

Research on Financial Risk Intelligent Monitoring and Early Warning Model Based on LSTM, Transformer, and Deep Learning

Yunan Song, Business School, Zhejiang Wanli University, China

Huaqing Du, International Business School, Hebei International Studies University, China

Tianyu Piao, School of Economics, Changchun University, China

Hongyu Shi, Business School, Henan Normal University, China*

ABSTRACT

As global financial markets continue to evolve and change, financial risk monitoring and early warning have become increasingly important. However, the complexity and diversity of financial markets have led to the emergence of multidimensional and multimodal data. Traditional risk monitoring methods face difficulties in handling such diverse data and adapting to the monitoring and early warning needs of emerging risk types. To address these issues, this article proposes a financial risk intelligent monitoring and early warning model that integrates deep learning to better cope with uncertainty and risk in the financial market. Firstly, the authors introduce an LSTM model in the initial approach, trained on historical financial market data, to capture long-term dependencies and trends in the data, enabling effective monitoring of financial risk. They also optimize the model architecture to improve its performance and prediction accuracy. Secondly, the authors further introduce a transformer model with self-attention mechanism to better handle sequential data.

KEYWORDS

Deep Learning, Early Warning, Financial Risk, LSTM, Monitoring, Transformer

1. INTRODUCTION

As an integral component of a country's economy, the financial market not only reflects the nation's competitiveness but also carries significant responsibilities in the context of the country's socioeconomic mission. With the rapid development of the socioeconomic landscape, the complexity and diversity of financial markets have been on the rise, leading to the accumulation of vast volumes of financial data. This has also raised higher demands for financial information, making the efficient extraction, analysis, and prediction of financial data a pressing challenge in both academia and industry. Therefore, research into intelligent monitoring and early warning models for financial risks holds substantial practical value. The financial sector generates a plethora of structured and unstructured

DOI: 10.4018/JOEUC.337607

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

data, including market trading data, news reports, economic indicators, company financial reports, among other information sources. These data not only come in massive quantities but also typically exhibit highly dynamic and diverse characteristics, reflecting the intricacies and uncertainties of financial markets. Traditional time series analysis methods find widespread application in the field of finance, including autoregressive models (AR) (Kaur, Parmar & Singh, 2023), moving average models (MA) (Xu et al., 2023), autoregressive Moving Average Models (ARMA) (Rapoo, Chanza & Motlhwe, 2023), and autoregressive integrated moving average models (ARIMA) (Wang et al., 2023a). Autoregressive (AR) models are advantageous for their simplicity, intuitiveness, and ease of understanding and implementation. They effectively capture the local patterns and trends in data, offering flexibility by adjusting the order to control model complexity. However, AR models, based on the assumption of linear relationships, may struggle to capture nonlinear dynamics and complex relationships. They are sensitive to initial values, require data stationarity, and may have limited effectiveness when dealing with non-stationary or complex data. Moving Average (MA) models, on the other hand, excel at adapting to short-term fluctuations in data. By considering the moving average of past observations, they reduce the impact of noise and random fluctuations, resulting in a smoother and more stable model. MA models are particularly effective in handling seasonal and periodic time series data. However, they have limitations in modeling trends and long-term dependencies, as they primarily focus on short-term average effects and may not fully capture long-term trends in time series. Additionally, MA models may perform poorly with long-term memory in noise, requiring a careful balance and selection based on the data's characteristics in practical applications. Autoregressive Moving Average (ARMA) models combine the strengths of both AR and MA components. They capture long-term dependencies and trends (via the AR part) while effectively handling short-term fluctuations and noise (via the MA part). ARMA model parameter estimation is relatively intuitive, exhibiting strong adaptability to time series data of different natures. However, ARMA models have limited capabilities in modeling nonlinearity and non-stationarity, requiring prior assurance of data stationarity. Additionally, parameter selection for the model may demand empirical and domain knowledge. Careful consideration is necessary when balancing model complexity and fitting performance, especially when dealing with high-order models to avoid overfitting. Autoregressive Integrated Moving Average (ARIMA) models are widely used, decomposing time series data into trend, seasonality, and residual components for predicting future trends. Although traditional methods perform well in certain situations, they often rely on strong domain knowledge and manual feature engineering. Their ability to handle nonlinear and non-stationary data is limited, and they typically depend on statistical models and rule-based systems, posing constraints when dealing with large-scale, multimodal, and high-dimensional data.

With the rapid advancement of deep learning technologies, particularly the emergence of models such as Long Short-Term Memory (LSTM) (Alizamir et al., 2023) and Transformer (Korthikanti et al., 2023), we have the opportunity to leverage these advanced methods to gain a better understanding of financial markets, capture non-linear relationships, handle multidimensional data, and predict potential risk events (Cheng, van Dongen, & van der Aalst, 2019). This paper aims to delve into deep learning-based models for intelligent monitoring and early warning of financial risks. We begin by introducing the LSTM model and applying it to historical data from financial markets (Gupta et al., 2022). Through training on financial time series data, this model can capture long-term dependencies and trends within the data. This is crucial for effective monitoring of financial risks since risk events in financial markets often exhibit temporal correlations. Subsequently, we optimize the architecture of the LSTM model to enhance its performance and prediction accuracy. Furthermore, we introduce the Transformer model to complement the LSTM model, further improving prediction accuracy. The Transformer model, with its self-attention mechanism (Hong, Zhang & Xu, 2023), possesses exceptional capabilities in modeling multidimensional time series data and adaptive feature extraction. It efficiently handles multidimensional financial data, automatically captures complex market trends, and long-term dependencies, thus enabling more accurate risk prediction and real-time monitoring (Liu,

Zeng, Cheng, Duan, & Cheng, 2021). Moreover, the Transformer model can also be utilized for anomaly detection and portfolio optimization, aiding in the identification of exceptional events and optimizing investment portfolios (Al Janabi, 2022). Compared to traditional statistical and linear models, the Transformer model excels at capturing non-linear relationships and complex market dynamics within the data, thereby enhancing the accuracy of financial risk prediction.

This study integrates deep learning methods such as LSTM and Transformer to harness their unique strengths in the field of financial risk monitoring. We will delve into how to construct a more comprehensive and efficient financial risk monitoring and early warning system by combining the outputs of these deep learning models. This fusion approach will leverage LSTM's ability to model long-term dependencies in time series data and Transformer's self-attention mechanism for capturing global correlations within sequences to better address the complexity and uncertainty of financial markets. We will also consider other factors such as market sentiment indicators, macroeconomic data, and interrelationships among assets to enrich the input information of the model and enhance its sensitivity to financial risk (Mba & Mai, 2022). To validate the performance of the proposed financial risk monitoring and early warning model, we will extensively utilize large-scale financial time series datasets for experiments. The experimental design will encompass multiple steps, including data preprocessing, model training, and performance evaluation, to ensure the model's effectiveness in real financial environments. We will employ various evaluation metrics such as accuracy, recall, F1 score, among others (Megdad, Abu-Naser & Abu-Nasser, 2022), to comprehensively assess the model and conduct in-depth comparisons with various methods to demonstrate its superiority and practicality. Through this research, we aim to provide a more comprehensive and flexible approach to financial risk monitoring and early warning, better equipped to handle market fluctuations and risks.

