



A data-driven deep learning approach incorporating investor sentiment and government interventions to predict post-crash stock return in China's A-share market

Weiran Lin^a, Haijing Yu^b , Liugen Wang^{c,*}

^a Department of Statistics and Data Science, Hangzhou City University, Hangzhou, 310015, China

^b Business School, Xiangtan University, Xiangtan, 411105, China

^c Zhejiang Shanyuan Hedge Fund Management CO., LTD, Zhejiang, 311121, China

ARTICLE INFO

JEL codes:

G17
G18
C45
C53

Keywords:

Stock market crash
Investor sentiment
Government intervention
Deep learning
Stock return prediction

ABSTRACT

Global financial markets frequently experience extreme volatility, which poses significant challenges in forecasting stock returns, particularly following market crashes. Traditional models often falter under these conditions due to heightened investor sentiment and strong regulatory interventions. Predicting individual stock returns after a crash is especially challenging in China's A-share market, which is characterized by high volatility and active government involvement. Although deep learning has advanced stock return forecasting, most studies have focused on general market conditions or relied solely on sentiments extracted from texts, leaving firm-level government intervention metrics largely unaddressed. To bridge this gap, we propose a novel deep learning framework that leverages historical post-crash data ("distant relative data") to forecast future stock returns. Unlike conventional methods that rely on recent pre-crash data—often overlooking government interventions—our approach leverages post-crash data, where investor sentiment and regulatory responses are already reflected, to model stable relationships between financial and momentum factors and subsequent returns, thereby implicitly integrating the effects of government interventions on investor behavior. We validate our framework using data from four distinct "thousand-stock limit-down" events in China's A-share market from 2018 to 2023. For the Fully Connected Neural Network (FCNN) model, training with close neighbor data yielded average F1-scores of 0.219 (2019), 0.106 (2020), and 0.282 (2022), whereas using distant relative data improved these to 0.571 (2019), 0.311 (2020), and 0.412 (2022). Notably, incorporating two distant relative datasets further boosted the FCNN F1-scores to 0.627 and 0.533 for 2020 and 2022, respectively. Additionally, Long Short-Term Memory (LSTM) networks consistently outperform FCNN models, underscoring their advantages in capturing temporal dependencies. Overall, our findings indicate that leveraging multiple historical crisis data sets significantly enhances post-crash stock return predictions. This data-driven approach, analogous to the stand-alone application of SMOTE for data balancing, offers a robust framework that can be integrated with other post-crisis models, thereby providing promising directions for future research and practical implementation.

Introduction

Global financial markets frequently experience extreme volatility, attracting significant attention from investors and regulators regarding post-crash trends (Heiden, Klein & Zwergel, 2013; Petit, Lafuente & Vieites, 2019; Herculano & Lutkebohmert, 2023). In emerging markets, such as China's A-share market—characterized by high turnover rates and pronounced volatility—investor sentiment and government interventions are crucial for maintaining price stability (Huang, Miao &

Wang, 2019; Zhou, Li, Zhang & Xiong, 2022; Cheng, Jin, Li & Lin, 2022). During major crises, ranging from the 2000 tech bubble crash to the 2020 COVID-19 outbreak, governments implemented various direct and indirect measures, sparking ongoing debates on moral hazards and the appropriate extent of state involvement in financial markets.

Although numerous studies have applied machine learning, deep learning, generative adversarial networks, and reinforcement learning to forecast stock returns (Li, Ni & Chang, 2019; Faraz & Khaloozadeh, 2020; Yu & Yan, 2020; Maqsood et al., 2020; Khattak et al., 2023; Liang,

* Corresponding author.

E-mail addresses: linwr@hzcu.edu.cn (W. Lin), wanglg@syinvestment.com (L. Wang).

<https://doi.org/10.1016/j.jik.2025.100704>

Received 10 November 2024; Accepted 9 April 2025

Available online 23 April 2025

2444-569X/© 2025 The Authors. Published by Elsevier España, S.L.U. on behalf of Journal of Innovation & Knowledge. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Xu, Hu & Du, 2023; Wu, Gao, Su, Liang & Du, 2023; Naveed, Yao, Memon, Ali, Alhussam & Sohu, 2023; Suárez-Cetrulo, Quintana & Cervantes, 2024; Liu et al., 2025), few have focused specifically on post-crash scenarios, where traditional models often fail due to the disruptive influence of government actions and heightened investor sentiment.

Although investor sentiment is effectively quantified via text analysis of social media platforms (e.g., Twitter, Weibo) and integrated into deep learning frameworks (Fu, Zhou, Liu & Wu, 2020; Lu, Liu & Chen, 2021; Liu, Huynh & Dai, 2021; Cheng et al., 2022; Maqsood et al., 2022; Yang, Wang & Li, 2022; Wang, Hu, Li, Ho & Cambria, 2023; Katsafados, Nikoloutsopoulos & Leledakis, 2023; Jena & Majhi, 2023; Cruz, Kinyua & Mutigwe, 2023; Liang, Xu, Hu & Du, 2023; Wu, Gao, Su, Liang & Du, 2023; Naveed, Yao, Memon, Ali, Alhussam & Sohu, 2023), firm-level government intervention metrics remain largely underexplored. While previous studies have examined government interventions for their effects on market efficiency during crises, such as COVID-19 (Huynh, Nguyen & Dao, 2021; Goel & Dash, 2021; Wang, Zhang, Gao & Yang, 2021; Nguyen, Wu, Ke & Liao, 2022; Yang, Abedin, Zhang, Weng & Hajek, 2023; Zaremba, Kizys, Tzouvanas, Aharon & Demir, 2021; Li, Chen, Xu & Zhang, 2023) and historical Chinese stock market crashes (Li, 2019; Cheng, Jin, Li & Lin, 2022; Zhou, Li, Zhang & Xiong, 2022; Li, Yuan & Jin, 2023, 2024; Shahab, Wang, Yeung & Zhou, 2024), most analyses are limited to industry or market-wide assessments.

At the firm level, Cheng, Jin, Li and Lin (2022)) employed a panel model that used a dummy variable for “national team” shareholding to assess government interventions during the 2015 crash. However, it is important to note that, during market crashes, the “national team” typically targets blue-chip stocks—those with high weightings in indices, such as the CSI 300 or SSE 50 (e.g., Industrial and Commercial Bank of China)—and most A-share stocks do not receive direct intervention. Consequently, using the national team shareholding ratio as a firm-level feature is impractical for the majority of stocks. Moreover, few studies have integrated government actions into firm-level return forecasting, particularly within deep learning frameworks, even though interventions often extend beyond direct stock purchases to include monetary and administrative measures. This complexity underscores the challenge of identifying suitable proxies for capturing the influence of government intervention on individual stocks in post-crash prediction models.

To address this gap, we propose a novel data-driven deep learning framework that integrates investor sentiments, government interventions, and their complex interrelations. The core idea is to leverage historical post-crash stock price data from previous “thousand-stock limit-down” events—data that inherently capture investor behavior and regulatory responses—to model current market conditions. We evaluate our approach by comparing it with a baseline model trained using data from the most recent crisis (close neighbor data) and incorporating over 100 traditional, quantifiable financial, and momentum factors. Notably, the close-neighbor data are from periods prior to a crisis, during which government interventions are typically absent; therefore, the baseline model does not capture the effects of such interventions. By contrast, our proposed framework employs the same factors but uses historical post-crash data as the training input, thereby implicitly incorporating the impact of government interventions on investor sentiment. In extreme market conditions, such as thousand stock limit-down events, government interventions and investor sentiment are critical drivers of stock returns. If our method significantly improves the prediction accuracy relative to the baseline, it would indicate that historical post-crash data contain valuable information, specifically regarding government intervention, which is missing from recent data. In other words, using historical post-crash (distant relative) data effectively captures the impact of government intervention on investor behavior.

Specifically, we analyzed four such events in China’s A-share market from 2018 to 2023, during which investors exhibited extreme

pessimism, prompting government intervention. These historical data were then incorporated into our deep learning models to forecast one-month stock returns following new crises. To ensure reliability, we experimented with multiple neural network architectures and parameter configurations, repeating each configuration 10 times, and averaging the results. Our approach is analogous to the SMOTE method for balancing data because it directly processes raw data. This strategy not only improves predictive accuracy, but also provides a new perspective on post-crisis research, enabling easier integration with other post-crash modeling frameworks.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature, Section 3 describes the data and methodology, Section 4 presents the empirical findings, and Section 5 concludes with the key takeaways.

Literature review

Investor sentiment and machine learning

Behavioral finance merges psychology and finance to illuminate the irrational elements that drive market behavior. Building on the prospect theory (Kahneman & Tversky, 1979), it contrasts with traditional models, such as the Capital Asset Pricing Model (CAPM), which often fail to capture post-crash dynamics (Shiller, 2003; Yalamova & McKelvey, 2011; Sharma & Kumar 2020; Kanzari, Gaies & Sahut, 2023). This shortfall underscores the need to delve deeper into both the economic and psychological roots of stock volatility, highlighting the rising importance of behavioral finance in understanding how investors’ psyche impacts market fluctuations. Studies have consistently shown that emotions and cognitive biases significantly influence stock returns and elucidate the mechanisms through which investors deviate from rational expectations (Barberis, Shleifer & Vishny, 1998; Daniel, Hirshleifer & Subrahmanyam, 1998; Baker & Wurgler, 2007; Hirshleifer, Lim & Teoh, 2009; Ni, Wang & Xue, 2015; Guo, Sun & Qian, 2017; Li & Li, 2021; Kim & Lee, 2022; Gong, Zhang, Wang & Wang, 2022; Song, Gong, Zhang & Yu, 2023; Li & Xing, 2023).

