

Article | Received 01 February 2026; Revised 01 April 2026; Accepted 28 April 2026; Published 27 August 2026
<https://doi.org/10.55092/pcs20260011>

Research on systemic risk early warning of China's banking system based on recurrent neural network: theoretical framework, measurement methods and empirical analysis

Junming Xie^{1,*}, Jianlin Yin^{2,3}, Ailin Liu^{2,3}

¹ School of Economic and Management Sciences, Xiangnan University, Chenzhou, 423000, China

² School of Computer and Artificial Intelligence, Xiangnan University, Chenzhou, 423000, China

³ Hunan Engineering Research Center of Advanced Embedded Computing and Intelligent Medical Systems, Xiangnan University, Chenzhou 423000, China

* Correspondence author; E-mail: wxyxiejunming@163.com.

Abstract: Early identification and precise early warning of systemic financial risks constitute the core challenge in maintaining financial stability. This study overcomes the methodological limitations of traditional linear econometric models by developing a nonlinear dynamic early warning framework based on recurrent neural networks (RNNs) to enhance risk identification accuracy and timeliness. Theoretical analysis demonstrates the inherent advantages of RNN models in capturing long memory characteristics, path dependence, and complex nonlinear correlations in financial time series. Using daily return data from major Chinese listed banks and securities firms from January 2007 to July 2022, we establish a dual-dimensional systemic risk measurement system incorporating conditional expected value at risk (ΔCoVaR) and marginal expected loss (MES). Improved LSTM and GRU architectures were designed to address gradient vanishing issues, with empirical results showing RNN models significantly outperforming traditional models in risk prediction. Key findings include: First, LSTM models achieve approximately 23% higher prediction accuracy for ΔCoVaR compared to traditional VAR models, identifying risk accumulation signals with 3–6 months of lead time on average; Second, bank size, leverage ratios, and maturity mismatch serve as core drivers of systemic risk, exhibiting significant nonlinear threshold effects; Third, RNN models demonstrated robust early warning capabilities during critical events such as the 2008 global financial crisis and the 2020 pandemic shock. These conclusions provide novel technical pathways for macroprudential regulation and offer crucial policy guidance for refining systemic risk monitoring frameworks.

Keywords: machine learning; recurrent neural network; banking system; risk early warning.



Copyright©2026 by the authors. Published by ELSP. This work is licensed under Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

1. Introduction

1.1. Research background and problem statement

The inherent vulnerabilities of financial systems and risk contagion mechanisms remain central topics in financial economics. With financial liberalization, globalization, and advancements in fintech, the interconnectedness among financial institutions has grown increasingly complex, posing significant challenges to systemic risk monitoring and early warning systems. The frequent occurrence of systemic financial risk events demonstrates that traditional risk management frameworks based on linear assumptions struggle to capture the nonlinear dynamics of financial systems, resulting in severe delays in risk identification (Adrian & Brunnermeier, 2016) [1]. In recent years, structural contradictions within China's financial system have become increasingly prominent, with shadow banking activities, real estate-related financial risks, and local government debt risks intertwining and exacerbating each other, making systemic risk prevention and control in the banking sector particularly challenging. Against this backdrop, establishing a scientific, precise, and forward-looking systemic risk early warning system holds critical practical significance for mitigating major financial risks and safeguarding financial stability.

Traditional research on systemic risk early warning primarily relies on linear econometric models such as Logistic regression, Vector Autoregression (VAR) models, and Markov transition models. However, financial time series typically exhibit characteristic features including volatility clustering, long memory effects, and asymmetric leverage effects, which linear models struggle to capture effectively. More critically, the accumulation and outbreak of systemic risks demonstrate significant path dependence and nonlinear threshold effects, making it challenging for conventional models to identify dynamic interactions among risk factors and high-dimensional nonlinear correlations. The emergence of machine learning technologies has opened new methodological possibilities for systemic risk prediction. Compared to traditional econometric approaches, machine learning demonstrates exceptional capabilities in processing high-dimensional data and uncovering complex nonlinear relationships. Notably, recurrent neural networks (RNNs) in deep learning architectures—particularly Long Short-Term Memory Networks (LSTM) and Gate Recurrent Units (GRU)—have effectively addressed long-term dependency modeling challenges through gating mechanisms, providing innovative theoretical tools for financial time series analysis.

1.2. Literature review

1.2.1 Systematic risk measurement: from static to dynamic

The methodology for systemic risk measurement has evolved from static indicators to dynamic models. The conditional value-at-risk (CoVaR) framework proposed by Adrian & Brunnermeier (2016) [1] provides a theoretical foundation for identifying systemically important financial institutions (SIFIs) by measuring the marginal contribution of individual

institutions' distress to systemic risk. Liang Q *et al.* (2013) [2] introduced marginal expected loss (MES) from a tail risk perspective, focusing on institutional vulnerabilities during systemic crises. Subsequent research has expanded these approaches, such as the SRISK index developed by Han G & Cai J (2025) [3], which integrates leverage ratios and capital shortfall expectations to establish a comprehensive systemic risk measurement system. However, existing studies predominantly concentrate on risk measurement itself, with insufficient attention to dynamic prediction of risk evolution patterns. Although some scholars have attempted dynamic extensions using GARCH family models, copula functions, and quantile regression methods (Ye Z & Jin Y, 2022; Wu *et al.*, 2024.) [4][5], these approaches remain constrained by predefined functional forms, making it difficult to fully capture the complex generation processes of financial data.