The contributions of this paper can be summarized in the following three aspects:

- (1) Our research introduces Long Short-Term Memory (LSTM) networks into the field of financial risk monitoring. This initiative not only adds innovation to our model but also expands the boundaries of traditional financial risk monitoring frameworks. Through the incorporation of LSTM, we are better equipped to capture long-term dependencies in time series data from financial markets. This aids in more accurately predicting various potential risk events, including market volatility and credit risks, thereby enhancing risk management in financial markets and safeguarding financial stability.
- (2) We introduce the Transformer model, emphasizing its sensitivity to non-linear relationships and complex dynamics in financial markets. Traditional statistical and linear models have limitations in capturing the non-linear characteristics of financial markets. The Transformer model, with its self-attention mechanism, is better at capturing non-linear relationships in the data, thereby increasing sensitivity to the diversity of financial market data. This enables financial risk monitoring models to more comprehensively understand the dynamic changes in financial markets.
- (3) We introduce deep learning techniques, combining traditional statistical and machine learning methods with deep learning. This integrated approach not only captures non-linear relationships in financial markets more effectively but also handles multidimensional data and long-term dependencies in time series. As a result, it improves the accuracy of financial risk prediction and enhances the sensitivity of monitoring models to complex market dynamics and uncertainty. This approach brings a new paradigm to financial market risk management, providing financial institutions with more reliable and comprehensive tools to address risks.

The organizational structure of this paper is as follows: Firstly, in the introduction, we emphasize the crucial role of financial markets in the national economy. Traditional methods such as AR, MA, ARMA, and ARIMA are noted for their limitations in handling non-linear and non-stationary data. By integrating deep learning models like LSTM and Transformer, we elucidate the paper's objective—to

enhance the efficiency of financial risk monitoring through deep learning models. In the literature review section, we review traditional financial risk monitoring methods including VaR, cointegration models, and rule-based systems, as well as the application of deep learning technologies such as CNN, RNN, GRU, and GAN. Given the challenges they face in dealing with dynamic markets and multi-source data, we emphasize the key models employed—LSTM and Transformer—to improve comprehensiveness and accuracy. Subsequently, in the methodology section, we provide a detailed overview of the comprehensive model based on LSTM, Transformer, and deep learning. This includes model architecture, data preprocessing strategies, and training methods. Additionally, we introduce considerations for model design, taking into account factors such as data quality, interpretability, and computational resources. In the experimental section, we construct a financial risk monitoring and warning model using high-performance computing servers and the Python programming language, incorporating multiple datasets such as Fama-French, CRSP, Compustat, and World Bank. Through a synthesis of metrics such as accuracy, precision, recall, and F1 score, we demonstrate the outstanding performance of the model, particularly when combining LSTM and Transformer, leading to significant improvements in accuracy and recall. Comparative analysis also indicates the competitiveness of our model in terms of parameters, inference time, and training time, showcasing efficiency and scalability. Furthermore, the discussion section delves into a thorough analysis of experimental results, highlighting the model's advantages in risk monitoring while addressing limitations and challenges such as interpretability, data quality, and resource consumption. External factors, such as macroeconomic changes and policy adjustments affecting risk, are also considered to enhance the study's comprehensiveness. Finally, the conclusion section summarizes the main contributions of the research, emphasizing the potential value of the comprehensive model in financial risk monitoring. Future research directions are outlined, including improving model efficiency, practical application in financial markets, and further investigation into interpretability. This organizational structure aims to provide readers with a comprehensive understanding of our research, delving into the issues, solutions, and future prospects while considering the multifaceted factors in the field of financial risk monitoring.

2. RELEVANT WORK

Financial risk monitoring and early warning have always been core tasks in the field of finance, and an excellent model can provide crucial decision support for financial institutions and investors. Traditional models for financial risk monitoring and early warning have made significant progress over the past few decades, primarily focused on the development of statistical and econometric models, as well as rule-based methods. For instance, (Behera et al., 2023) introduced the Value at Risk (VaR) model, a risk measurement method based on statistical approaches aimed at estimating potential losses in a portfolio or asset. It typically uses historical data and probability distributions to calculate the maximum possible loss at a certain confidence level. By leveraging historical data and probability distributions, the model is capable of quantifying various risk levels, providing investors with decision-making references. However, due to its sensitivity to market assumptions, this may result in inaccurate estimations of real market conditions and relative difficulties in handling extreme events and dynamic market changes. (Gianfreda et al., 2023) proposed a cointegration model based on time series analysis methods, used to study long-term relationships among multiple related variables. Its core concept is that although these variables may be non-stationary, there exists a stationary linear combination among them, known as cointegration, indicating their long-term association. This provides researchers with a more in-depth understanding of the dynamic relationships between variables, particularly with significant applications in the fields of economics and finance. However, the model's identification of cointegration relationships requires thorough testing of the data and is susceptible to factors such as sample period and data quality in empirical studies. Therefore, caution should be exercised in its application, taking these limitations into account to ensure the reliability of the results. Traditional financial institutions often use rule-based systems, as described in (Hassan et al.,

2023), to monitor potential risk signals. These systems rely on manually defined rules and thresholds to detect situations such as abnormal transactions, credit defaults, or market volatility. However, these systems often struggle to cope with complex market conditions and emerging risks. Traditional time series analysis methods, such as the ARIMA (AutoRegressive Integrated Moving Average) model introduced in (Mgammal, Al-Matari & Alruwaili, 2023), aim to capture trends, seasonality, and randomness in time series data. It combines an AutoRegressive (AR) model to describe the correlation between the current value and past values and a Moving Average (MA) model to describe the correlation between the current value and white noise errors. It also includes a differencing operation to transform non-stationary time series into stationary ones for better application of the AR and MA models. This model finds widespread use in modeling and forecasting time series data in fields such as economics and finance. However, this model has relatively high data requirements, necessitating a certain level of stationarity and recognizability of trends. Additionally, its treatment of seasonality is relatively simplified. In practical applications, researchers should be mindful of the characteristics of the data to ensure the accuracy and effectiveness of the model. (Chen, Huang & Liang, 2023) discusses GARCH (Generalized Autoregressive Conditional Heteroskedasticity), a statistical model used for modeling volatility in time series data, particularly suitable for the financial domain. This model allows volatility to change over time and predicts future volatility based on past observations. This makes it crucial in risk management and asset pricing within the financial domain, particularly in modeling and forecasting financial market volatility. However, the limitation of GARCH models lies in their assumption that the conditional heteroscedasticity of volatility is stationary, potentially overlooking some nonlinear features. Predictive performance may be relatively limited in scenarios involving extreme events and fat-tail distributions. Therefore, it is essential to carefully consider the model's assumptions and applicability in practical applications. Although traditional methods have played a crucial role in financial risk management, they often face limitations in adapting to non-linear relationships and handling large-scale data, as well as challenges in real-time monitoring and early warning. With the emergence of deep learning technologies, we have the opportunity to explore new approaches, leveraging the capabilities of neural networks to capture complex relationships in data and enhance the accuracy and efficiency of risk prediction.