Technological revolution, particularly in machine learning, has significantly advanced the analysis and prediction of market trends. Loughran and McDonald (2011) pioneered the use of natural language processing (NLP) to decipher corporate sentiment in financial reports, marking the early integration of technology in behavioral finance. Building on this foundation, Huang, Lehmann and Zhu (2017) demonstrated the efficacy of machine learning models in identifying investor emotions and cognitive biases, offering valuable insights for market trend forecasting. This trajectory of innovation continued with Buehlmaier and Whited (2018), who applied machine-learning techniques to social media data, thereby enhancing the capacity to predict market fluctuations. Viebig (2020) conducted a notable study that uncovered a relationship between high equity market valuations and subsequent low returns, utilizing machine learning tools, such as support vector machines, to detect “clustering patterns” that signal irrational market exuberance. Extending the scope of this research, Gupta, Nel and Pierdzioch (2023) used random forest algorithms to highlight the significant predictive value of investor confidence, especially its uncertainty aspect, in determining the volatility of the US stock market from 2001 to 2020, providing key insights for both investors and policymakers in navigating market complexities.

Deep learning, particularly long short-term memory (LSTM) networks, has played a vital role in advancing stock market forecasting. Studies demonstrate that integrating behavioral factors and sentiment analysis improves model accuracy and interpretability. Chen et al. (2014) and Fischer and Krauss (2018) emphasized the effectiveness of LSTM in capturing complex market patterns. Recent studies, such as those by Srivastava, Zhang and Eachempati (2021); Bhuyan, Jaiswal and Cherif (2023); and Kanzari, Gaies and Sahut (2023), have further leveraged machine learning algorithms, including support vector

machines and deep neural networks, to predict market trends and financial system stability. Utilizing diverse data sources, ranging from technical indicators to sentiment indices and EEG data, these studies demonstrate the efficacy of deep and machine learning in various aspects of market analysis, ranging from the emotional impact on day traders to macroprudential policymaking.

Government intervention and stock market

Government interventions in financial markets have garnered considerable research interest owing to their profound impact on market stability and efficiency. During crises, such as COVID-19, a variety of policy actions—ranging from monetary easing to administrative controls—have been deployed to contain panic and restore investor confidence (Huynh, Nguyen & Dao, 2021; Goel & Dash, 2021; Wang, Zhang, Gao & Yang, 2021; Nguyen, Wu, Ke & Liao, 2022; Yang, Abedin, Zhang, Weng & Hajek, 2023). These studies typically examine macro-level indicators, such as aggregate returns, volatility indices, or sector-wide trends to assess how interventions influence market outcomes (Zaremba, Kizys, Tzouvanas, Aharon & Demir, 2021; Li, Chen, Xu & Zhang, 2023). In particular, government rescue measures have been scrutinized for their ability to avert systemic collapse in global equity markets, with machine learning methods (including panel regression and factor analysis) increasingly employed to quantify these policy effects (Hale, Petherick, Phillips & Webster, 2020; Zaremba, Kizys, Tzouvanas, Aharon & Demir, 2021).

Beyond COVID-19, research on historical crashes in China's stock market has further illuminated how government intervention influences investor behavior. For example, Li (2019) and Li, Yuan, Jin and Yuan (2024) investigated how regulatory and administrative measures affect both short- and long-term price efficiency, demonstrating that government-led “bailout” strategies can provide market stability but also introduce potential distortions. In a similar vein, analyses of the 2015 Chinese stock market crash highlight the critical role of “national team” funds in bolstering liquidity and mitigating crash risk (Cheng, Jin, Li & Lin, 2022; Zhou, Li, Zhang & Xiong, 2022; Li, Yuan & Jin, 2023; Shahab, Wang, Yeung & Zhou, 2024). Although these investigations often center on industry-level or overall market indicators, demonstrating that targeted government purchases and liquidity injections can stabilize short-term sentiment, few have delved into how policy measures affect individual firms. This omission is noteworthy, because firm heterogeneity may significantly alter the efficacy of government interventions.

A few exceptions explore firm-level dimensions of state involvement. For example, Cheng, Jin, Li, & Lin (2022)) utilized a panel model with a dummy variable to indicate “national team” holdings at quarter-end, thus capturing how direct share purchases affect specific firms during crashes. However, such a binary indicator cannot fully reflect the diverse intervention channels, including administrative bans, trading restrictions, and monetary stimuli. Consequently, proxies for government intervention at the firm level remain limited, posing challenges to constructing accurate post-crash prediction models.

This shortfall is particularly pertinent in the era of advanced analytics, where machine and deep learning techniques offer new ways to integrate multifaceted data-spanning macroeconomic indicators, sentiment measures, and firm-level fundamentals. Although some scholars have employed machine learning to assess industry-wide or cross-country policy impacts (Zaremba, Kizys, Tzouvanas, Aharon & Demir, 2021; Yang, Abedin, Zhang, Weng & Hajek, 2023), few have extended these techniques to examine how government interventions at the firm level interact with investor behavior to influence future stock returns. Exploring this intersection of firm-level intervention proxies, investor sentiment, and machine-learning frameworks represents a critical gap in the literature and a compelling avenue for further research.

Deep learning approaches applied during market crash periods

Research on stock return forecasting using deep learning has advanced rapidly, driven by the critical need for accurate predictions in risk management and investment strategies (Hanaki, Akiyama & Ishikawa, 2018; Meier & Mello, 2020; Jagannathan et al., 2022; Ulku et al., 2023). Although deep learning models have been widely applied to stock prediction, portfolio optimization, and risk assessment, most studies have not focused specifically on periods of market crash. Several studies have employed techniques, such as deep reinforcement learning with CNNs on candlestick images, transformer-based architectures combined with U-Net frameworks, and hybrid CNN-LSTM models for general stock trading; however, these address their performance during extreme market downturns only tangentially (Cheong, Wu & Huang, 2021; Song et al., 2022; Singh, Jha, Sharaf, El-Meligy & Gadekallu, 2023; Yang, Liang, Xiong & Zhong, 2023; Brim & Flann, 2022).

A smaller body of research directly targets market crash risk. For instance, studies like “Volatility Spillovers and Contagion During Major Crises: An Early Warning Approach Based on a Deep Learning Model” and “Forecasting Stock Market Crisis Events Using Deep and Statistical Machine Learning Techniques” (Chatzis, Siakoulis, Petropoulos, Stavroulakis & Vlachogiannakis, 2018; Sahiner, 2023) focus on modeling volatility and contagion effects during crises. Ghasemieh and Kashef (2023) developed an enhanced Wasserstein generative adversarial network with Gramian Angular Fields, specifically for predicting stock market behavior during crash periods.

Overall, although deep learning models have demonstrated promise in stock market predictions, their targeted application in forecasting returns during market crashes remains underexplored. This gap highlights the need for further research that leverages deep learning techniques to address the challenges posed by extreme market conditions.

Methodologies and data

The basic algorithm

Several approaches can be used to define a stock market crash. Ma and Lin (2023)) used a machine learning (ML) indicator based on historical data, identifying a crisis when the ML drops >2.5 standard deviations below the mean, consistent with prior methods (Fu, Zhou, Liu & Wu, 2020; Coudert & Gex, 2008). In contrast, this study focuses on the “thousand-stock crash” day in the Chinese A-share market, chosen due to its extreme volatility, heightened investor risk, significant government interventions, and shifts in sentiment—providing a unique perspective on market dynamics under stress.

We hypothesize that post-crash stock data capture the combined effects of government intervention and investor sentiment, potentially revealing recurring historical patterns. If our model, trained on data from past crashes, performs better than traditional models using pre-crash data, this would validate our hypotheses.

Fig. 1 shows the timeline of four ‘thousand-stock’ crashes in the Chinese A-share market from 2015 to 2023, followed by an explanation of our basic algorithm.

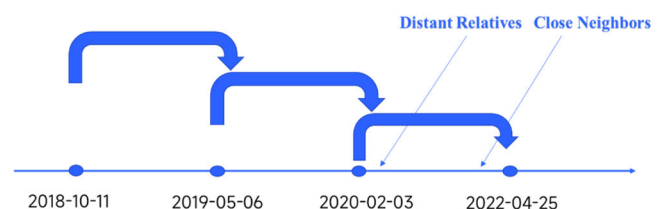


Fig. 1. The dates of thousand shares-lower limit from 2015 to 2023 in China.

- A thousand stock crash occurs on date A. Identify the most recent stock market crash date B before date A.
- Target model: Use market-wide stock data for a period after date B as training data to train the neural network model.
- Benchmark model: Use market-wide stock data for the period before date A as training data to train the neural network model.
- Use the two trained models to predict stock returns after the thousand stock crash on date A.
- Compare and analyze the performance of the two models.

Taking the event on April 25, 2022, as an example, traditional methods would train models using data prior to this date, referred to as 'neighbor' data in this paper. However, this study suggests that the data following the previous crisis on February 3, 2020, termed 'distant relative' data, contain vital information about government intervention and investor sentiment. Considering the challenge of creating proxy variables for government intervention, this study utilized 'distant relative' data for model training and compared this model with those trained on neighboring data. The test dataset comprised data following the current crisis for predictive analysis. Table 1 shows the data required for this study.

Design of neural network models

In this study, we initially trained the Fully Connected Neural Network (FCNN) model and validated its effectiveness using the LSTM model. Given that the predictive performance of neural network models is highly influenced by hyperparameters, it is essential to design the neural network models used in this study and test the performance of the basic algorithm under various hyperparameters. This approach ensures that the conclusions are reliable and robust.

FCNN model

A fully connected neural network (FCNN) consists of a singular input layer, an output layer, and multiple hidden layers. Each neuron in a layer connects to every neuron in the preceding layer, with neurons distributed across these different layers. Fig. 2 illustrates an FCNN wherein the input layer is on the extreme left, the output layer, on the far right, and the hidden layers, between them. In both, the hidden and output layers, the neurons incorporate activation functions. These functions are crucial for converting the weighted sum of the inputs into nonlinear outputs, which are essential for the network's predictive capabilities. The network architecture was designed for unidirectional signal flow with no feedback loops, allowing the signals to move seamlessly from the input to the output layer.