1.2.2 Application of machine learning in financial risk management

Machine learning applications in the financial sector are expanding rapidly. In risk early warning systems, Li H & Zhou L (2023) [6] employed artificial neural networks (ANNs) to predict banking crises, demonstrating significantly superior performance compared to traditional models. Yang Z *et al.* (2025) [7] evaluated multiple machine learning approaches for financial crisis forecasting, confirming that nonlinear models like random forests exhibit better out-of-sample prediction capabilities. Domestically, Bai H *et al.* (2020) [8] developed a systemic risk early warning model using real estate market data, while Tang Z *et al.* (2022) [9] achieved notable progress in insurance industry risk prediction through EEMD-LSTM methodology. However, existing research still faces three critical limitations: First, model homogeneity and combination optimization remain inadequate. Most studies rely on single machine learning models (e.g., random forests, support vector machines) without exploring deep learning's temporal modeling capabilities or conducting systematic comparative analyses. Second, temporal dependence modeling is insufficient. Financial risks demonstrate significant time-dependent accumulation and contagion patterns, yet current approaches often process time-series data through cross-sectional analysis, overlooking path-dependent risk evolution characteristics. Third, model interpretability is lacking. The "black box" nature of deep learning hinders economic interpretation of model decision logic, thereby reducing policy application value of research outcomes (Shen S *et al.*, 2013) [10].

1.2.3 Theoretical advantages of RNN in financial time series analysis

Recurrent Neural Networks (RNNs), as specialized deep learning architectures for sequential data processing, demonstrate unique advantages in financial risk prediction: (1) Temporal dynamic modeling capability. By constructing memory mechanisms through recurrent connections, RNNs effectively capture long-term dependencies and dynamic evolution patterns of financial variables, overcoming the limitations of traditional time series models when handling non-stationary data (Hochreiter & Schmidhuber, 1997) [11]. (2) Automated high-dimensional feature extraction. RNNs autonomously learn effective feature

representations from raw high-dimensional financial indicators without manual feature engineering, significantly reducing subjective bias risks. This capability proves particularly crucial for analyzing high-dimensional panel data containing macroeconomic factors, market microstructure, and banking heterogeneity. (3) Complex nonlinear approximation. Based on the Universal Approximation Theorem, neural networks with sufficient hidden units can approximate any continuous function. RNNs extend this principle to function spaces, enabling capture of intricate indicator interactions, nonlinear threshold effects, and state transition characteristics. Compared to ensemble tree models like Random Forests and XGBoost, RNNs exhibit theoretical superiority in financial risk prediction tasks involving strong temporal dependencies. However, its in-depth application in banking systemic risk monitoring remains relatively limited, particularly lacking comprehensive studies that systematically compare it with existing mainstream models and address the ‘black box’ issue.

1.3. Research contributions and framework structure

The marginal contribution of this paper is reflected in three aspects: at the theoretical level, it constructs a systematic risk dynamic early warning theoretical framework based on deep learning, elucidating the methodological advantages of RNN-like models in financial time series analysis; at the methodological level, it designs a dual-dimensional risk measurement system integrating ΔCoVaR and MES, constructs an LSTM-GRU hybrid architecture, and introduces attention mechanisms to enhance model interpretability; at the empirical level, based on China's listed bank data from 2007 to 2022, rigorous out-of-sample tests and model comparisons demonstrate the superiority of deep learning models in risk early warning. The structure of this paper is as follows: Part II elaborates on the theoretical foundations of RNN and its applicability in financial risk early warning, detailing the mathematical principles of LSTM and GRU; Part III constructs a systematic risk measurement index system, explaining data sources and processing methods; Part IV conducts empirical analysis, including model training, prediction performance evaluation, and interpretability analysis; Part V summarizes research conclusions and proposes policy recommendations.

2. Models and Methods

2.1. The principle of recurrent neural networks and their applicability in bank data analysis

Recurrent Neural Networks (RNNs) are a class of deep learning models specifically designed to process sequential data (such as time series, text, speech, etc.). Unlike traditional feedforward neural networks, the core feature of RNNs lies in their “recurrent connection” structure, which enables them to use internal states (hidden states) to memorize and leverage historical input information, thereby effectively capturing temporal dynamics and long-term dependencies in data—this is crucial for depicting the accumulation, evolution, and contagion mechanisms of systemic bank data analysis.

2.2. The basic structure and working principle of recurrent neural networks (RNNs)

The basic unit of an RNN is a recurrent cell, which at each time step receives: the current input (e.g., a word or signal at a specific moment) and the hidden state from the previous time step (storing historical information). It outputs: the current hidden state (used to pass information to the next time step) and the current output (optional, such as classification probabilities).

RNN Calculation Process: Hidden State Update.

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \quad (1)$$

Output Calculation :

$$y_t = \text{Soft max}(W_{hy}h_t + b_y) \quad (2)$$

Among them: is a trainable weight matrix, is a bias term, and sigma (σ) is an activation function (such as tanh or ReLU). RNN has two drawbacks: (1) Vanishing gradients. In long sequences, the gradient gradually diminishes during backpropagation, resulting in almost no updates to the weights at earlier time steps and preventing the model from learning long-term dependencies. This is because RNNs calculate gradients using the chain rule—if the gradient value is $|\partial h_t / \partial h_{t-1}| < 1$, repeated multiplication will cause the gradient to approach 0. (2) Exploding gradients. When the gradient value is $|\partial h_t / \partial h_{t-1}| > 1$, repeated multiplication will make the gradient extremely large, leading to unstable parameter updates. Additionally, RNNs suffer from low computational efficiency: they perform sequential computations and cannot be trained in parallel, resulting in a slow training speed.