The application of deep learning techniques in the field of financial risk monitoring has garnered widespread attention and research. These models leverage deep neural network architectures to automatically extract crucial features from large-scale financial data, enabling accurate monitoring and early warning of potential risks. For example, (Mousapour Mamoudan et al., 2023) introduces a financial risk monitoring model based on Convolutional Neural Networks (CNNs). This model autonomously learns and extracts complex features from financial market data, such as stock price trends and heatmaps, by employing multi-level feature extraction and pooling operations in the convolutional layers. It can capture local patterns and trends in time series data, making it highly effective for monitoring short-term volatility. (Ashtiani & Raahmei, 2023) presents a Recurrent Neural Network (RNN) model designed for handling time series data such as stock prices, exchange rates, and interest rates. This model harnesses the sequential modeling capability of RNNs to better capture temporal and complex dynamics in financial market data, making it perform well in long-term risk monitoring and prediction. However, traditional RNN models suffer from the vanishing gradient problem. Therefore, more advanced variants like Gated Recurrent Units (GRUs) address this issue effectively, offering fewer parameters and faster training speeds. (Hu, Chang & Yan, 2023) introduces an innovative financial risk monitoring model that utilizes GRUs as its core structure to model historical market data, capturing dynamic market features. With its built-in gating mechanisms, it efficiently handles data at different time scales and captures both long-term and short-term dependencies, enabling the model to better identify and predict potential risk signals. Additionally, this model offers faster training speeds, making it suitable for high-frequency trading and real-time decision-making scenarios. (Vuletić, Prenzel & Cucuringu, 2023) proposes a model based on Generative Adversarial Networks (GANs) for synthesizing financial data and improving data quality

to enhance the performance of monitoring models. By generating realistic synthetic data, GANs help expand the training dataset, mitigate overfitting issues, and improve model generalization. This model can monitor new data in real-time and provide alerts upon detecting abnormal behavior, facilitating timely risk mitigation. (Fuchs & Horvath, 2023) introduces a financial risk monitoring model based on Wasserstein Generative Adversarial Networks (WGANs). This model uses Wasserstein distance to enhance data quality and training stability, employing it to measure the distance between generated data and real data for anomaly detection, thereby improving the accuracy of anomaly detection. These applications of deep learning in financial risk monitoring demonstrate the potential of neural networks in capturing complex relationships and enhancing the accuracy and efficiency of risk prediction.

When it comes to financial risk monitoring, despite the significant achievements of deep learning methods, they still face some challenges and issues. Firstly, the dynamism and complexity of financial markets may impose limitations on traditional deep learning models in capturing long-term dependencies and nonlinear features in time-series data. This introduces our first model: the LSTM model, which excels at memorizing time-series data and can better handle the temporal and dynamic nature of the market. By introducing this model, we can more accurately model historical financial market data and predict future risk trends. On the other hand, financial markets involve a large amount of multisource data, including market trading data, news events, social media sentiment, and more. Traditional deep learning models may face challenges in integrating and jointly modeling multisource data. Therefore, we introduce the second model: the Transformer model, renowned for its self-attention mechanism's ability to effectively handle correlation learning and feature extraction between different data sources. By introducing this model, we can more comprehensively leverage information from multisource data, enhancing the comprehensiveness and accuracy of risk monitoring. Most importantly, while the introduction of deep learning methods can improve the efficiency and automation of risk monitoring, it still requires a large amount of training data and computational resources. Additionally, robustness and generalization need to be considered to ensure reliability in different market environments. The application of these models not only contributes to enhancing the stability and efficiency of financial markets but also provides better tools for investors, financial institutions, and regulatory authorities to manage financial risks. However, given the complexity and risks of financial markets, the research and application of these models still require continuous improvement and validation. We will continue to explore how to better integrate the strengths of different deep learning models, enhance model robustness and practicality, to better address the challenges and changes in financial markets.

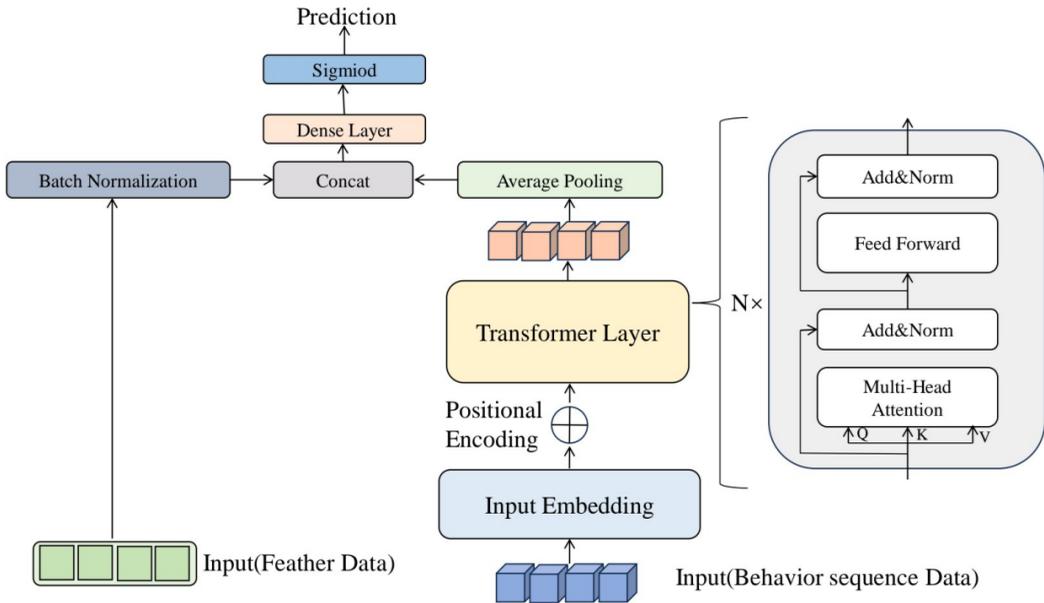
3. METHOD

The overall flowchart of the algorithm in this article is shown in Figure 1.

3.1 LSTM Architecture

Long Short-Term Memory (LSTM) is a variant of Recurrent Neural Network (RNN) (Lazcano, Herrera & Monge, 2023) widely employed for handling sequential data, particularly in domains such as natural language processing, time series analysis, and financial forecasting. This architecture, distinguished by its unique structure, excels at effectively capturing long-term dependencies and retaining information within sequences, rendering it a potent tool for processing time series data. Its role in intelligent financial risk monitoring and early warning is also significant. To begin with, one of the core advantages of this network is its outstanding ability for time series modeling. Financial market data often exhibits pronounced temporal correlations and long-term dependencies, as seen in stock prices, exchange rates, and interest rates (Adebayo, Akadiri & Rjoub, 2022). LSTM, equipped with its internal gate units and memory cells, efficiently captures these long-term dependencies, resulting in more precise predictions of future market trends and risk events. Secondly, this architecture is adept at addressing the nonlinear characteristics prevalent in financial markets. Financial market behavior

Figure 1. Overall algorithm flowchart



is influenced by various factors, and these factors often exhibit intricate nonlinear relationships that traditional linear models struggle to encapsulate. Leveraging its multi-layer neural network structure and activation functions, LSTM can adeptly model the nonlinear features inherent in financial data, thereby enhancing the comprehension and prediction of market behavior (Ali et al., 2023). Furthermore, LSTM excels in handling sequence data, which is ubiquitous in financial datasets, encompassing historical prices, trading volumes, financial indicators, and more. It efficiently captures the correlations between different time points, facilitating a more comprehensive analysis of market conditions. It can also be employed to establish sequence-to-sequence models, such as mapping a series of historical data to future predictions, a technique with extensive applications in financial risk early warning systems.

This architecture consists of multiple units, with each unit containing three gate units: the Forget Gate, the Input Gate, and the Output Gate. These gate units regulate the flow of information through a Sigmoid function and a dot product operation (Liu et al., 2023). The model diagram of the LSTM is shown in Figure 2.

