Forecasting Stock Prices Changes Using Long-Short Term Memory Neural Network with Symbolic Genetic Algorithm

Qi Li
University of Technology Malaysia
Norshaliza Kamaruddin (norshaliza.k@utm.my)
University of Technology Malaysia
Hamdan Amer Ali Al-Jaifi
Taylor's University

Article

Keywords: symbolic genetic algorithm (SGA), deep learning, multilayer perceptron (MLP), Long-Short Term Memory Neural Network (LSTM), cross-sectional stock return prediction, feature engineering

Posted Date: August 30th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3284486/v1

Abstract

This paper presents an enhanced Long-Short Term Memory Neural Network (LSTM) framework that combines Symbolic Genetic Algorithm (SGA) to predict cross-sectional price returns for 4500 listed stock in China from 2014 to 2022. Using the S&P Alpha Pool Dataset for China, the framework incorporates data augmentation and feature selection techniques. The study demonstrates significant improvements in Rank Information coefficient (Rank IC) and IC information ratio (ICIR) by 1128% and 5360% respectively when applied to fundamental indicators. For technical indicators, the hybrid model achieves a 206% increase in Rank IC and an impressive surge of 2752% in ICIR. Furthermore, a simple rule-based strategy based on the proposed hybrid SGA-LSTM model outperforms major Chinese stock indexes, generating average annualized excess returns of 31.00%, 24.48%, and 17.38% compared to the CSI 300 index, CSI 500 index, and the average portfolio, respectively. These findings highlight the effectiveness of LSTM with SGA in optimizing the accuracy of cross-sectional stock return predictions and provide valuable insights for fund managers, traders, and financial analysts.

1. Introduction

Predicting the Stock return is a challenging endeavour, given the nonlinear nature of the stock market and the different approaches to predict the stock change. Though, advancements in artificial intelligence and other superior models have been used to increase forecasting accuracy, the prediction accuracy rate is still an unresolved issues [1].

Enormous amount of attention in the empirical asset pricing literature has been directed to answer the questions of what drives the stock prices [2] and what input features play major role in generating accurate results. In early years, researchers started with the price trend itself, using technical indicators and found that technical indicators were efficient in predicting the market in the past [3]. Fama proposed in a weak-form market, people can make abnormal returns by mastering fundamental information, such as financial statements[4]. However, many scholars doubt the financial ratios do not consistently outperform the historical average benchmark forecast out of sample[3]. Few researchers have adopted a fundamental-driven approach and fed financial ratios into machine learning models to beat the market [5]. Besides the above anomalies, more recent evidence also shows that return are predictable by macroeconomic variables as well[6].

In the stock return prediction literature, the mainstream model to predict stock return is through supervised learning [7]. In the early days, parametric statistical models, such as ARMA, ARIMA, and vector autoregression, have been commonly used to explore linear relationships [8]. However, linear statistical models have limitations in capturing the nonlinear nature in financial time series data, and assuming that the series data is non-noisy and has constant variance. Due to such inconsistency, Scholars started to apply non-linear techniques such as machine learning technologies to enhance anomalies. Empirical

results show those results have further been enhanced by applying a diverse set of machine learning approaches[9,10].

Prior studies researchers have compared the prediction results of linear regression, machine learning model with Deep Learning Models (DNN) with the consensus that being that DNN models show significant improvements in forecasting accuracy. The typical DNN models used for stock prediction are CNN, LSTM, GRU, and attention mechanism [11]. Comparing to machine learning, DNN models can extract features from a large set of raw data. However, it often suffers from overfitting and weak generation power and initial feature selection usually must be carried out manually based on domain knowledge [12,13]

Despite DNN network being used as a powerful tool in pattern recognition and price change predictions, there are many drawbacks including the data integration and feature engineering. As there are at least two types of data sources available. However, there is a lack of a framework to integrate all types of information together. In terms of the feature engineering, most of the existing works in this field have limited themselves to a set of technical indicators, and the initial features should be selected manually heavily rely on previous domain knowledge [12,13]

This paper aims to enhance the accuracy of stock return prediction by improving the DNN framework. In our approach, we propose and develop a symbolic genetic programming (SGP) to fill the gap of feature engineering. The SGP is utilized to generate input, replacing traditional feature engineering techniques. In addition, we develop different LSTM models tailored to the characteristics of the dataset. Through this approach, we have achieved remarkable improvements in the Rank Information Coefficient (IC) and Information Ratio of IC (ICIR).

Moreover, we present a hybrid LSTM model integrated with SGP to incorporate all available raw dataset to predict cross-sectional short-term changes in stock prices. This aim to synthesize the findings of our study into a simple and rule-based strategy for a complete active index fund strategy for selecting winning and losing stocks, compared with the benchmark.

Our hybrid model exhibits superior performance compared to the CSI 300 and CSI 500 indexes. Notably, our strategy consistently outperforms these indexes by an average of 31% and 24.48% per year, respectively. Additionally, it surpasses the average returns of the entire market by 17.38% annually. We also calculate the information ratio of the strategy, and it is found that it is 2.49, and this further highlighting its effectiveness.

The remaining sections of this paper are organized as follows: Section 2 will cover related works, including existing DNN models and their combinations with Genetic Algorithms. In Section 3, we provide an in-depth discussion of the methodology, including enhanced SGA for new features, the proposed architecture of the symbolic genetic algorithm (SGA-DNN model), input data descriptions, forecasting horizon, segmentation predictions method and the trading strategy setting. Section 4 will focus on the experiments. Section 5 will cover result and discussion. Finally, Section 5 will conclude the paper.

2. Related Works

Previously, several studies have been conducted from the perspectives of statistical methods and AI techniques in the prediction of the stock return data. The earliest study on applying machine learning in the stock domain can be traced back to 2006, where an accurate event weighting method and an automated event extraction system were presented [14]. The random forest model for technical features is also proved to have the ability to rank the cross-sectional stocks[10]. However, there are several limitations to machine learning models. The challenges come from the employed dataset. Traditional machine learning models are best suited for small or medium-sized datasets and have limitations in processing high-dimensional datasets. They are prone to encountering the curse of dimensionality, especially for big or massive datasets, such as high-frequency or unstructured data [15].

Comparing with machine learning algorithms, the Deep Neural Networks (DNNs) have significant advantages when it comes to handling large sets of time series data. Since 2017, the most state-of-the-art DNN algorithms for stock prediction are sequence models, specifically Recurrent Neural Networks (RNNs). Among the RNNS model, LSTM is the most used model and advantageous over the conventional RNN due to the reason that it overcomes the problems of gradient vanishing or exploding. In 2015, Chen et al. built an LSTM-based model for the China stock market [16]. However, the most referenced paper for LSTM in the application in finance data was done by Thomas Fischer and Benedikt Kraus. They were the first to deploy the LSTM network on large-scale financial time series data and explained the source of the black box, which is high volatility, below-mean momentum, and extremal directional movement related stocks in the recent trading days[17]. Fisher's attempt of LSTM is single LSTM module, and the attributes of overfitting was challenged by other scholars due to the limited availability of data points. Yujin presented a novel data augmentation approach to avoid the overfitting and propose ModAugNet Framework including two modules, one is overfitting prevention LSTM module, and another is prediction LSTM module. The data augmentation approach only acts in overfitting prevention LSTM module [18].

After the success of LSTM and LSTM variants models, many scholars started to combine LSTM or Bi-Directional LSTM and CNN as an integral network. As for CNN part, research attempt to use multi-filter to extract the features map to replace the traditional feature engineering and also keep the LSTM part to catch the sequential trend and pattern.[19–21]. Deep CNN with reinforcement LSTM model was also conducted to obtain better prediction [22].

Besides the single DNN application, the combination of Genetic Algorithm (GA) and Deep Neural Network (DNN) or other Machine Learning models has been utilized by many researchers to improve prediction accuracy. The key factor that drives evolution in Genetic Algorithms (GA) is the fitness function, which is used to evaluate the performance of models. In empirical studies involving stock prediction, Mean Squared Error (MSE) or Pearson correlation is typically used as the objective function for optimization. These metrics are used to measure the accuracy of the model's predictions and guide the search for optimal hyperparameters. For the application of GA in conjunction with Deep Neural Networks (DNNs), two main applications can be observed: hyperparameter tuning and feature selection. Hyperparameter

tuning is a crucial aspect that needs to be addressed in the optimization process, including parameters setting such as the number of layers, nodes per layer, and number of time lags.

GA is frequently employed to search for optimal hyperparameters for DNN. In 2018, Chung and Shin employed GA to identify the optimal number of time lags and LSTM units for hidden layers in stock prediction models [23]. In a similar study in 2019, Chung and Shin optimized the kernel size, kernel window, and pooling window size for CNN [24]. In addition, GA has been used to determine appropriate hyperparameters and input data sizes for Generative Adversarial Networks (GANs) in stock prediction by He and Kita in 2021[25]. These studies demonstrate the effectiveness of GA in optimizing the hyperparameters of various deep learning models for stock prediction.

As for the feature selection, many researchers combine GA and other DNN model to reduce input variables and enhance calculation speed by selecting appropriate factors from a large pool of candidate variables. For instance, Chen and Zhou used GA to rank factor importance and select features for a Long Short-Term Memory (LSTM) model, while Milad employed GA as a heuristic approach for selecting relevant features for an Artificial Neural Network (ANN)[26,27]. Li utilized a multilayer GA to select features and reduce high dimensionality in a stock dividend dataset [28]. Recently, Yun revised GA-based selection methods to a two-stage process, using a wrapper method to select important features to avoid the curse of dimensionality, followed by the filter method to select more critical factors [29].

