.0427	0.0514	0.0443
	5-percent VaR	0.0207	0.0189	0.0188	0.0178	0.0207	0.0188
	5-percent CVaR	0.0306	0.0282	0.0285	0.0271	0.0318	0.0277
	Maximum drawdown	2.5277	2.4182	2.9685	2.2441	2.5612	2.1990
C	Mean return	0.1367	0.0913	0.1022	0.0743	0.0963	0.0676
	Standard deviation	0.2125	0.1919	0.1963	0.1844	0.2176	0.1878
	Sharpe ratio	0.5845	0.4105	0.4569	0.3354	0.3852	0.2932
	Sortino ratio	13.7078	9.3844	10.4918	7.6352	8.8549	6.6693

Table 7
Performance characteristics for portfolios including transaction cost (0.05 bps).

		LSTM+MV	LSTM+1/N	SVM+MV	SVM+1/N	RAF+MV	RAF+1/N
A	Mean return	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
	Standard deviation	0.0125	0.0139	0.0155	0.0125	0.0139	0.0134
	Maximum	0.1152	0.1397	0.1535	0.1031	0.1186	0.1302
	Minimum	−0.1027	−0.1746	−0.2575	−0.1322	−0.1387	−0.1698
B	1-percent VaR	0.0341	0.0368	0.0408	0.0337	0.0384	0.0365
	1-percent CVaR	0.0472	0.2119	0.0598	0.1897	0.0529	0.2047
	5-percent VaR	0.0196	0.0210	0.0232	0.0196	0.0209	0.0202
	5-percent CVaR	0.0293	0.0424	0.0352	0.0379	0.0321	0.0409
	Maximum drawdown	2.3442	2.5068	2.7753	2.9920	3.6550	2.3043
C	Mean return	0.0765	0.0789	0.0792	0.0780	0.0691	0.0630
	Standard deviation	0.1988	0.2203	0.2462	0.1988	0.2207	0.2129
	Sharpe ratio	0.3218	0.2978	0.2710	0.3294	0.2567	0.2374
	Sortino ratio	0.0906	0.0834	0.0753	0.0926	0.0720	0.0667

Table 8
Performance characteristics for portfolios including transaction cost (0.1 bps).

		LSTM+MV	LSTM+1/N	SVM+MV	SVM+1/N	RAF+MV	RAF+1/N
A	Mean return	0.0003	0.0003	0.0003	0.0003	0.0002	0.0003
	Standard deviation	0.0149	0.0140	0.0153	0.0126	0.0155	0.0135
	Maximum	0.1524	0.1411	0.1445	0.1041	0.1658	0.1314
	Minimum	−0.2124	−0.1762	−0.2340	−0.1334	−0.2541	−0.1714
B	1-percent VaR	0.0388	0.0372	0.0405	0.0340	0.0408	0.0368
	1-percent CVaR	0.0582	0.0212	0.0600	0.1915	0.0612	0.2066
	5-percent VaR	0.0222	0.2140	0.0228	0.0198	0.0229	0.0204
	5-percent CVaR	0.0344	0.0428	0.0352	0.0383	0.0356	0.0413
	Maximum drawdown	2.7861	2.5068	3.0653	2.9920	2.5326	2.3043
C	Mean return	0.0763	0.0787	0.0705	0.0787	0.0616	0.0636
	Standard deviation	0.2366	0.2224	0.2434	0.2007	0.2468	0.2149
	Sharpe ratio	0.2697	0.2975	0.2384	0.3300	0.1990	0.2379
	Sortino ratio	0.0752	0.0833	0.0662	0.0924	0.0553	0.0665

Fig. 6. cumulative return without transaction cost.

Risk characteristics: In panel B of Table 6, Table 7 and Table 8, we can see a mixed picture corresponding to risk characteristics. Before transaction cost, SVM+1/N achieved the best place with a 1-percent VaR of 0.0327, 5-percent VaR of 0.0178, 1-percent CVaR of 0.0427 and 5-percent CVaR of 0.0271. After including transaction cost 0.05 bps, the LSTM+MV performs better, with 1-percent CVaR of 0.0472, 5-percent VaR of 0.0196 and 5-percent CVaR of 0.0293. After including transaction cost 0.1 bps, in terms of 1-percent VaR, SVM+1/N model has the lowest value. LSTM+1/N achieves the lowest 1-percent VaR, SVM+1/N performs best for 5-percent VaR.

Annualised risk-return metrics: In panel C of Table 6, Table 7 and Table 8, we discuss risk-return metrics on an annualised basis. It is clear that the LSTM+MV achieves the highest annualised returns of 0.1367 without transaction costs, followed by the SVM+MV (0.1022). SVM+MV and SVM+1/N perform best in terms of annualised mean returns with transaction cost 0.05 bps and 0.1 bps. The Sharpe ratio measures excess return using standard deviation and can be explained as the return per unit of risk. We find that the LSTM+MV achieves the highest level of 0.5845, with the SVM+MV coming in second with 0.4569. After transaction cost 0.1 bps and 0.05 bps, SVM+1/N gets the highest Sharpe ratio at 0.3294 and 0.3300 respectively. In addition, SVM+1/N achieves the first place in terms of standard deviation and Sortino ratio, followed by LSTM+MV (0.05 bps) and LSTM+1/N (0.1 bps) respectively.