Let $\alpha^{(0)} = x$, and the fully connected neural network iteratively transmits information through Formulae (1) and (2):

$$z^{(l)} = W^{(l)}\alpha^{(l-1)} + b^{(l)} \quad (1)$$

$$\alpha^{(l)} = f_l(z^{(l)}) \quad (2)$$

Table 1

The date period of close neighbors, distant relatives, and current crisis.

Crisis Date	Train A: Close Neighbors	Train B: Distant Relatives	Test C: Current Crisis
6 May 2019	30 April 2019 and the previous 19 trading days	12 October 2018 and the next 19 trading days	7 May 2019 and the next 19 trading days
3 February 2020	23 January 2020 and the previous 19 trading days	7 May 2019 and the next 19 trading days	4 February 2020 and the next 19 trading days
25 April 2022	22 April 2022 and the previous 19 trading days	4 February 2020 and the next 19 trading days	26 April 2022 and the next 19 trading days

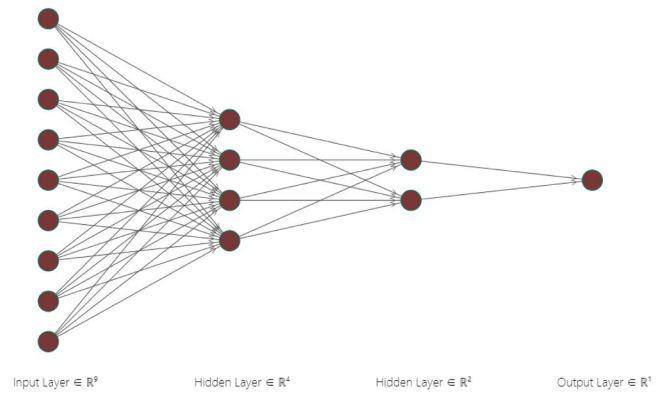


Fig. 2. Diagram of the FCNN.

Note: There are two hidden layers with neuron parameters (4,2). The input layer has nine neurons representing nine features, and the output layer has one neuron, which, in a binary classification model, represents the probability of classification for Category 1.

Where W and b represent the connection weights and biases of all layers in the network.

Our model design encompasses three key aspects:

- Overall Process Design:** The experiment was conducted using Python 3.6. A fully connected neural network (FCNN) model was developed using the Google TensorFlow network. The Smote algorithm (Blagus & Lusa, 2013) was used to address the imbalance between positive and negative samples in the training set. The key hyper-parameters of the FCNN model were set as follows: epoch = 100, batchsize = 2000, learningrate = 0.005, and validationsplit = 0.2. The FCNN algorithm structure included (1) an input layer with the 'relu' activation function; (2) a hidden layer, also using 'relu,' with a dropout probability of 0.5, to periodically deactivate neurons, aiding in reducing overfitting; and (3) an output layer employing the 'sigmoid' activation function.
- Stock Sample Selection Strategy:** To address the issue of limited historical data, which often restrict the number of training samples in machine learning for quantitative investment, we adopted two key strategies to ensure sufficient data for training deep learning models, particularly for medium- to long-term investment strategies, where the scarcity of independent sample points can affect model effectiveness. First, we included all stocks from the Shanghai and Shenzhen A-share markets (over 3000 stocks) to maximize the sample size. Second, we constructed the training dataset by selecting all A-share stocks following a historical "thousand-stock limit-down" event occurring at time t_0 . Each stock i was represented by its feature set $F_{i,t}$ at time t , whereas the label $y_{i,t}$ was assigned, based on the stock's future price movement over the next 20 trading days (encoded as 1 if the stock rose and 0, otherwise). Here, $i = 1, 2, 3, \dots$, approximately 3000, and $t = t_0 + 1, t_0 + 2, \dots, t_0 + 20$. This approach ultimately generated a training set with over 60,000 samples, thereby meeting the data volume requirements for deep learning. Additionally, the test set was constructed, using daily stock data from the 20 trading days following the current crisis (denoted as $t_0^{current}$), to evaluate the model's predictive capability. Specifically, the test set comprised feature sets $F_{i,t}$ for each stock i (approximately 3000 stocks) over trading days $t = t_0^{current} + 1, t_0^{current} + 2, \dots, t_0^{current} + 20$, resulting in a total of over 60,000 samples.
- Hidden Layers Design:** The number and structure of hidden layers play a critical role in determining the performance of the FCNN model. Because financial research datasets tend to be much smaller than those used in image or natural language processing, where data volumes can reach hundreds of millions of samples,

our training set of approximately 60,000 instances required a comparatively shallow network architecture. We found that three to six layers were sufficient for this sample size; adding more layers introduced a large number of parameters and increased the risk of overfitting. Consequently, this study focused on models with three, four, five, or six hidden layers. To ensure model robustness, each variant was tested 10 times and the averages were recorded. Additionally, while typical deep learning studies determine the optimal parameters for each network structure using criteria, such as information criteria or validation set performance, our objective was to validate our hypothesis across four thousand distinct stock limit-down crises. If we optimized the parameters separately for each crisis (e.g., selecting parameter A for 2018 data, parameter B for 2019 data, and parameter C for 2020 data), it would be unclear whether the observed differences were due to parameter optimization rather than the proposed approach, thereby introducing confounding factors. Consequently, we fixed specific parameters across all crisis events to facilitate a direct comparison. Generally, the number of neurons in hidden layers is designed to decrease with network depth, as reducing the neuron counts in deeper layers helps capture abstract, higher-level features that are more representative and discriminative, a strategy akin to data compression that enhances model performance in complex tasks. Given that our feature set comprises 128 variables, we started with 200 neurons in the first hidden layer for robustness. As shown in Table 2, the neurons in the hidden layers of FCNN follow a descending order that sufficiently reflects reasonable parameter design and reinforces our confidence in the conclusions.

LSTM model

The long short-term memory (LSTM) network, a variant of the recurrent neural network (RNN), was first introduced by Hochreiter and Schmidhuber (1997) and further refined in subsequent years by Gers and Schmidhuber (2000) and Gers, Schmidhuber and Cummins (2000). LSTM networks incorporate internal cell states and gating mechanisms, effectively addressing the issues of vanishing or exploding gradients, which are often encountered in traditional RNNs.

The LSTM cell is composed of forget gate f_t , input gate i_t , and output gate o_t . The mathematical principles and operational formulas are shown in Eqs. (3)–(8), respectively.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$\hat{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (6)$$

$$c_t = f_t * c_{t-1} + i_t * \hat{c}_t \quad (7)$$

$$h_t = o_t * \tanh(c_t) \quad (8)$$

Where W represents the corresponding weight matrix for each unit, b represents the corresponding bias vector, and $*$ denotes the Hadamard

product of the matrices.

Fig. 3 illustrates a multilayer LSTM architecture consisting of m layers. As highlighted by Karpathy and Li (2017), the distinguishing feature of a multi-layer LSTM lies in its layered structure, where the input for the l -th layer is informed not only by its own processing, but also by output from the $(l - 1)$ -th layer. This design enables each successive layer to refine and build upon the information processed by the preceding layer, thereby enhancing the capacity of the network to learn intricate and abstract data patterns. The multilayered approach is particularly beneficial for complex tasks that require an in-depth understanding and recognition of long-term dependencies in sequential data.

This study's LSTM model is methodically structured as follows:

- (a) **Overall Process Design: (1). Initial Dense Layer:** The model begins with a fully connected (dense) layer, aimed at managing a large number of input variables. This layer is integral to feature-dimensionality reduction and is particularly beneficial for handling a multitude of variables. The reduced neuron count in this layer condenses the input data into principal components, thus simplifying the dataset for more efficient processing. **(2). Stacked LSTM Layers:** The core of the model comprises two LSTM layers strategically stacked to enhance the model's capacity for learning. For most applications, stacking up to approximately four LSTM layers is sufficient. In the context of this study, considering the volume of available data, we stacked two layers. **(3). Final Dense Layer:** The final segment of the model was another dense layer designed for binary classification. Employing a sigmoid activation function, this layer translated the output of the model into a probability value ranging from zero to one. This probability reflects the likelihood of classifying input data into two predefined categories.
- (b) **Stock Sample Selection Strategy:** The stock sample selection strategy for the LSTM model aligns closely with that of the FCNN model. In our training dataset covering 20 days, each day featured over 3000 stocks from the Shanghai and Shenzhen stock exchanges, resulting in over 6000 stock samples in both, the training and testing datasets.
- (c) **Key Parameters:** The critical parameters of the model included the number of units in the first dense layer (unit1) and the number of hidden units in the subsequent second (unit2) and third (unit3) LSTM layers, as shown in Table 3. The model also considered time steps, incorporating both, a short-term lag of one week (five trading days) and a longer-term span of the following month (20 trading days). To verify the stability and reliability of the model, each parameter set was tested across ten separate experiments, and the average of these trials constituted the reported findings.

Evaluation metrics

The confusion matrix categorizes the samples into four scenarios: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), as shown in Table 4. Using the confusion matrix, common evaluation standards, such as accuracy, precision, recall, and F1-score, can be calculated.

Table 2
Hidden layers in the FCNN model.