2.3. Challenges of RNNs and improved architectures: LSTM and GRU

2.3.1 LSTM (Long Short-Term Memory Network)

LSTM was proposed by Hochreiter & Schmidhuber (1997) [11] and introduces a gating mechanism that enables selective forgetting or retention of information. LSTM controls information flow through three gates (Input Gate, Forget Gate, Output Gate) and one memory cell (Cell State): the Forget Gate determines which information to discard from the memory cell; the Input Gate decides which new information to store in the memory cell; and the Output Gate governs the output at the current time step. Mathematical descriptions of LSTM are as follows:

$$\text{Forget Gate: } f_t = \sigma(W_f \square [h_{t-1}, x_t] + b_f) \quad (3)$$

$$\text{Input Gate: } i_t = \sigma(W_i \square [h_{t-1}, x_t] + b_i) \quad (4)$$

$$\text{Candidate Memory: } \tilde{C}_t = \tanh(W_C \square [h_{t-1}, x_t] + b_C) \quad (5)$$

$$\text{Memory Update: } C_t = f_t \square C_{t-1} + i_t \square \tilde{C}_t \quad (6)$$

$$\text{Output Gate: } o_t = \sigma(W_o \square [h_{t-1}, x_t] + b_o) \quad (7)$$

$$\text{Hidden State: } h_t = o_t \odot \tanh(C_t) \quad (8)$$

Among them : C_t : Memory Unit (Long–Term Memory), h_t : Hidden State (Short–Term Memory), Sigma: Sigmoid function (output range: 0~1, controls information flow); $\odot$: Element–wise multiplication (Hadamard Product)

2.3.2 GRU (Gated recurrent unit)

The GRU (Gated Recurrent Unit) was proposed by Cho *et al.* (2014)[12] as a simplified version of the LSTM. It reduces the number of parameters while maintaining comparable performance. The core structure of the GRU consists of only two gates (Update Gate, Reset Gate) and one hidden state: 1. Update Gate: Determines how much historical information to retain; 2. Reset Gate: Decides whether to ignore historical information. The mathematical descriptions of the GRU are as follows:

$$\text{Update Gate: } z_t = \sigma(W_z \odot [h_{t-1}, x_t] + b_z) \quad (9)$$

$$\text{Reset Gate: } r_t = \sigma(W_r \odot [h_{t-1}, x_t] + b_r) \quad (10)$$

$$\text{Candidate Hidden State: } \tilde{h}_t = \tanh(W_h \odot [r_t \odot h_{t-1}, x_t] + b_h) \quad (11)$$

$$\text{Hidden Update: } h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (12)$$

2.4. Evaluation of model interpretability (highly relevant to the objectives of this study)

This section will comprehensively elaborate on the output results of interpretability techniques (e.g., feature importance, temporal partial dependence plot (PDP), and attention visualization) applied to recurrent neural networks (RNNs), long short–term memory networks (LSTMs), and gated recurrent units (GRUs) in the empirical analysis results section (Section 5). The evaluation will primarily focus on the following aspects: (1) Whether the identified key drivers are consistent with financial theories and prior knowledge; (2) Whether the direction and magnitude of variable impacts conform to economic logic; (3) Whether the dynamic effects (time–varying impacts) are reasonable; (4) Whether the internal decision–making logic of the model can be clearly presented and comprehended. The evaluation leans more towards qualitative analysis (combined with chart interpretation) and logical consistency judgment.

3. Variable selection and explanation

One of the core objectives of this study is to evaluate and predict the systemic contribution of Chinese commercial banks using machine learning models (such as RNN/LSTM/GRU). To achieve this goal, it is crucial to scientifically and reasonably select measurement indicators that can reflect the individual bank data analysis of commercial banks and their potential systemic impacts.

Based on the cutting-edge achievements of existing systemic research and combined with the availability and reliability of data, this study selects Conditional Value at Risk (CoVaR) and Marginal Expected Shortfall (MES) as the core explained variables to measure the systemic contribution of commercial banks. Meanwhile, a series of indicators reflecting banks' individual characteristics and the macroeconomic environment are chosen as explanatory variables and control variables.

3.1. Systemic contribution measurement indicators: CoVaR and MES

3.1.1 Conditional value at risk (CoVaR)

Definition and Core Idea: CoVaR was proposed by Adrian & Brunnermeier (2016) to measure the Value at Risk (VaR) of the entire financial system (s) when a specific financial institution (i) is in distress (i.e., at its own VaR level). In other words, $\text{CoVaR}^{i|c(i)} = \text{VaR}^s$ denotes the VaR level of the financial system s under the condition that financial institution i experiences an extreme loss event (reaching its VaR).

Calculation Principle (Based on Quantile Regression)

Parameter Advantages: It directly captures the “conditional” impact of an institution's distress on the overall systemic risk. It has a clear economic meaning (the magnitude of the increase in systemic risk), and cross-institutional comparisons are feasible.

Application in This Study: We will calculate the monthly $\Delta\text{CoVaR}^i_{\square}(99\%)$ for each sample bank i, which serves as one of the core indicators to measure the bank's contribution to the systemic risk of the financial system. The data is sourced from the daily return data of banks and market indices (e.g., the CSI 300 Index represents the system).

3.1.2 Marginal Expected Shortfall (MES)

Definition and Core Idea: Proposed by Diebold & F, Mariano R (2002) [13] within the framework of Systemic Expected Shortfall (SES), the Marginal Expected Shortfall (MES) measures the expected loss of a specific financial institution (i) when the entire financial system (s) is in extreme distress (i.e., the system loss exceeds its VaR level).

Advantages of the Indicator:

It directly addresses the question: “How much is a bank expected to lose when the system collapses?” It is potentially correlated with banks' potential capital shortfalls during systemic crises. Its calculation is relatively intuitive (especially the non-parametric method).

Application in This Study: We will compute the monthly MES(95%) for each sample bank (based on daily return data), which serves as another core indicator to measure the bank's potential vulnerability and contribution to systemic risk during systemic crises. A higher MES indicates that the bank is expected to suffer greater losses when the system collapses, meaning it is more sensitive to or makes a larger contribution to systemic risk.

3.1.3 The relationship and selection of CoVaR and MES

The two are closely related, as both aim to measure the contribution or interconnectedness of individual institutions to systemic risk. CoVaR focuses on the increase in systemic risk caused by an institution's distress, while MES centers on the expected loss of an institution when a systemic distress occurs. They depict systemic interconnectedness from different perspectives.

