# Integrating Symbolic Genetic Programming With Lstm for **Forecasting Cross-Sectional Price Returns: A Comparative Analysis** of Chinese And Japanese Stock Market

Li Oi

Faculty Of Artificial Intelligence UTM, Malaysia

Norshaliza Kamaruddin

Faculty Of Artificial Intelligence UTM, Malaysia

li.qi@graduate.utm.my

norshaliza.k@utm.my

2981

# Xun Gong

Asset Mgt Department Ping An Property Casualty Insurance China

# **Chen Peng**

Frontier Technology Team Ping An Technology, China.

Corresponding Author: Norshaliza Kamaruddin

Copyright © 2024 Li Qi, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Abstract

This paper introduces a novel deep learning framework and integrates symbolic genetic programming (SGP) to predict cross-sectional stock return rankings. The data of this framework covers more than 4,600 Chinese listed stocks and Japan from 2014 to 2022 for comparison. Leveraging the S&P Dataset for both countries, the hybrid model absorbs the algorithm of data augmentation and novel DNN framework. The research showcases substantial enhancements in Rank IC by 588.03% in China and 194.27% in Japan, respectively. Moreover, a simple rule-based investment strategy based on the application of the SGPLSTM model in China achieved an annualized return of 22.35% exceeding the index (CSI 300) and 16.26% compared to CSI 500 index in China, respectively, and surpasses the N225 index by 4.56% and the TPX index by 5.10% in Japan. These results fully demonstrate that SGP combined with the LSTM model can greatly improve the precision and accuracy of cross-sectional stock selection and also provide very valuable guidance to financial analysts, fund managers and traders.

**Keywords:** Ssymbolic Genetic Programming (SGP), Long-Short Term Memory (LSTM), Neural Network, Data augmentation, Feature engineering, Chinese and Japanese stock, Cross-sectional stock return forecasting rankings.

# **1. INTRODUCTION**

As a crucial element in the process of decision-making for investment, "stock market forecast" observes the market trends and stock fluctuation patterns [1]. However, this is not always straightforward, easy, and simple to understand because of the complex and constantly-changing dynamics [2–4], inherent in the stock market [5], which can be observed and summarized as five key characteristics by financial practitioners including non-linearity, non-stationarity, lead-lag effect, market microstructure and volatility [2–4].

The most frequently referenced paper on Long Short-Term Memory (LSTM) was authored by Fisher in 2018. In this paper, the price change was used as the sole feature and fed into a single LSTM model. The research observed a significant alpha effect from 1993 to 2009. However, the model did not perform well after 2010 [6], and Ghosh, P. supplemented Fisher's original model with two additional features, which altered the alpha effect from 2010 to 2015. However, after 2015, the model also failed for alpha [7]. However, after 2015, the model also failed to perform well [7]. These examples show the existing problems for deep neural network (DNN) model in stock prediction e.g., overfitting and feature limitations. In addition, the feature numbers of limits for DNN also restricts the model's abilities to capture and retain pattern for the purpose of accurate prediction.

In response to this, the proposed hybrid model integrates deep neural network (DNN) with symbolic regression genetic programming (SGP) for data augmentation. This helps to improve stock market prediction by leveraging GA, this new technique seeks to apply GA for generate more features and select higher quality ones to feed to DNN framework, filling the gap of feature number limitations.

Our SGP-LSTM model on Chinese stocks achieved an annualized return of 22.35% exceeding the index (CSI 300) and 16.26% compared to CSI 500 index in China, respectively, and surpasses the N225 index by 4.56% and the TPX index by 5.10% in Japan. We also calculate the performance of the strategy in Japan, and it is found that the proposed SGP-LSTM model also showed superior performance, surpassing the N225 index by 4.56% and the TPX index by 5.10%., and this further highlighting its effectiveness and robustness of model across the region.

The subsequent parts of this paper are arranged like followings: Section 2 will explore a comprehensive literature review, examining LSTM models and its combinations with Genetic Algorithms. Section 3 will discuss how to use the enhanced version of the SGP algorithm to generate factors, and how to effectively combine LSTM with SGP for prediction. Section 4 will explain the experimental results. Finally, Section 5 will offer conclusions, summarizing key findings and implications derived from the study.

# **2. LITERATURE REVIEW**

Deep learning originated from the last generation model of machine learning, artificial neural networks (ANN), but it has evolved with additional layers, neurons, and more complex network designs. Of course, the support of larger databases and more powerful computing power is also a core reason for its development. A DNN model consists of three components: input, weights, and bias terms. Unlike traditional linear models, each neuron in a DNN model incorporates activation function. The application and promotion of deep learning have attracted great attention in both public fund management and hedge fund management.

The earliest application of LSTM can be dated back to 2017 when it was applied to three specific stocks. The research compared the impacts of LSTM with CNN, and concluded that CNN outperformed LSTM [8]. The most classical Artificial Intelligence Applied to Stock Market Trading were demonstrated for cross-sectional stock selection basing on return series [6], and the model was extended from one return related factors to three ones [7].

The majority of recent stock forecasting papers focus on time series forecasting for specific equities or stock indexes, with significantly less emphasis on cross-sectional stock price forecasting. Typically, these studies use the price return-related factors as input and feed them directly into the DNN network. However, this approach has notable limitations.

Firstly price series is originally autocorrelated and even if the predicted correlation coefficient is high, it cannot be put into practice for actual trading strategy. Secondly, because the number of factors is too small, the robustness of the model is not strong, and it is very easy to overfit to the small number of factors.

To address the challenges posed by limited data points, the authors Baek and Kim (2018) [9], offered data augmentation techniques for two modules in ModAugNet. One module was designed for preventive LSTM, while the other module was intended for prediction [9].