From a financial perspective, we can find that the LSTM+MV, SVM+MV, LSTM+1/N and SVM+1/N outperform the RAF+MV and RAF+1/N in terms of the return, risk or risk-return metrics. In order to compare these models further, we are thus able to choose these four more competitive strategies to visualize performance over time, i.e., from March 1994 to March 2019.

4.2.3. Visualization on financial performance

In this section, we select 4 models, LSTM+MV, SVM+MV, LSTM+1/N, and SVM+1/N, that perform better in the previous section to display their performance for further comparisons. Besides, we also consider Random+MV and Random+1/N as comparison models to examine the necessity of using machine learning for asset pre-selection and further verify whether our proposed method is effective compared with other portfolio data sets. Fig. 6 presents the cumulative return for each model without transaction cost. The LSTM+MV model has an obviously higher result and achieves

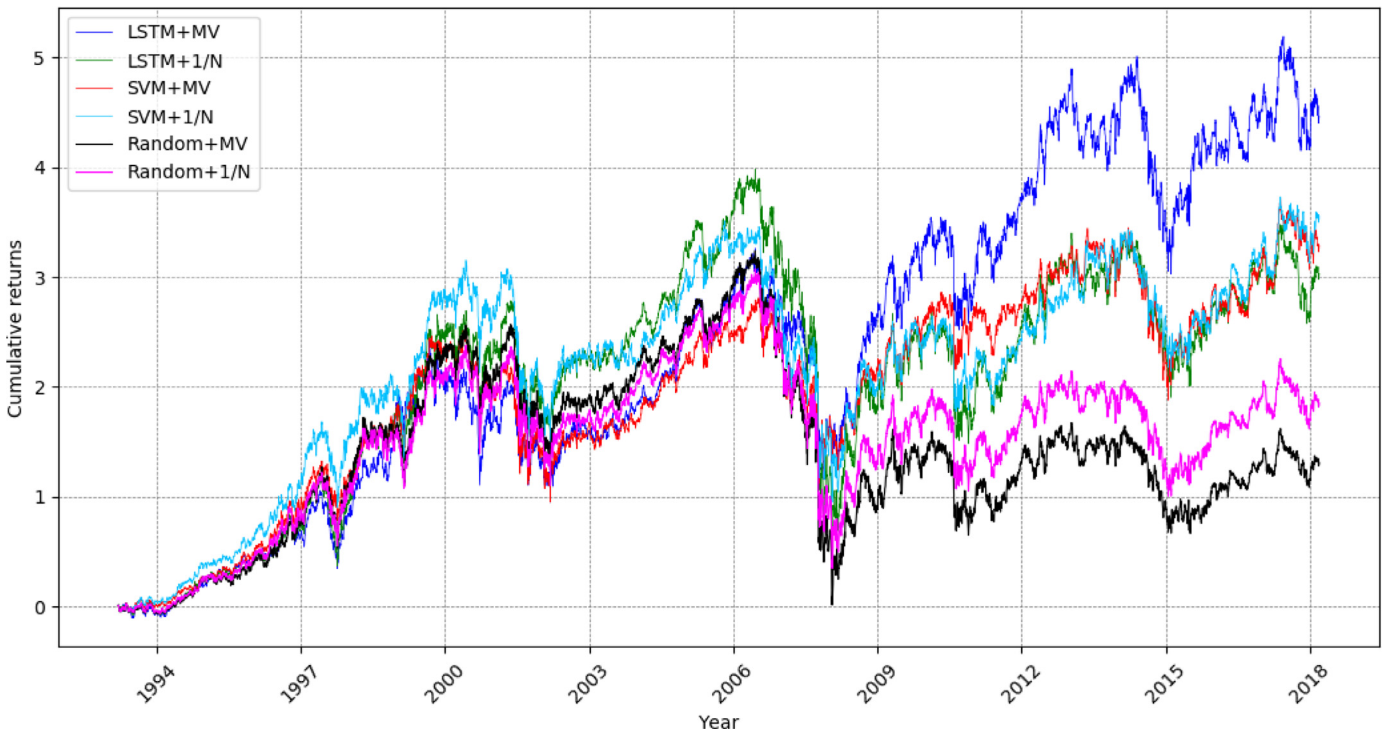

Fig. 7. cumulative return including transaction cost (0.05 bps).