3 Hidden Layers	4 Hidden Layers	5 Hidden Layers	6 Hidden Layers
(200,120,80)	(200,160,120,80)	(200,160,120,80,60)	(200,180,160,120,80,60)
(120,80,60)	(160,120,80,60)	(160,120,80,60,40)	(180,160,120,80,60,40)
(80,60,20)	(120,80,60,40)	(120,80,60,40,20)	(160,120,80,60,40,20)
(60,20,10)	(80,60,40,20)	(80,60,40,20,10)	(120,80,60,40,20,10)
(20,10,5)	(60,40,20,10)	(60,40,20,10,5)	(80,60,40,20,10,4)

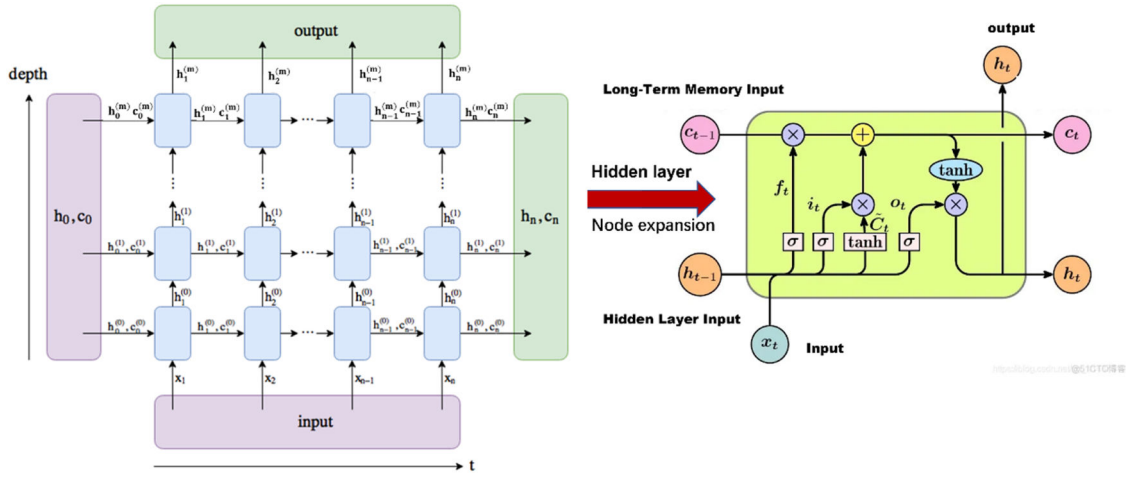


Fig. 3. Diagram of the Multiple-layer LSTM.

Table 3

Crucial parameters in the LSTM model.

(time_step,unit1,unit2,unit3)	(5,50,100,20)	(5,80,80,20)	(5,50,80,20)	(5,80,50,10)	(5,50,20,5)
(time_step,unit1,unit2,unit3)	(20,50,100,20)	(20,80,80,20)	(20,50,80,20)	(20,80,50,10)	(20,50,20,5)

Table 4

Confusion matrix for classification results.

	Predicted Positive	Predicted Negative
Actual Positive	True positives (TP)	False negatives (FN)
Actual Negative	False positives (FP)	True negatives (TN)

Note: True positives (TP): number of samples correctly predicted as positive by the model. False negatives (FN): The number of positive samples incorrectly predicted as negative. False positives (FP): the number of negative samples incorrectly predicted as positive. True negatives (TN): the number of samples correctly predicted as negative by the model.

The three common evaluation metrics— Precision (P), Recall (R), and F1-Score— were defined and calculated as follows:

Precision (P): Precision measures the proportion of samples correctly predicted as positive among all samples predicted as positive,

$$P = \frac{TP}{TP + FP} \quad (9)$$

Recall (R): Recall measures the proportion of samples correctly predicted as positive from all actual positive samples:

$$R = \frac{TP}{TP + FN} \quad (10)$$

F1-Score: The F1-Score is the harmonic mean of Precision and Recall, and is used to measure the model's overall performance. It seeks a balance between Precision and Recall and is particularly useful for imbalanced classification problems:

$$F1_Score = \frac{2TP}{2TP + FN + FP} \quad (11)$$

Data description

Our research data were gathered exclusively from the SuperMind stock trading platform (<https://quant.10jqka.com.cn/view/>), a part of Tonghuashun's quantitative investment platform, accessed on September 1, 2023. This platform offers a comprehensive analysis of 128 variables for all stocks in the Shanghai and Shenzhen A-share markets, encompassing both, technical and financial dimensions. Technical

factors, used as momentum proxy variables, were divided into five categories: overbought/oversold indicators, volume-price indicators, trend indicators, energy indicators, and support/resistance indicators. Financial factors were classified as profitability, solvency, operational efficiency, growth potential, and valuation. Additionally, the significant influence of industry sectors was considered by including stock industry classification as a crucial variable.

Key aspects of our data preparation included the following:

- Stock Selection:** We focused on stocks listed on the Shanghai and Shenzhen A-share markets as of January 1, 2020, and excluded any stocks that were IPOed after that date.
- Industry Classification:** We used Tonghuashun's industry classification from SuperMind, as of January 1, 2020, as a predictive sub-variable.
- Feature Set Compilation:** The final training feature set was derived by intersecting the indicators from both, the training and test sets. First, factors with >80 % missing values were removed from the raw data. We then eliminated stock records containing missing data, resulting in a clean and robust training set.
- Data Standardization:** We standardized the data by normalizing the indicators in the validation and test sets using the maximum and minimum values from the training set, as follows:

$$\text{normalized value} = \frac{(x - \min)}{(\max - \min)}.$$

Findings

In China's A-share market, the thousand stock limit-down events in 2018, 2019, 2020, and 2022 had unique causes: global turbulence and domestic slowdown in 2018, escalating trade frictions and tighter financing in 2019, the COVID-19 outbreak in 2020, and stricter macro-control policies in 2022. These events highlight the varied responses of the market to external shocks and internal pressures. Under these different conditions, we examined whether using post-crisis data from a previous crash ("distant relative data") as training data can improve one-month stock return predictions as compared to using data from the most

recent crisis. To test this hypothesis, we designed multiple deep learning model architectures and parameter configurations by calculating Precision, Recall, and F1score for each scenario. We then conducted a Wilcoxon test with the null hypothesis $H_0 : M_0 = M_1$ and alternative hypothesis $H_0 : M_0 < M_1$. A significantly low p-value implies that training with distant relative data can substantially enhance the predictive accuracy of deep learning models in forecasting post-crash stock returns.

FCNN model

Based on Table 1 and the FCNN model configuration, we trained the FCNN model, using two types of training data: "close neighbor" data (collected from near the crisis day) and "distant relative" data (obtained from a previous market crash). As detailed in Table 2, we experimented with three to six hidden layers and set five parameter groups for each configuration. Each parameter set was run ten times, and the average results were recorded. Tables 5–7 present the FCNN test set results for the 2019, 2020, and 2022 stock market crashes, and Figs. 4–6 illustrate the F1-scores across the various parameter groups.

Prediction after the 2019 A-share market crash

Table 5 shows that, for the 2019 crash, the FCNN model built using factors from Section 3.3, when trained with close neighbor data, achieved an average Precision of approximately 0.406, Recall of 0.187, and F1-score of 0.219. In contrast, when trained with distant relative data from the 2018 crash, the model produced an average Precision of about 0.403, Recall of 0.982, and F1-score of 0.571. The Wilcoxon test ($p \approx 0$) confirms that training with distant relative data significantly improves performance. Fig. 4 visually demonstrates that using post-2018 data, rather than pre-2019 data, yields superior results, indicating that historical crisis data contain repeatable patterns.

Prediction after the 2020 A-share market crash

The 2019 crash was triggered by escalating trade frictions and

tightened financing conditions, whereas the 2020 crisis occurred at the onset of the COVID-19 outbreak, driven by international uncertainty and liquidity shortages. Despite these differing backgrounds, we tested whether using post-2019 distant relative data as training data could enhance predictions of the 2020 crisis. As shown in Table 6, the model trained with close-neighbor data achieved an average Precision of 0.358, Recall of 0.066, and F1-score of 0.106, whereas using distant relative data improved these to a Precision of 0.381, Recall of 0.306, and F1-score of 0.312. The Wilcoxon test ($p \approx 0$) confirms a significant improvement, and Fig. 5 further supports that distant relative data lead to better performance across most parameter groups.

Prediction after the 2022 A-share market crash

The 2022 crash, influenced by tightened macro-control policies and structural adjustments, presented another market environment. We evaluated whether training with post-2020 distant relative data could enhance predictions. As detailed in Table 7, the model trained with close neighbor data achieved an average Precision of 0.809, Recall of 0.202, and F1-score of 0.282, whereas training with distant relative data yielded improved averages of 0.872, 0.285, and F1-score 0.412. The Wilcoxon test ($p < 0$) confirms the superior performance of the distant relative data approach. Fig. 6 illustrates that, except for one parameter group, using distant relative data consistently outperforms the close neighbor data, further validating our findings of Sections 4.1.1 and 4.1.2.

LSTM model

To further validate the findings of the FCNN model, we employed Long Short-Term Memory (LSTM) models, which are particularly effective in capturing temporal dependencies in stock pricing. This allowed us to compare the predictive performance of distant relative data with that of close-neighbor data for forecasting post-crisis stock returns. We evaluated the performance of the LSTM model using both, monthly (20 trading days) and weekly (5 trading days) time steps. Table 8 presents the LSTM model predictions for the 2019, 2020, and

Table 5
Results for the test set of 7 May 2019 and the subsequent 19 trading days.