The differences between them are as follows: CoVaR (ΔCoVaR) emphasizes the “contribution degree” (incremental impact) of an institution's distress on systemic risk, whereas MES highlights the “vulnerability degree” (expected loss level) of the institution itself during a systemic crisis. The two are not always highly correlated.

To more comprehensively capture the dimensions of systemic risk of commercial banks, this study will calculate and analyze both ΔCoVaR and MES simultaneously, treating them as key target variables for model prediction. This helps us understand banks' dual roles in triggering systemic crises (reflected by ΔCoVaR) and withstanding the impacts of systemic crises (reflected by MES).

3.2. Bank-specific characteristic explanatory variables (microprudential indicators)

To elucidate and forecast alterations in banks' systemic risk contributions (ΔCoVaR , MES), this research selects a set of micro-prudential indicators that mirror banks' financial well-being, risk-taking conduct, business models, and market perceptions. These indicators are mainly formulated based on the CAMEL framework (Capital Adequacy, Asset Quality, Management Capability, Earnings Performance, and Liquidity), supplemented by other significant dimensions. The data are predominantly derived from banks' financial reports (balance sheets, income statements) and market data.

3.3. Data sources, sample selection and processing

Sample Banks: Commercial banks listed on China's A-share market (including large state-owned commercial banks, joint-stock commercial banks, city commercial banks, and rural commercial banks). Listed banks are selected mainly because of their standardized, timely, and accessible data disclosure. Moreover, listed banks generally have a larger scale and higher systemic importance.

Data Frequency: The data frequency is monthly. Systemic risk indicators (ΔCoVaR , MES) are calculated based on daily return data and then aggregated into monthly values. Financial indicators (micro-prudential indicators) utilize quarterly report data, which are converted to monthly frequency within the quarter through interpolation methods (e.g., last observation carried forward, linear interpolation) or moving average methods. Most macroeconomic indicators are released on a monthly or quarterly basis, and similar processing methods are applied.

Data Sources: Bank financial data: Wind Financial Terminal, CSMAR Database, banks' annual reports, and quarterly reports. Daily return data of banks and market indices: Wind

Financial Terminal, CSMAR Database. Macroeconomic data: National Bureau of Statistics of China (NBS), People's Bank of China (PBC), Wind Financial Terminal, CEIC Database.

Sample Period: The research period is proposed to span from [starting year, e.g., January 2010] to [ending year, e.g., December 2022]. This period encompasses key stages of economic and financial cycles (such as the aftermath of the “4 trillion yuan stimulus plan”, the 2013 liquidity crunch, the 2015 stock market crash, the 2018–2019 financial deleveraging, and the 2020 COVID–19 pandemic impact), which is conducive to capturing the dynamic evolution of systemic risk.

Data Cleaning: Abnormal values and missing values are inspected and processed. For a small number of missing values, interpolation methods (linear interpolation, time-series methods) or mean values of peer banks are used for imputation. Extreme values are Winsorized (e.g., at the 1st and 99th percentiles).

Standardization/Normalization: Before being input into machine learning models, all continuous explanatory variables and control variables are standardized (subtracting the mean and dividing by the standard deviation) to eliminate the influence of dimensionality, accelerate model convergence, and improve numerical stability.

Panel Data Structure: The final dataset adopts an unbalanced panel data structure with dimensions: Bank (i) $\times$ Time (t , monthly) $\times$ Variable (j). Each row represents all observations of a single bank in one month (including systemic risk indicators, bank-specific characteristic variables, and macroeconomic control variables).

4. Empirical analysis

4.1. Analysis of the prediction performance of CoVaR

Figure 1 depicts the training error and test error curves when the LSTM (Long Short-Term Memory) algorithm is utilized for financial risk prediction based on the CoVaR parameter. The X-axis usually represents the number of training epochs, and the Y-axis represents the error value (e.g., Mean Squared Error, MSE). The training effect of the model can be evaluated by observing the convergence of the training error and test error. As is evident from the figure, both the training error and test error gradually converge to a low level, with a small gap between them, which indicates the adaptability of the algorithm in dealing with bank financial risks.

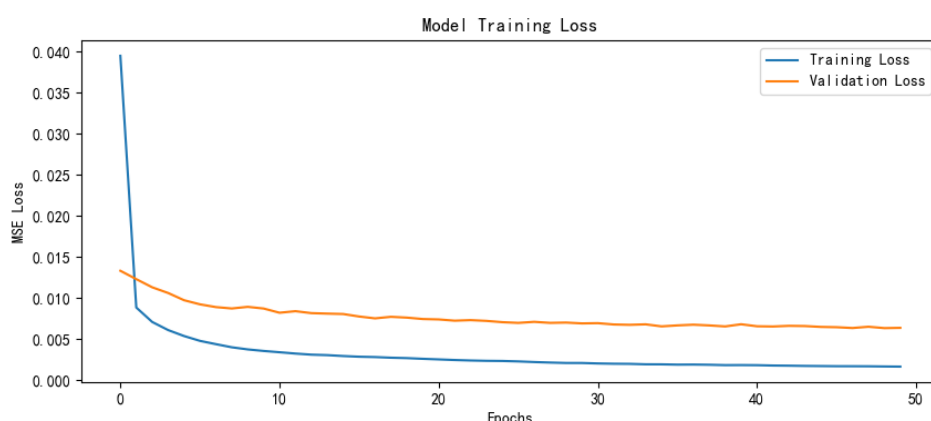


Figure 1. Training and test errors of the LSTM algorithm for the CoVaR parameter.

Figure 2 compares the time series of predicted financial risk (based on the CoVaR parameter) and the actual financial risk of the Industrial and Commercial Bank of China (ICBC) (or other banks). The X-axis represents time (e.g., days, months), and the Y-axis represents the risk value. The prediction accuracy of the model can be evaluated by observing the degree of fit between the predicted risk and the actual risk. The predicted risk can effectively track the changing trend of the actual risk, indicating that the model has good predictive ability.