In the field of computer science, simple data augmentation algorithm can be applied to generate the effective data points through methods such as data flipping and data extraction, ultimately enhancing the generalization capability of DNN models and preventing overfitting. In the field of investment, this method has been expanded by many scholars. The highly cited article by Fisher initially faced overfitting challenges due to limited data in the LSTM model. To address this issue, Yujin proposed a novel data augmentation method aimed at preventing overfitting. They introduced the ModAugNet framework, which includes two modules: one for preventing overfitting and another for using LSTM for prediction. This method significantly increases the number of data points, up to 252 times, increasing effective data points [9]. On the other hand, Shen used phase space reconstruction (PSR) [10], for data augmentation and applied the results to financial time series prediction.

The previous article introduced many methods of data amplification. However, these methods only increase the quatity of the data but cannot improve the quality of the data itself. The GA algorithm has also historically been considered as a method to increase data, because it can reorganize chromosomes to form new genes through three methods: crossover, selection, and mutation. Historically, more scientists have used the GA algorithm for factor selection [11, 12], in addition Li et al. (2024) [13], not only leveraged GA to select features but also reduce high dimension.

The gap in the literature is the lack of methods to generate new factors through GA to amplify data. At the same time, in addition to GA data augumentation based on common technical and fundamental factors, rolling windows and formulas can also be considered as chromosomes for symbolic genetic programming. The specific SGP algorithm will be explained in detail in the methodology section.

# **3. THE PROPOSED DEEP NEURAL NETWORK**

The network structure of our proposed SGP-DNN usually includes four phases, as shown in FIG-URE 1, followed by data preprocessing, data augmentation, feature selection, and investment benchmarking.



Figure 1: The proposed Hybrid SPG-DNN Framework

## **Phase 1: Data Preparation**

It includes four steps, which are data acquisition, data cleansing, and data pre-processing and data splitting. Typically, researchers gather data through methods such as conducting surveys or retrieving datasets from third-party providers.

The second step is to clean the data. Typical procedures are conducted such as removing noisy data or outliers, filling in missing data, and eliminating duplicate entries during this process to ensure the dataset's quality and integrity.

Data pre-processing is the third step in data preparation, involving procedures such as standardizing data through z-score normalization, categorizing data, and transforming it to the appropriate scale. This is particularly important to ensure compatibility with the various activation functions of deep learning models.

The last step involves data splitting, which includes dividing the dataset into training set, validation set and testing sets typically in a ratio of 7:2:1. This segmentation is crucial for facilitating subsequent model training and evaluating performance.

## Phase 2: Data Augmentation

In this phase, Hybrid DNN models are used to predict short-term stock price movements (e.g., over 5 days), The important step in prediction is data augmentation basing on SGP for prepared data from phase 1.

Data augmentation phase is for generating more high quality features. In this phase, the GA is used to generate the needed data. In this research, the GA was combined with Symbolic Regression which was called Symbolic Genetic Programming (SGP). In finance, SGP treats input features, operations, and a sliding window of time series data as critical genetic chromosomes, refining models for effective predictive performance. Finally, SGP creates good expressions by using selected features, adding the right windows according to customized fitness function, and using a custom filter system to get high-quality input features that can be fed into good DNN models for prediction. The elaborated methodology for DNN with SGP was demonstrated in **Section 3.2**.

## Phase 3: Feature Selection

In the feature selection stage, we usually choose MLP or LSTM network selection based on different data characteristics. The main purpose of building hybrid DNN models is to adjust the network structure according to different data characteristics. Traditionally, the Multilayer Perceptron (MLP), a key supervised learning method featuring multiple neuron layers, has been used. Nonetheless, MLPs struggle with managing sequence or time-series data, which are important for predicting stock returns that depend on historical trends. Long Short-Term Memory (LSTM) are designed to handle sequence data, thereby addressing the limitations of MLPs. The metrics like Rank IC and ICIR will be used to decide which DNN framework is more suitable according to different dataset characteristics like as shown in FIGURE 2. The main objective across all network settings is to minimize cross entropy.

The specific DNN models structures for MLP and LSTM can be demonstrated from FIGURE 3, as followings and configurations are described from TABLE 1.



Figure 2: Feature Selection: LSTM vs MLP

# **Phase 4: Investment and Benchmarking**

In this phase, building on the results of the proposed hybrid prediction model with optimized hyperparameters, a Long-short strategy or a long-only strategy is implemented based on these predictions to generate profit. Benchmarks are constructed to serve as the foundation for the research. the models developed are compared with baseline models, such as traditional single MLP or LSTM



Augmented LSTM Model



Tal	ble	1:	DNN	model	configu	rations
-----	-----	----	-----	-------	---------	---------

MLP		LSTM		
Layes	2	Layes	2	
Layer one nodes	100	Layer one nodes	100	
Layer two nodes	30	Layer two nodes	100	
Learning Rate	0.01	Learning Rate	0.01	
Dropout	0.6	Dropout	0.6	

models. Additionally, performance metrics of the DNN models including accuracy, precision, recall, Rank IC, and Rank ICIR are evaluated to assess the effectiveness and superiority of the hybrid prediction model by equations 1-5 [13].

This research primarily focuses on classification rather than regression analysis. Accuracy, precision, recall, Rank IC and Information ICIR were used as metrics. Rank IC is an important evaluation metric for constructing stock portfolios and making informed investment decisions basing on predicting cross-sectional stock returns either for alpha or relative return in your portfolio. The fluctuation range of Rank IC is between -1 and +1, but when applied to cross-sectional stock selection, -0.1 to +0.1 is the usual range. The closer it is to the upper and lower bounds, the more predictive the factor is. Similarly, ICIR, like the Sharpe ratio, is used to measure quality of Rank IC over its volatility. Usually, ICIR is between 0.4 and 0.6. When the two indicators are integrated, Rank IC with an absolute value of around 0.05 and ICIR with an absolute value of more than 0.5 will be regarded as good factor selection criteria.

$$Accuracy = \frac{Number of correct predictions}{Total number of predictions}$$
(1)

$$Precision = \frac{True Positives}{True Positives + False Positives}$$
(2)

$$Recall = \frac{True \, Positives}{True \, Positives + False \, Negatives} \tag{3}$$

$$Rank \ IC = \frac{\sum_{i=1}^{n} \left( Rx_i - \overline{Rx} \right) \left( Ry_i - \overline{Ry} \right)}{\sqrt{1 - \frac{1}{2}}}$$
(4)

$$\sqrt{\sum_{i=1}^{n} \left( Rx_i - \overline{Rx} \right)^2} \sqrt{\sum_{i=1}^{n} \left( Ry_i - \overline{Ry} \right)^2}$$

$$Information\ ratio\ of\ IC\ (ICIR) = \frac{IC}{Standadized\ Deviation\ of\ IC}$$
(5)

### 3.1 Dataset, Software and Hardware

In this research, we employed six categories of fundamental indicators in our experiments. This comprehensive collection encompasses explainable factors for all A-listed stocks, which comprise more than 4,700 stocks in China, as well as over 4657 listed stocks in the Tokyo Stock Exchange (TSE). This dataset includes al fundamental indicators.