Number	Train Dataset	Train A: Close Neighbors			Train B: Distant Relatives		
	3 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
1	(200,120,80)	0.415836132	0.071193916	0.117605245	0.402583459	0.98591482	0.571713954
2	(120,80,60)	0.404253066	0.107547864	0.160665899	0.402698207	0.985444456	0.571749122
3	(80,60,20)	0.404439712	0.159989711	0.211572572	0.402929577	0.977323341	0.570594819
4	(60,20,10)	0.403652167	0.267522875	0.310201109	0.403469974	0.979439968	0.571487432
5	(20,10,5)	0.40727818	0.534362251	0.429039652	0.402929512	0.958093554	0.567209889
	4 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
6	(200,160,120,80)	0.42152929	0.106353581	0.145093437	0.402802053	0.985565722	0.571875911
7	(160,120,80,60)	0.425444454	0.053198838	0.090406421	0.402734703	0.987116462	0.57206891
8	(120,80,60,40)	0.404804841	0.161018629	0.205386018	0.403082949	0.984481686	0.571970945
9	(80,60,40,20)	0.408366445	0.145544406	0.18716123	0.402496499	0.979880935	0.570601692
10	(60,40,20,10)	0.395959246	0.226189688	0.260839489	0.403440306	0.972854894	0.570345198
	5 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
11	(200,160,120,80,60)	0.410998768	0.09639878	0.146816224	0.402948898	0.982339317	0.571476173
12	(160,120,80,60,40)	0.401170707	0.190115023	0.216753993	0.403046542	0.984882236	0.572007263
13	(120,80,60,40,20)	0.396382475	0.141649212	0.194006323	0.402818304	0.987667662	0.572242804
14	(80,60,40,20,10)	0.389432284	0.119711169	0.176169407	0.403018466	0.981211179	0.571349594
15	(60,40,20,10,5)	0.410996616	0.240763604	0.274293604	0.402891326	0.976224601	0.570349513
	6 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
16	(200,180,160,120,80,60)	0.398417351	0.219292251	0.264786899	0.403049272	0.986454999	0.572271903
17	(180,160,120,80,60,40)	0.395683539	0.20428839	0.255323759	0.40258044	0.991809058	0.572696308
18	(160,120,80,60,40,20)	0.422299507	0.164752141	0.181739323	0.402499953	0.991680449	0.572592934
19	(120,80,60,40,20,10)	0.414435276	0.214360781	0.243474186	0.402672225	0.984915298	0.571624131
20	(80,60,40,20,10,4)	0.397218409	0.313449453	0.311237585	0.402877721	0.978197926	0.570665374
	Arithmetic mean	0.406429923	0.186885128	0.219128619	0.402878519	0.982074928	0.571344693

Table 6

Results for the test set of 4 February 2020 and the subsequent 19 trading days.

Number	Train Dataset	Train A: Close Neighbors			Train B: Distant Relatives		
	3 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
1	(200,120,80)	0.351542035	0.054065387	0.092452176	0.420619741	0.39554649	0.391812886
2	(120,80,60)	0.371771774	0.045611649	0.08046519	0.432591617	0.430962096	0.420134962
3	(80,60,20)	0.382197714	0.048583489	0.085717315	0.445905426	0.485987872	0.457733289
4	(60,20,10)	0.363089111	0.105940846	0.157451485	0.394659984	0.396345402	0.357055279
5	(20,10,5)	0.391751921	0.192605817	0.240915977	0.336838546	0.16176554	0.211605262
	4 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
6	(200,160,120,80)	0.347770959	0.047421951	0.083048072	0.418844369	0.393042098	0.393277137
7	(160,120,80,60)	0.35431076	0.051535497	0.089235149	0.437463579	0.437928496	0.425936717
8	(120,80,60,40)	0.357129386	0.044382117	0.078494658	0.412114218	0.385222958	0.385190543
9	(80,60,40,20)	0.352517202	0.057490509	0.098086771	0.405249032	0.389716129	0.375516662
10	(60,40,20,10)	0.364187372	0.08006686	0.1295189	0.376607513	0.298577826	0.300898958
	5 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
11	(200,160,120,80,60)	0.360371011	0.042203525	0.075194365	0.439845145	0.475049576	0.450894666
12	(160,120,80,60,40)	0.344375655	0.049028274	0.08514967	0.395538676	0.382953715	0.365779466
13	(120,80,60,40,20)	0.34908767	0.050847073	0.087386598	0.375633329	0.309921245	0.312956306
14	(80,60,40,20,10)	0.351561755	0.072664173	0.117810322	0.327564003	0.158861692	0.184509807
15	(60,40,20,10,5)	0.368021736	0.106360134	0.159980594	0.347907346	0.151184201	0.202971494
	6 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
16	(200,180,160,120,80,60)	0.3451065	0.046968666	0.082248753	0.331819558	0.189509322	0.202314178
17	(180,160,120,80,60,40)	0.349451056	0.049875347	0.086067478	0.296705647	0.118933086	0.148698215
18	(160,120,80,60,40,20)	0.356109086	0.045186696	0.079668636	0.33766166	0.22852853	0.226276386
19	(120,80,60,40,20,10)	0.356672186	0.055422403	0.094457685	0.340746585	0.159555783	0.205711337
20	(80,60,40,20,10,4)	0.338237122	0.070924699	0.112275537	0.351139504	0.16774888	0.214622639
	Arithmetic mean	0.357763101	0.065859256	0.105781267	0.381272774	0.305867047	0.311694809

Table 7

Results for the test set of 26 April 2022 and the subsequent 19 trading days.

Train Dataset		Train A: Close Neighbors			Train B: Distant Relatives		
Number	3 hidden layers	Precision	Recall	F1_score	Precision	Recall	F1_score
1	(200,120,80)	0.816202098	0.1596479	0.243021	0.8693099	0.3054602	0.4410279
2	(120,80,60)	0.803548342	0.138672	0.2146602	0.8607947	0.2726542	0.3852927
3	(80,60,20)	0.779680568	0.1222556	0.1960671	0.8817572	0.2541591	0.3862964
4	(60,20,10)	0.808649415	0.1998588	0.3122536	0.8793392	0.2721269	0.3891123
5	(20,10,5)	0.815161014	0.7224218	0.7436218	0.8798516	0.1808919	0.2742091
4 hidden layers		Precision	Recall	F1_score	Precision	Recall	F1_score
6	(200,160,120,80)	0.818923557	0.1221752	0.1895686	0.8702669	0.2683914	0.3996911
7	(160,120,80,60)	0.813246727	0.1055979	0.1826618	0.8621389	0.2438088	0.3629463
8	(120,80,60,40)	0.792267519	0.1127239	0.1900652	0.8679847	0.3223771	0.4590234
9	(80,60,40,20)	0.782802427	0.1649848	0.2291402	0.8706475	0.2463539	0.373088
10	(60,40,20,10)	0.808100104	0.228202	0.3298787	0.8778254	0.2166345	0.3380754
5 hidden layers		Precision	Recall	F1_score	Precision	Recall	F1_score
11	(200,160,120,80,60)	0.79557904	0.096513	0.1603203	0.8665392	0.2866953	0.4147439
12	(160,120,80,60,40)	0.814412129	0.1644379	0.2525481	0.8775144	0.2997337	0.4366313
13	(120,80,60,40,20)	0.809006983	0.1744147	0.2610201	0.8761429	0.3240089	0.4643393
14	(80,60,40,20,10)	0.807811618	0.1992082	0.2782282	0.8838005	0.2548222	0.3905473
15	(60,40,20,10,5)	0.807501215	0.1676586	0.2434793	0.873719	0.2386631	0.3634739
6 hidden layers		Precision	Recall	F1_score	Precision	Recall	F1_score
16	(200,180,160,120,80,60)	0.811377436	0.1559321	0.2537706	0.8717316	0.3175961	0.4640737
17	(180,160,120,80,60,40)	0.820307112	0.2481019	0.3553607	0.863058	0.3635657	0.4919565
18	(160,120,80,60,40,20)	0.823057067	0.2093155	0.2962642	0.8625058	0.4119267	0.5276555
19	(120,80,60,40,20,10)	0.81096037	0.2885809	0.3775229	0.871104	0.3180518	0.4524035
20	(80,60,40,20,10,4)	0.838740307	0.2543986	0.327803	0.8736905	0.2992189	0.4227628
Arithmetic mean		0.808866752	0.201755065	0.28186278	0.871986095	0.28485702	0.411867515

2022 market crashes, and Figs. 7–9 illustrate the corresponding F1-scores.

As shown in Table 8, the Wilcoxon test yielded p-values close to 0 for the 2019 and 2022 crashes, indicating that using distant relative, rather than close neighbor, data significantly improves prediction performance. Specifically, the F1-scores for the A-share market post-crisis predictions, using close neighbor data, were approximately 0.288, 0.389, and 0.393 for 2019, 2020, and 2022, respectively. In contrast,

using distant relative data yielded F1-scores of approximately 0.611, 0.386, and 0.630, respectively. These results suggest that, for the 2019 and 2022 crises, the LSTM model benefitted from historical crash data, which is consistent with the findings of the FCNN model. However, the 2020 results are less conclusive, likely because of the drastic market shift on February 3, 2020, which marked the first trading day after the COVID-19 outbreak, leading to a very different market environment compared to the pre-pandemic period.

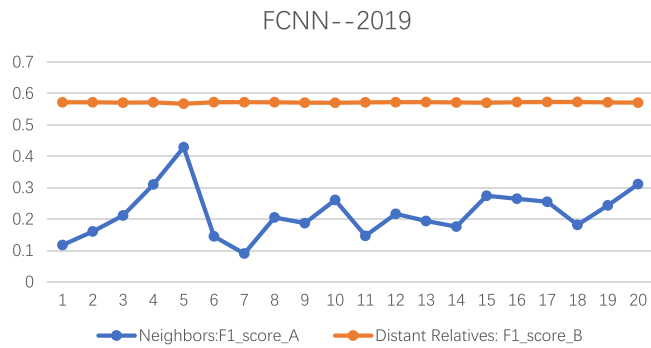


Fig. 4. FCNN-F1-score for the test set of 7 May 2019 and the subsequent 19 trading days.

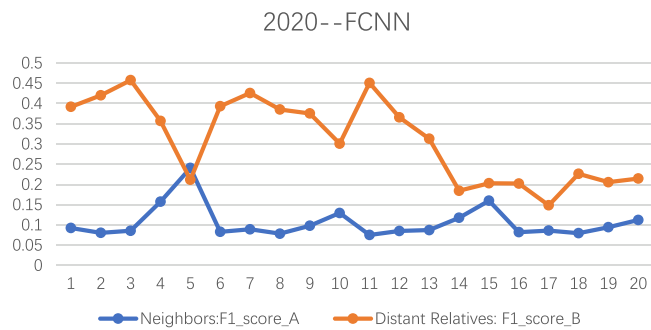


Fig. 5. FCNN-F1-score for the test set of 4 February 2020 and the subsequent 19 trading days.