Tuning a DNN model seems theoretically feasible, but in practice, most DNN models have an excessive number of parameters. Even with the use of genetic algorithms, the computational workload and time required for calculations are immense. As for the features selection, the genetic algorithm described above only helps in reducing the total number of factors. However, if the genetic algorithm can continuously evolve towards effective factors and generate new ones like the symbolic based ‘magic’ factors mentioned before, it could become a more promising direction. In methodology part, the corresponding data augmentation method basing on GA will be proposed.

3. The Proposed Deep Neural Network

In Artificial Intelligence (AI), Deep Neural Network (DNN) falls under the subset of Machine Learning and Neural Network[30]. DNN is based on the artificial neural network (ANN) which contained one or several layers between the input and output layers. In each layer it consists of the same components, and they are neurons, synapses, weights, biases, and functions [31]. Generally, our proposed DNN framework consist of two main phases, and they are data augmentation phase and feature selection phase. Data augmentation phase is responsible in generating or producing the genes or data of the stock return data. In this phase, the Symbolic Genetic Algorithm is utilized to produce the needed data. In this study, we propose to integrate the GA with Symbolic Regression and named it as Symbolic Genetic Algorithm (SGA). Details explanation on SGA will be discussed in section 3.2. While the feature selection phase involved the utilization of Long Short Term Memory method to find the non-linear pattern in order to optimize the accuracy of the stock return prediction. However, in the feature selection phase, we also

experiment the raw data with Multi-Layer Perceptron (MLP). The objective is to observe whether LSTM or MLP could handle the raw data in the extracting the features of the stock price return data. The raw data mentioned here is consisted of the fundamental indicator and the technical indicator. Figure 1 below illustrates the phases involved.

The accuracy of the prediction from phase I and phase II is measured using Mean-square deviation (MSE), Rank Information Coefficient(Rank IC) and Information Ratio of IC (ICIR) as performance metrics for cross-sectional price change prediction, as demonstrated in equations 1 to 3 In the next Section, we present the discussion on the dataset, software and hardware used in this study, as well as the elaboration on phase I, data augmentation and phase II, feature selection.

MSE = $\frac{1}{N}\sum_{t=1}^{N}\left(y_t-\widehat{y_t}\right)^2$ Eq. 1

Rank Information Coefficient(Rank IC) = $\frac{\sum_{i=1}^{n}(Rx_i-\overline{Rx})(Ry_i-\overline{Ry})}{\sqrt{\sum_{i=1}^{n}\left(Rx_i-\overline{Rx}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Ry_i-\overline{Ry}\right)^2}}$ Eq. 2

(R denoted Rank)

$\mathrm{InformationratioofIC\ (ICIR)}=\frac{\mathrm{IC}}{\mathrm{StandadizedDeviationofIC}}$ Eq. 3

3.1 Dataset, Software and Hardware

In this study, two types of data were utilized during the experiments: fundamental indicators and technical indicators. Fundamental indicators comprise data derived from three types of financial statements, namely the balance sheet, profit and loss report, and cash flow report. On the other hand, technical indicators are based on price and volume, providing users with patterns of momentum and reversal. Prior to processing the data using the proposed method, an analysis based on Rank IC was conducted. Rank IC describes the correlation between predicted and actual stock returns, thereby indicating the degree of alignment between the analyst's fundamental and technical forecasts and the actual financial results. The Information Coefficient (Rank IC) is a numerical measure that ranges from 1.0 to -1.0. A value of -1 indicates a perfect negative relationship between the analyst's forecasts and the actual results, while a value of 1 indicates a perfect positive match between the forecasts and the actual results. This metric is highly important when making informed investment decisions, especially in the evaluation of cross-sectional stock returns forecasting. Typically, an information ratio of IC (ICIR) within the range of 0.40 to 0.60, and Rank IC values exceeding 5% in absolute terms, are considered highly favorable in this context.

The data used in this study is dataset of The Alpha Factor Library by S&P Global Market Intelligence[32], which includes explainable factors for all A-listed stocks (around 4500 listed companies) in the Shanghai and Shenzhen Stock Exchange Market, including fundamental and technical indicators. The appendix contains a comprehensive description of both types of quantitative indicators (304) and their corresponding Rank IC values from 2015 to 2022. Table 1(a) presents the average Rank IC (Information

Coefficient) of two specific type of quantitative indicators, while Table 1(b) illustrates the ICIR (Information Coefficient Information Ratio) of these indicators.

Table 1
(a): Rank IC mean of the dataset.

	IC Mean of two types of datasets								
name of datasets	2015	2016	2017	2018	2019	2020	2021	2022	mean
fundamental indicators	0.92%	1.06%	1.63%	0.79%	1.28%	1.36%	1.09%	1.16%	0.65%
technical indicators	4.33%	3.75%	2.34%	2.66%	2.81%	1.99%	3.44%	3.73%	2.82%

Table 1
(b): ICIR mean of the dataset.

	ICIR Mean of two types of datasets								
name of datasets	2015	2016	2017	2018	2019	2020	2021	2022	mean
fundamental indicators	0.14	0.17	0.26	0.12	0.20	0.20	0.14	0.14	0.10
technical indicators	0.34	0.34	0.19	0.22	0.26	0.16	0.27	0.29	0.23

For the data preparation and pre-processing, Python 3.8 was employed along with the numpy and pandas packages. The design of DNN models, including LSTM and MLP, was achieved using KERAS 2.4, a package based on Google TensorFlow 2.4. The Symbolic Genetic Algorithm (SGA) was implemented using the gplearn 0.0.2 package in Python. While the DNN network was trained on NVIDIA GPUs, the remaining models, such as SGA part, were trained on a CPU cluster. Detailed information regarding the software and hardware specifications utilized can be found in Table 2.

Table 2
Descriptions on the software and hardware

Item	Descriptions	Numbers
CPU	Intel(R) Xeon(R) Gold 6248R CPU @ 3.00GHz	96
RAM	503G	
GPU	GeForce RTX 3090	2
System	Ubuntu 20.04.2 LTS	
Python Version	Python 3.8.5	
Keras Version	2.4.3	
gplearn Version	0.0.2	
Tensorflow Version	2.4.0	

The primary objective of this study is to forecast and predict cross-sectional stock price changes. The target variable is defined as the logarithmic return of individual stocks over a specified duration. The study examines standard periods commonly employed for stock predictions, encompassing short-term intervals such as 5 days (one week), 10 days (two weeks), and long-term intervals such as 20 days (one month). In this study, both short-term features for technical indicators and long-term features for fundamental indicators are incorporated. As such, the 5-day forecasting period is selected to explore the accuracy of predictions. By integrating these diverse features, the study endeavours to offer precise forecasts of cross-sectional stock price changes within the designated 5-day period.

3.2 Data Augmentation: Symbolic Genetic Algorithm

The first step of the proposed DNN framework is to investigate the Genetic Algorithm (GA) in the data augmentation phase. Based on literature, Genetic Algorithms are a type of learning algorithm, that would result in a better neural network by crossing over the weights of two good neural networks. This algorithm could also generate and evaluates consecutive generations of humans in order to achieve optimization objectives. The algorithm creates mutation from the stock related indicators by randomly changing the chromosomes or genes of the individual parents. In this situation, GA can be complicated and costly when implemented on the stock related indicators which is nonlinear and having lots of noise or outliers. Therefore, to solve the problem of nonlinear type of data, the Symbolic Genetic Algorithm (SGA) is employed in this study. SGA has several advantages as it evolve by building blocks. In SGA, it employed the regression analysis which is more robust to search the space in finding the best model to fit the given stock return data. Different from GA, SGA find an intrinsic relationship between two or more variables which is hidden. Typically, there are two types of genes that contribute to the generations.

The first type in the study refers to the input features, while the second type represents the processing operators, encompassing mathematical functions like addition, subtraction, division, and multiplication.

Predicting stock price data can be a daunting task, given its complex, dynamic, and non-linear nature. To tackle this challenge, mainstream hedge funds like World Quant, Cubist, and Menelia employ various heuristic operators such as correlation, covariance, and variance. These operators help them analyze and interpret the data, enabling them to make informed investment decisions [33], as depicted in Table 3, to enhance the analysis and prediction of stock price data. In this study, an improved Symbolic Genetic Algorithm (SGA) is proposed, which utilizes symbolic tree expressions to handle and solve complex optimization problems, providing greater flexibility. The four-step approach outlined in Fig. 2 is applied to enhance the performance of the SGA.

Table 3
Heuristic Operators

The Heuristic Operators				
'decay_linear'	'rank_add',	'rank_sub',	rank_mul'	'rank_div'
'ts_max'	'ts_min'	'ts_nanmean'	'ts_prod'	'ts_rank'
'ts_stddev'	'ts_sum'	'ts_corr'	'ts_cov'	'delta'
sign'	'ts_skewness'	'ts_kurtosis'	'ts_max_diff'	'ts_min_diff'
'ts_zscore'	'ts_scale'	'ts_min_max_cps'	'ts_ir'	'ts_median'
'winsorize'	'zscore'	'ts_argmax'	'ts_argmin'	'rank'
'delay'	'sigmoid'	'ts_return'		

The first step in our proposed SGP, is to initiate the population of the genes. Here, we introduce the heuristic operators like the Table 3 shows in the reproduction of the genes. To guide the evolution of the genes, we set certain parameters. For instance, we established a probability of 40% for crossover, which involves exchanging genes between two individuals in the population. Additionally, we set a 40% probability for replacement, which involves copying an individual gene in the population. Finally, we assigned a very low probability for three types of mutation to prevent an excessive influx of new input features, which could lead to unpredictability. This helps maintain stability in the incorporation of new genetic material into the population.