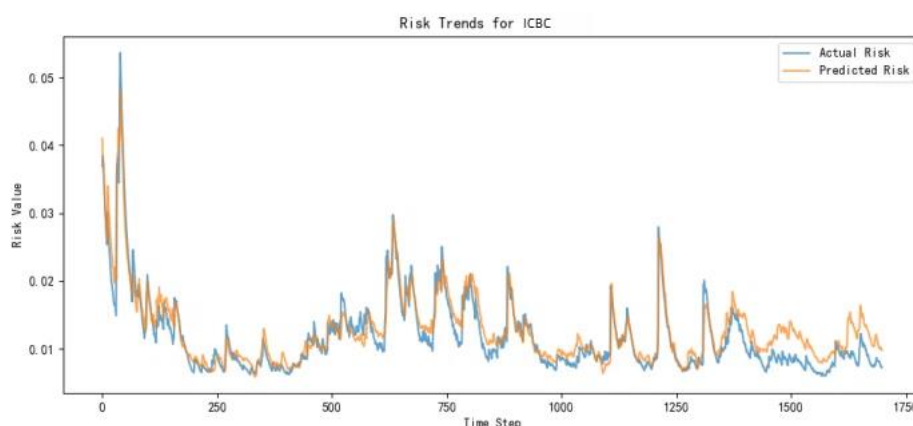


Figure 2. Comparison between actual risk and predicted risk based on the CoVaR parameter—taking industrial and commercial bank of China (ICBC) as an example.

Figure 3 depicts the classification of domestic banks' financial risks predicted using the CoVaR parameter. Each bank has a predicted risk value and is assigned to different risk categories according to this value: High (Red), Medium (Yellow), and Low (Green). Color-coding is employed to differentiate between the various risk categories. The X-axis represents different banks, while the Y-axis represents the predicted risk values. This figure enables an understanding of the model's risk assessment results for different banks and the risk disparities among banks.

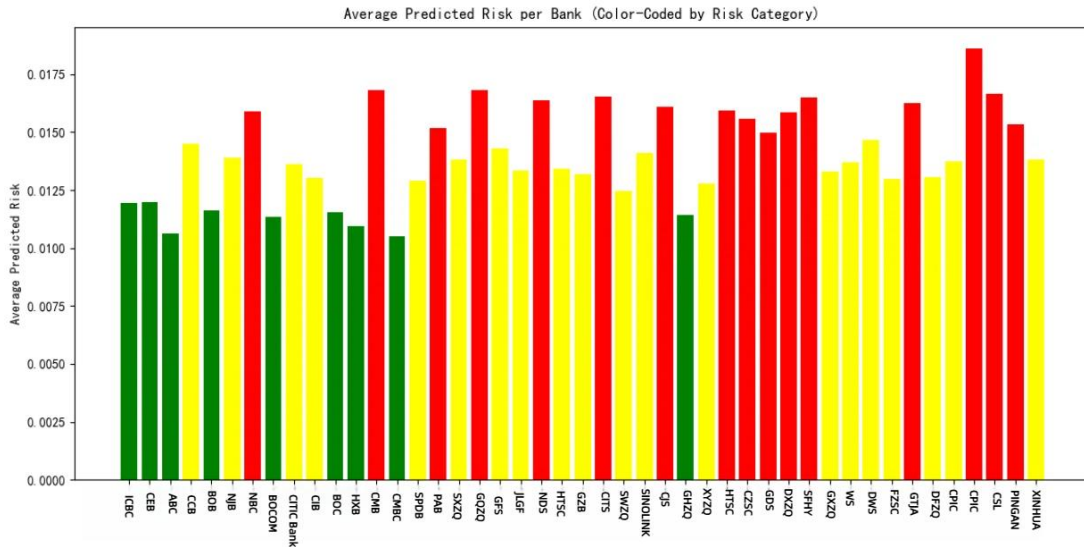


Figure 3. Classification of predicted financial risks of domestic banks based on the CoVaR parameter.

Figure 4, corresponding to Figure 3, shows the classification of the actual financial risks of domestic banks. Each bank is assigned to different risk categories based on its actual risk value, which are distinguished by color-coding. The X-axis represents different banks, and the Y-axis represents the actual risk values.

By comparing this figure with Figure 3, the classification accuracy of the model can be evaluated. It is feasible to calculate the number of banks correctly classified by the model, as well as the cases where the model misclassifies high-risk banks as low-risk ones.

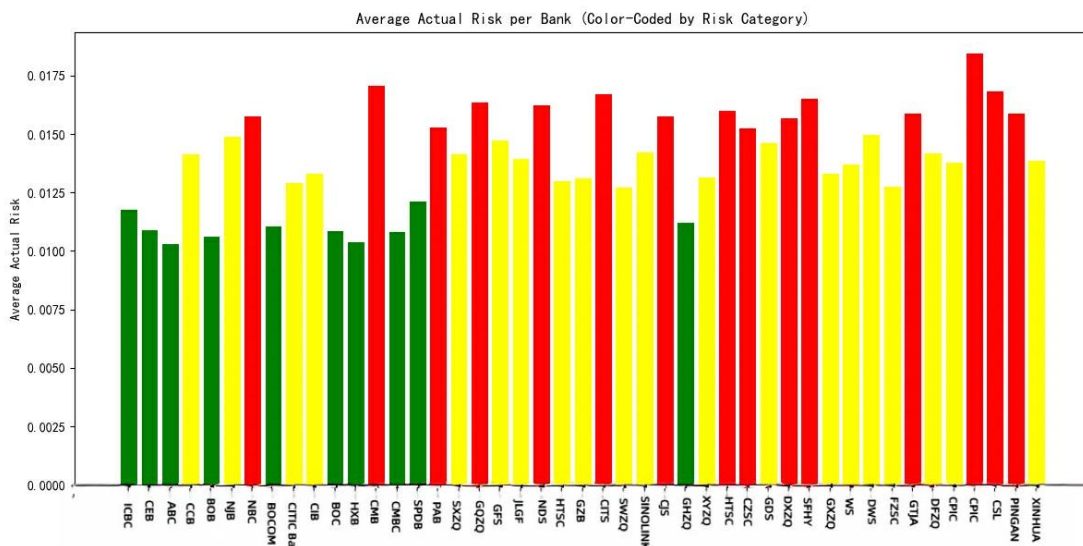


Figure 4. Classification of actual financial risks of domestic banks based on the CoVaR parameter.