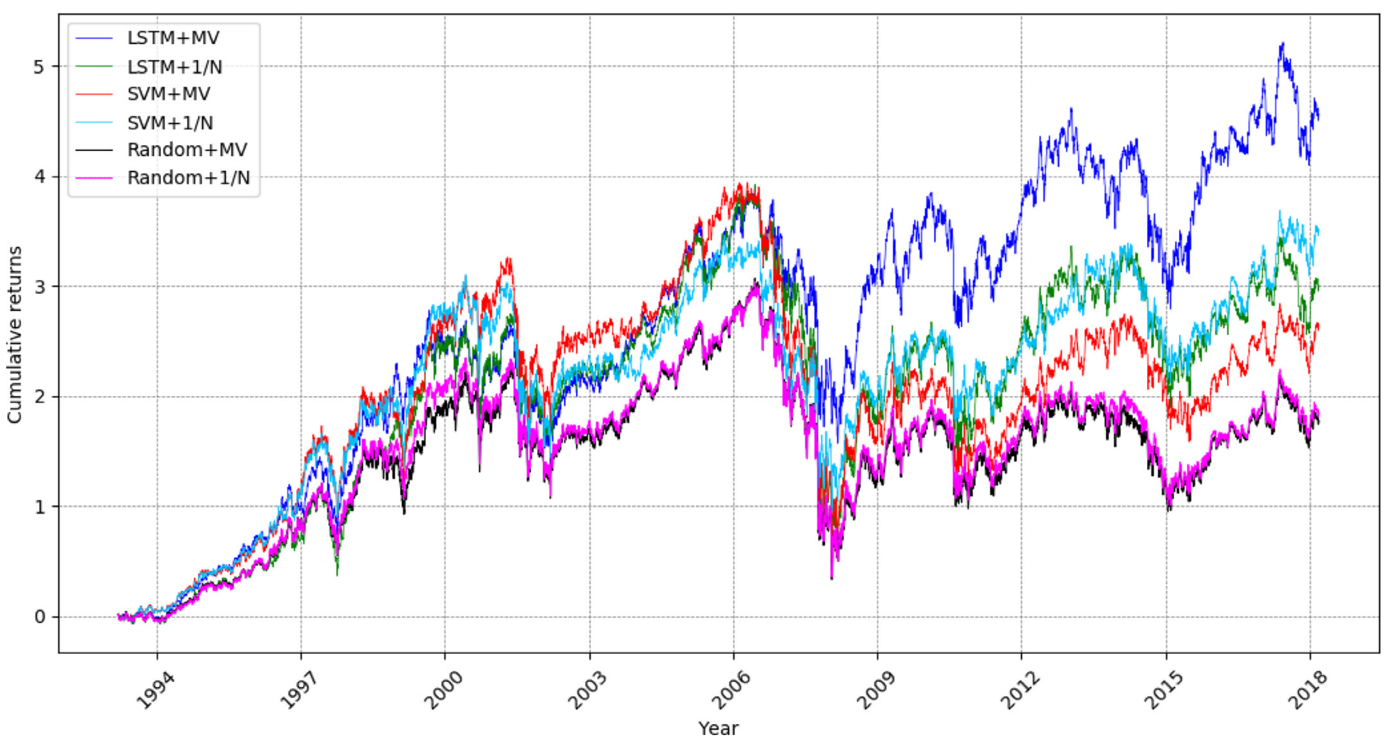

Fig. 8. cumulative return including transaction cost (0.1 bps).

cumulative return of 15.9 approximately. The profitability of the LSTM+1/N model follows, with 5.7, and then the SVM+MV, with 5.5. And the Random+MV and the SVM+1/N keep similar at about 3.3, the Random+1/N is the lowest, at 2.5. Furthermore, we should also figure out how LSTM+MV and other models behave at different levels of transaction costs.

Figs. 7 and 8 depict the simulations of the cumulative returns considering transaction costs of 0.05 bps and 0.10 bps, respectively, and the accumulated returns are strongly decreased. But in general, the LSTM+MV model still maintains a better accumulated return. The cumulative return with a transaction cost of 0.05 bps is about 4.6, while for a transaction cost of 0.10 bps it is 4.5.

From the comparison of the cumulative returns between the LSTM+MV model and the other baseline strategies, we can discover that LSTM+MV performs much better than other baselines in terms of return metrics. Another idea which is inspired by this