Second, we designed and added rolling windows for all heuristic operators to the original SGA. To this end, we randomly generate rolling window seeds between 3–20 for rolling window to produce additional symbolic formulas. The third step is to design the fitness function. In this study, calculations are performed to determine the fitness target. In addition to using the original Pearson correlation (Rank IC) between the value of the symbolic formula and future price change as the fitness target, a combined formula will be used. This combined formula takes into consideration both the relatively high cumulated return of the bottom group among all cross-sectional stocks and the maintenance of monotonicity in the cumulated return of k groups based on the order of values in the symbolic formula. By incorporating

these factors, the fitness target aims to optimize the performance of the symbolic formula in predicting stock price changes.

The formula is shown from Eq. 4 to Eq. 7 below:

1. Top_R = $\max(TopR - mean\left(totalR\right), FlopR - mean\left(totalR\right))$ Eq. 4
2. Monotonicity = $\max\left(\frac{1}{N}\sum_{k=1}^{N}\max(0, Sign(R_k - R_{k+1}\right)), \frac{1}{N}\sum_{k=1}^{N}\max(0, Sign(R_{k+1} - R_k)\ ))$ Eq. 5
3. Rank Information Coefficient= $\frac{\sum_{i=1}^{n}(Rx_i - \overline{Rx})(Ry_i - \overline{Ry})}{\sqrt{\sum_{i=1}^{n}(Rx_i - \overline{Rx})^2}\sqrt{\sum_{i=1}^{n}(Ry_i - \overline{Ry})^2}}$ Eq. 6
4. Fitness=$\mathrm{Top_R} + \lambda_1 \times \mathrm{Monotonicity} + \lambda_2 \times \mathrm{InformationCoefficient}$ Eq. 7

$$(\mathrm{Default}\lambda_1 = 0.4 \mathrm{Default}\lambda_2 = 2)$$

After obtaining many symbolic formulas based on the above algorithms, the final amendment for SGA is the filter system for the outcomes. The success ratio of Pearson correlation (Rank IC) and the profit and loss ratio(P&L ratio) of Pearson correlation from Eq. 8 to 9 will be employed to select the final synthetic symbolic formulas generated by the SGA model. These above two ratios will also be used as metrics for the experiment part

1. Success Ratio of Rank IC (IC success Ratio) = $\frac{NumbersofCorrectPearsonIC}{Totalnumof PearsonIC}$ Eq. 8
2. Profit and Loss Ratio (IC PNL) = Eq. 9 $\frac{Mean(|PearsonIC|)}{Standarddevation(PearsonIC)}$

3.3 Feature Selection: LSTM vs MLP

The second step in the proposed hybrid DNN framework involves extracting features from the augmented selected data obtained through the SGA process. Feature selection is carried out by creating a Hybrid DNN model that accommodates individual data sources based on their specific characteristics.

Since the development of DNN, the Multiple Layer Perceptron (MLP) was initially introduced as a basic supervised learning algorithm with multiple layers, each consisting of several neurons. However, MLPs have a significant drawback in their ability to handle sequence or time series data effectively. This limitation poses a crucial challenge in stock return forecasting, which heavily relies on the historical states of stocks, following a Markov Chain. To address this issue, a more suitable approach is to utilize the LSTM (Long Short-Term Memory) model, which falls under the category of Recurrent Neural Networks (RNN). LSTMs are specifically designed for sequence modelling tasks and overcome the limitations of MLP. Both LSTM and MLP models are chosen for comparisons, as shown in Fig. 3.

In the first step, the original indicators are either inputted into the SGA model (as depicted in Fig. 3) to obtain selected features, which are then fed to the MLP or LSTM model. Alternatively, the original indicators can be directly fed into the MLP or LSTM model for comparison.

The performance of the four experiments is evaluated using metrics such as Rank IC and ICIR to determine the best model based on the dataset's unique characteristics. The optimization goal for all network settings is to minimize Mean Squared Error (MSE), while the performance quality is assessed using Rank IC and ICIR as metric indicators.

Finally, the trained network is used to recognize feature patterns, and based on the Enhanced SGP-DNN Framework, simple trading rules suitable for the stock market are formulated. These rules are then 'backtested' in stock trading scenarios.

To ensure simulating the real stock investing and considering the 'backtest'[34], the forward rolling window and the segmentation prediction method were followed, the specific details are illustrated in Fig. 4. The whole sample period will be divided into three parts, in the training part, the dataset length is 1020 days which is used to update the model parameters. As for validation part, we use 160 days for tunning and the test part is 20 days and as a result the rolling window is also set as 20 days. The ratio of training set, validation set is taken as 8.5:1 and the real test days is 720 days from 2019-11-30 to 2022-12-31- resulting in a total of 36 non-overlapping trading periods.

3.4 Forecasting, ranking, and trading

The SGA-DNN model utilizes available information prior to time t to forecast the future price change of each stock. Its objective is for each stock to surpass the average price changes observed in the cross-sectional market during the subsequent period t + 1. To achieve this, the model ranks all cross-sectional stocks (4500 in total) in ascending order based on the predicted return by SGA-DNN for the next period. The highest-ranked stocks form the top group, and historically, we have divided the entire cross-sectional stocks into 10 groups, each containing 450 stocks. This ranking score serves as a basis for long only portfolio construction.

Long-Only Portfolio Strategy: The Long-Only Portfolio Strategy focuses on taking long positions in the top k stock portfolios, which are then held for a single period (t + 1). To gauge the effectiveness of this strategy, we will compare its performance against the CSI 300 and CSI 500 benchmarks (denoted as Relative R above 300 and Relative R above 500). These benchmarks represent broad-based indexes in the Chinese stock market. Moreover, we will also consider the average performance of an equal-weighted portfolio as a third performance benchmark (denoted as Relative R above average), the sharp ratio of Relative R above average(Sharp Ratio) will be also measured as the metrics in experiment part.

4. Experiments and Setting

As mentioned earlier, in this study two type of raw data is used to observe the performance of Neural Network. Therefore, we would like to experiment with both the fundamental and technical indicators, to observe whether they produce a different result. During the execution of the experiments, the performance

of the Neural Network model is observed based on two categories; directly processed the data without integrating with SGA and secondly, process the data using integration of SGA and LSTM or MLP.

- First the performance will be observed when the raw data (fundamental and technical indicators) is processed directly using Multi-Layer Protocol (MLP) method and Long Short-Term Memory (LSTM) method, we named this as MLP and LSTM respectively.
- Next the performance of the raw data will also be observed when MLP and LSTM is integrated using SGA. The objective is to see the effectiveness of the SGA when it is integrated with MLP and LSTM. These methods we called it as SGA-MLP and SGA-LSTM respectively.

The experiment will be divided into two main section, **Section 4.1** will be explaining on the experiments using the fundamental indicator while Section 4.2 will be explaining on experiment using technical indicator.

4.1 Experiment with fundamental indicator

We execute the experiment with the fundamental indicator. This experiment is to observe 8 metrics of the cross-sectional stock return prediction based on the fundamental indicator, whether the integration of SGA give improvement or vice versa. First, we experiment the fundamental indicator directly using the MLP method. Then followed by experimenting it using LSTM method. This experiment is without integration of SGA. To observe the capability of SGA, we executed an experiment based on the using the LSTM and MLP respectively with the integration of SGA method. Table 4(a), illustrates the results of the experiments conducted, where the LSTM or MLP is integrated with SGP, is labelled as SGA-MLP and SGA-LSTM respectively. While the results obtained without the integration of SGA is shown in the column labelled as MLP and LSTM respectively.

Table 4
(a):The metrics of fundamental indicators for DNN with MLP or LSTM

Fundamental Indicators					
2020	Metric	SGA-MLP	SGA-LSTM	MLP	LSTM
	Rank IC	-7.15%	-6.70%	-2.95%	-2.46%
	ICIR	-4.20	-3.90	-2.79	-2.25
	IC-Success Ratio	71.43%	73.47%	57.14%	65.31%
	IC-PNL	1.85	1.73	2.14	1.19
	Relative R above 300	-7.87%	-5.83%	1.39%	-3.53%
	Relative R above 500	-3.35%	-1.20%	6.28%	1.15%
	Relative R above average	0.27%	2.39%	10.01%	4.84%
	Sharp ratio	0.04	0.34	1.84	1.01
2021	Rank IC	-7.13%	-7.25%	-1.04%	-2.13%
	ICIR	-4.24	-5.58	-0.80	-2.72
	IC-Success Ratio	73.47%	75.51%	55.10%	61.22%
	IC-PNL	1.77	2.65	1.07	1.74
	Relative R above 300	45.71%	41.29%	31.96%	23.85%
	Relative R above 500	21.40%	17.49%	9.92%	2.91%
	Relative R above average	12.77%	8.83%	1.62%	-4.61%
	Sharp ratio	1.67	1.38	0.27	-0.91
2022	Rank IC	-9.86%	-9.99%	-1.58%	-4.07%
	ICIR	-6.11	-7.32	-1.26	-5.08
	IC-Success Ratio	87.76%	87.76%	55.10%	75.51%
	IC-PNL	1.44	2.28	**1.28**	**2.26**
	Relative R above 300	35.99%	34.59%	20.57%	20.60%
	Relative R above 500	34.54%	33.12%	18.96%	19.16%
	Relative R above average	19.05%	17.84%	5.27%	5.46%
	Sharp ratio	2.77	3.09	0.69	1.26

Table 4
(b): The metrics from 2020–2022 for fundamental indicators

Average of mean from 2020 to 2022	Rank IC	-8.05%	-7.98%	-1.85%	-2.88%
	ICIR	-4.88	**-5.46**	-1.55	-3.22
	IC-Success Ratio	77.55%	**78.91%**	55.78%	67.35%
	IC-PNL	1.73	**2.13**	1.37	1.57
	Relative R above 300	23.16%	22.35%	18.00%	13.48%
	Relative R above 500	17.12%	16.26%	12.07%	7.74%
	Relative R above average	10.84%	9.89%	5.80%	1.86%
	Sharp ratio	1.49	1.47	0.86	0.38

Table 4(a) shows the results executed from the experiment for data in the year of 2020 to 2022. Whereas Table 4(b) summarize the data from 2020 to 2022 based on its average mean. Based on the results shown in Table 4(b) above, the results indicate that when the raw fundamental indicators were used as input for LSTM or MLP models, the average IC values were − 1.85% and − 2.88%, respectively. The average value of IC in this situation is considered low whereby the ideal average value should be above 8%. While the average value for ICIR were − 1.55 and − 3.22, respectively. This value for cross-sectional stock price change prediction is considered average or acceptable. The ideal value for ICIR is above 3. However, after integrating the models with the SGA algorithm, the IC absolute values increased to 8.05% for MLP and 7.98% for LSTM which is considered as ideal outcome.