To maintain consistency, the year 2014 was chosen as the starting point for testing, aligning with the significant revision of China's "Accounting Standards for Business Enterprises." As indicated in TABLE 9, the S&P Alpha Pool taset comprises six types of fundamental features, including growth, profitability, value, size, analyst, and efficiency, totaling 244 daily fundamental factors.

The characteristics of the Alpha pool dataset for Chinese listed stocks can be illustrated through TABLE 2, to TABLE 3. It shows the average Rank IC and ICIR of six distinct categories of quantitative indicators, respectively in China and Japan.

name of datasets	China	Japan
Value	0.0329	-0.0348
Profitability	0.0101	0.0158
Efficiency	0.0090	0.0114
Growth	0.0071	0.0034
Size	0.0068	-0.0283
Analyst Expectation	0.0043	0.0003
Fundamental Average	0.0117	0.0157

Table 2: Rank IC mean in China and Japan

name of datasets	China	Japan
Value	0.290	0.260
Profitability	0.150	-0.220
Efficiency	0.140	-0.150
Growth	0.130	-0.050
Size	0.110	-0.140
Analyst Expectation	0.050	0.000
Fundamental Average	0.140	0.140

Table 3: ICIR mean in China and J	apan
-----------------------------------	------

After examining, it is apparent that while IC fundamental indicators of Japan are slightly higher than those in China, their absolute values remain relatively low. Specifically, the mean IC for fundamental indicators is 1.17% in China and 1.57% in Japan. Similarly, the mean ICIR for fundamental indicators in both countries is 0.14.

The main goal of this research is to forecast classification of cross-sectional stocks. Our approach deviates from traditional classification methods that categorize stock returns into two groups: positive returns as '1' and negative returns as '0', which can lead to an unbalanced distribution in training data. Instead, we define our target variable by dividing it into two categories based on the median of cross-sectional stock returns. A stock return that exceeds the median is labelled as '1', indicating a higher-than-average return, while a return below the median is marked as '0', indicating a lower-than-average return. This methodology aims to provide a more balanced framework for predicting stock price movements over short periods.

Furthermore, experiments are planned to be conducted in both the Chinese and Japanese stock markets to evaluate the model's robustness across regions. Despite the similarity in the magnitude of raw indicator characteristics between the two markets, it is assumed that alpha and pattern to be discovery remains relatively consistent. This comparative analysis across regions contributes to a comprehensive understanding of the model's effectiveness in different market environments.

To give a brief explanation, numpy and pandas are used for data preprocessing and preparation. Pytorch 2.2.2 is used to build the DNN network, and GPlearn 0.0.2 is used as the implementation tool of the SGP algorithm for factor generation. Overall, the DNN network part will use the NVIDIA GPU part, and other calculations will use the CPU module.

#### 3.2 Data Augmentation: Symbolic Genetic Programming

In second phase is, SGP was used to generate new factors, and these factors were eventually feed to the DNN network after selection. In contrast to GA, which primarily concentrate on individual feature selection or parameter adjustment, SGP aims to reveal inherent relationships among individual features that may be hidden by combining generations and windows to identify patterns. Subsequently, an explanatory rationale is provided following the application of SGP. Within SGP, three primary types of chromosomes contribute to the generations: the features themselves, their potential generators, and windows.

In this research, all fundamental indicators are utilized as the basic genetic elements within the chromosome. These indicators served as the raw genetic material guiding our evolutionary process. Furthermore, Essential genes sourced from the gplearn library are integrated, consisting of 21basic mathematical functions such as addition, subtraction, division, and multiplication. Alongside these, 33 heuristic formulas are incorporated as described in TABLE 9, inspired by practitioners in finance, drawing from expert knowledge and industry practices. For example, famous hedge funds like World Quant, Cubist, and Millennium used various heuristic operators combining with raw features to generate signal to beat the market.

**Flowchart of SGP**: In the research discussed, This approach enhances adaptability and offers innovative solutions in the field of genetic programming. To improve the performance of SGP, we adopt a four-step methodology as described in FIGURE 3 [13].



Figure 3: The proposed Symbolic Genetic Programming

The initialization of the gene population is the first step in our proposed Symbolic Genetic Programming (SGP). In TABLE 6, heuristic operations, as part of the original chromosome, will also enter inheritance and promote the formation of genes.

In the modification stage of SGP approach, rolling windows are introduced for all heuristic operators, randomly set to enhance the genetic pool's diversity.