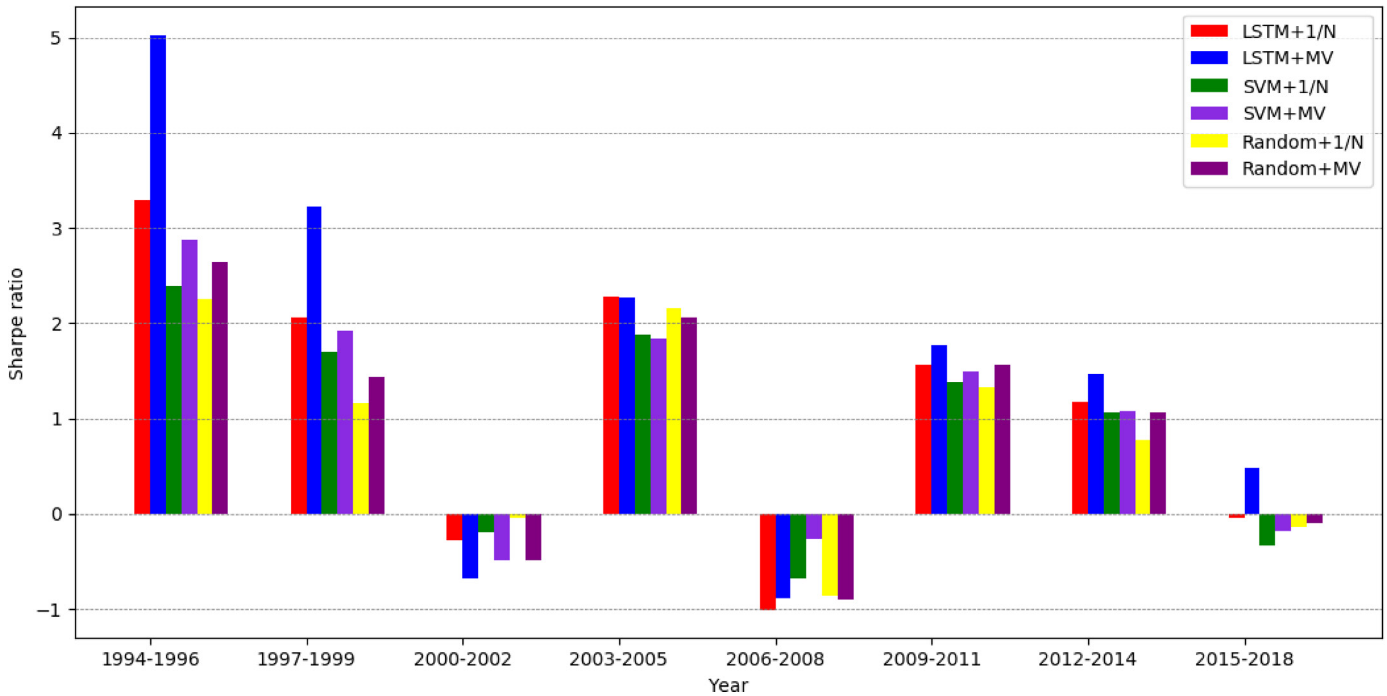

Fig. 9. Sharpe ratio of each triennium without transaction costs.

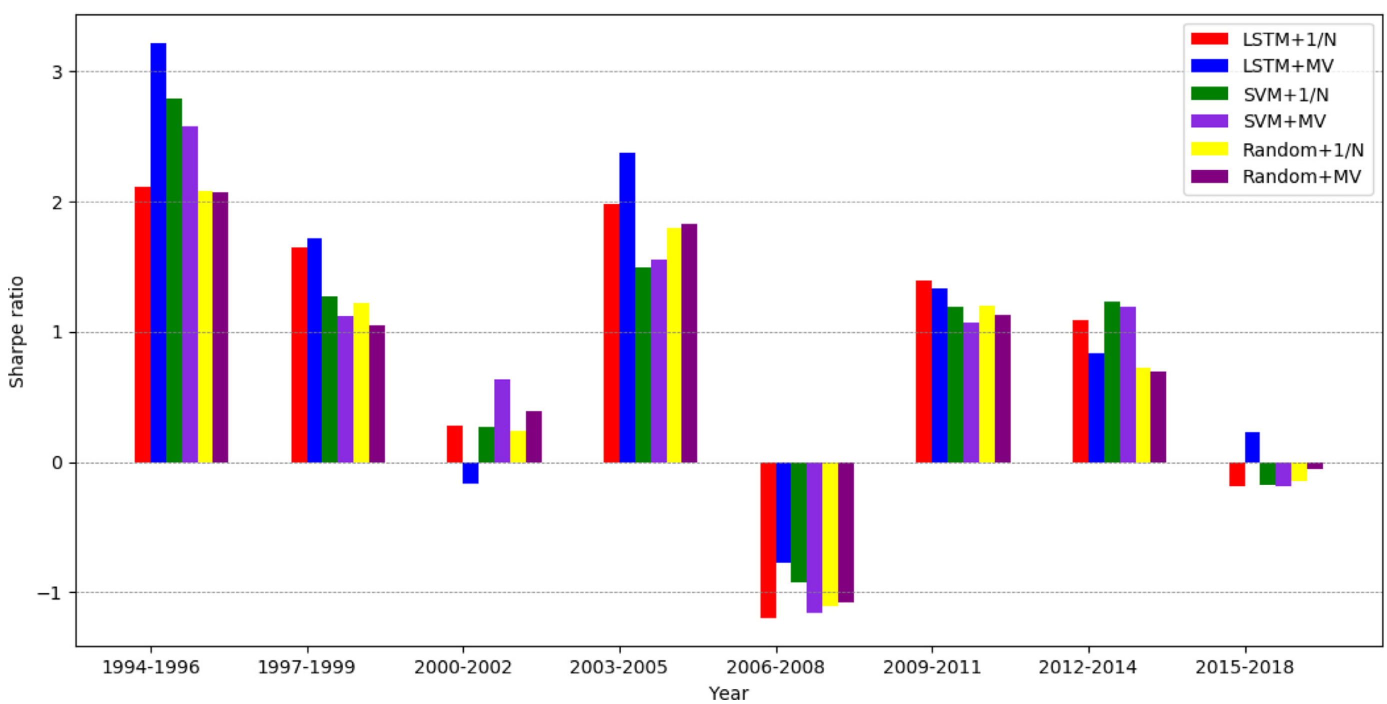

Fig. 10. Sharpe ratio of each triennium including transaction costs (0.05 bps).

is that we would like to see the results when integrating risks and whether the good performance only occurs during a certain period time. As shown in Fig. 9, we use the Sharpe ratio performance of each model every three years. We can observe that, of the eight surveyed triennia, six of them show that the Sharpe ratio of the LSTM+MV model has a better result than other models during the corresponding periods. Figs. 10 and 11 present the Sharpe ratio per triennium with transaction costs. The LSTM+MV model, with transaction costs of 0.05 bps, behaves better. Specifically, among the eight surveyed triennia, five of them have a higher Sharpe ratio in LSTM+MV model than other models. After including transaction cost 0.1 bps, only half of the surveyed period shows a greater result of the LSTM+MV model.