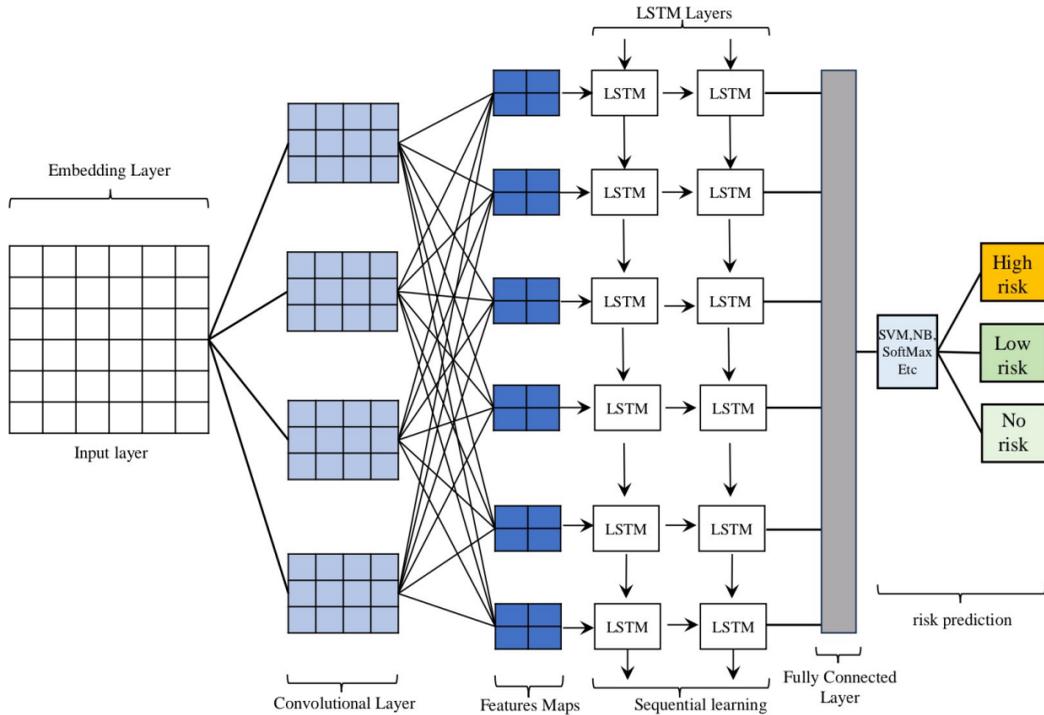
The Forget Gate determines which information to discard from the cell state. Its output ranges from 0 to 1, where 0 means complete forgetting, and 1 means complete retention. The output of the Forget Gate can be represented by the following formula:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

where f_t is the output of the Forget Gate, W_f and b_f are the weights and bias of the Forget Gate, h_{t-1} is the previous time step's hidden state, and x_t is the input at the current time step.

The Input Gate determines how much information to add from the new input to the cell state. It also uses a sigmoid function to decide which values to update. The Input Gate can be represented by the following formula:

Figure 2. Model diagram of the LSTM



$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

where i_t is the output of the Input Gate, \tilde{C}_t is the new candidate cell state, W_i, W_c, b_i , and b_c are the respective weights and biases.

The cell state is the core of LSTM and is updated through the Forget Gate and Input Gate. The cell state can be represented using the following formula:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

Among them, C_t is the new cell state, and C_{t-1} is the cell state of the previous time step.

The output gate determines how much information is output from the cell state. It can be expressed by the following formula:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

Among them, h_t is the hidden state of the current time step, o_t is the output of the output gate, W_o and b_o are the weights and biases of the output gate.

3.2 Transformer Architecture

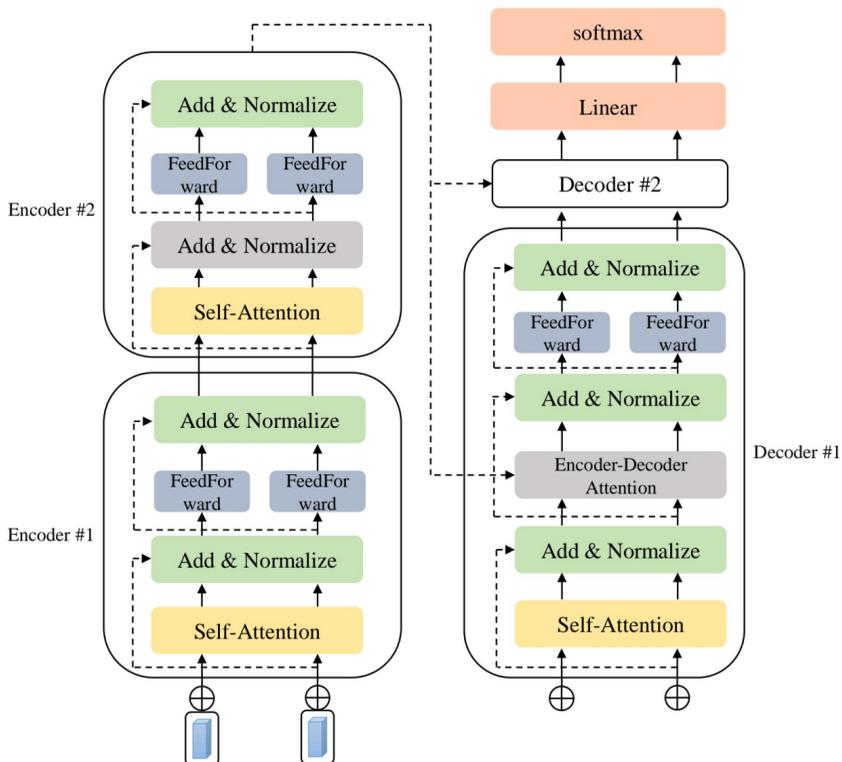
The Transformer architecture is a deep learning model used for sequence-to-sequence tasks (Chen et al., 2023). Its design goal is to overcome the computational bottleneck issues present in recurrent neural networks (RNNs) and handle longer sequence data. This architecture introduces self-attention mechanisms, allowing the model to compute representations for all positions in the sequence in parallel, without the need for sequential processing like RNNs. Therefore, it is more suitable for tasks such as financial risk monitoring and forecasting. The model diagram of the Transformer is shown in Figure 3.

One of the core ideas of the Transformer is the self-attention mechanism, which allows the model to learn dependencies between each position in a sequence and other positions. The self-attention mechanism computes a weighted sum of representations, where each position is weighted based on its relationship with other positions. The self-attention mechanism can be represented using the following formula:

$$Attention(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q, K, and V are representations of Query, Key, and Value, and d_k is the dimension of the query and key.

Figure 3. Model diagram of the transformer



To enhance the model's representational capacity, the architecture introduces multi-head self-attention mechanism (Wang et al., 2023b). This is a core component of the architecture that allows the model to simultaneously focus on information from different positions in the input sequence to capture various types of relationships and dependencies. The introduction of this mechanism enables the model to better handle sequence data, including natural language text, time series, and financial data, among others. The multi-head attention mechanism can be represented by the following formula:

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \dots, head_n)W^O$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

where W_i^Q , W_i^K , and W_i^V are the weight matrices for each head.