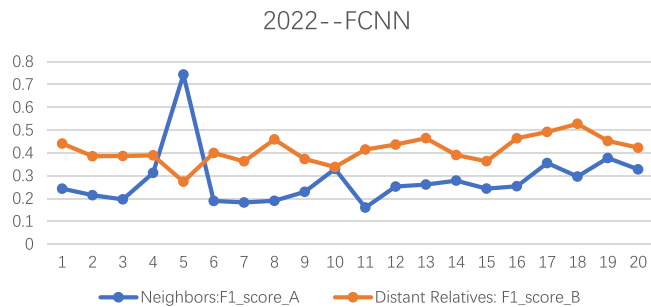


Fig. 6. FCNN-F1-score for the test set of 26 April 2022 and the subsequent 19 trading days.

Overall, the performance of the LSTM model using distant relative data is superior or comparable to that using close neighbor data, reinforcing the robustness of our approach.

Additionally, as shown in Tables 5, 6, 7 and 8, the average F1-scores for the FCNN model using distant relative data were 0.571, 0.312, and 0.412 for the 2019, 2020, and 2022 crises, respectively. By contrast shows that the LSTM model achieved higher average F1-scores of 0.611, 0.386, and 0.630 for the same periods. These findings underscore the advantage of the LSTM model in leveraging temporal memory to provide more accurate predictions than the FCNN model.

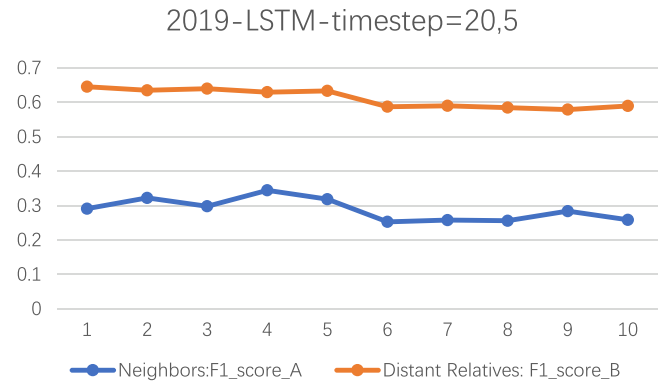


Fig. 7. LSTM-F1-score for the test set of 7 May 2019 and the subsequent 19 trading days.

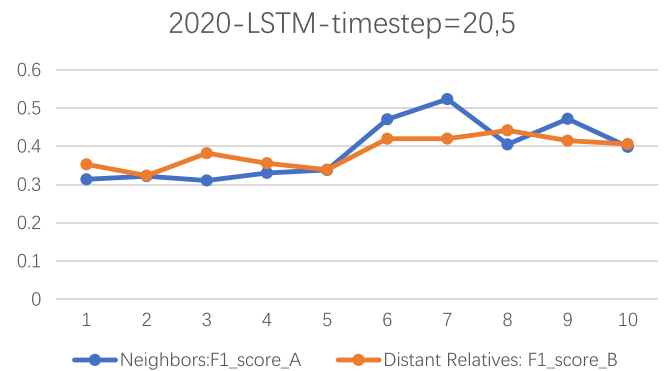


Fig. 8. LSTM-F1-score for the test set of 4 February 2020 and the subsequent 19 trading days.

Table 8

LSTM test set results following the 2019, 2020 and 2022 stock market crashes.

Time_Step	2019		2020		2022	
	Neighbors: F1_Score_A	Distant Relatives: F1_Score_B	Neighbors: F1_Score_A	Distant Relatives: F1_Score_B	Neighbors: F1_Score_A	Distant Relatives: F1_Score_B
20	0.290988227	0.645534898	0.313906727	0.353134317	0.433668958	0.622149165
20	0.322314461	0.634939485	0.322184185	0.323405905	0.331341011	0.620356789
20	0.297990795	0.639745247	0.310830552	0.382184931	0.446118965	0.615951464
20	0.344552935	0.629465929	0.330546719	0.355917514	0.4447285	0.64386031
20	0.318637882	0.633338693	0.338287273	0.338987949	0.497958943	0.650947716
5	0.252877025	0.587269886	0.470739248	0.420075365	0.355112641	0.627359613
5	0.257910477	0.589856103	0.523578822	0.42022891	0.358028946	0.622174036
5	0.256071046	0.5846716	0.405267799	0.441905547	0.360911875	0.630011225
5	0.283925383	0.579042272	0.472104509	0.414968682	0.359863657	0.638405426
5	0.25856645	0.589529797	0.398874967	0.406138546	0.343954326	0.630708871
Arithmetic mean	0.288383468	0.611339391	0.38863208	0.385694767	0.393168782	0.630192462

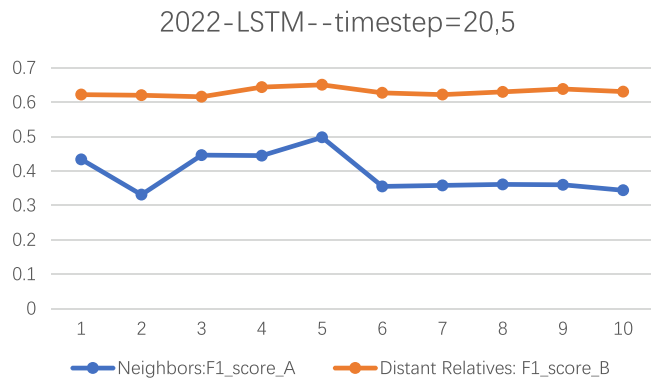


Fig. 9. LSTM-F1-score for the test set of 26 April 2022 and the subsequent 19 trading days.

Effective utilization of distant relative data

In Section 4.1, we observed that the average F1 Scores for predicting post-crash stock returns using the FCNN model with distant relative data were 0.571 for 2019, 0.311 for 2020 and 0.411 for 2022. The F1 Scores for 2020 and 2022 were below 0.5, raising concerns regarding their practical utility. Deep learning models are particularly effective when trained on large datasets, and distant relative data can provide more comprehensive post-crash information, thereby improving prediction accuracy. To address this, we expanded the training approach for the 2020 and 2022 crises by incorporating two distant relative datasets instead of just one, as described in Section 4.1. Specifically, we used post-2019 and post-2018 data for the 2020 crisis, and post-2020 and post-2019 data for the 2022 crisis. The results showed significant improvements: for the 2020 crisis, the average F1 Score increased from 0.105 (using close neighbor data) to 0.311 (using a single distant relative dataset) and further rose to 0.627 (using two distant relative datasets). Similarly, for the 2022 crisis, the average F1 Score improved from 0.282 (close neighbor data) to 0.412 (single distant relative dataset) and reached 0.533 with two distant relative datasets, just above the 0.5 threshold, indicating practical value.

These results demonstrate that training with multiple distant relative datasets significantly outperforms that with a single distant relative dataset, thus highlighting the importance of combining multiple historical crash events. Additionally, we explored the use of transfer learning by integrating close neighbor data with distant relative data. The model was first trained on two distant relative datasets and subsequently fine-tuned using close neighbor data. The results show that, for the 2020 crisis, the FCNN model, trained with transfer learning, achieved an average F1 Score of 0.448, and, for the 2022 crisis, a score of

0.395. Although transfer learning improved the performance compared to using a single distant relative dataset, it still did not surpass the accuracy of the model trained with both distant relative datasets.

Figs. 10 and 11 further illustrate that using multiple distant relative datasets for training significantly improves prediction accuracy, with F1 Scores consistently above 0.5, across all 20 parameter sets. In contrast, the model trained using traditional close neighbor data performed the worst across all configurations. Overall, our findings suggest that, for post-crisis stock return predictions, training deep learning models with a combination of multiple distant relative datasets offer substantial practical value and enhance predictive performance.

Conclusion

This study demonstrates the effectiveness of using historical crash data (referred to as "distant relative data") for forecasting post-crash stock returns in China's A-share market. The results consistently show that utilizing distant relative data significantly improves predictive accuracy compared to traditional close neighbor data. This finding holds true for both, FCNN and LSTM models.

For the FCNN model, the average F1 scores for predicting post-crash stock returns using distant relative data are 0.571, 0.311, and 0.411 for the 2019, 2020, and 2022 crash periods, respectively. Notably, the F1 scores for the 2020 and 2022 crises were below 0.5, indicating limited practical utility. However, by incorporating data from two distant relative events, the F1 score for the 2020 crisis increased from 0.105 (using close neighbor data) to 0.311 (with a single distant relative dataset) and further to 0.627 (with two distant relative datasets). Similarly, for the 2022 crisis, the F1 score improved from 0.282 (using close neighbor data) to 0.533 (with two distant relative datasets), thus surpassing the threshold of 0.5, and providing a more practically useful model.

Additionally, the LSTM model outperformed the FCNN model, with F1 scores of 0.611, 0.386, and 0.630 for the 2019, 2020, and 2022 crashes, respectively, compared to the latter's F1 scores of 0.571, 0.312, and 0.412, respectively. This demonstrates the advantage of LSTM in leveraging temporal dependencies to enhance the predictive performance.

Moreover, the use of transfer learning, which integrates close neighbor data with distant relative data, improved the model's performance, but did not surpass the accuracy of the model trained with both distant relative datasets. Specifically, transfer learning led to F1 scores of 0.448 and 0.395 for the 2020 and 2022 crises, respectively, which were lower than the F1 scores achieved by the model trained with two distant relative datasets (0.627 and 0.533, respectively).