Fig. 12 depicts the result of average return to the risk per month of each triennium per model without transaction costs. Apparently, the LSTM+MV model obtains a remarkable performance for the return-risk ratio during the most study period. We also discover the average results as followings: 0.2670 for the LSTM+MV model, 0.1966 for the LSTM+1/N model, 0.1808 for the SVM+MV, 0.1581 for the SVM+1/N model, 0.1593 for the Random+MV, and 0.1458

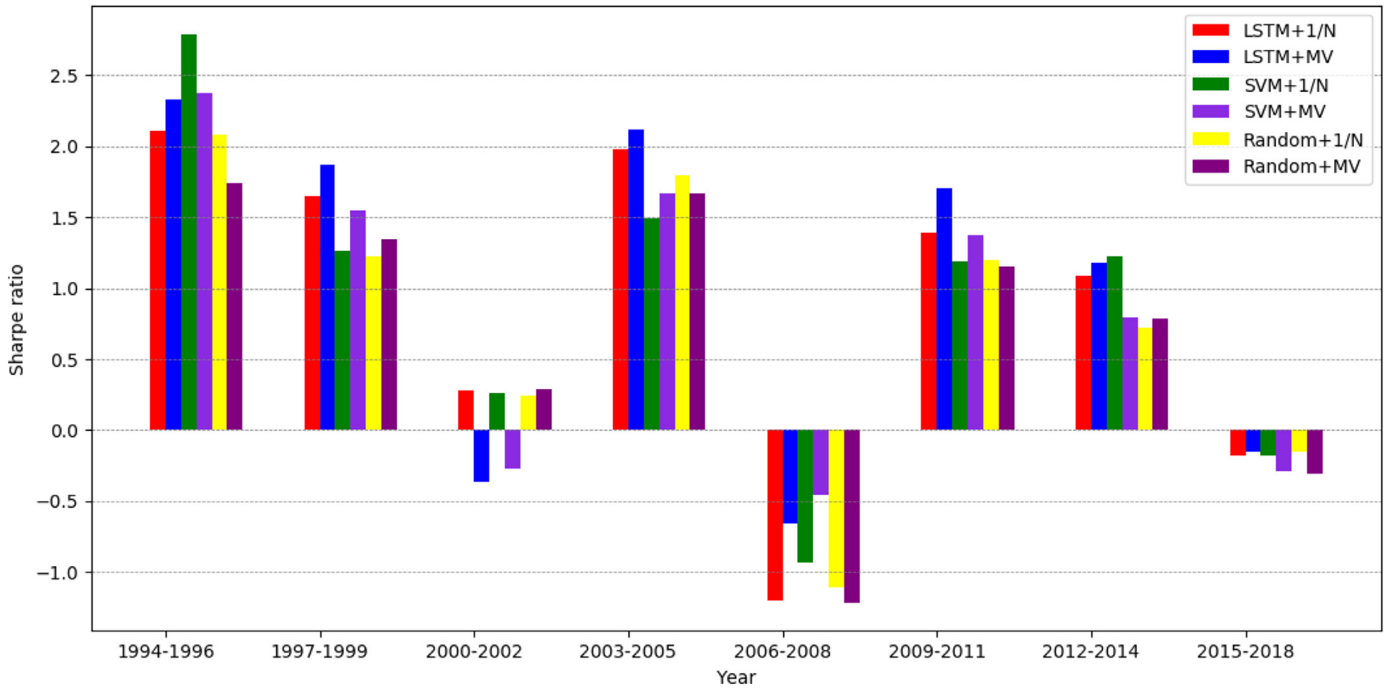

Fig. 11. Sharpe ratio of each triennium including transaction costs (0.1 bps).