Positional encoding is typically represented as a matrix with the same dimensions as the input data. Its values vary based on position to provide a unique encoding for each position. Positional encoding can be expressed using the following formula:

$$PE(pos, 2i) = \sin(pos / 10000^{2i/d_{model}})$$

$$PE(pos, 2i + 1) = \cos(pos / 10000^{2i/d_{model}})$$

where “pos” is the position, “i” is the dimension, and “ d_{model} ” is the model's dimension.

The architecture consists of encoders and decoders, where encoders handle input sequences, and decoders generate output sequences. Each encoder and decoder layer includes multi-head self-attention and feedforward neural network layers (Dai, 2022). This can be represented using the following formula:

$$E(x) = M(S(x) + x) + F(x)$$

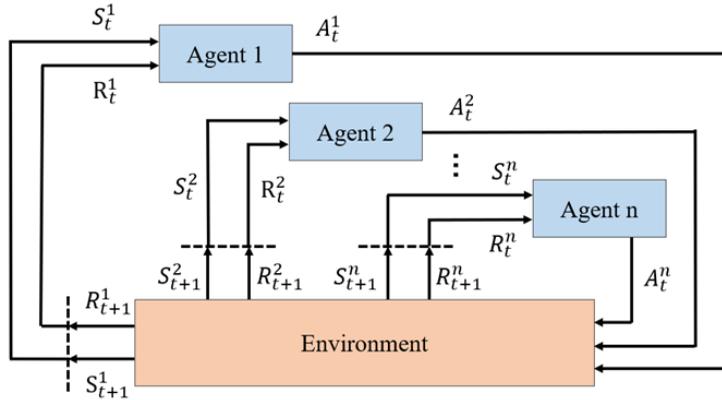
$$D(y, enc_output) = M(S(y) + y) + M(ED(y, enc_output)) + y + F(y)$$

Here, x represents the input to the encoder. In this process, the encoder first captures dependencies within the input sequence using the Multi Head Self-Attention mechanism. The result is then added to the input x using residual connections, and finally processed through a Feed Forward neural network layer. Y represents the input to the decoder, and enc_output is the output from the encoder. In the decoder, a similar process occurs. It starts with capturing self-dependencies within the input sequence using Multi Head Self-Attention, followed by capturing relationships between the input sequence and the decoder input using the Multi Head Encoder-Decoder Attention mechanism. Finally, the data is processed through a Feed Forward neural network layer.

3.3 Deep Learning Model

Deep learning models, with multi-layer neural networks at their core, have the capability to automatically extract high-level features and representations from data. In the financial domain, deep learning models find extensive applications in risk monitoring and warning tasks, including credit risk assessment, market risk analysis, and fraud detection, among others (Wang & Han, 2021). These models typically consist of multiple neural network layers, including input layers, hidden layers, and output layers. Each hidden layer contains multiple neurons, which communicate information through weighted connections. The model diagram of the deep learning is shown in Figure 4.

Figure 4. The model diagram of the deep learning



The training process of a deep learning model involves multiple steps, including data preparation, selection of loss functions, and the use of optimization algorithms. In deep learning, proper data preparation is a crucial first step. This involves steps such as data cleaning, standardization, and splitting into training, validation, and test sets. Ensuring the quality and diversity of data is crucial for the performance of deep learning models. The loss function is used to measure the difference between the model's predictions and the actual values. During training, the goal is to minimize the loss function. Common loss functions in the financial domain include Mean Squared Error and Cross-Entropy, with the specific choice depending on the nature of the task. In the forward propagation stage, activation functions perform non-linear transformations on the output of neurons. Common activation functions include ReLU, Sigmoid, and Tanh. Choosing an appropriate activation function helps the model better learn complex data representations. Optimization algorithms are used to adjust the model's weights to minimize the loss function. Common optimization algorithms include Stochastic Gradient Descent (SGD), Adam, and RMSprop. Selecting the right optimization algorithm is crucial for the convergence speed and performance of the model. During the training process, forward propagation calculates the model's output, and then backpropagation computes the gradient of the loss function. Finally, the model's weights are updated using gradient information. This process iterates multiple times until the model converges to a satisfactory state.

Forward propagation is the process in deep learning models used to compute the output by passing input data through the layers of the network, ultimately producing the model's predictions or outputs. It can be represented using the following formula:

$$z_i = \sum_{j=1}^n (w_{ij} \cdot x_j + b_i)$$

$$a_i = f(z_i)$$

In this equation, z_i represents the weighted input of neuron i , x_j is the input data, w_{ij} is the weight, b_i is the bias, a_i is the activation function output of neuron i , and $f()$ is the activation function.

Backpropagation is the process in deep learning models used to update weights in order to minimize the loss function. It calculates gradients and updates weights in the direction of the gradient. It can be represented using the following formula:

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial a_i} \cdot \frac{\partial a_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_{ij}}$$

where L is the loss function, and $\frac{\partial L}{\partial w_{ij}}$ represents the gradient of weight w_{ij} .

We incorporate LSTM and Transformer architectures into the model for handling financial risk monitoring tasks through deep learning. LSTM, a variant of Recurrent Neural Network (RNN), is specifically designed for processing and learning sequential data. Its core concept introduces gated units, including forget gates, input gates, and output gates, to better capture long-term dependencies in time series data. The input gate determines the information to be updated, the forget gate decides the information to be discarded, and the output gate determines the information to be outputted. Through these gating mechanisms, LSTM effectively addresses issues such as gradient vanishing and exploding, enabling it to capture long-term dependencies in time series more effectively. In financial risk monitoring, LSTM finds extensive applications in modeling time series data, such as stock prices, exchange rates, and interest rates. Its characteristics of long-term memory enable better capturing of the complex dynamics in financial markets, aiding in improving the accuracy of future risk predictions. Transformer is an architecture based on the self-attention mechanism initially used for natural language processing tasks. Its core idea is to establish connections between different positions through self-attention, allowing the model to simultaneously consider all positions in the input sequence. The self-attention mechanism enables the model to dynamically focus on different parts of the input sequence without being constrained by a fixed window size. Transformers consist of encoders for extracting features from the input sequence and decoders for generating output sequences. In financial risk monitoring, the nonlinear relationship modeling capability of Transformer makes it suitable for handling complex, nonlinear relationships in financial markets. It can more flexibly capture features in market data, including correlations and nonlinear dynamics between different assets, providing a more comprehensive perspective for risk monitoring.

The combined deep learning models incorporating LSTM and Transformer architectures not only demonstrate improved predictive performance in the field of intelligent financial risk monitoring and early warning but also play a crucial role in the following aspects. Firstly, these models are better equipped to handle the complexity of the financial market, including factors such as market volatility, asset price changes, and investor sentiment, thereby enhancing their sensitivity to market risks and aiding in the early detection of potential issues. Secondly, deep learning models exhibit superior generalization capabilities, enabling them to adapt to different types of financial data and market conditions, thus extending their applicability across various financial domains, including stock markets, bond markets, and forex markets, among others (Wang, Zhao & Qiu, 2022). Additionally, these models provide financial practitioners with enhanced decision support, assisting them in risk management, portfolio optimization, and strategic planning. Most importantly, deep learning models offer a novel approach to financial risk monitoring, better equipped to address the uncertainty and complexity of financial markets, thereby providing new insights and opportunities for future research and practice in the financial field.

The pseudocode of the algorithm in this paper is shown in Algorithm 1.