The SGA-LSTM model attained the highest average value of -5.46 for ICIR, surpassing the performance of other models. It exhibited superior results in terms of IC-success ratio and IC-PNL, with values of 78.91% and 2.13, respectively. Furthermore, both the SGP-LSTM and SGP-MLP models showcased notable advantages over the single DNN models by employing a straightforward rule-based strategy for a long-only approach. Specifically, the SGP-LSTM model demonstrated a relative R exceeding the CSI 300 index by 22.35% and surpassing the CSI 500 index by 16.26%. Moreover, it achieved a relative R above the average by 9.89% per year, positioning it among the top 10% of mutual fund managers in China.

4.2 Experiment using technical indicator

In contrast, according to the findings presented in Table 5(a) and 5(b), SGP-MLP or SGP-LSTM does not demonstrate significant advantages over single DNN models when it comes to technical indicators. On average, the single LSTM model for technical indicators produced the best results in terms of normal metrics such as IC, ICIR, and IC-success, with percentages of -8.64%, -6.561, and 85.71% respectively (the original IC mean of technical indicators is 2.82% and ICIR mean is 0.23 from Table 1(a) and 1(b),). When comparing the performance of the two single DNN models in relation to a simple rule-based strategy, the LSTM model outperformed the MLP model. This could be attributed to the fact that technical indicators

represent sequential time series data, which is better suited for the LSTM model, as explained in the methodology section. Notably, when considering a long-only strategy, the LSTM model exhibited a significantly higher relative R above average at 13.28%, compared to the MLP model's 3.94%.

Table 5
(a) The metrics of technical indicators for DNN with MLP or LSTM

Technical Indicators					
2020	Metric	SGA-MLP	SGA-LSTM	MLP	LSTM
	Rank IC	-8.07%	-7.68%	-8.08%	-9.49%
	ICIR	-4.575	-5.313	-4.345	-6.970
	IC-Success Ratio	77.55%	79.59%	73.47%	87.76%
	IC-PNL	1.494	1.664	1.762	2.051
	Relative R above 300	-5.06%	-4.28%	-7.65%	8.98%
	Relative R above 500	-0.41%	0.59%	-3.22%	14.25%
	Relative R above average	3.32%	4.05%	0.48%	18.60%
	Sharp ratio	0.520	0.701	0.072	3.098
2021	Rank IC	-7.76%	-8.14%	-7.84%	-7.74%
	ICIR	-4.741	-5.684	-4.978	-5.876
	IC-Success Ratio	77.55%	81.63%	77.55%	83.67%
	IC-PNL	1.945	1.903	2.404	1.738
	Relative R above 300	41.55%	45.18%	35.25%	42.83%
	Relative R above 500	17.58%	20.61%	12.29%	18.61%
	Relative R above average	9.25%	11.76%	4.41%	9.92%
	Sharp ratio	1.315	2.062	0.586	1.591
2022	Rank IC	-9.26%	-8.92%	-9.49%	-8.68%
	ICIR	-6.033	-5.899	-5.541	-6.540
	IC-Success Ratio	83.67%	77.55%	75.51%	85.71%
	IC-PNL	2.048	2.374	2.367	1.504
	Relative R above 300	26.00%	31.07%	22.11%	25.87%
	Relative R above 500	24.55%	29.54%	20.44%	24.45%
	Relative R above average	10.15%	14.39%	6.56%	9.97%
	Sharp ratio	1.551	2.268	0.920	1.612

Table 5
(b) The metrics of technical indicators for DNN with MLP or LSTM

mean	Rank IC	-8.47%	-8.64%	-8.36%	-8.25%
	ICIR	-5.002	**-6.561**	-5.156	-5.726
	IC-Success Ratio	75.51%	85.71%	79.59%	79.59%
	IC-PNL	2.092	1.735	1.763	1.957
	Relative R above 300	15.73%	26.21%	19.99%	23.07%
	Relative R above 500	9.77%	19.83%	13.95%	16.94%
	Relative R above average	3.94%	13.28%	7.83%	10.38%
	Sharp ratio	0.531	2.076	1.130	1.675

Based on the experiments conducted earlier, we could summarize that the fundamental indicator will achieve the best result, when the indicators are fed into SGA algorithm, while the technical indicator will achieve the best result without integrating the SGA but directly through LSTM technique. Therefore, we design a new DNN framework that could work well with both fundamental and technical indicators. Figure 5 below illustrates the proposed DNN framework where both fundamental and technical indicators are fed as the raw data. The explanation on the experiment on this proposed framework will be discussed in the next section.

Based on Fig. 5 above, the fundamental indicators are first fed as the raw data into the framework. As mentioned earlier, the results shows better when SGA is integrated with LSTM or MLP. Therefore, the fundamental indicators are processed based on SGA and the output is being an input for the Phase I, the augmentation phase. The output for the augmentation phase is combined with the technical indicators to be an input for the Phase II, the feature selection. Here, only LSTM is utilized as from the experiment executed earlier, LSTM outperformed the MLP in terms of a better results. Two layers of LSTM are performed with 100 nodes each, where the final feature selection is only 30 notes for the price changes prediction. In Section 4.3, we present the results based on the experiments conducted using the new proposed framework as shown in Fig. 5.

4.3 Experiment using fundamental and technical indicator

Table 6
(a): The metrics based on the proposed SGA-LSTM framework

	original Rank IC	**hybrid model IC**	**original ICIR**	**hybrid model ICIR**
Fundamental indicators	0.65%	7.98%	0.1	5.46
Technical indicators	2.82%	8.64%	0.23	6.56
proposed SGP-DNN		9.24%		7.24

Table 6
(b): The metrics based on the proposed SGA-DNN framework

Year	Metric	hybrid model for quantitative indicators	hybrid model for quantitative indicators individually	
2020	Metric	Hybrid SGA-LSTM for both fundamental and technical indicators	SGA-LSTM for fundamental indicators	LSTM for Technical Indicators
	Rank IC	-9.64%	-6.70%	-9.49%
	ICIR	-7.1	-3.9	-6.97
	IC-Success Ratio	87.76%	73.47%	87.76%
	IC-PNL	2.79	1.73	2.05
	Relative R above 300	13.53%	-5.83%	8.98%
	Relative R above 500	19.13%	-1.20%	14.25%
	Relative R above average	23.10%	2.39%	18.60%
	Sharp ratio	3.04	0.34	3.1
2021	Rank IC	-8.57%	-7.25%	-7.74%
	ICIR	-6.58	-5.58	-5.88
	IC-Success Ratio	83.67%	75.51%	83.67%
	IC-PNL	2.64	2.65	1.74
	Relative R above 300	52.46%	41.29%	42.83%
	Relative R above 500	26.71%	17.49%	18.61%
	Relative R above average	17.08%	8.83%	9.92%
	Sharp ratio	2.92	1.38	1.59
2022	Rank IC	-9.51%	-9.99%	-8.68%
	ICIR	-7.69	-7.32	-6.54
	IC-Success Ratio	89.80%	87.76%	85.71%

Year	Metric	hybrid model for quantitative indicators	hybrid model for quantitative indicators individually	
	IC-PNL	1.65	2.28	1.5
	Relative R above 300	26.04%	34.59%	25.87%
	Relative R above 500	24.72%	33.12%	24.45%
	Relative R above average	10.22%	17.84%	9.97%
	Sharp ratio	1.57	3.09	1.61

Table 6
(c): The metrics based on the proposed SGA-LSTM framework for the average mean data

average metrics from 2020 to 2022	Rank IC	-9.24%	-7.98%	-8.64%
	ICIR	**-7.24**	-5.48	-6.58
	IC-Success Ratio	87.07%	78.91%	85.71%
	IC-PNL	2.32	2.13	1.74
	Relative R above 300	**31.00%**	22.35%	26.21%
	Relative R above 500	**24.48%**	16.26%	19.83%
	Relative R above average	**17.38%**	9.89%	13.28%
	Sharp ratio	**2.49**	1.47	2.08

According to Table 6(a) our hybrid model showcased a significant improvement of 1128% in information coefficient (IC) and an impressive surge of 5360% in IC information ratio (ICIR) when applied to fundamental indicators. For technical indicators, the hybrid model achieved a commendable 206% increase in IC and a remarkable surge of 2752% in ICIR. According to Table 6(b) and 6(c), the proposed SGP-LSTM model attained an rank IC value of 9.24% and an ICIR of 7.24 for a five-day prediction horizon.

5. Results and Discussion

In the stock trading experiment, we conducted ‘backtests’ on three portfolios mentioned in Table 6(a) and 6(b) over a period of 720 days. The prediction horizon was set at 5 days, and the rolling cycle was set at 20 days. This means that every 20 days, the hybrid model optimized its parameters based on the previously mentioned 1180 data points. The model parameters remained unchanged for the next 20 days. Additionally, every 5 days, the model's stock prediction values were ranked, and the top 10% (450

stocks) were selected for portfolio construction using equal weights for buying and holding. Limit stocks were excluded to account for trading issues.