Our findings indicate that training deep learning models with distant

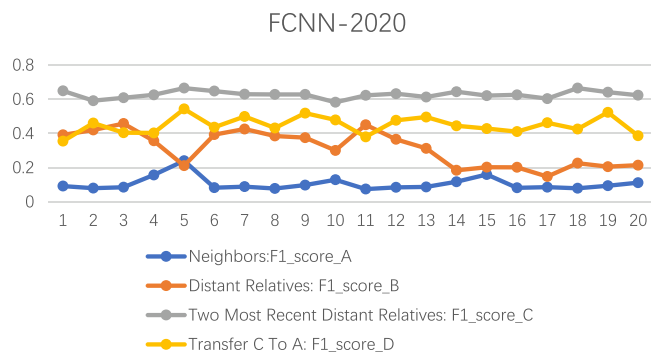


Fig. 10. FCNN-2020 model's F1 score with dual distant relative data and the transfer learning technique.

Note: Transferring from C to A involved training the deep learning model on two recent post-disaster datasets (C), followed by additional training on the pre-disaster dataset (A) of the current stock crisis.

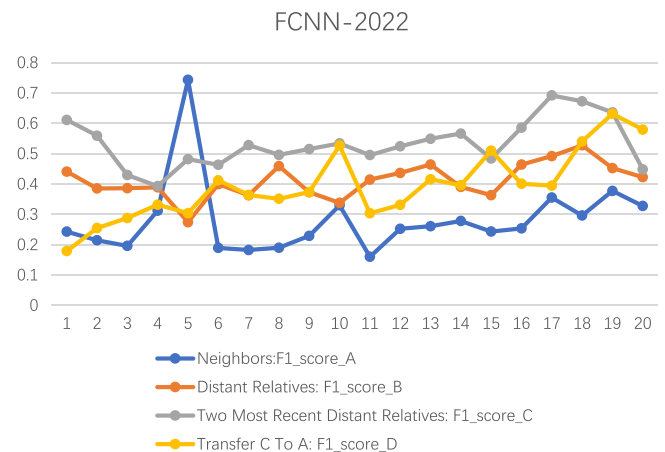


Fig. 11. FCNN-2022 model's F1 score with dual distant relative data and the transfer learning technique.

relative datasets significantly improves post-crisis stock return predictions, regardless of whether the model is based on the FCNN or LSTM architecture. Using a combination of multiple distant relative datasets, the trained FCNN model achieved F1-scores above 0.5 for both the 2020 and 2022 crises, demonstrating its practical applicability. Moreover, under the same settings, the LSTM models consistently outperformed the FCNN models, underscoring the advantage of LSTM in leveraging temporal dependencies.

A recent study by Ghasemieh and Kashef (2023) introduced robust deep learning models, such as the GAF-EWGAN stacking ensemble, for next-day price prediction during market crashes. Their study did not report explicit prediction accuracy; instead, it focused on evaluating post-crisis stock strategies, achieving an average annual return of 16.49 % across 20 stocks. In contrast, our approach analyzes the directional effects of stock returns across the entire A-share market (over 3000 stocks) following a crash.

In future research, one promising approach would be to build on our method-training models on historical post-crash data and integrate it with advanced deep learning architectures, such as those proposed by Ghasemieh and Kashef (2023), to further enhance post-crisis stock return predictions. Another direction is to extend our methodology by employing statistical breakpoint techniques to detect structural shifts in market indices, such as the CSI 300. Once a new breakpoint is identified, the historical data following previous breakpoints can be used to train a deep learning model to forecast individual stock returns after the event. A significant challenge with this method is that accurate breakpoint detection requires data both, before and after an event, and timely identification remains a hurdle. Finally, our model was designed for emerging markets with frequent government intervention. Future research could also explore cross-market transfer learning, specifically by training a model on A-share data and testing its performance on Hong Kong stocks or other emerging markets.

CRedit authorship contribution statement

Weiran Lin: Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Formal analysis, Data curation. **Haijing Yu:** Visualization, Software, Data curation. **Liugen Wang:** Supervision, Formal analysis, Conceptualization.

Acknowledgements

The author would like to thank the reviewers for their constructive comments. This research was supported by the Zhejiang Provincial Social Science Planning Special Project titled “Interpretation of the Spirit of the Third Plenary Session of the 20th CPC Central Committee and the Fifth Plenary Session of the 15th Zhejiang Provincial Committee.” Additionally, we gratefully acknowledge the advanced computing resources provided by the Supercomputing Center of Hangzhou City University.

References

- Barberis, N., Shleifer, A., & Vishny, R. A. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307–343.
- Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129–152.
- Buehlmaier, M. M., & Whited, T. M. (2018). Are financial constraints priced? Evidence from textual analysis. *Review of Financial Studies*, 31(6), 2342–2385.
- Bhuyan, B. P., Jaiswal, V., & Cherif, A. R. (2023). A knowledge representation system for the Indian stock market. *Computers*, 12(5), 90.
- L. Blagus, R., & Lusa, L. (2013). Smote for high-dimensional class-imbalanced data *BMC Bioinformatics*, 14, 106
- Brim, A., & Flann, N. S. (2022). Deep reinforcement learning stock market trading, utilizing a CNN with candlestick images. *PLOS ONE*, 17(2), Article e0263181.
- Cruz, R., Kinyua, J., & Mutigwe, C. (2023). Analysis of social media impact on stock price movements using machine learning anomaly detection. *Intelligent Automation and Soft Computing*, 36(3), 3405–3423.
- Cheng, W. K., Bea, K. T., Leow, S. M. H., Chan, J. Y. L., Hong, Z. W., Hong, Z. W., & Chen, Y. L. (2022a). A review of sentiment, semantic and event-extraction-based approaches in stock forecasting. *Mathematics*, 10(14), 2437.
- Chen, H. L., De, P., Hu, Y., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social Media. *Review of Financial Studies*, 27(5), 1367–1403.
- Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications*, 112, 353–371.
- Cheng, M. Y., Jin, L., Li, Z. S., & Lin, B. X. (2022b). The effectiveness of government stock purchase during market crash: Evidence from China. *Pacific-Basin Finance Journal*, 71, Article 101706.
- Coudert, V., & Gex, M. (2008). Does risk aversion drive financial crises? Testing the predictive power of empirical indicators. *Journal of Empirical Finance*, 15(2), 167–184.
- Cheong, S. P., Wu, M. C., & Huang, S. H. (2021). Interpretable stock anomaly detection based on spatio-temporal relation networks with genetic algorithm. *IEEE access : Practical innovations, open solutions*, 9, 68302–68319.
- Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, 53(6), 1839–1885.
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *Quantitative Finance*, 18(11), 1933–1945.
- Fu, J. H., Zhou, Q. L., Liu, Y. F., & Wu, X. (2020). Predicting stock market crises using daily stock market valuation and investor sentiment indicators. *North American Journal of Economics and Finance*, 51, Article 100905.
- Faraz, M., & Khaloozadeh, H. (2020). Multi-step-ahead stock market prediction based on Least Squares Generative Adversarial network. In *2020 28th Iranian Conference on Electrical Engineering (ICEE)* (pp. 1–6). IEEE. <https://doi.org/10.1109/ICEE50131.2020.9260858>.
- Goel, G., & Dash, S. R. (2021). Investor sentiment and government policy interventions: Evidence from COVID-19 spread. *Journal of Financial Economic Policy*, 14(2), 242–267.
- Gupta, R., Nel, J., & Pierdzioch, C. (2023). Investor confidence and forecastability of US stock market realized volatility: Evidence from machine learning. *Journal of Behavioral Finance*, 24(1), 111–122.
- Gong, X., Zhang, W. G., Wang, J. B., & Wang, C. (2022). Investor sentiment and stock volatility: New evidence. *International Review of Financial Analysis*, 80, Article 102028.
- Gers, F. A., & Schmidhuber, J. (2000). Recurrent nets that time and count. In , 3. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN 2000)* (pp. 189–194).
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471.
- Guo, K., Sun, Y., & Qian, X. (2017). Can investor sentiment be used to predict the stock price? Dynamic analysis based on China stock market. *Physica A: Statistical Mechanics and its Applications*, 469(1), 390–396.
- Ghasemieh, A., & Kashef, R. (2023). An enhanced Wasserstein generative adversarial network with Gramian Angular Fields for efficient stock market prediction during market crash periods. *Applied Intelligence*, 53, 28479–28500.
- Herculano, M. C., & Lutkebohmert, E. (2023). Investor sentiment and global economic conditions. *Journal of Empirical Finance*, 73, 134–152.
- Heiden, S., Klein, C., & Zwergel, B. (2013). Beyond fundamentals: Investor sentiment and exchange rate forecasting. *European Financial Management*, 19(3), 558–578.
- Hanaki, N., Akiyama, E., & Ishikawa, R. (2018). Behavioral uncertainty and the dynamics of traders' confidence in their price forecasts. *Journal of Economic Dynamics & Control*, 88, 121–136.
- Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance*, 64(5), 2289–2325.
- Huang, W., Lehmann, B. N., & Zhu, Y. (2017). Machine learning in behavioral finance. *Financial Analysts Journal*, 73(3), 70–82.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hale, T., Petherick, A., Phillips, T., & Webster, S. (2020). *Variation in government responses to covid-19* (p. 31). Blavatnik School of Government Working Paper.
- Huang, Y., Miao, J., & Wang, P. F. (2019). Saving China's stock market? *IMF Economic Review*, 67(2), 349–394.
- Huynh, N., Nguyen, D., & Dao, A. (2021). Sectoral performance and the government interventions during COVID-19 pandemic: Australian evidence. *Journal of Risk and Financial Management*, 14(4), 178.
- Jagannathan, R., Pelizzon, L., Schaumburg, E., Sherman, M. G., & Yuferova, D. (2022). Recovery from fast crashes: Role of mutual funds. *Journal of Financial Markets*, 59, Article 100646. Part B.
- Jena, P. R., & Majhi, R. (2023). Are Twitter sentiments during COVID-19 pandemic a critical determinant to predict stock market movements? A machine learning approach. *Scientific African*, 19, Article e01480.
- Kanzari, D., Gaies, M. S. B., & Sahut, J. M. (2023). Predicting macro-financial instability-how relevant is sentiment? Evidence from long short-term memory networks. *Research in International Business and Finance*, 65, Article 101912.
- Katsafados, A. G., Nikoloutsopoulos, S., & Leledakis, G. N. (2023). Twitter sentiment and stock market: A COVID-19 analysis. *Journal of Economic Studies*, 50(8), 1866–1888.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica : Journal of the Econometric Society*, 47(2), 263–291.
- Kim, Y., & Lee, K. Y. (2022). Impact of investor sentiment on stock returns. *Asia-Pacific Journal of Financial Studies*, 1(51), 132–162.
- Khattak, B. H. A., Shafi, I., Khan, A. S., Flores, E. S., Lara, R. G., Samad, M. A., & Ashraf, I. (2023). A systematic survey of AI models in financial market forecasting for