4. EXPERIMENT

The experimental flow chart of this paper is shown in Figure 5.

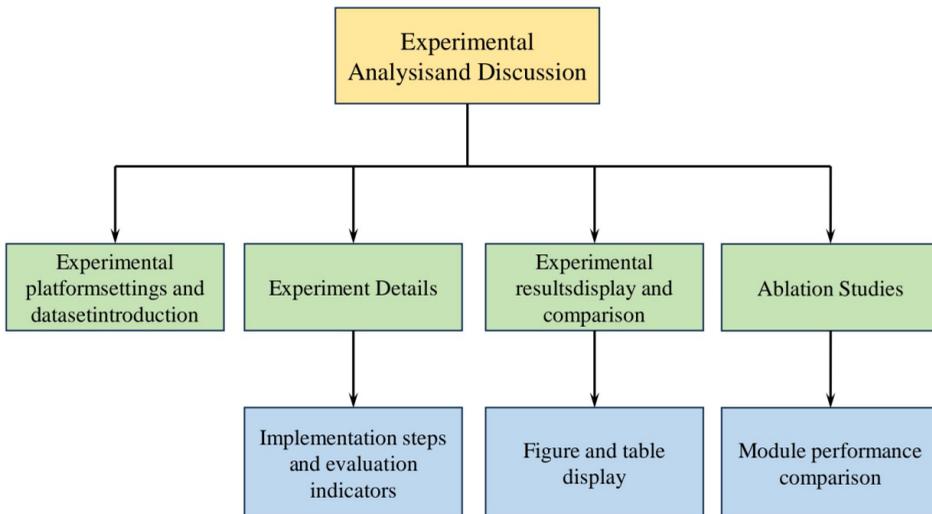
Algorithm 1. LSTM: Transformer training

```

Input : Financial datasets (Fama-French Three-Factor, CRSP, Compustat, World Bank)
Output : Trained LSTM-Transformer model with evaluation metrics

Initialize LSTM and Transformer model parameters;
Load financial datasets;
Preprocess datasets: Normalize, feature engineering, split into training and testing sets;
Define evaluation metrics: Accuracy, Precision, Recall, F1-Score;
while Not converged or max epochs not reached do
    foreach mini-batch in training set do
        Sample mini-batch data;
        Compute LSTM hidden states for sequential data;
        Compute Transformer embeddings for non-sequential data;
        Concatenate LSTM and Transformer outputs;
        Pass concatenated data through fully connected layers;
        Compute loss based on predictions and ground truth labels;
        Perform backpropagation to update model weights;
    end
    Calculate evaluation metrics on validation set;
    Check for early stopping based on validation metrics;
end
Evaluate the trained model on the testing set;
Calculate final evaluation metrics: Accuracy, Precision, Recall, F1-Score;
return Trained LSTM-Transformer model with evaluation metrics;
    
```

Figure 5. Experiment flow chart



4.1 Experimental Environment

4.1.1 Hardware Environment

This experiment utilized a high-performance computing server, which provided excellent computational and storage capabilities to support research on financial risk monitoring and warning models. The server was equipped with an Intel Xeon E5-2690 v4 @ 2.60GHz CPU, a high-performance multi-core processor that offered powerful computational capabilities suitable for deep learning tasks. With 512GB of RAM, it ensured ample memory resources for model training and data processing,

contributing to improved experimental efficiency. The server was equipped with eight Nvidia Tesla P100 16GB GPUs, which excelled in deep learning tasks, significantly accelerating model training and inference processes. These GPUs provided researchers with robust data processing capabilities, allowing the model to converge faster and make more accurate predictions in the realm of financial risk.

4.1.2 Software Environment

In this research, we chose Python as the primary programming language and PyTorch as the deep learning framework to explore efficient approaches to financial risk monitoring and warning models. Leveraging the powerful capabilities of deep learning, our aim was to enhance the performance and efficiency of financial risk monitoring and warning tasks. Taking full advantage of Python's convenience and flexibility, we swiftly constructed intelligent risk control models based on deep learning. PyTorch, as our preferred deep learning framework, provided us with a rich set of tools and algorithm libraries, greatly simplifying the process of model development and training. By utilizing PyTorch's dynamic computation graph mechanism and built-in automatic differentiation functionality, we were able to more easily build, optimize, and fine-tune models to achieve more precise financial risk monitoring and warning results.

4.2 Experimental Data

4.2.1 Fama-French Three-Factor Dataset

This dataset is an essential resource for researching financial markets and asset pricing. The dataset is named after economists Eugene F. Fama and Kenneth R. French, who proposed the famous three-factor model to explain the volatility of stock returns. The dataset originates from historical data of the U.S. stock market. It contains a vast amount of financial market indicators and stock data, spanning multiple years, and even decades. The dataset includes monthly, quarterly, or yearly data, typically encompassing stock returns, market capitalization, price-to-book ratios, and more. These data allow researchers to analyze the performance of stocks and portfolios under different time periods and market conditions. This data can be used to validate risk models, explore the behavior of financial markets, assess the risk and return of portfolios, and develop intelligent risk monitoring and warning systems.

4.2.2 CRSP Dataset

This dataset is maintained and provided by the Center for Research in Security Prices (CRSP) at the University of Chicago. It originates from the U.S. stock market, covering multiple exchanges, including the New York Stock Exchange (NYSE), NASDAQ, and various types of financial assets. The dataset contains multidimensional financial market data, including stock opening prices, closing prices, high prices, low prices, and other price-related information. It also records daily trading volumes, dividend payments, and stock splits, which help analyze trading activity and liquidity and make timely adjustments to stock prices and returns. Additionally, it includes historical data for various market indices, such as the S&P 500 index, which is used to study overall market performance. This dataset finds wide applications in fields such as finance, asset pricing, portfolio management, and market behavior research. It can be used for tasks such as stock price analysis, portfolio construction, and the development of risk models.

4.2.3 Compustat Dataset

The Compustat dataset is a global financial and accounting data resource provided by S&P Global Market Intelligence. It includes rich financial and accounting information for both public and private companies from different countries and industries. This dataset comprises financial statements, accounting metrics, company information, and stock market data, covering various aspects of a company's financial condition, operational performance, and market value. The Compustat dataset finds extensive applications in fields such as finance, corporate analysis, investment decision-

making, supporting researchers and investment professionals in tasks such as company valuation, risk assessment, and portfolio construction.

4.2.4 World Bank Dataset

The World Bank Dataset is a comprehensive resource managed and maintained by the World Bank, which extensively collects and provides macroeconomic, social, environmental, and development data from countries and regions worldwide. This dataset includes multidimensional information such as a country's Gross Domestic Product (GDP), population statistics, education levels, poverty rates, environmental indicators, and offers a range of specialized thematic data in areas like infrastructure, agriculture, urban development, and more. It aims to support research, policymaking, and international development efforts. Researchers and policymakers can use the World Bank dataset to analyze global development trends, conduct socio-economic research, assess policies, and formulate international cooperation projects, thereby enhancing their ability to predict and assess financial risks effectively.

4.3 Evaluation Metrics

Multiple evaluation metrics are used in the study to comprehensively assess the performance of financial risk intelligence monitoring and warning models, ensuring that the models achieve the expected performance levels across various aspects. These metrics include accuracy, precision, recall, and F1 score. By taking into account the results of these metrics in combination, it is possible to assess and compare the performance of different models more comprehensively, thereby providing a more reliable basis for financial risk assessment.