The fitness of these formulas is evaluated by their ability to solve the optimization issue, using a custom-designed fitness function that fits the specific context of the problem. Additionally, this research incorporates a specialized novel formula as metric, alongside the traditional Rank

(8)

IC, which measures the correlation between symbolic formula predictions and actual future price movements. This new formula, detailed as function 4.3, captures the high cumulative returns of the top performing stocks in cross-sectional analysis and also ensures that the returns of grouped stocks are monotonically aligned with their formula-derived ranks. The formula is shown from equation 6 to equation 8 below [13]:

$$Top_{R} = \max\left(TopR - mean\left(totalR\right), FlopR - mean\left(totalR\right)\right)$$
(6)

$$Monotonicity = max \left( \frac{1}{N} \sum_{k=1}^{N} \max(0, Sign(R_k - R_{k+1})) \right), \frac{1}{N} \sum_{k=1}^{N} \max(0, Sign(R_{k+1} - R_k)))$$
(7)

Fitness Function = 
$$Top_{\rm R} + \lambda_1 \times Monotonicity + \lambda_2 \times Information$$
  
(Default  $\lambda_1 = 0.4$  Default  $\lambda_2 = 2$ )

In our evolutionary process, Regarding the parameters of the genetic algorithm, we selected 40% likelihood as the crossover parameter. At the same time, a 40% probability is given to selection to directly copy the chromosomes of the parent. For mutations, very small probabilities are given to prevent very strange factors from being generated. This careful approach guarantees a controlled and balanced integration of new genetic material into the population, promoting stability throughout the evolutionary process. To avoid the overfitting of SGP method, the following 2 strategies have been employed as it shows in TABLE 4:

Names	numbers
generations	10
hall of fame	4000
tournament size	50
init depth	1,3
metric	'rank ic'
crossover	0.6
mutation	0.13
n_jobs	30
crit	0.04
n_jobs	30
crit	0.04
population_size	20000
n_components	50

Table 4: Parameter Settings

#### • Simplified Model Architectures:

The complexity of SGP is limited to generations (10) and depth in the model, ensuring a focus on capturing meaningful patterns rather than overfitting to noise

#### Noise-Reduction Preprocessing

We employed data smoothing and outlier filtering steps for data preprocessing to reduce the noise level for features generated by SGP while preserving meaningful patterns.

Once the previously stated algorithms have produced many symbolic formulas, the last improvement to the SGP is to apply a filtering mechanism to the results. The equation 9 to equation 10 [13], were used to choose for selecting high quality features.

Sucess Ratio of Rank IC = 
$$\frac{Numbers of Correct Pearson IC}{Total num of Pearson IC}$$
(9)  

$$IC PNL = \frac{Mean (|Pearson IC|)}{Standard devation (Pearson IC)}$$
(10)

## 3.3 Forecasting and rules of Picking Up Stocks

In the backtest phrase, the data is segmented into three phases like FIGURE 4: the training phase uses 1020 days to adjust the model parameters, the validation phase utilizes 160 days for finetuning, and the test phase is brief, consisting of only 20 days. The ratio of training to validation data is firmly set at 8.5:1. A 20-day rolling window is applied across all phases to ensure continuity and relevance in the data being analyzed. The comprehensive testing period covers 720 days, extending from November 30, 2019, to December 31, 2022, which is divided into 36 distinct trading periods. This division ensures that the model is rigorously evaluated under various market conditions, and it underscores that the data segmentation is exclusively for the purpose of validating the trading strategy.



Figure 4: Train/validation/test set

The SGP-DNN model forecasts each stock's future price change by utilising information that is currently accessible as of time t. Its main goal is for each stock to do better in the next period

t+1 than the average seen in the cross-sectional market. To do this, the model uses the anticipated return using SGP-DNN in ascending order. The stocks with the highest ranks constitute the top group, which was considered as the basis for constructing long-only portfolios in China, as shorting stocks is either restricted or associated with high fees. Conversely, in Japan, shorting is permissible, and a long-short strategy was employed to aim for absolute returns along with long-only strategy for relative returns.

The Long-Only portfolio strategy means that in each individual rebalance time period, the top k% stocks are selected according to the cross-sectional ranking value, and they are bought and held with equal weight. To evaluate this strategy's effectiveness in China, the CSI 300 and CSI 500 indices serve as the primary benchmarks which represents major 300 largest and most liquid A-share stocks and 500 small and mid-cap A share listed stocks. Both of these two indexes are weighted on market capitalization and alphas upon CSI 300 and CSI 500 were referred to respectively as Relative Return over 300 and Relative Return over 500. CSI 300 and CSI 500 are both China's broad-based stock indexes

As for Japan, DSE as a broad index as the benchmark for long-only strategy. By contrast, in Japan, the TPX and Nikkei 225 were used as broad benchmark, where Nikkei 225 includes 225 large publicly traded companies on TSE including technology, finance, automotive and retail which is price weighted while TPX includes more than 2000 companies listed in TSE which is based on market capitalization weighted. Nikkei 225 was more considered by outside-of-Japan investors. And alphas in Japan upon TPX and Nikkei225 were referred as Excess R over TPX and Excess R over Nikkei225.

Additionally, the strategy's performance was compared against an equal-weighted portfolio, termed Excess R over average, which serves as a third benchmark both in China and Japan. A critical metric to be analyzed in this research is the Sharpe Ratio of Excess R over average, which help quantify the risk-adjusted return of the portfolio.

The experiment was divided into two main section, **Section 4.1** explained on the general outcome of experiments using the fundamental indicator while **Section 4.2** explained on detailed analysis for SGP outcome. **Section 4.3** demonstrated the detailed experiment outcome from perspective of metrics for both DNN and Strategy.

## **3.4 General Outcome of Experiment**

Originally, an experiment was undertaken using fundamental indicators to investigate six metrics associated with forecasting cross-sectional stock returns. The research sought to determine whether integrating the SGP method would improve the results. Initially, the process involved applying the MLP (Multilayer Perceptron) method directly with fundamental indicators, then converting to the LSTM (Long Short-Term Memory) method. These preliminary experiments were conducted without incorporating the SGP method. Furthermore, the filter system conditions like section 3.2 described for SGP for fundamental indicators corresponded to those outlined in TABLE 5, for China and Japan:

Typically, a low average value of Information Coefficient (IC) is observed in such scenarios for individual features, with the optimal average value being above 6%. In this research, the average

Filter Items	China	Japan
Rank IC	7%	6%
Success Ratio	75%	70%
IC PNL	1.8	1.65

Table 5:	Filter	Settings	for	SGP
----------	--------	----------	-----	-----

values Success Ratio is more than 70%, rank IC being above 6% and IC PNL more than 1.5 are considered acceptable for predicting cross-sectional stock selection all over the countries, and specific filter items was provided in TABLE 5, for China and Japan individually due to its characteristics of its raw features.