From Fig. 6, it is apparent that the performance monotonicity of the three long-only portfolios can be compared. The figure represents the accumulated returns of 10 groups comprising 4,500 stocks in the China A-Share stock market from 2020 to 2022. Notably, the proposed SGA-LSTM framework denoted as "t_gaf_features_hybrid" portfolio demonstrates the most consistent performance and has best monotonicity comparing with the SGA-LSTM for fundamental indicators and LSTM for technical indicators. Additionally, Fig. 7 presents the accumulated performance comparisons of relative returns over the average return of total stocks among the 3 long-only portfolios. Our proposed enhanced SGA-DNN model, known as the "t_gaf_features_hybrid" model, yields the best outcomes. Throughout the three-year out-of-sample period, it achieves a relative annual return of 17.38% and accumulates a total return of 61.72%.

Figure 8 presents a comparison of the cumulative return curves for the proposed SGD-DNN portfolio and two broad-based indices, as well as the average portfolio, during the period of 2020–2022. The results clearly demonstrate that the proposed model outperformed the average portfolio, as well as the CSI 300 and CSI 500 indices. Notably, the SGD-DNN hybrid model exhibited significant outperformance compared to the CSI 300 index, the CSI 500 index, and the average portfolio, as shown in Table 6(a) and 6(b). Over a three-year timeframe, the model generated excess returns of 124.80%, 92.89%, and 61.72%, respectively, with average annualized excess returns of 31.00%, 24.48%, and 17.38%.

6. Conclusion

This paper introduced a methodology to enhance the cross-sectional stock return prediction by utilizing Symbolic Genetic Algorithm (SGA) for input generation and integrating it with Deep Neural Network (DNN) models. The study demonstrated significant improvements in prediction, outperforming popular market indexes. A hybrid model combining SGA with Long Short-Term Memory (LSTM) showcased superior performance, consistently surpassing market returns a simple rule-based strategy based on the proposed hybrid SGA-LSTM model outperforms major Chinese stock indexes, generating average annualized excess returns of 31.00%, 24.48%, and 17.38% compared to the CSI 300 index, CSI 500 index, and the average portfolio, respectively. The findings highlight the potential of the proposed approach in generating profitable investment strategies and provide insights into addressing challenges in data integration and feature engineering.

This study focused solely on financial time series data, which is known for its high autocorrelation. However, recent research has explored the incorporation of diverse data sources such as social media data, news, macroeconomic data, and high-frequency data. Moreover, the proposed hybrid SGA-DNN model could benefit from additional optimization targets, such as relative return of top groups or monotonicity of ten groups of target stocks, instead of solely relying on MSE as the optimization goal. Additionally, recent advancements in reinforcement learning or generative adversarial networks (GANs),

such as ChartGPT application, have been suggested to be combined with hybrid DNN models. Therefore, it could be worthwhile to consider supplementing the suggested hybrid SGA-DNN model with GANs or reinforcement learning techniques to leverage multi-source information and improve prediction performance.

Declarations

Availability of Data and Materials

The data that support the findings of this study are available from S&P Global but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of S&P Global. Any further related information can be found in the link https://www.marketplace.spglobal.com/en/datasets/alpha-factor-library-(3).

References

1. Sharma, A., Bhuriya, D. & Singh, U. Survey of Stock Market Prediction Using Machine Learning Approach. *2017 International Conference of Electronics, Communication and Aerospace Technology (Iceca), Vol 2*, 506-509 (2017).
2. Yu, D., Huang, D. & Chen, L. Stock return predictability and cyclical movements in valuation ratios. *Journal of Empirical Finance* **72**, 36-53, doi:10.1016/j.jempfin.2023.02.004 (2023).
3. Wing-Keung Wong, M. M. B.-K. C. How rewarding is technical analysis? Evidence from Singapore stock market. *Applied Financial Economics* **Volume 13, 2003 - Issue 7**, Pages 543-551, doi:10.1080/0960310022000020906 (2010).
4. Fama, E. F. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance* **Vol. 25, No. 2, Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, 1969 (May, 1970), pp. 383-417 (35 pages)**, doi:10.2307/2325486 (1969).
5. Iltuzer, Z. Predicting stock returns with financial ratios: A new methodology incorporating machine learning techniques to beat the market. *Asia-Pac J Account E*, doi:10.1080/16081625.2021.2007408 (2021).
6. Giovannelli, A., Massacci, D. & Soccorsi, S. Forecasting stock returns with large dimensional factor models. *Journal of Empirical Finance* **63**, 252-269, doi:10.1016/j.jempfin.2021.07.009 (2021).
7. Rajput, V. S. & Dubey, S. M. Stock Market Sentiment Analysis Based On Machine Learning. *Proceedings on 2016 2nd International Conference on Next Generation Computing Technologies (Ngct)*, 506-510 (2016).
8. Box, G. E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. Time Series Analysis: Forecasting and Control. *John Wiley & Sons, Hoboken.* (2015).

9. Azevedo, V. & Hoegner, C. Enhancing stock market anomalies with machine learning. *Rev Quant Financ Acc* **60**, 195-230, doi:10.1007/s11156-022-01099-z (2023).
10. Breitung, C. Automated stock picking using random forests. *Journal of Empirical Finance* **72**, 532-556, doi:10.1016/j.jempfin.2023.05.001 (2023).
11. Fister, D., Mun, J. C., Jagric, V. & Jagric, T. Deep Learning for Stock Market Trading: A Superior Trading Strategy? *Neural Netw World* **29**, 151-171, doi:10.14311/Nnw.2019.29.011 (2019).
12. Samarakoon, P. A. & Athukorala, D. A. S. System Abnormality Detection in Stock Market Complex Trading Systems Using Machine Learning Techniques. *2017 National Information Technology Conference (Nitc)*, 125-130 (2017).
13. Nabipour, M., Nayyeri, P., Jabani, H., Shahab, S. & Mosavi, A. Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *Ieee Access* **8**, 150199-150212, doi:10.1109/Access.2020.3015966 (2020).
14. Yoo, P. D., Kim, M. H. & Jan, T. Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. *International Conference on Computational Intelligence for Modelling, Control & Automation Jointly with International Conference on Intelligent Agents, Web Technologies & Internet Commerce, Vol 2, Proceedings*, 835-+ (2006).
15. L'Heureux, A., Grolinger, K., Elyamany, H. F. & Capretz, M. A. M. Machine Learning With Big Data: Challenges and Approaches. *IEEE Access* **5**, 7776-7797, doi:10.1109/access.2017.2696365 (2017).
16. Chen, K., Zhou, Y. & Dai, F. Y. A LSTM-based method for stock returns prediction : A case study of China stock market. *Proceedings 2015 Ieee International Conference on Big Data*, 2823-2824 (2015).
17. Fischer, T. & Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* **270**, 654-669, doi:10.1016/j.ejor.2017.11.054 (2018).
18. Baek, Y. & Kim, H. Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Syst Appl* **113**, 457-480, doi:10.1016/j.eswa.2018.07.019 (2018).
19. Eapen, J., Verma, A. & Bein, D. Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction. *2019 Ieee 9th Annual Computing and Communication Workshop and Conference (Ccwc)*, 264-270 (2019).
20. Long, W., Lu, Z. & Cui, L. Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems* **164**, 163-173, doi:10.1016/j.knosys.2018.10.034 (2019).
21. Zhou, X. R. Stock Price Prediction using Combined LSTM-CNN Model. *2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (Mlbdbi 2021)*, 67-71, doi:10.1109/Mlbdbi54094.2021.00020 (2021).
22. Ishwarappa & Anuradha, J. Big data based stock trend prediction using deep CNN with reinforcement-LSTM model. *Int J Syst Assur Eng*, doi:10.1007/s13198-021-01074-2 (2021).
23. Chung, H. & Shin, K.-s. Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction. *Sustainability* **10**, doi:10.3390/su10103765 (2018).

24. Chung, H. & Shin, K.-s. Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. *Neural Computing and Applications* **32**, 7897-7914, doi:10.1007/s00521-019-04236-3 (2019).
25. He, B. & Kita, E. in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)* 199-202 (2021).
26. Chen, S. & Zhou, C. Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network. *IEEE Access* **9**, 9066-9072, doi:10.1109/access.2020.3047109 (2021).
27. Shahvaroughi Farahani, M. & Razavi Hajiagha, S. H. Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models. *Soft Comput* **25**, 8483-8513, doi:10.1007/s00500-021-05775-5 (2021).
28. Li, X., Yu, Q., Tang, C., Lu, Z. & Yang, Y. Application of Feature Selection Based on Multilayer GA in Stock Prediction. *Symmetry* **14**, doi:10.3390/sym14071415 (2022).
29. Yun, K. K., Yoon, S. W. & Won, D. Interpretable stock price forecasting model using genetic algorithm-machine learning regressions and best feature subset selection. *Expert Syst Appl* **213**, doi:10.1016/j.eswa.2022.118803 (2023).
30. Leung, C. K. S., MacKinnon, R. K. & Wang, Y. A Machine Learning Approach for Stock Price Prediction. *Proceedings of the 18th International Database Engineering and Applications Symposium (Ideas14)*, 274-277, doi:10.1145/2628194.2628211 (2014).
31. Yoshua Bengio, Y. L., Geoffrey Hinton. Deep Learning for AI. *Communications of the ACM* **Vol. 64 No. 7**, 58-65, doi:10.1145/3448250 (2021).
32. Global, S. P. in *S&P Global* (2022). https://www.marketplace.spglobal.com/en/datasets/alpha-factor-library-(3)
33. Kakushadze, Z. alpha101-formulars. *Free University of Tbilisi, Business School & School of Physics 240, David Agmashenebeli Alley, Tbilisi, 0159, Georgia (2015).*
34. Jui-Sheng Chou, T.-K. N. Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine Learning Regression. *IEEE Transactions on Industrial Informatics* **14(7)**, 3132-3142, doi:10.1109/TII.2018.2794389 (2018).