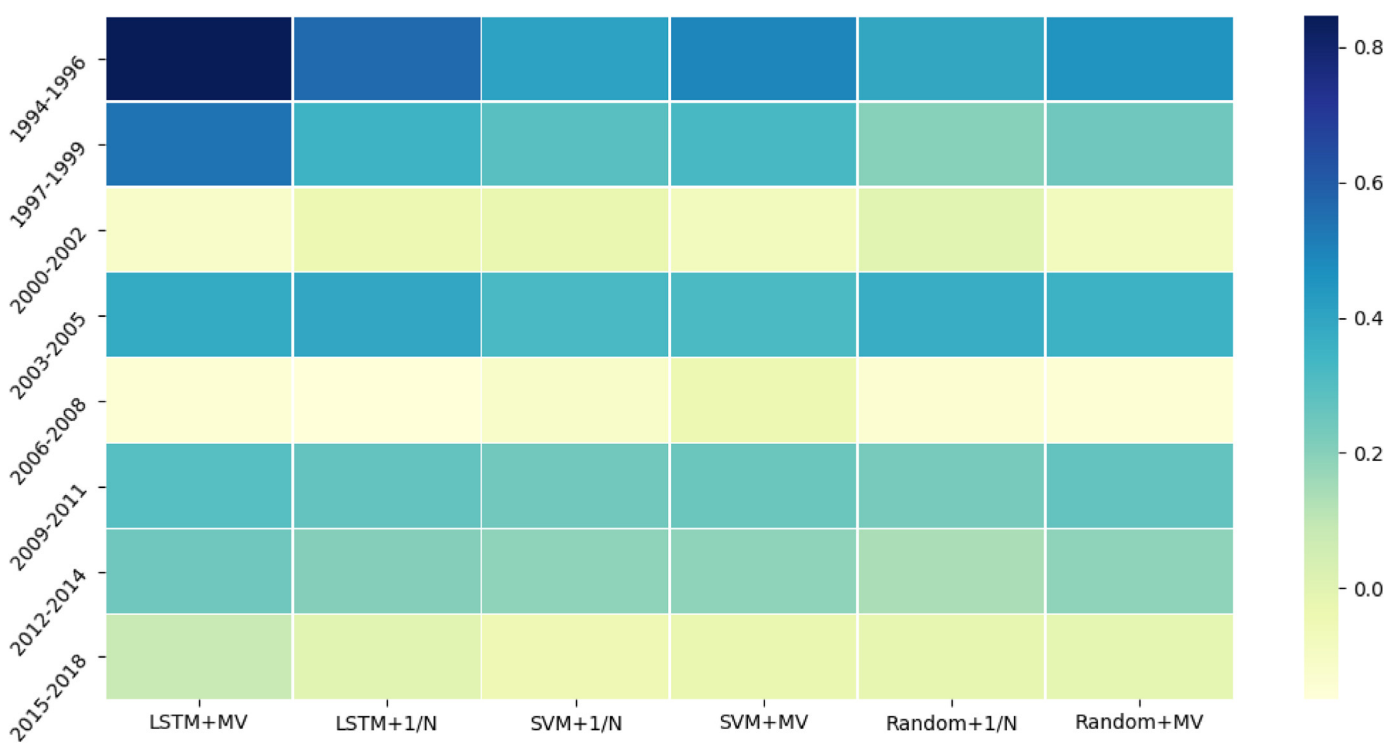

Fig. 12. Average return to the risk per month of each triennium without transaction costs.

for the Random+1/N model. The LSTM+MV model stops having the highest value during the period 2006–2008, and this result coincides with the financial crisis and troubled political.

5. Discussion and conclusions

5.1. Discussion for key findings

This paper puts forward an investment decision model entitled LSTM+MV. Based on the LSTM method, predict and select assets with a higher daily return of gain, then integrate this prediction with the MV diversification method to compose the optimal portfolio. Our study results in several important findings.

First of all, LSTM networks are applied to achieve the financial time-series prediction empirical application on big data volume. Specifically, we create an appropriate prediction task, divide whole sample data set into 22 overlapping training-testing sets, normalize the input features in order to facilitate model training, find an appropriate LSTM architecture for forecasting. After comparing the outcomes of the LSTM against SVM, RAF, DNN as well as ARIMA, we discover the LSTM networks are appropriate for financial time-series forecasting, to beat the other early machine learning models and the statistics model by a very clear margin.

Secondly, for individual investors, holding 10 assets is realistic and helps them maintain better returns with the same level of risk. In this case, the LSTM+MV, SVM+MV, LSTM+1/N, and SVM+1/N outperform the Random+MV and Random+1/N in terms of the return, risk or risk-return metrics. Among these results, we further display their performance in accordance with cumulative return per year, Sharpe ratio per triennium as well as average return to the risk per month of each triennium.

Finally, for cumulative return performance without transaction costs, the LSTM+MV model is significantly better than the other baseline models. A three-year Sharpe ratio experiment also confirms the better performance of the LSTM+MV model. After including transaction costs, the LSTM+MV model still outperforms the other models with a better outcome. In that case, the applicability of the model's implementation may depend on the amount of money invested by investors.

5.2. Theoretical implications

This research enriches the theoretical literature on stock return prediction and portfolio management. First of all, the portfolio formation method proposed in this paper is able to capture the long-term dependences of financial time-series data fluctuation, which fills the gap in corresponding portfolio optimisation researches paying insufficient attention to the continuity and memory characteristics of financial time-series data. To be specific, this paper compares the forecasting outcomes of the LSTM with SVM, RAF, DNN as well as ARIMA to demonstrate the accuracy and feasibility of LSTM networks in predicting financial time-series more convincingly.

Second, the preselection process of assets is incorporated into optimal portfolio formation. Instead changing and improving the Markowitz' MV model, this paper puts effort into the preliminary phase of portfolio construction to ensure that the portfolio is composed of assets with high-return in the beginning. Specifically, our study demonstrates that the proposed model LSTM+MV is able to help individual investors obtain remarkable outcomes for the cumulative returns as well as risk-adjusted return for the majority of periods. The merger of the return forecasting and portfolio optimisation processes may provide a new perspective for research in fintech area.