It is evident that SGP proves significant advantages for fundamental indicators. Following the SGP process, the FIGURE 5, and FIGURE 6, provides a concise comparison of Rank IC across the four models over the same period in China and two models in Japan, based on their mean averages with raw values. Additionally, FIGURE 7 and FIGURE 8, described the Information Coefficient Information Ratio (ICIR) for these models in both countries.



Figure 5: Rank IC in China

The insights from FIGURE 8, reveal that in China utilizing raw fundamental indicators as inputs for LSTM or MLP models yielded an original Rank IC for fundamental indicators of -1.17%, serving as the baseline. As can be seen from FIGURE 6, the Rank IC of separate LSTM and MLP is only -2.89% and -1.86%, and the MLP is even lower, only -1.86%. However, after integrating SGP, the Rank IC of SGP-LSTM and SGP-MLP correspond to -7.98% and -8.05%, respectively, which are far stronger than the independent DNN model. The similar trend can also be observed in Japan from FIGURE 6, the Rank IC decreased to -4.62% after SGP with LSTM comparing with the -1.57%.

The result demonstrates significant improvements in Rank Information coefficient (Rank IC) by 588.03% in China and 194.27% in Japan respectively when applied to fundamental indicators,



Figure 6: Rank IC in Japan



Figure 7: ICIR for 4 Models in China

however, the magnitude of Rank IC of China is bigger than that in Japan which means less alpha in long-only or long-short strategy later.

This trend is similarly reflected in the ICIR shown in FIGURE 7, and FIGURE 8, the original ICIR for both countries are -0.14 from TABLE 3, and TABLE 4, where the two DNN models integrated with SGP outperform the two individual DNN models. Specifically, in terms of ICIR, SGP-LSTM demonstrates superior performance compared to LSTM (-5.6 vs. -3.35) in China and LSTM model (-3.25 vs. -1.31) in Japan although the magnitude of China is also bigger than that in Japan.



Figure 8: ICIR in Japan

# 3.5 SGP Analysis

After comparison, we confirmed the effectiveness of SGP for the DNN model. In this section, the SGP procedure was elaborated to understand why and how SGP contributes to improving prediction power.

TABLE 6 and TABLE 7, displays the top 9 selected features obtained through SGP in China and Japan. A total of 1893 synthetic features were selected to be fed into the DNN network in China and a total of 880 synthetic features for Japan. Analysis of TABLE 6, to 7, reveals the structures of these selected features, primarily comprised of three components: the raw feature, generators, and rolling windows, as detailed in the methodology section.

As for Rank IC and Success Raito, this expectation is confirmed by the average Rank IC exceeding 8.7% in TABLE 6, compared to the original rank IC of only 1.17% for the original fundamental indicators in China and average Rank IC 6.2% in Japan comparing with the original 1.57% from TABLE 7. For average the SGP is more applicable for China than that in Japan from the perspective of Rank IC and Success Ratio.

TABLE 8 outlines the most frequently used operations for the selected features, indicating that nearly all predominant operations are heuristic. Many common generators are both selected both for China and Japan such as rank\_mul, zscore, rank\_add which are all heuristic generators.

TABLE 9 identifies the most utilized features, including style attributes such as size (LogMktCap, LogMktCapCubed), value (BP), profitability (AdjEBITP and CFOIC) and analyst-related metrics (AdjEPSNumRevFY2C, RevMagFY1C, etc.). These findings suggest that SGP aims to construct optimal features by combining value, size and analyst factors in China and combining profitabity (AdjEBITP, REToAst), size (LogMktCapCubed) and analyst (AdjEPSNumRevFY2C) factors in

China like normal anomalies [14–16], but in a nonlinear manner, surpassing the original fundamental indicators. Finally, TABLE 10 reveals that the frequency of a 5-day window constitutes over 30% in both countries, indicating its significance as a forecast window, aligning with a standard 5-day working period.

Formula	Rank IC	Success Ratio	PNL
ts_max(ts_min_diff(rank_mul(rank(BP),	8.93%	75.32%	2.23
LogMktCapCubed), RevMagFY1C), 5)			
ts_max(ts_min_diff(rank_mul(rank(BP),	8.93%	75.32%	2.23
LogMktCapCubed), 16), 5)			
ts_max(ts_min_diff(rank_mul(BP, LogMktCapCubed),	8.93%	75.32%	2.23
RevMagFY1C), AdjEPSNumRevFY2C)			
ts_max(ts_min_diff(rank_mul(BP, LogMktCapCubed),	8.93%	75.32%	2.23
RevMagFY1C), 5)			
ts_min_diff(rank_mul(rank(BP), LogMktCapCubed),	8.87%	78.45%	1.92
IndRel_LTDE)			
ts_stddev(ts_min_diff(rank_mul(BP, LogMktCapCubed),	8.72%	75.32%	1.91
RevMagFY1C), 10)			
ts stddev(ts min diff(rank mul(rank(BP),	8.72%	75.32%	1.91
LogMktCapCubed), 16), 10)			
ts_stddev(ts_min_diff(rank_mul(BP, LogMktCapCubed),	8.72%	75.32%	1.91
RevMagFY1C), EPSEstDispFY2C)			
ts max(ts min diff(rank mul(rank mul(rank(BP),	8.71%	77.35%	2.27
LogMktCapCubed), LogMktCapCubed), 16), 9)			

Table 6: High	Quality	<sup>y</sup> Synthetic	Factors	From	SGP	in	China
---------------	---------	------------------------	---------	------	-----	----	-------

## 3.6 Results and Discussion

Subsequently, to evaluate the effectiveness of the SGP method, further experiments were conducted using the LSTM and MLP methods in conjunction with SGP. The results of these experiments in China and Japan are presented in TABLE 11, and TABLE 12, the case in China was conducted intensively with the integration of LSTM or MLP with SGP denoted as SGP-LSTM and SGP-MLP, respectively and the experiment in Japan followed the way of Chinese case only focusing on SGP-LSTM for comparison. The outcomes obtained in China without incorporating the SGP method are displayed in the columns labelled as LSTM and MLP.