Figures

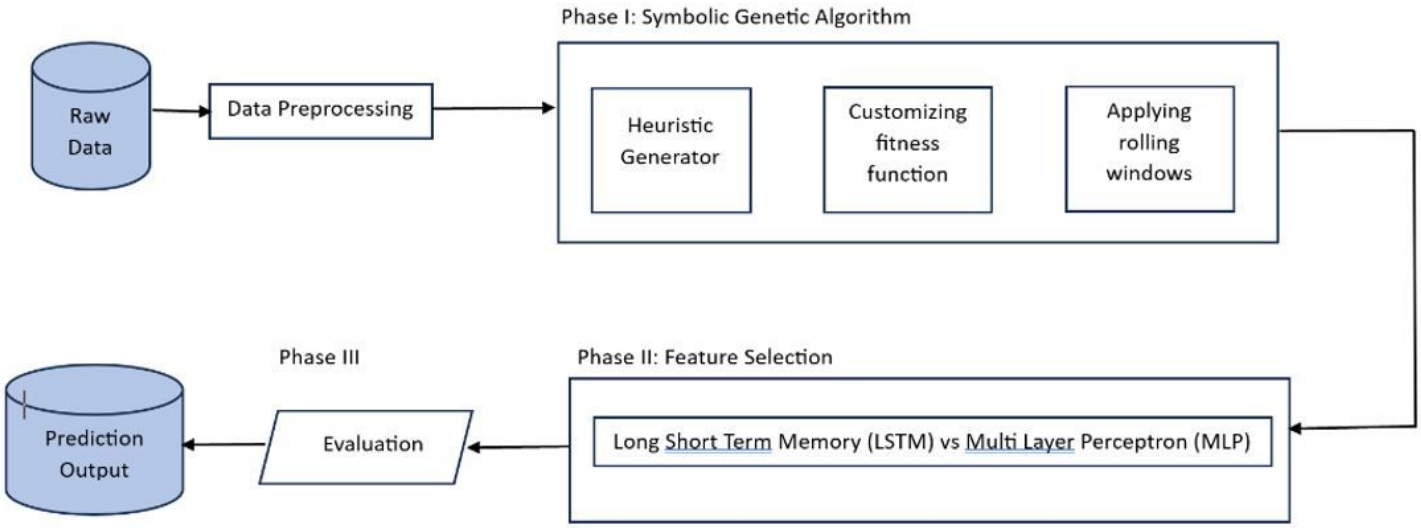

Figure 1

Illustration of the proposed Deep Neural Network Framework

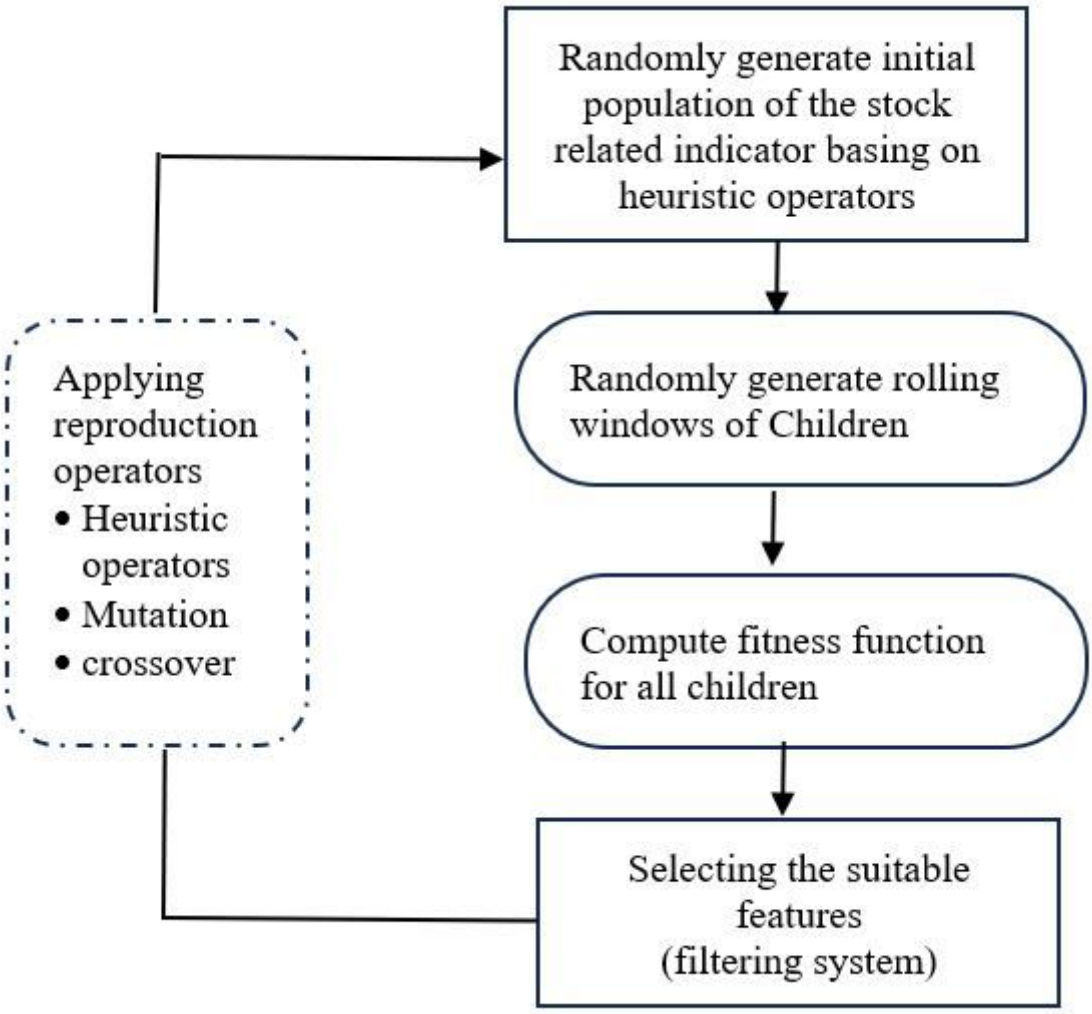

Figure 2

The structure of the proposed Symbolic Genetic Algorithm

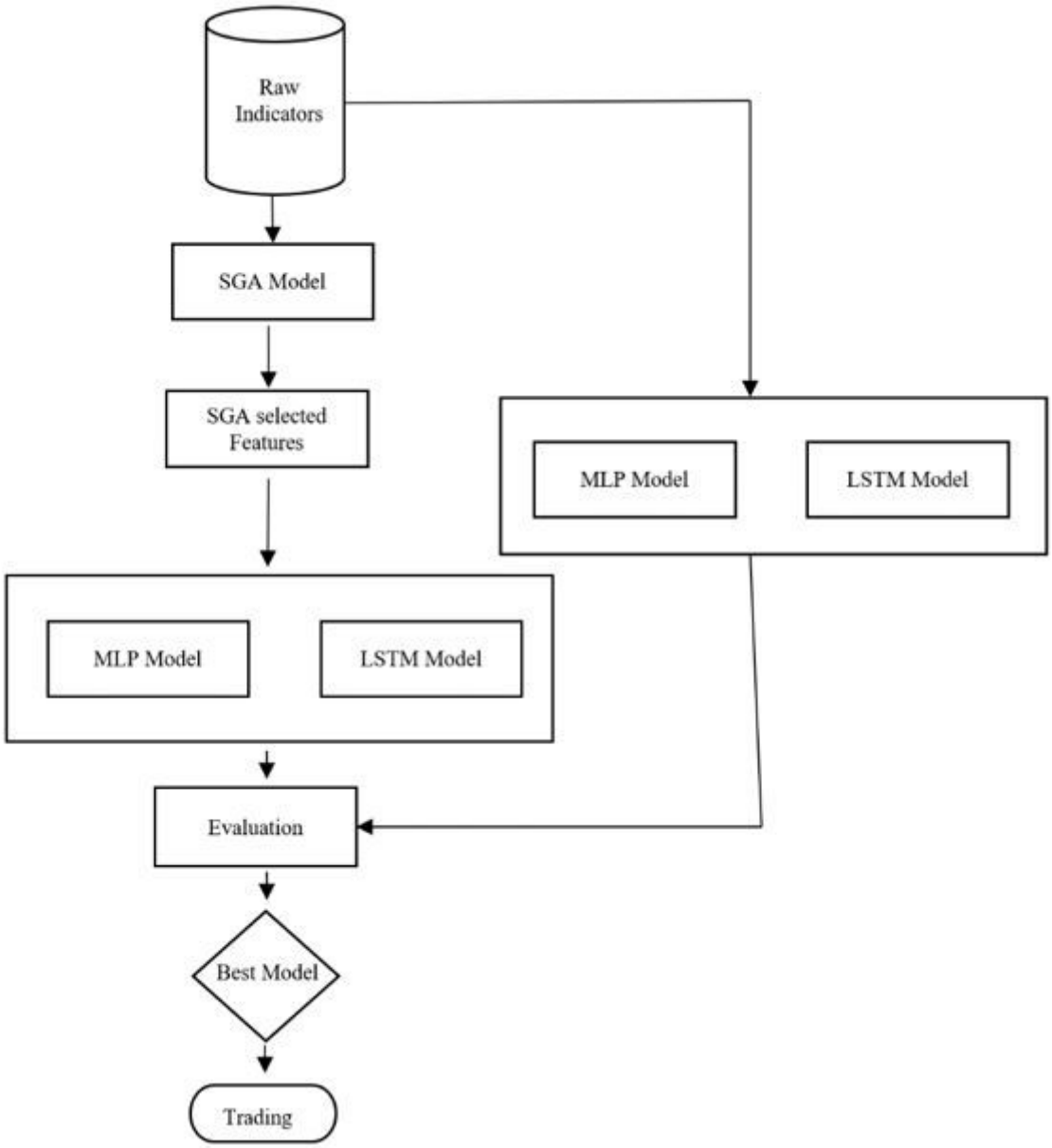

Figure 3