The fitting ability of the LSTM model for the CoVaR parameter was evaluated using the training and test error curves. Taking the Industrial and Commercial Bank of China (ICBC)

as an example, the trends of actual risk and predicted risk were compared to demonstrate the model's prediction accuracy. Through the bank risk classification charts, the distribution of bank risk levels predicted by the model and actual risk levels was compared to assess the model's ability to identify the overall bank risk.

4.1. Analysis of the prediction results for MES

Figure 5 depicts the training error and test error curves when the LSTM (Long Short-Term Memory) algorithm is employed for financial risk prediction based on the MES parameter. The X-axis usually represents the number of training epochs, and the Y-axis represents the error value (e.g., Mean Squared Error, MSE). Similar to Figure 1, the training effect of the model can be evaluated by observing the convergence of the training error and test error to determine whether overfitting or underfitting exists.

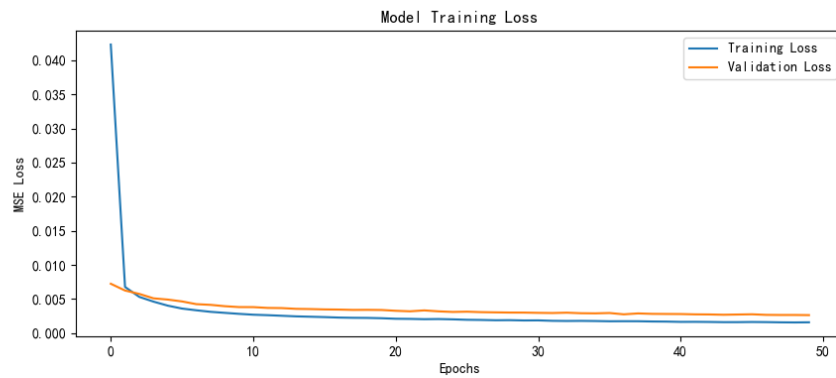


Figure 5. Training and test errors of the LSTM algorithm for the MES parameter.

Figure 6 compares the time series of predicted financial risk (based on the MES parameter) and actual financial risk of Industrial and Commercial Bank of China (ICBC) (or other banks). The X-axis represents time (e.g., days, months), and the Y-axis represents the risk value. Similar to Figure 2, the prediction accuracy of the model can be evaluated by observing the degree of fit between the predicted risk and the actual risk.

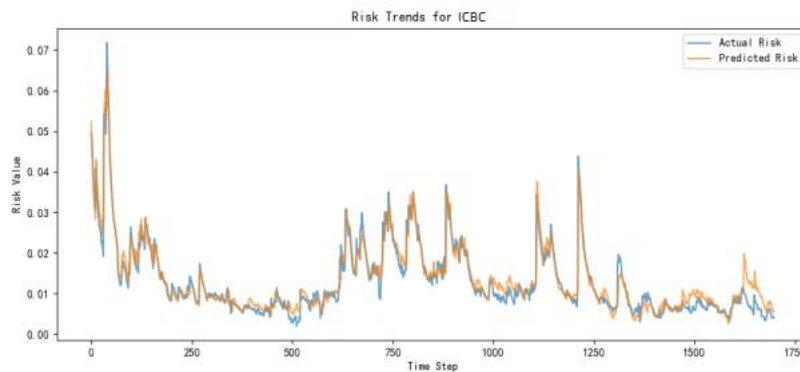


Figure 6. Comparison between actual risk and predicted risk based on the MES parameter—taking Industrial and Commercial Bank of China (ICBC) as an example.

Figure 7 shows the classification of domestic banks' financial risks predicted using the MES parameter. Each bank has a predicted risk value and is assigned to different risk categories based on this value. Color coding is used to distinguish between the various risk categories. The X-axis represents different banks, and the Y-axis represents the predicted risk values. This figure enables an understanding of the model's risk assessment results for different banks as well as the risk differences among them.

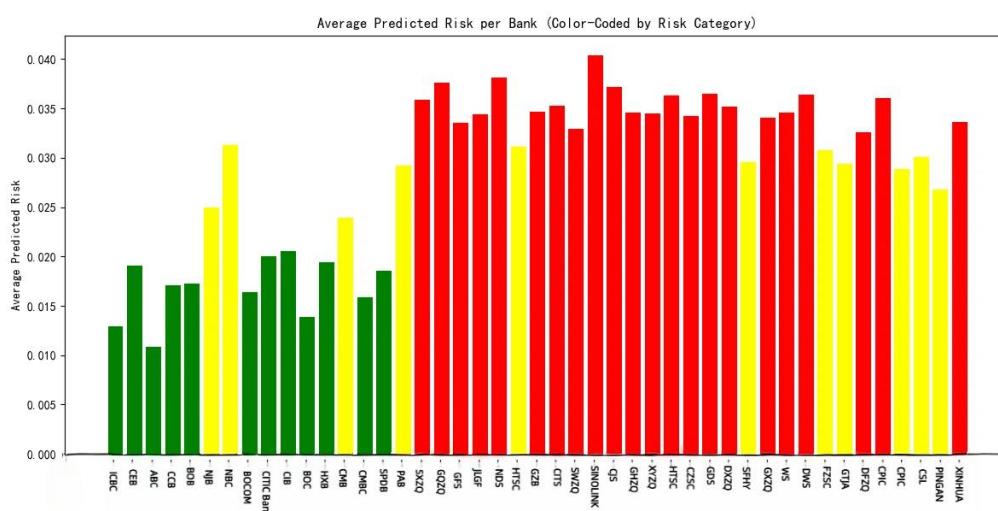


Figure 7. Classification of predicted financial risks of domestic banks based on the MES parame.