TABLE 13 summarizes the outcomes of an experiment conducted using data from 2020 to 2022 for both the Chinese and Japanese markets. Additionally, FIGURE 9 and FIGURE 10, depict the Excess R values for three benchmarks in each country, providing significant insights for investment applications. The study examines performance across two broad indices and the mean performance of stocks in China (average, HS300, and CSI500) and Japan (average, N225, and TPX). In China, the SGP-MLP model delivered exceptional returns, outperforming the CSI 300 index by 24.61% and the CSI 500 index by 17.53%. This result also achieved an excellent performance that exceeded the average portfolio by 10.70%. Such achievements have already ranked in the top 10% of China's

Formula	Rank IC	Success ratio	PNL
sub(ts_min_diff(LogMktCapCubed, 9),	6.43%	71.31%	1.83281
rank_mul(rank_mul(ts_return(AdjEBITP,			
AdjEPSNumRevFY1C),			
<pre>sub(ts_max_diff(neg(LogMktCap), CFROIC),</pre>			
ROEStddev20Q)), AdjEBITP))			
sub(ts_min_diff(LogMktCapCubed, EPSEstDispFY1C),	6.36%	70.77%	1.88268
rank_mul(rank_mul(ts_return(AdjEBITP, 4),			
rank_mul(ts_max_diff(sub(ts_min_diff(LogMktCapCubed,			
Revinage Y IC), sigmoid(RE IOAst)), CFROIC),			
KEIOASL_2)), AUJEBIIP)) sub(ta_min_diff(LagMlstCanCubad_EDSEstDisnEV1C)	6 2 2 0/	70 640/	1 9600
rank mul(rank mul(te return(AdiEBITP 4)	0.32%	/0.04%	1.8009
rank_mul(ts_max_diff(sub(ts_min_diff(LogMktCanCubed			
EPSEstDisnEV1C) sigmoid(REToAst)) CEROIC)			
RETOAST 2)) AdiEBITP))			
add(sub(add(ROEStddev200.	6.26%	71.72%	1.66422
ts min diff(LogMktCapCubed, AdjRevMagC)),			
AdjEBITP), ts min diff(LogMktCapCubed, 11))			
<pre>sub(sub(ts_min_diff(LogMktCapCubed, RevMagFY1C),</pre>	6.25%	72.12%	1.75898
rank_mul(ts_return(AdjEBITP, 4),			
<pre>sub(ts_max_diff(neg(LogMktCap), 27),</pre>			
ROEStddev20Q))), AdjEBITP)			
sub(ts_min_diff(LogMktCapCubed, EPSEstDispFY1C),	6.25%	71.58%	1.75523
rank_mul(rank_mul(delta(CashP, AdjEPSNumRevFY2C),			
sub(ts_max_diff(neg(LogMktCap), CFROIC),			
ROEStddev20Q)), AdjEBITP))	C 100/	71 100/	1 72010
add(ROEStddev20Q, sub(ts_min_diff(LogMktCapCubed,	6.18%	/1.18%	1./3818
EPSESIDISPFYIC), renk mul(renk mul(te return(AdiEDITD 4)			
rank_mul(ta_may_diff(neg(LogMktCan)_CEPOIC)			
RFToAst 2)) AdjERITP)))			
rank div(rank mults return(AdiEBITP 4)	6 16%	73 88%	1 85976
rank mul(neg(ts min diff(LogMktCap 11)) REToAst))	0.1070	/5.00/0	1.00970
sub(ts min diff(LogMktCapCubed, 9), AdiEBITP))			
add(ROEStddev20Q, sub(ts min diff(LogMktCapCubed,	6.16%	72.12%	1.64714
9), rank mul(ts return(AdjEBITP,			
AdjEPSNumRevFY1C), rank_sub(rank_div(AdjEBITP,			
ts_min_diff(LogMktCap, 11)), ROEStddev20Q))))			

Table 7: High Quality Synthetic Factors From SGP in Japan

public funds. Similarly, in Japan, the SGP-MLP model surpassed TPX by 5.10%, N225 by 4.56%, and the average by 4.33%, as detailed in TABLE 13.

China		Japan	n		
Operator	count	operator	count		
ts_stddev	2527	rank_mul	417		
ts_min_diff	1866	neg	353		
zscore	1557	ts_return	292		
rank	1481	ts_max_diff	273		
rank_mul	956	rank_add	223		
ts_max	528	ts_min_diff	215		
ts_nanmean	29	zscore	180		
sigmoid	22	sigmoid	123		
rank_add	22	delta	116		
ts_sum	13	sub	74		
winsorize	11	rank_sub	50		
ts_median	4	add	46		
ts_min_max_cps	3	rank_div	24		
neg	2	ts_min_max_cps	14		

Table 8: Operator Summary from SGP

Table 9: Features Summary from SGP

China			Japan		
Features	count	Rank IC	feature	count	Rank IC
LogMktCap	2257	2.94%	AdjEBITP	416	3.58%
LogMktCapCubed	1996	2.84%	REToAst_2	327	2.81%
AdjEPSNumRevFY2C	606	-0.30%	LogMktCap	309	-1.77%
BP	588	-3.46%	CFROIC	134	0.94%
RevMagFY1C	415	-0.98%	AdjEPSNumRevFY2C	120	0.09%
BuyToSellRecLess3MSMA	266	-0.80%	LogMktCapCubed	111	-1.72%
ROIC	262	-0.27%	REToAst	75	2.81%
AE-style	251	0.43%	ROEStddev20Q	61	-2.77%

As shown in TABLE 13, the SGP-DNN models demonstrated superior performance compared to individual MLP and LSTM models without data augmentation in China, highlighting the effectiveness of the SGP methodology in enhancing investment strategy performance metrics. Similarly, the SGP-DNN models outperformed the single DNN models in terms of the information ratio in China. Despite incorporating more variables, the Excess R for SGP-LSTM did not surpass that of SGP-MLP, likely due to fundamental indicators experiencing minimal changes over 10-day periods as lags. In the Japanese market, the Excess R for the suggested SGA-LSTM model was 4.33%. Notably, despite a notable difference in Excess R magnitude between China and Japan, the information ratio in Japan (1.27) compares favourably to that in China (1.49), indicating consistent performance relative to risk-adjusted returns across both markets. These findings highlight the effectiveness of the proposed models in enhancing investment strategies in diverse market contexts.