Feature Selection: LSTM vs MLP

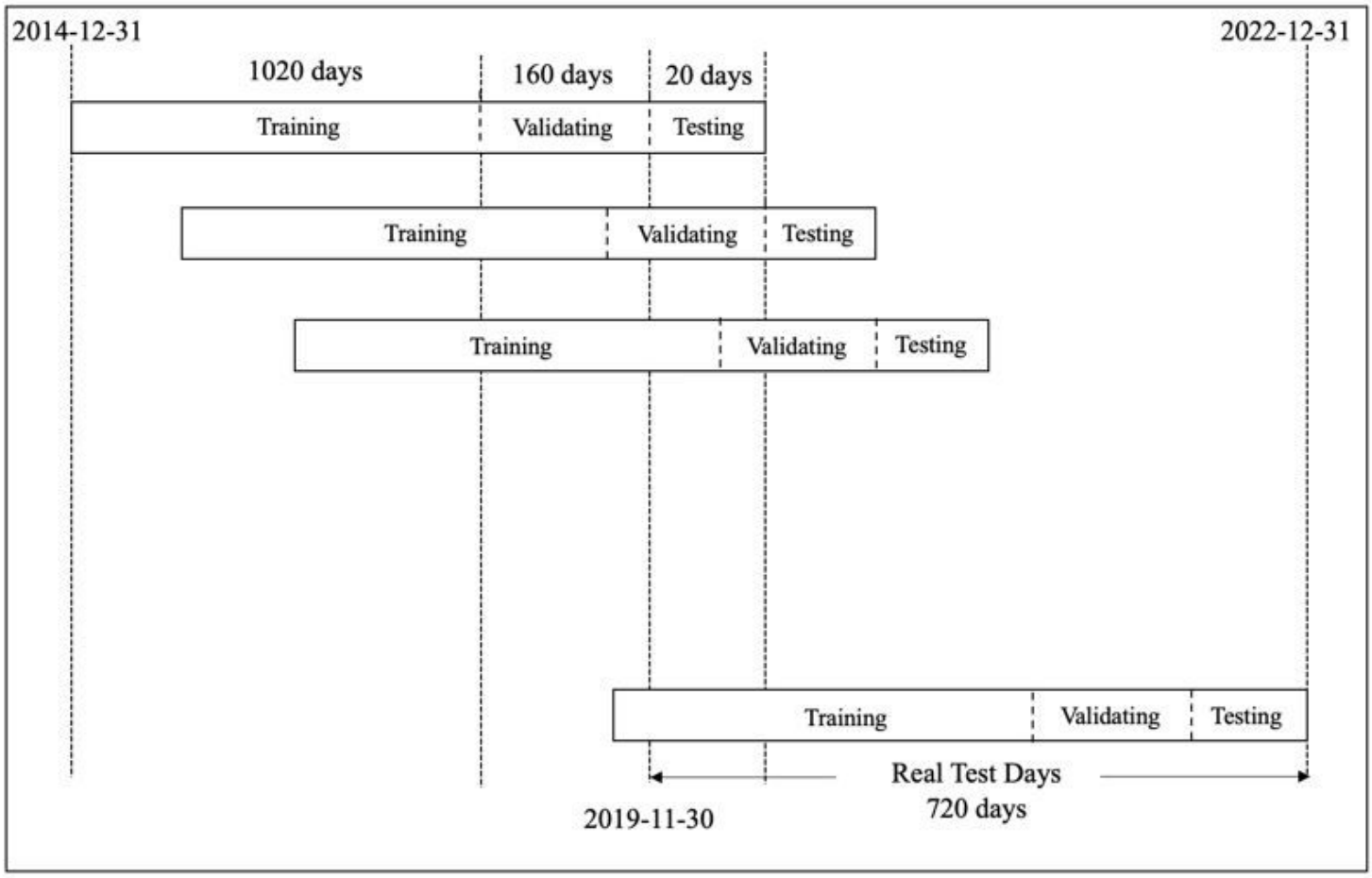

Figure 4

Train/validation/test set

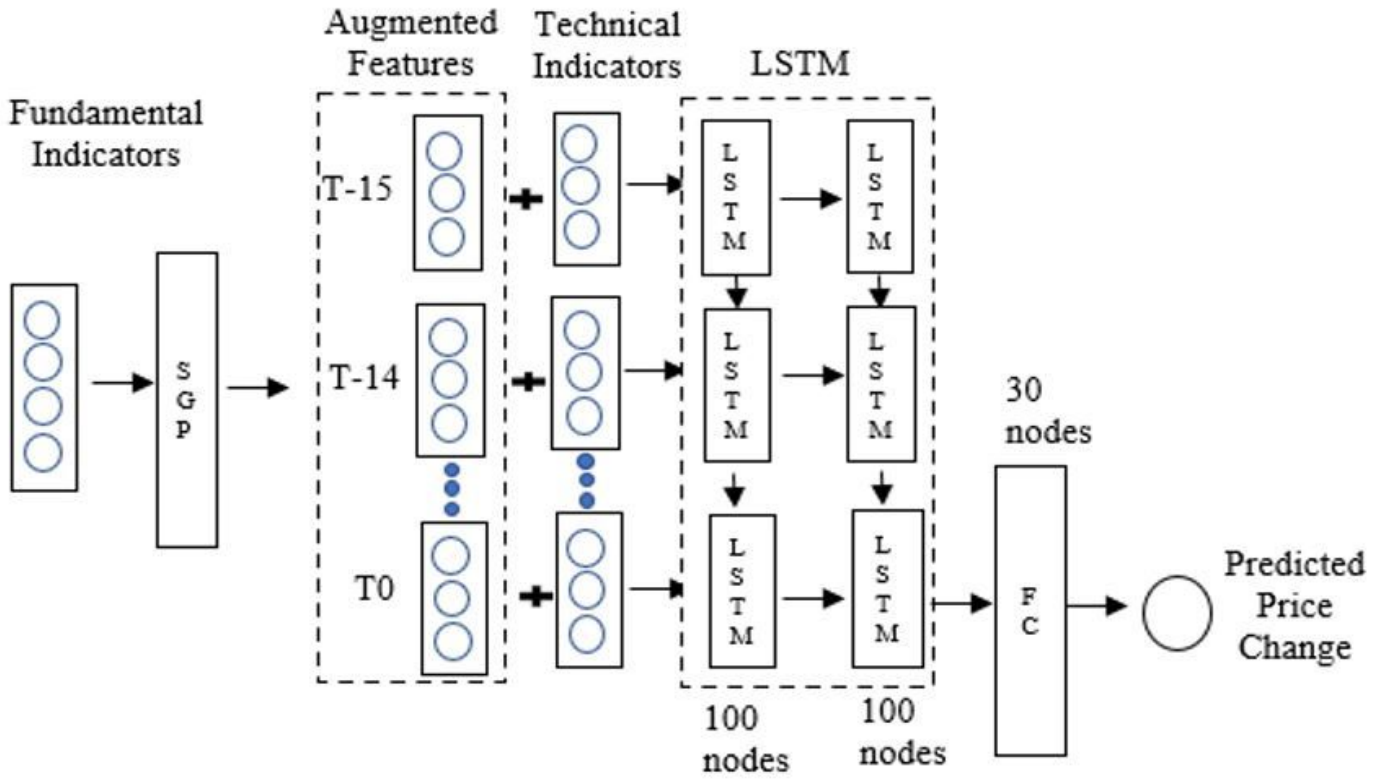

Figure 5

The proposed SGA-DNN framework for both fundamental and technical indicators

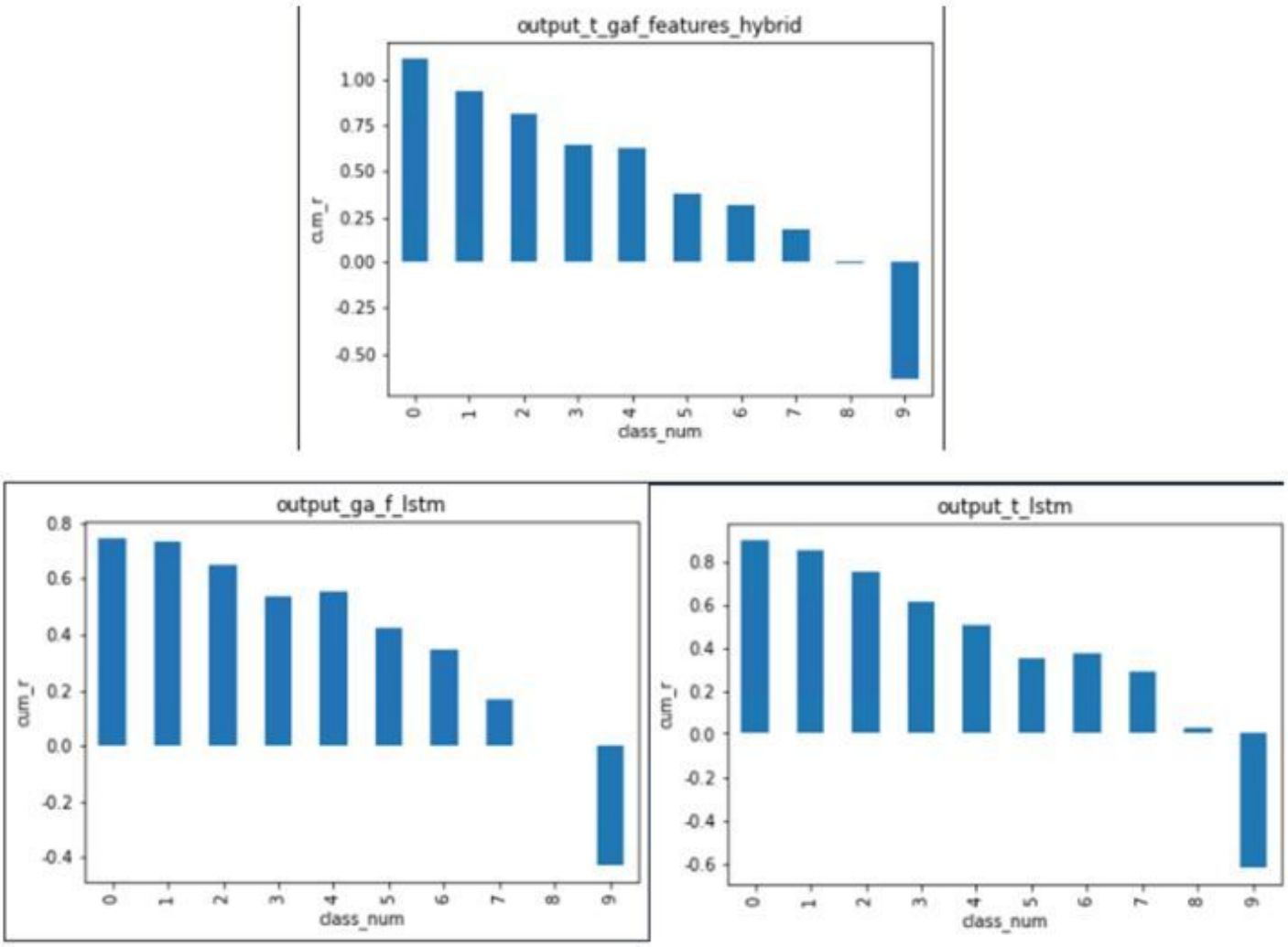

Figure 6