4.3.1 Accuracy

Accuracy is used to evaluate the performance of a model in classification tasks. It measures the proportion of correctly classified samples by the model out of the total number of samples, typically expressed as a percentage. Specifically, accuracy represents the model's ability to correctly classify samples into their respective categories. The formula for calculating accuracy is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

In this context, TP represents the number of risk events correctly identified as risk events, TN represents the number of normal cases correctly identified as normal cases, FP represents the number of normal cases incorrectly identified as risk events, and FN represents the number of risk events incorrectly identified as normal cases.

4.3.2 Precision

Precision focuses on the accuracy of a model's predictions for the positive class (e.g., risk events). Precision represents the proportion of samples correctly predicted as the positive class out of all samples predicted as the positive class by the model. Its formula for calculation is as follows:

$$Precision = \frac{TP}{TP + FP}$$

Precision's higher values indicate greater accuracy in the model's predictions for the positive class. A high precision implies that the model rarely misclassifies negative cases as positive, making it crucial in applications like financial risk monitoring to avoid incorrect risk alerts.

4.3.3 Recall

Recall measures the model's ability to successfully identify the positive class (such as risk events). It represents the proportion of samples that the model correctly predicts as the positive class out of all the actual positive class samples. The formula for calculating recall is as follows:

$$\text{Recall} = \frac{TP}{TP + FN}$$

where FN represents the number of samples that the model incorrectly predicts as the negative class, i.e., the number of risk events incorrectly identified as normal cases.

Higher values of recall indicate better performance by the model in capturing actual positive class samples. A high recall means that the model can effectively discover potential risk events. However, in some cases, a high recall may come with a lower precision. Therefore, in practical applications, it is necessary to strike a balance between recall and precision to meet specific task requirements.

4.3.4 F1-Score

The F1 score (F1-Score) is a comprehensive performance metric for classification models that combines precision and recall. It is used to provide a holistic assessment of a model's performance. The F1 score aims to balance a model's prediction accuracy for the positive class (precision) and its ability to capture the positive class (recall), making it particularly useful for addressing class imbalance issues. The formula for calculating the F1 score is as follows:

$$F1\text{Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Higher values of the F1 score indicate that the model performs better in balancing precision and recall. When both precision and recall of the model are high, the F1 score will also be high, and vice versa. In applications like financial risk assessment, researchers may place greater emphasis on the F1 score because it provides a comprehensive performance evaluation by considering both model errors and omissions.

4.4 Experimental Comparison and Analysis

In the research of intelligent monitoring and warning models for financial risk, we utilized four crucial financial datasets: the Fama-French Three-Factor Dataset, CRSP Dataset, Compustat Dataset, and World Bank Dataset. These datasets contain extensive information on financial markets and companies, covering areas such as stock markets, financial data, and macroeconomic indicators, providing us with rich data resources. To evaluate the performance of our model in financial risk monitoring tasks, we employed four core evaluation metrics: Accuracy, Precision, Recall, and F1-Score. These metrics have different strengths and can help us gain a comprehensive understanding of the model's performance, thereby providing valuable insights for financial risk management and decision-making.

In this section, first, we compared the performance of our approach with methods proposed by Du, Peng et al., Li, Xuetao et al., and others. We presented the results of the experimental comparisons and analyses using tables and visualizations. Additionally, we conducted comparative analyses of different models in terms of parameter count, inference time, and training time.

From the results in Table 1, we can analyze and compare the performance of different models on the Fama-French Three-Factor Dataset and CRSP Dataset. First, on the Fama-French Three-Factor Dataset, our model excels in all four metrics: Accuracy, Precision, Recall, and F1-Score, achieving 94.37%, 93.21%, 91.65%, and 92.42%, respectively. Specifically, Li, Xuetao et al.'s method also

Table 1. Comparison of accuracy and other aspects between the method in this article and other methods under Fama-French three-factor dataset and CRSP dataset

Model	Datasets							
	Fama-French Three-Factor Dataset				CRSP Dataset			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Du, Peng et al.	88.19	84.59	86.49	85.53	92.67	91.58	89.96	90.76
Li, Xuetao et al.	91.45	92.97	87.54	90.17	93.12	91.68	85.61	88.54
Regin, R et al.	91.67	91.93	84.39	88.00	85.71	84.91	86.68	85.79
Lei, Yang et al.	90.09	91.83	91.51	91.67	89.2	91.34	89.15	90.23
Wang, Lei et al.	87.34	90.74	84.86	87.70	90.05	88.99	86.95	87.96
VenkateswaraRao, M et al.	87.45	85.47	91.27	88.27	88.53	85.9	84.4	85.14
Ours	94.37	93.21	91.65	92.42	93.84	94.62	91.27	92.91

performs well on this dataset, with high precision and recall. Compared to Du, Peng et al.’s method, our approach shows improvements of 5.16% in recall and 6.89% in F1-Score. Next, on the CRSP Dataset, our model also outperforms others, with the highest values in all metrics, including an accuracy of 93.84% and an F1-Score of 92.91%. This indicates that our method exhibits excellent performance on this dataset as well. Li, Xuetao et al.’s model also performs well on the CRSP Dataset, especially in terms of precision and F1-Score. Regin, R et al.’s model shows relatively lower performance on this dataset, particularly in precision. By comparing the metrics on both datasets, it is evident that our method outperforms other models in terms of overall performance. Additionally, we have visualized the results from Table 1 for comparison, as shown in the following Figure 6.

Table 2 displays the performance of different experimental methods on the Compustat Dataset and World Bank Dataset. On the Compustat Dataset, our method achieves significantly better results in terms of accuracy, precision, recall, and F1-Score, with values of 95.13%, 93.42%, 91.49%, and 92.44%, respectively, outperforming other methods. Compared to Li, Xuetao et al.’s method, our approach improves accuracy and F1-Score by 8.91% and 4.03%, respectively. Overall, our model exhibits remarkable performance improvements on the Compustat Dataset, demonstrating higher classification accuracy and overall performance. On the World Bank Dataset, our method also performs exceptionally well. It achieves an accuracy of 94.34%, which is 6.54% higher than the average performance of other methods. Precision is 92.54%, surpassing the average performance of other methods. Recall is 92.31%, which is 5.36% higher than other methods, and the F1-Score reaches

Figure 6. Under the Fama-French three-factor dataset and CRSP dataset, the comparison of the accuracy of this method and other methods is visually displayed

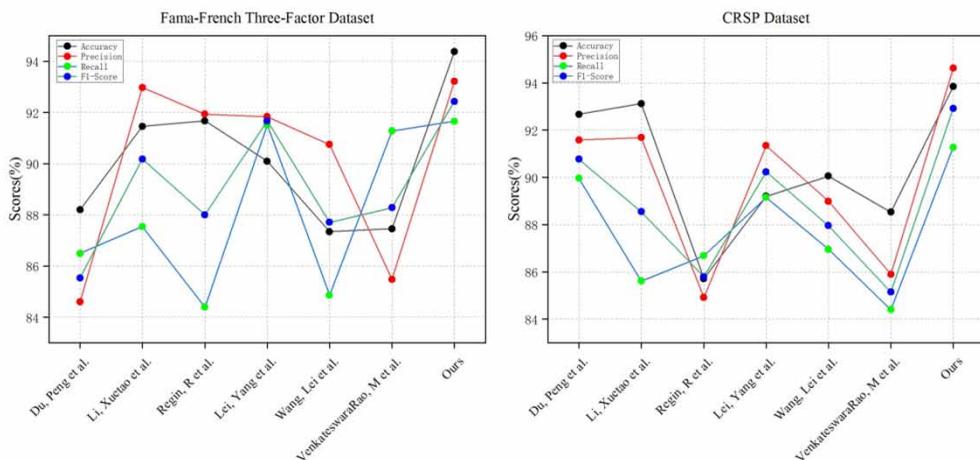


Table 2. Comparison of accuracy and other aspects between the method in this article and other methods under Compustat Dataset and World Bank Dataset