	China			Japan	
Windows	Frequency	Percentage	window	frequency	percentage
5	847	30.76%	5	181	33.64%
3	385	13.98%	27	131	24.35%
8	329	11.95%	4	76	14.13%
16	316	11.47%	11	69	12.83%
21	297	10.78%	9	30	5.58%
13	174	6.32%	6	18	3.35%
6	164	5.95%	28	13	2.42%
7	77	2.80%	8	4	0.74%
19	42	1.53%	3	4	0.74%
9	39	1.42%	16	3	0.56%
27	24	0.87%	7	3	0.56%
4	23	0.84%	17	3	0.56%
10	22	0.80%	26	3	0.56%

Table 10: Rolling Windows summary for SGP

# Table 11: Metric Summary for SGP in China

Fundamental Indicators						
	Metric	SGP-MLP	SGP-LSTM	MLP	LSTM	
	Rank IC	-0.0715	-0.0670	-0.0295	-0.0246	
	ICIR	-4.2000	-3.9000	-2.7900	-2.2500	
2020	Excess R above 300	-0.0787	-0.0583	0.0139	-0.0353	
2020	Excess R above 500	-0.0335	-0.0120	0.0628	0.0115	
	Excess R above average	0.0027	0.0239	0.1001	0.0484	
	Sharp ratio	0.0400	0.3400	1.8400	1.0100	
	Rank IC	-0.0713	-0.0725	-0.0104	-0.0213	
	ICIR	-4.2400	-5.5800	-0.8000	-2.7200	
2021	Excess R above 300	0.4571	0.4129	0.3196	0.2385	
2021	Excess R above 500	0.2140	0.1749	0.0992	0.0291	
	Excess R above average	0.1277	0.0883	0.0162	-0.0461	
	Sharp ratio	1.6700	1.3800	0.2700	-0.9100	
	Rank IC	-0.0986	-0.0999	-0.0158	-0.0407	
	ICIR	-6.1100	-7.3200	-1.2600	-5.0800	
2022	Excess R above 300	0.3599	0.3459	0.2057	0.2060	
2022	Excess R above 500	0.3454	0.3312	0.1896	0.1916	
	Excess R above average	0.1905	0.1784	0.0527	0.0546	
	Sharp ratio	2.7700	3.0900	0.6900	1.2600	

In addition, as TABLE 14, shows over three-year period, the accuracy and precision of single LSTM model has been improved from (51.2%, 50.80%) which was recorded as 51.4% in Fisher's model

	Metric	SGP-LSTM
	Rank IC	-0.0189
	ICIR	-1.1711
	Excess R above Average	0.0322
2020	Long-Short (Absolute R)	0.0476
	Excess R above TPX	(0.0109)
	Excess R above N225	(0.1152)
	Sharp ratio	0.55
	Rank IC	-0.0301
	ICIR	-1.9421
	Excess R above Average	0.0593
2021	Long-Short (Absolute R)	0.1209
	Excess R above TPX	0.0359
	Excess R above N225	0.0804
	Sharp ratio	2.10
	Rank IC	-0.0227
	ICIR	-1.3260
	Excess R above Average	0.0383
2022	Long-Short (Absolute R)	0.0705
	Excess R above TPX	0.1280
	Excess R above N225	0.1716
	Sharp ratio	1.15

Table 12: Metric Summary for SGP in Japan

Table 13: Metric Summary for SGP in China and Japan

		С	hina			Japa	in
	Metric	SGA-MLP	SGA-LSTM	MLP	LSTM	Metric	SGA-LSTM
	Excess R above 300	24.61%	22.35%	18.00%	13.48%	Excess R above TPX	5.10%
Average of mean from 2020 to 2022	Excess R above 500	17.53%	16.26%	12.07%	7.74%	Excess R above N225	4.56%
	Excess R above average	10.70%	9.89%	5.80%	1.86%	Excess R above average	4.33%
	information ratio	1.49	1.47	0.86	0.38	information ratio	1.27

[6], in US to (52.80%, 53.40%) in China and rank IC was improved dramatically from 1.63% to 8.07%, as a result the Excess Return was improved from negative figure to 14.32% in China which beat the Fisher's model dramatically without rolling windows. After rolling windows the excess R above average was 9.89% and information ratio is 1.47 like , shown in TABLE 13.

In contrast, TABLE 14 and FIGURE 11, describe the performance and metrics of LSTM with SGP in Japan. TABLE 15 compares a single LSTM model without rolling to the proposed SGP-LSTM



Figure 9: Excess R Comparison in China



Figure 10: Excess R Comparison in Japan

Metric	Single LSTM Model without Rolling	Proposed SGP-LSTM Model without Rolling
Rank IC	1.63%	8.07%
Accuracy	51.2%	52.80%
Precision	50.8%	53.4%
Recall	15.1%	31.3%
Excess Return	-0.79%	14.32%
Information Ratio	-0.11	2.33

Table 14. Methe Summary for SOT in China	Table	14:	Metric	Summary	for	SGP	in	China
--	-------	-----	--------	---------	-----	-----	----	-------

model with and without rolling. It is evident from TABLE 18, that metrics such as Accuracy, Recall, and Rank IC have significantly improved using the SGP methodology. Consequently, both the

Li Qi, et al.

Excess R and information ratio have shown substantial improvements. Additionally, the robustness of the model with a rolling window and without a rolling window can be observed, as demonstrated for two sliding designing in FIGURE 2, and FIGURE 7.