comparisons of monotonicity of forecasted value

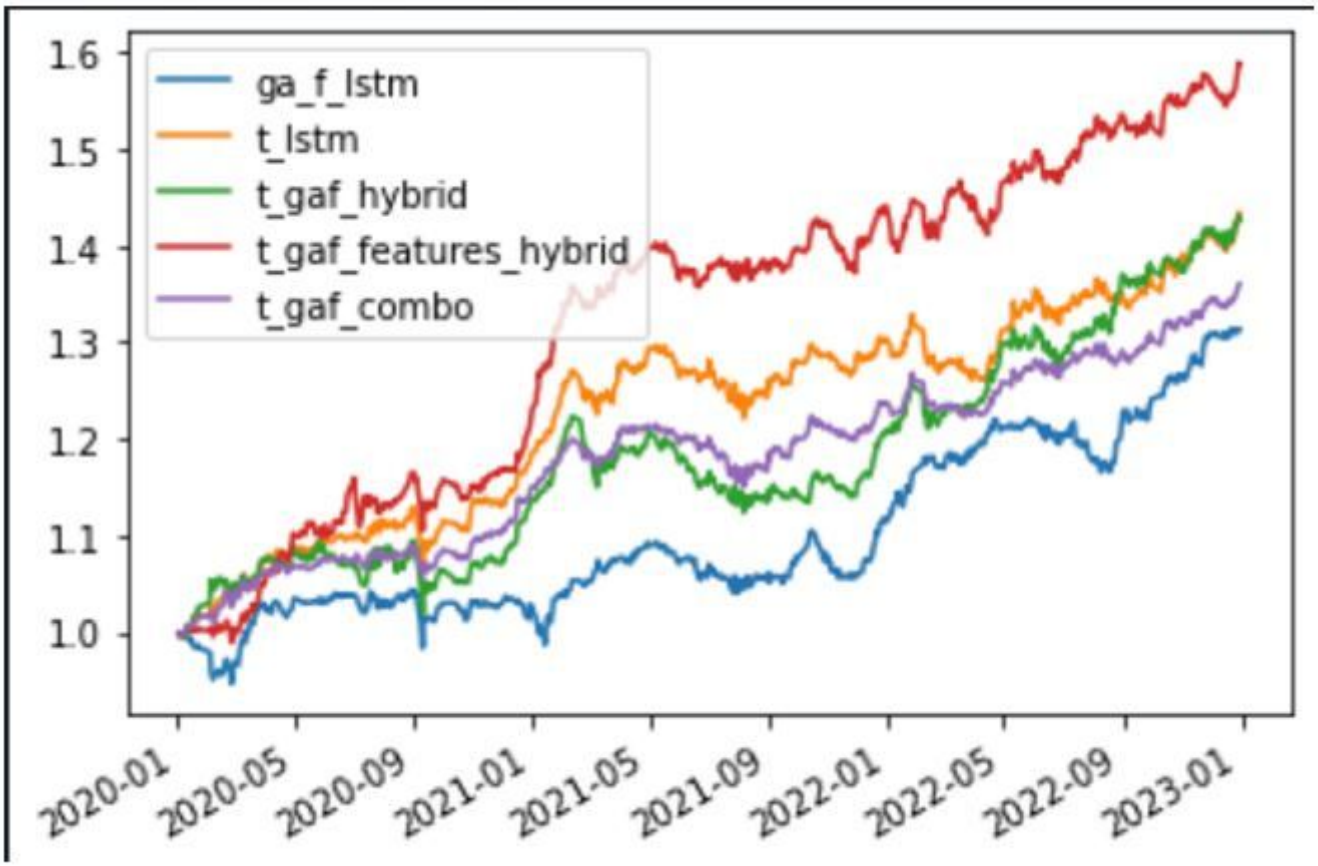

Figure 7

comparisons of relative R above average

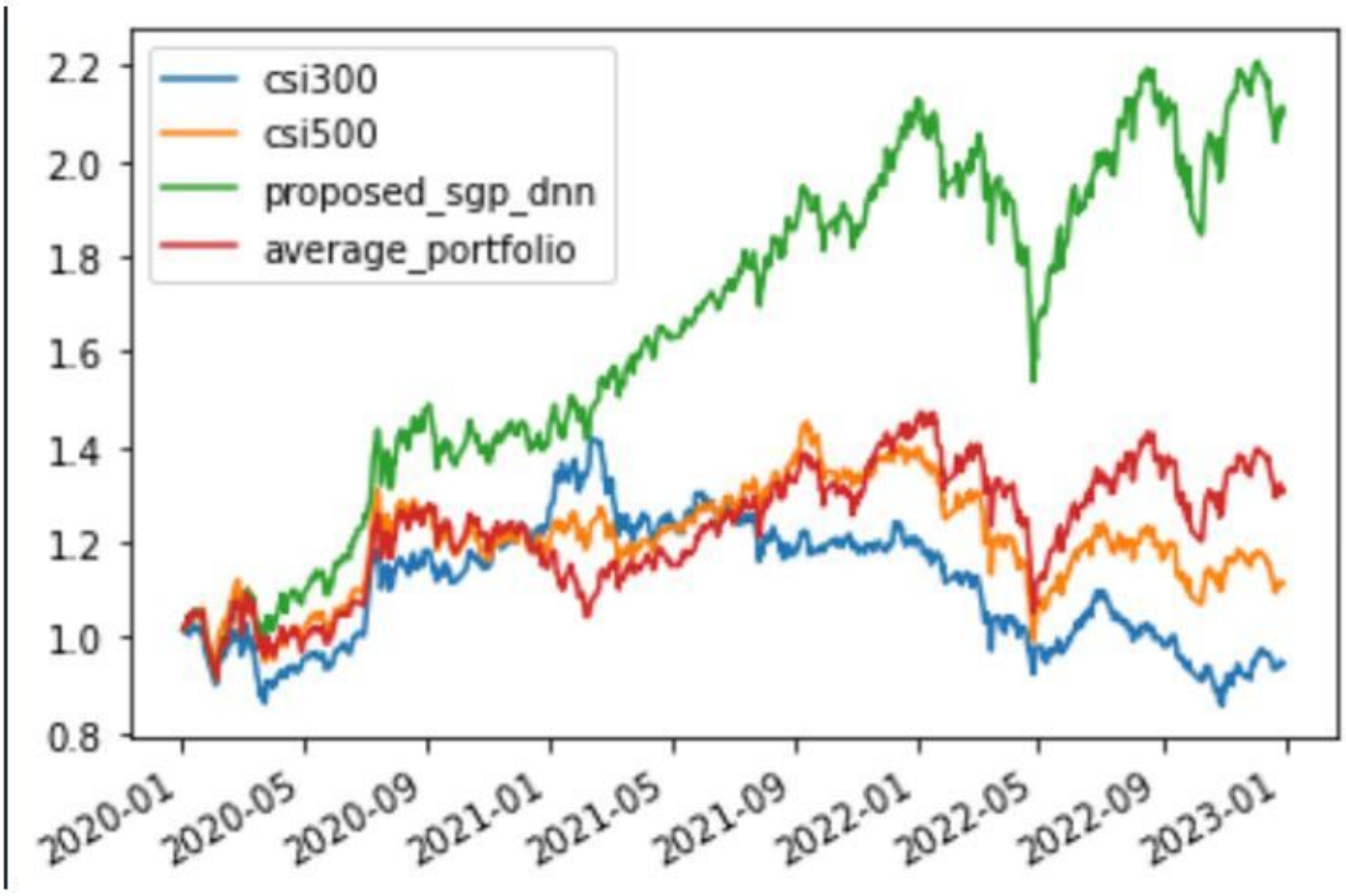

Figure 8

comparison of the cumulative return curves

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix.docx