- profitability analysis. *IEEE access : Practical innovations, open solutions*, 11, 125359–125380.
- Karpathy, A., & Li, F. F. (2017). Deep visual-semantic alignments for generating image descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 664–676.
- Li, Y. (2019). Multifractal view on China's stock market crashes. *Physica A-Statistical Mechanics and Its Applications*, 536, Article 122591.
- Liang, D. W., Xu, Y., Hu, Y., & Du, Q. Q. (2023). Intraday return forecasts and high-frequency trading of stock index futures: A hybrid wavelet-deep learning approach. *Emerging Markets Finance and Trade*, 59(7), 2118–2128.
- Liu, Z. F., Huynh, T. L. D., & Dai, P. F. (2021). The impact of COVID-19 on the stock market crash risk in China. *Research in International Business and Finance*, 57, Article 101419.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *Journal of Finance*, 66(1), 35–65.
- Liu, Y., Huang, X. H., Xiong, L. W., Chang, R. Y., Wang, W. J., & Chen, L. (2025). Stock price prediction with attentive temporal convolution-based generative adversarial network. *ARRAY*, 25, Article 100374.
- Li, Y., Ni, P., & Chang, V. (2019). An empirical research on the investment strategy of stock market based on deep reinforcement learning model. In *COMPLEXIS 2019 - Proceedings of the 4th International Conference on Complexity, Future Information Systems and Risk* (pp. 52–58).
- Li, P. R., Yuan, X. H., Jin, L. W., & Yuan, J. (2024). Pay inequality for stability? Investment preference of the national team intervention in Chinese stock market. *International Review of Financial Analysis*, 96, Article 103633.
- Li, X., & Xing, Y. (2023). When stock return synchronicity meets investor sentiment. *Finance Research Letters*, 53, Article 103655.
- Li, P. R., Yuan, X. H., & Jin, L. W. (2023). Government direct intervention and stock market concentration. *Applied Economics*, 55(25), 2863–2874.
- Li, Y., & Li, W. P. (2021). Firm-specific investor sentiment for the Chinese stock market. *Economic Modelling*, 97, 231–246.
- Lu, S., Liu, C. H., & Chen, Z. S. (2021). Predicting stock market crisis via market indicators and mixed frequency investor sentiments. *Expert Systems with Applications*, 186, Article 115844.
- Li, H. L., Chen, H. Q., Xu, G. L., & Zhang, W. G. (2023c). COVID-19, various government interventions and stock market performance. *China Finance Review International*, 13(4), 621–632.
- Meier, C., & De Mello, L. (2020). Investor overconfidence in experimental asset markets across market states. *Journal of Behavioral Finance*, 21(4), 369–384.
- Maqsood, H., Maqsood, M., Yasmin, S., Mehmood, I., Moon, J., & Rho, S. (2022). Analyzing the stock exchange markets of EU nations: A case study of Brexit social media sentiment. *Systems*, 10(2), 24.
- Ma, X. H., & Lin, H. (2023). Predicting stock market crises using stock index derivatives: Evidence from China. *Emerging Markets Finance and Trade*, 60, 576–597.
- Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M. M., & Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management*, 50, 432–451.
- Naveed, H. M., Yao, H. X., Memon, B. A., Ali, S., Alhussam, M. I., & Sohu, J. M. (2023). Artificial neural network (ANN)-based estimation of the influence of COVID-19 pandemic on dynamic and emerging financial markets. *Technological Forecasting and Social Change*, 190, Article 122470.
- Nguyen, T. T., Wu, Y. C., Ke, M. C., & Liao, T. L. (2022). Can direct government intervention save the stock market? *Quarterly Review of Economics and Finance*, 84, 271–284.
- Ni, Z. X., Wang, D. Z., & Xue, W. J. (2015). Investor sentiment and its nonlinear effect on stock returns—New evidence from the Chinese stock market based on panel quantile regression model. *Economic Modelling*, 50, 266–274.
- Petit, J. J. G., Lafuente, E. V., & Vieites, A. R. (2019). How information technologies shape investor sentiment: A web-based investor sentiment index. *Borsa Istanbul Review*, 19(2), 95–105.
- Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1), 83–104.
- Sharma, A., & Kumar, A. (2020). A review paper on behavioral finance: Study of emerging trends. *Qualitative Research in Financial Markets*, 12(2), 137–157.
- Srivastava, P. R., Zhang, Z. P., & Eachempati, P. (2021). Deep neural network and time series approach for finance systems: Predicting the movement of the Indian stock market. *Journal of Organizational and End User Computing*, 33(5), 204–226.
- Song, Z. Y., Gong, X. M., Zhang, C., & Yu, C. R. (2023). Investor sentiment based on scaled PCA method: A powerful predictor of realized volatility in the Chinese stock market. *International Review of Economics & Finance*, 83, 528–545.
- Singh, P., Jha, M., Sharaf, M., El-Meligy, M. A., & Gadekallu, T. R. (2023). Harnessing a hybrid CNN-LSTM model for portfolio performance: A case study on stock selection and optimization. *IEEE access : Practical innovations, open solutions*, 11, 104000–104015.
- Song, Z. T., Wang, Y. N., Qian, P., Song, S. F., Coenen, F., Jiang, Z. Y., & Su, J. L. (2022). From deterministic to stochastic: An interpretable stochastic model-free reinforcement learning framework for portfolio optimization. *Applied Intelligence*, 53(12), 15188–15203.
- Sahiner, M. (2023). Volatility spillovers and contagion during major crises: An early warning approach based on a deep learning model. *Computational Economics*, 63(6), 2435–2499.
- Shahab, Y., Wang, C. X., Yeung, E., & Zhou, J. N. (2024). The national team: Stock market interventions and corporate catering behavior. *International Review of Financial Analysis*, 93, Article 103203.
- Suárez-Cetrulo, A. L., Quintana, D., & Cervantes, A. (2024). Machine learning for financial prediction under regime change using technical analysis: A systematic review. *International Journal of Interactive Multimedia and Artificial Intelligence*, 9(1). <https://doi.org/10.9781/ijimai.2023.06.003>
- Ulku, N., Ali, F., Saydumarov, S., & Izkizlerli, D. (2023). COVID caused a negative bubble. Who profited? Who lost? How stock markets changed? *Pacific-Basin Finance Journal*, 79, Article 102044.
- Viebig, J. (2020). Exuberance in financial markets: Evidence from machine learning algorithms. *Journal of Behavioral Finance*, 21(2), 128–135.
- Wang, Y., Zhang, H. W., Gao, W., & Yang, C. (2021). COVID-19-related government interventions and travel and leisure stock. *Journal of Hospitality and Tourism Management*, 49, 189–194.
- Wu, L. W., Gao, D., Su, W. X., Liang, D. W., & Du, Q. Q. (2023). A trading strategy based on analysts' industry analyses - evidence from textual analyses of analyst reports in Chinese stock market. *Emerging Markets Finance and Trade*, 59(10), 3378–3389.
- Wang, Z. X., Hu, Z. D., Li, F., Ho, S. B., & Cambria, B. (2023). Learning-based stock trending prediction by incorporating technical indicators and social Media sentiment. *Cognitive Computation*, 15(3), 1092–1102.
- Yalamova, R., & McKelvey, B. (2011). Explaining what leads up to stock market crashes: A phase transition model and scalability dynamics. *Journal of Behavioral Finance*, 12(3), 169–182.
- Yang, J. W., Wang, Y. M., & Li, X. (2022). Prediction of stock price direction using the LASSO-LSTM model combines technical indicators and financial sentiment analysis. *PeerJ Computer Science*, 8, e1148.
- Yang, C., Abedin, M. Z., Zhang, H. W., Weng, F. T., & Hajek, P. (2023a). An interpretable system for predicting the impact of COVID-19 government interventions on stock market sectors. *Annals of operations research*. <https://doi.org/10.1007/s10479-023-05311-8>
- Yu, P. F., & Yan, X. S. (2020). Stock price prediction based on deep neural networks. *Neural Computing & Applications*, 32(6), 1609–1628.
- Yang, B., Liang, T., Xiong, J., & Zhong, C. (2023b). Deep reinforcement learning based on transformer and U-net framework for stock trading. *Knowledge-Based Systems*, 263, Article 110211.
- Zhou, Z. Q., Li, J., Zhang, W., & Xiong, X. (2022). Government intervention model based on behavioral heterogeneity for China's stock market. *Financial Innovation*, 8(1), 9.
- Zaremba, A., Kizys, R., Tzouvanas, P., Aharon, D. Y., & Demir, E. (2021). The quest for multidimensional financial immunity to the COVID-19 pandemic: Evidence from international stock markets. *Journal of International Financial Markets Institutions & Money*, 71, Article 101284.