Metric	Single LSTM Model without Rolling	Proposed SGP-LSTM Model without Rolling	Proposed SGP-LSTM Model with Rolling
Rank IC	1.66%	4.62%	2.39%
Accuracy	52.23%	55.28%	54.22%
Precision	48.74%	48.55%	49.38%
Recall	42.89%	58.03%	67.83%
Excess Return	1.55%	4.56%	4.33%
Information Ratio	0.31	1.40	1.26

Table 15: Metric Summary for SGP in Japan with and without Rolling

The metrics of DNN model can be observed that as shown in TABLE 15, in Japan, the original metric like accuracy with single LSTM without SGP and rolling windows is 52.23% and its information ratio of long-only strategy is 0.31, after with SGP, the accuracy of proposed SGP-LSTM model without rolling and with rolling is improved to 55.28% and 54.22% respectively and information ratio is improved to 1.40 and 1.26 individually



Figure 11: The cumulative return curves of Proposed Model VS TPX&N225

Over this three-year period, the SGP-LSTM model in Japan witnessed a total return of 27.66%, as shown in FIGURE 11. This performance compares favourably with TPX achieving 10.15% and N225 registering 10.82%. These results underscore the effectiveness of the proposed SGP-LSTM model in capturing investment opportunities in the Japanese market.

# 4. CONCLUSION

The research introduced a novel approach aimed at improving the prediction of cross-sectional stock returns through the utilization of SGP for generating high-quality features and integrating it with DNN models. The findings indicated significant improvements in prediction accuracy and Rank IC. Notably, a hybrid model that combined SGP with LSTM consistently outperformed market returns based on a simplistic rule-based strategy. In comparison to the CSI 300 & CSI500 and TPX & N225, and average portfolio, the hybrid SGP-LSTM model that was suggested produced average annualised excess returns of 10.70 percent, 24.61%, and 17.53 percent in China and annualised excess returns of 4.56 percent, 5.10%, and 4.33 percent, respectively. These results highlight how well the suggested methodology can be used to create profitable investment strategies and provide guidance on how to overcome obstacles with data integration and feature selection.

While the initial research primarily concentrated on financial time series data, more recent studies have broadened the scope to incorporate a variety of sources such as technical indicators, fundamental indicators, and price series data. Furthermore, by combining various data sources, hybrid DNN models for sentiment characteristics[17], could be merged. Finally, the primary limitation of this paper is that the synthesized features through SGP lack economic or financial significance. To address this in the future research, on the one hand, we can impose scenario-based restrictions, such as prohibiting chromosomes with certain non-meaningful feature or partial formulas from progressing to the next generation of evolution, so as to ensure economic interpretability. At the same time, we can add screening methods to filter out factors with high predictability and financial or economic relevance into the DNN network. Finally, we substitute the synthetic factors into historical scenarios with stress tests to verify their economic implications under extreme scenarios.

### References

- [1] Kumbure MM, Lohrmann C, Luukka P, Porras J. Machine Learning Techniques and Data for Stock Market Forecasting: A Literature Review. Expert Syst. Appl. 2022;197:116659.
- [2] Bowden RJ. Non-linearity and Non-stationarity in Dynamic Econometric Models. Rev Econ Stud. 1974;41:173-179.
- [3] Munir Q, Ching KS, Furouka F, Mansur K. The Efficient Market Hypothesis Revisited: Evidence From the Five Small Open ASEAN Stock Markets. Singap Econ Rev. 2012;57:1250021.
- [4] Zhang XD, Li A, Pan R. Stock Trend Prediction Based on a New Status Box Method and Adaboost Probabilistic Support Vector Machine. Appl Soft Comput. 2016;49:385-98.
- [5] Levy M, Persky N, Solomon S. The Complex Dynamics of a Simple Stock Market Model. Int J High Speed Comput. 1996;8:93-113.

- [6] Fischer T, Krauss C. Deep Learning With Long Short-Term Memory Networks for Financial Market Predictions. Eur J Oper Res. 2018;270:654-669.
- [7] Ghosh P, Neufeld A, Sahoo JK. Forecasting Directional Movements of Stock Prices for Intraday Trading Using LSTM and Random Forests. Fin Res Lett. 2022;46:102280.
- [8] Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock Price Prediction Using LSTM, RNN and CNN-Sliding Window Model. In 2017 international conference on advances in computing, communications and informatics (ICACCI).IEEE. 2017:1643-1647.
- [9] Baek Y, Kim HY. Modaugnet: A New Forecasting Framework for Stock Market Index Value With an Overfitting Prevention LSTM Module and a Prediction LSTM Module. Expert Syst. Appl. 2018;113:457-480.
- [10] Shen J, Shafiq MO. Short-Term Stock Market Price Trend Prediction Using a Comprehensive Deep Learning System. J. Big Data. 2020;7:1-33.
- [11] Chen S, Zhou C. Stock Prediction Based on Genetic Algorithm Feature Selection and Long Short-Term Memory Neural Network. IEEE Access. 2020;9:9066-9072.
- [12] Shahvaroughi Farahani M, Razavi Hajiagha SH. Forecasting Stock Price Using Integrated Artificial Neural Network and Metaheuristic Algorithms Compared to Time Series Models. Soft Comput. 2021;25:8483-8513.
- [13] Li Q, Kamaruddin N, Yuhaniz SS, Al-Jaifi HA. Forecasting Stock Prices Changes Using Long-Short Term Memory Neural Network With Symbolic Genetic Programming. Sci Rep. 2024;14:422.
- [14] Ang JS, Ciccone SJ. Analyst Forecasts and Stock Returns. Available at SSRN : https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=271713
- [15] Berk JB. A Critique of Size-Related Anomalies. The review of financial studies. 1995;8:275-286.
- [16] Ohlson J. Earnings, Book Values, and Dividends in Equity Valuation. Contemp Account Res. 1995;11:661-687.
- [17] Swathi T, Kasiviswanath N, Rao AA. An Optimal Deep Learning-Based Lstm for Stock Price Prediction Using Twitter Sentiment Analysis. Appl Intell. 2022;52:13675-13688.