5.3. Practical implications

The study also provides several practical implications. For portfolio managers, this paper puts forward a practical method for optimal portfolio selection that can help improve day investments. Following this model, managers can pick assets with higher return based on the predicting results in the real market, and then apply MV model to reduce risk level so that keep investments safe and beneficial. For individual investors, this method is able to systematically help them to make decisions for investing. In another word, tell them which assets they should hold and how much to invest in each asset to achieve the goal of maximal potential return with minimal risk.

5.4. Limitations and future work

Although this research provides useful insights, there are some limitations in this study, which provide opportunities for further research. First, five technical indicators and fifteen lagged variables are used as input features to predict the return in the future, however, there are some other external environmental factors, such as government policies, interest rates, public events and so forth that have an impact on financial market can also be considered as the input indicators to the models (Christou, Cunado, Gupta, & Hassapis, 2017). In addition, the study uses the asset data in only one country of the UK. Due to the different political environment and economic backgrounds, we cannot ensure whether the proposed method is suitable for the stock markets from other countries. Thus, in future research, asset data from more countries should be used for experiments and comparisons to further testify the applicability and establish the boundaries of the proposed model.

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.

Credit authorship contribution statement

Wuyu Wang: Conceptualization, Writing - original draft, Formal analysis. **Weizi Li:** Supervision. **Ning Zhang:** Writing - review & editing. **Kecheng Liu:** Conceptualization, Writing - review & editing.

Acknowledgements

Wuyu Wang and Kecheng Liu conceptualised the research. Under the guidance of Weizi Li, Wuyu Wang collected and analysed the data, and conducted the experiment. Wuyu Wang wrote the paper; and Kecheng Liu and Ning Zhang reviewed and edited the paper.

References

Aggarwal, S., & Aggarwal, S. (2017). Deep investment in financial markets using deep learning models. *International Journal of Computer Applications, 162*(2), 0975–8887.

Alexander, D. L. J., Tropsha, A., & Winkler, D. A. (2015). Beware of R^2: Simple, unambiguous assessment of the prediction accuracy of qsar and qspr models. *Journal of Chemical Information and Modeling, 55*, 1316–1322.

Almahdi, S., & Yang, S. Y. (2017). An adaptive portfolio trading system: A risk-return portfolio optimisation using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems With Applications, 87*, 267–279.

Arévalo, A., Niño, J., Hernández, G., & Sandoval, J. (2016). High-frequency trading strategy based on deep neural networks. *International conference on intelligent computing*, 424–436.

Baek, Y., & Kim, H. Y. (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention lstm module and a prediction LSTM module. *Expert Systems With Applications, 113*, 457–480.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one, 12*(7), 1–24.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, (3), 281–305.

Bodnar, T., Mazur, S., & Okhrin, Y. (2017). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research, 256*, 292–307.

Bodnar, T., Parolya, T., & Schmid, W. (2018). Estimation of the global minimum variance portfolio in high dimensions. *European Journal of Operational Research, 266*, 371–390.

Brooks, C., Rew, A. G., & Ritson, S. (2001). A trading strategy based on the lead-lag relationship between the spot index and futures contract for the ftse 100. *International Journal of Forecasting, 17*, 31–44.

Brown, D. B., & Smith, J. E. (2011). Dynamic portfolio optimisation with transaction costs: Heuristics and dual bounds. *Management Science, 57*(10), 1752–1770.

Chen, Y. J., & Hao, Y. J. (2018). Integrating principle component analysis and weighted support vector machine for stock trading signals prediction. *Neurocomputing*, (321), 381–402.

Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems With Applications*, (83), 187–205.

Chourmouziadis, K., & Chatzoglou, P. D. (2016). An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. *Expert Systems with Applications*, (43), 298–311.

Christou, C., Cunado, J., Gupta, R., & Hassapis, C. (2017). Economic policy uncertainty and stock market returns in Pacificrim countries: Evidence based on a bayesian panel var model. *Journal of Multinational Financial Management*, (40), 92–102.