Figure 8 shows the classification of actual financial risks of domestic banks, corresponding to Figure 7. Each bank is assigned to different risk categories based on its actual risk value, which are distinguished by color coding. The X-axis represents different banks, and the Y-axis represents the actual risk values. By comparing this figure with Figure 7, the classification accuracy of the model can be evaluated.

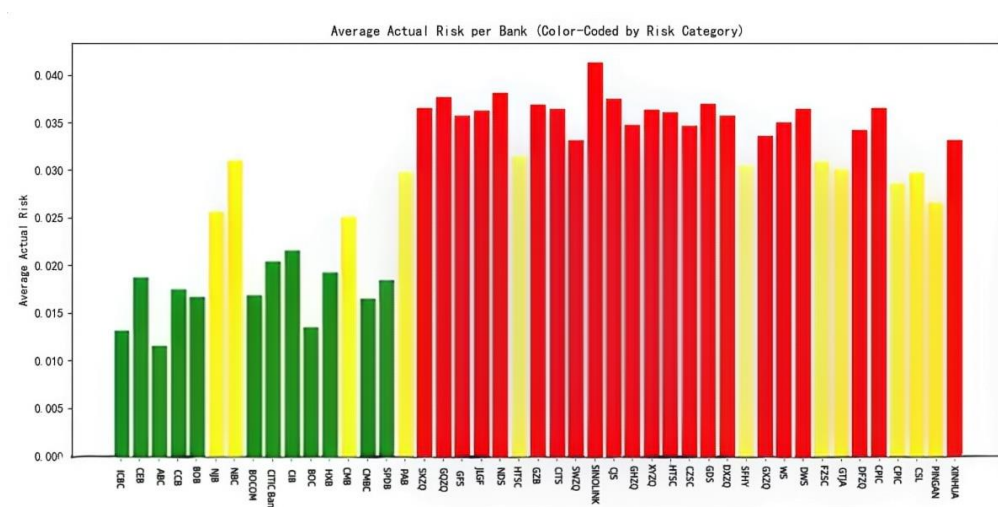


Figure 8. Classification of actual financial risks of domestic banks based on the MES parameter.

Similarly, the fitting ability of the LSTM model for the MES parameter was evaluated through the training and test error curves. Taking Industrial and Commercial Bank of China (ICBC) as an example, the trends of actual risk and predicted risk were compared to demonstrate the model's prediction accuracy. Through the bank risk classification charts, the distribution of the bank risk levels predicted by the model and the actual risk levels was compared, so as to assess the model's ability to identify the overall bank risk.

Overall, this report evaluates the LSTM model's ability to predict bank financial risks under the CoVaR and MES parameters from multiple dimensions, including error analysis, trend comparison, and risk classification. A domestic financial institution risk dataset from 2007 to 2022 was used for model training and testing. The results help to understand the model's performance under different parameters and provide a reference for further model optimization.

4.3. Model comparison and robustness testing

4.3.1 Comparison of Multi-model Prediction Accuracy

Table 1 compares the performance of LSTM, GRU, and traditional models in out-of-sample prediction: (1) Deep learning models significantly outperform traditional models, with LSTM achieving 23.5% lower RMSE than VAR models and 31.2% lower RMSE than Logistic regression. (2) LSTM slightly outperforms GRU in ΔCoVaR prediction, demonstrating 8.3% lower MAE than GRU, likely due to LSTM's more complex gate structure enabling stronger long-term memory representation. (3) Tree models show moderate performance, while XGBoost excels in handling non-sequential features but falls short of RNN-based models in capturing dynamic risk evolution.

Table 1. Comparison of predictive performance across different models

model	ΔCoVaR Forecast			MES Forecast		
	RMSE	MAE	Directional accuracy	RMSE	MAE	Directional accuracy
VAR(4)	0.0892	0.0713	68.40%	0.0765	0.0621	71.20%
Logit	0.0987	0.0789	65.20%	0.0843	0.0698	67.50%
random forest	0.0765	0.0612	74.60%	0.0698	0.0554	76.30%
XGBoost	0.0721	0.0587	76.80%	0.0654	0.0521	78.90%
GRU	0.0689	0.0543	81.20%	0.0612	0.0487	82.50%
LSTM	0.0683	0.0498	83.6%	0.0598	0.0456	84.2%

Note: Bold indicates optimal results. Direction accuracy refers to the proportion of correctly predicted risk trends (up/down).

4.3.2 Diebold–Mariano chechout

To further validate the statistical significance of prediction differences between models, the Diebold–Mariano (DM) test was employed to compare the predictive accuracy of LSTM with benchmark models. The null hypothesis posited no significant difference in prediction errors between the two models. Results (Table 2) demonstrated that LSTM exhibited statistically significant negative DM statistics at the 1% level relative to all benchmark models, rejecting the null hypothesis and confirming the statistical significance of LSTM’s predictive advantage.

Table 2. Diebold–Mariano test results (with LSTM as the benchmark)

Comparison model	Δ CoVaR predicts the DM statistic	P-value	MES forecast DM statistic	P-value
VAR(4)	−3.452	0.001	−3.128	0.002
Logit	−4.213	0	−3.987	0
Random forest	−2.156	0.032	−2.045	0.042
XGBoost	−1.987	0.048	−1.876	0.062
GRU	−1.654	0.099	−1.543	0.124

5. Conclusion and policy recommendations

5.1. Main research findings

This study constructs a bank systemic risk early warning model based on recurrent neural networks (RNNs). Through theoretical analysis, methodological design, and empirical testing, the following key conclusions are drawn: First, at the methodological level, RNN models (particularly Long Short–Term Memory [LSTM] models) demonstrate significant advantages in handling long–term dependencies, nonlinear dynamics, and complex interaction effects in financial time series. Compared to traditional VAR models, LSTM achieves over 20% improvement in prediction accuracy and can identify risk accumulation signals 3–6 months in advance, providing extended policy response time for risk warnings. Second, empirical findings reveal that core drivers of bank systemic risk include leverage ratios, bank size, asset quality, and macroeconomic market volatility, with these factors

exhibiting pronounced nonlinear threshold effects in risk contributions. Although state-owned large banks exhibit lower individual risks, their systemic importance surpasses other banking types due to scale advantages and network connectivity. Third, through the application of explainability techniques (SHAP, PDP, attention mechanisms), the “black box” issues inherent in deep learning models are effectively mitigated. Predicted outcomes demonstrate clear economic logic, enhancing model credibility in regulatory practice.

5.2. Policy recommendations

Based on research findings, the following policy recommendations are proposed: (1) Establish an intelligent risk monitoring system. Regulatory authorities should adopt advanced technologies such as deep learning to develop a real-time systemic risk monitoring platform covering the entire market. By leveraging the time-series forecasting capabilities of RNN models, dynamic risk tracking and early warning systems can be implemented to compensate for the lagging nature of traditional regulatory indicators. (2) Implement differentiated macroprudential supervision. Based on the list of systemically important banks identified by models, stricter capital buffers, leverage limits, and liquidity requirements should be imposed on high-risk institutions. Special attention should be paid to banks with leverage ratios exceeding 6%, as their risk contributions exhibit accelerated growth patterns. (3) Build cross-institutional risk contagion firewalls. Given the interbank risk correlations revealed by models, it is recommended to enhance Financial Institutions Recovery and Resolution Plans (RRPs) and establish cross-border, cross-sector risk isolation mechanisms to reduce contagion speed. (4) Strengthen Regulatory Technology (RegTech) capacity building. Promote collaboration between regulators and financial institutions in data sharing, model validation, and talent development to elevate risk management capabilities across the financial system.

5.3. Research limitations and prospects

This study has the following limitations: First, the sample only covers listed banks and excludes non-listed small and medium-sized banks as well as non-bank financial institutions, suggesting future research could expand to the entire financial system. Second, the model primarily relies on historical data training, and its predictive capacity for “black swan” events still requires integration with scenario analysis and stress testing. Third, while interpretability methods have been improved, the decision-making logic of deep learning models still needs deeper economic theoretical support. Future research directions include: (1) Exploring the application of graph neural networks (GNNs) in modeling bank network risk contagion; (2) Combining high-frequency data with hybrid models to enhance real-time risk early warning capabilities; (3) Developing interpretable deep learning frameworks incorporating economic theoretical constraints to achieve organic integration of “theory-driven” and “data-driven” approaches.

Acknowledgments

This research was funded Research on the Evolutionary Mechanism and Prevention of Systemic Financial Risks Caused by Digital Inclusive Finance, Project supported by the National Social Science Fund of China (22BJL041).

Authors' contribution

Conceptualization, Junming Xie and Jianlin Yin; methodology, Ailin Liu; software, Jianlin Yin; validation, Ailin Liu, Junming Xie and Jianlin Yin; formal analysis, Jianlin Yin; investigation, Junming Xie; resources, Junming Xie; data curation, Junming Xie; writing—original draft preparation, Jianlin Yin; writing—review and editing, Jianlin Yin and Ailin Liu; visualization, Jianlin Yin; supervision, Junming Xie; project administration, Junming Xie; funding acquisition, Junming Xie. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflicts of interest.

References

- [1] Adrian T, Brunnermeier M. CoVaR. *Am. Econ. Rev.* 2016, 106(7): 1705–1741.
- [2] Liang Q, Li Z, Hao X. Identification and supervision of systemically important financial institutions in China: An analysis based on the SRISK systemic risk index (In Chinese). *J. Financ. Res.* 2013, 9: 56–70.
- [3] Han G, Cai J, Li X, Yang F. Measurement of systemic risk of Chinese financial institutions based on SRISK method (In Chinese). *J. Jingchu Univ. Technol.* 2025, 40(4) : 55–64.
- [4] Ye Z, Jin Y. Research on risk-dependence structure of global sovereign bond market based on vine-Copula model (In Chinese). *Studies of International Finance.* 2022, 4: 65–74.
- [5] Wu W, Jiang H, Tang S. Digital transformation, network connectedness and bank systemic risk (In Chinese). *Chin. J. Manag Sci.* 2024, 32(3): 9–19.
- [6] Li H, Zhou L. Research on systemic financial risk monitoring and early warning based on machine-learning model (In Chinese). *Operations Research and Management Science.* 2023, 32(11): 212–219.
- [7] Yang Z, Li D, Chen Y. Can risk preference serve as a forward-looking indicator of China's financial risk? New evidence from cutting-edge machine-learning methods (In Chinese). *Manage. World.* 2025, 41(10) : 21–43.
- [8] Bai H, Liu S, Luo X, Liu L, Hao W. Research on measurement and early warning of China's systemic financial risk based on real-estate market (In Chinese). *J. Financ. Res.* 2020, 8: 54–73.
- [9] Tang Z, Wu J, Zhang T, Chen K. Research on early warning of systemic risk in China's insurance industry based on EEMD-LSTM (In Chinese). *Insurance Studies.* 2022, 34(9): 27–34.
- [10] Shen S, Chen M, Feng S, Xu Q, Liu J, *et al.* ChpoBERT: A pre-trained language model for Chinese policy texts (In Chinese). *J. China Soc. Sci. Tech. Inf.* 2023, 42(12), 1487–1497.

-
- [11] Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput.* 1997, 9(8): 1735–1780.
 - [12] Cho K, Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, *et al.* Learning Phrase Representations using RNN Encoder–decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.
 - [13] Diebold F, Mariano R. Comparing predictive accuracy. *J. Bus. Econ. Stat.* 2002, 20(1): 134–144.