Deng, S. J., & Min, X. (2013). Applied optimization in global efficient portfolio construction using earning forecasts. *Journal of Investing, 22*(4), 104–114.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*.
Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems, 9*, 155–161.
Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*, 654–669.
Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review, 34*, 100190.
Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies, 19*(3), 797–827.
Grauer, R. R., & Hakansson, N. H. (1993). On the use of mean-variance and quadratic approximations in implementing dynamic investment strategies: A comparison of returns and investment policies. *Management Science, 39*(7), 856–871.
Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks, 18*(5), 602–610.
Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems, 28*(10), 2222–22232.
Guerard, J. B., Jr, Markowitz, H., & Xu, G. L. (2015). Earnings forecasting in a global stock selection model and efficient portfolio construction and management. *International Journal of Forecasting, 31*, 550–560.
Hinton, G. (2009). Deep belief networks. *Scholarpedia, 4*(5), 5947.
Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science (New York, N.Y.), 313*(5786), 504–507.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735–1780.
Huang, C. F. (2012). A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing, 12*, 807–818.
Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). An empirical exploration of recurrent network architectures. In *Proceedings of the 32nd International Conference on Machine Learning* (pp. 2342–2350).
Kao, L. J., Chiu, C. C., Lu, C. J., & Yang, J. L. (2013). Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing, 99*, 534–542.
Kara, Y., Boyacioglu, M. A., & Baykan, O. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the istanbul stock exchange. *Expert Systems with Applications*, (38), 5311–5319.
Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, arXiv:1412.6980.
Kocuk, B., & Cornuéjols, G. (2018). Incorporating black-litterman views in portfolio construction when stock returns are a mixture of normals, *omega*, In press.
Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 Years of portfolio optimisation: Practical challenges and current trends. *European Journal of Operational Research, 356*-(234), 371.
Kraus, M., & Feuerriegel, S. (2017). Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems*, (104), 38–48.
Lachiheba, O., & Gouiderb, M. S. (2018). A hierarchical deep neural network design for stock returns prediction. *Procedia Computer Science*, (126), 264–272.
Längkvist, M., Karlsson, L. &., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time - series Modeling. *Pattern Recognition Letters*, (42), 11–24.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
Li, J., & Xu, J. (2013). Multi-objective portfolio selection model with fuzzy random returns and a compromise approach-based genetic algorithm. *Information Sciences, 220*, 507–521.
Li, T., Zhang, W., & Xu, W. (2013). Fuzzy possibilistic portfolio selection model with var constraint and risk-free investment. *Economic Modelling, 31*, 12–17.
Li, T., Zhang, W., & Xu, W. (2015). A fuzzy portfolio selection model with background risk. *Applied Mathematics and Computation, 256*, 505–513.
Li, Y., Yuan, Y. (2017). Convergence analysis of two-layer neural networks with relu activation. *Neural Information Processing Systems Conference* 1–46.
Liu, H., & Loewenstein, M. (2002). Optimal portfolio selection with transaction costs and finite horizons. *The Review of Financial Studies Summer, 15*(3), 805–835.
Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, (164), 163–173.
Loureiro, A. L. D., Miguéis, V. L., & Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems, 114*, 81–93.
Malkiel, B. G. (2007). A random walk down wall street: The time-tested strategy for successful investing WW Norton & Company.
Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77–91.
Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics, 51*(3), 247–257.
Paiva, F. D. A, Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision–making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, (115), 635–655.
Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, (42), 259–268.
Pektas, A. O., & Cigizoglu, H. K. (2013). ANN hybrid model versus ARIMA and ARIMAX models of runoff coefficient. *Journal of Hydrology, 500*, 21–36.
Ranguelova, E. (2001). Disposition effect and firm size: New evidence on individual investor trading activity. *SSRN eLibrary.*.
Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications, 42*, 3234–3241.
Reimers, N., & Gurevych, I. (2017). Optimal hyperparameters for deep LSTM-Networks for sequence labeling tasks. *CoRR*, arXiv:1707.06799.
Sadaei, H. J., Enayatifar, R., Lee, M. H., & Mahmud, M. (2016). A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting. *Applied Soft Computing, 40*, 132–149.
Santos, A. A. P., & Tessari, C. (2012). Quantitative portfolio optimisation techniques applied to the Brazilian stock market. *Revista Brasileira de Finanças, 10*, 369–394.
Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science, 9*(2), 277–293.
Shen, F., Chao, J., & Zhao, J. (2015). Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing, 167*, 243–253.
Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310..
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.
Tanaka, H., Guoa, P., & Turksen, B. (2000). Portfolio selection based on fuzzy probabilitiesand possibility distributions. *Fuzzy Sets and Systems, 111*, 387–397.
Ticknor, J. L. (2013). A bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications, 40*(14), 5501–5506.
Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies, 25*(2), 65–86.
Tu, J., & Zhou, G. (2010). Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertain. *Journal of Financial and Quantitative Analysis, 45*(4), 959–986.
Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short--term stock prices using ensemble methods and online data sources. *Expert Systems With Applications, 112*, 258–273.
Yunusoglu, M. G., & Selim, H. (2013). A fuzzy rule based expert system for stock evaluation and portfolio construction: An application to istanbul stock exchange. *Expert Systems with Applications*, (40), 908–920.
Zbikowski, K. (2015). Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy. *Expert Systems with Applications, 42*, 1797–1805.
Zhang, Y., Li, X., & Guo, S. (2018). Portfolio selection problems with Markowitz's mean-variance framework: A review of literature. *Fuzzy Optimal Decision Making, 17*, 125–158.

Wuyu Wang is a PhD candidate in Applied Economics in School of Information, Central University of Finance and Economics, China. She is also a visiting PhD researcher in Henley Business School, University of Reading, UK. Her research interest is related to financial technology and sharing economy.

Weizi Li, Associate Professor of Informatics, Deputy Director in Informatics Research Centre, Henley Business School, University of Reading, UK. Her research focuses on digital health, integrated systems, artificial intelligence and machine learning applications.

Ning Zhang received her PhD from Beijing Institute of Technology, China. She is currently a professor at Central University of Finance and Economics. Her research focuses on personal privacy protection and financial technology.

Kecheng Liu is a full professor and holds a chair of Applied Informatics, Henley Business School, University of Reading, UK. His research interests and publications span from organisational semiotics, applied informatics in business, digital strategy and architecture, digital business ecosystems and transformation, and intelligent spaces for working and living.