Model	Datasets							
	Compustat Dataset				World Bank Dataset			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Du, Peng et al.	93.17	84.01	85.59	84.79	86.87	85.81	89.49	87.61
Li, Xuetao et al.	86.22	85.89	91.09	88.41	87.77	84.81	90.17	87.41
Regin, R et al.	94.38	84.01	85.29	84.65	88.7	86.87	84.85	85.85
Lei, Yang et al.	90.9	89.22	86.65	87.92	94.02	92.42	85.94	89.06
Wang, Lei et al.	92.23	85.5	89.37	87.39	88.59	85.81	88.39	87.08
VenkateswaraRao, M et al.	92.03	87.85	87.43	87.64	86.9	87.41	88.07	87.74
Ours	95.13	93.42	91.49	92.44	94.34	92.54	92.31	92.42

92.42%, 4.01% higher than the average performance of other methods. These results indicate that our model demonstrates excellent stability and generalization across different financial datasets, further validating the reliability of our approach. Similarly, we have visualized the results from Table 2 for comparison, as shown in Figure 7.

From the data in Table 3, it can be observed that we conducted a comparative analysis of parameter count, inference time, and training time between other methods and our method on four datasets. First, on the Fama-French Three-Factor Dataset, our model has a relatively low parameter count (336.34M), indicating its efficiency in terms of model storage and computational resources, making it capable of running in resource-constrained environments. Compared to other methods, our model demonstrates excellent performance in inference time, requiring only 213.58ms, showcasing its potential for real-time

Figure 7. Under the Compustat Dataset and World Bank Dataset, the comparison between the accuracy and other aspects of this method and other methods is visually displayed

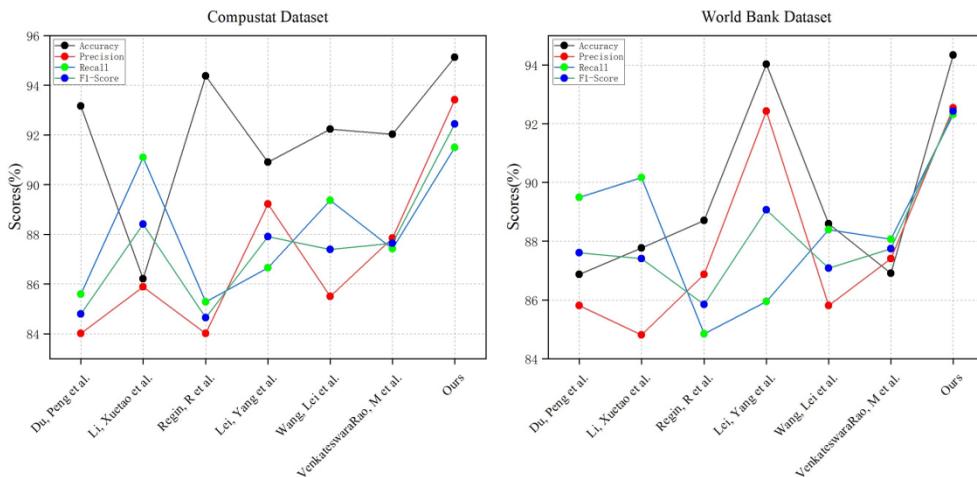


Table 3. Under the four data sets, the method in this article is compared with other methods in terms of parameters and other indicators

Model	Dataset											
	Fama-French Three-Factor Dataset			CRSP Dataset			Compustat Dataset			World Bank Dataset		
	Parameters(M)	Inference Time(ms)	Training Time(s)	Parameters(M)	Inference Time(ms)	Training Time(s)	Parameters(M)	Inference Time(ms)	Training Time(s)	Parameters(M)	Inference Time(ms)	Training Time(s)
Du, Peng et al.	380.04	246.21	162.55	487.80	288.34	182.47	384.31	304.14	162.04	472.95	266.24	195.88
Li, Xuetao et al.	454.09	301.24	183.43	365.28	310.24	196.83	352.37	241.87	174.74	441.02	256.94	181.08
Regin, R et al.	489.85	309.72	185.65	454.36	263.89	154.73	445.59	233.98	182.49	484.07	295.08	199.84
Lei, Yang et al.	439.29	314.38	156.25	440.25	240.53	172.46	483.19	268.89	152.80	474.31	243.70	198.40
Wang, Lei et al.	348.38	276.02	165.27	380.49	275.09	167.71	244.06	305.07	160.01	468.48	248.01	195.06
Rao, M et al.	383.01	297.66	190.29	497.83	305.53	154.31	368.64	296.65	186.96	391.73	249.59	194.78
Ours	336.34	213.58	147.24	347.26	236.73	137.91	338.71	224.83	146.85	341.83	218.53	143.59

monitoring and warning. In terms of training time, our model is also competitive, completing the entire training process in just 147.24 seconds, which accelerates model iteration and optimization. On the CRSP Dataset, Compustat Dataset, and World Bank Dataset, our model also performs exceptionally well. On these datasets, our model has a relatively low parameter count, and both inference time and training time are significantly shorter than other methods. Particularly on the CRSP Dataset, our model achieves the best training time of 137.91 seconds. These results indicate that our intelligent monitoring and warning model for financial risk exhibits efficiency and scalability across different financial datasets. These advantages make our model well-suited for practical financial monitoring and warning tasks, especially when analyzing and applying it across multiple datasets. Similarly, we have visualized the results from Table 3 for comparison, as shown in Figure 8.

In Table 4, we compared the performance metrics, including precision, recall, and F1-Score, of different models on the Fama-French Three-Factor Dataset and CRSP Dataset. On the Fama-French Three-Factor Dataset, the baseline model has a precision of 76.82%, recall of 74.92%, and an F1-Score of 75.86%. With the addition of an LSTM layer, there is a slight improvement in performance, with precision increasing to 81.73%, recall to 80.25%, and F1-Score to 80.98%. Further incorporating the Transformer, the performance is enhanced, with precision reaching 87.92%, recall at 86.92%, and an F1-Score of 87.42%. Finally, by combining LSTM and Transformer, the model achieves the best performance on the Fama-French Three-Factor Dataset, with precision at 93.21%, recall at 91.65%, and an F1-Score of 92.42%. On the CRSP Dataset, a similar trend is observed. The baseline model performs relatively poorly, with precision at 74.94%, recall at 72.42%, and an F1-Score of 73.66%. Adding LSTM and Transformer individually improves performance, but the best performance is still achieved when combining LSTM and Transformer, with precision at 94.62%, recall at 91.27%, and an F1-Score of 92.91%. These results indicate that on both datasets, the model that combines LSTM and Transformer achieves the best performance in financial risk monitoring, playing a crucial role

Figure 8. Under four data sets, the comparison of parameters and other indicators between this method and other methods is visually displayed

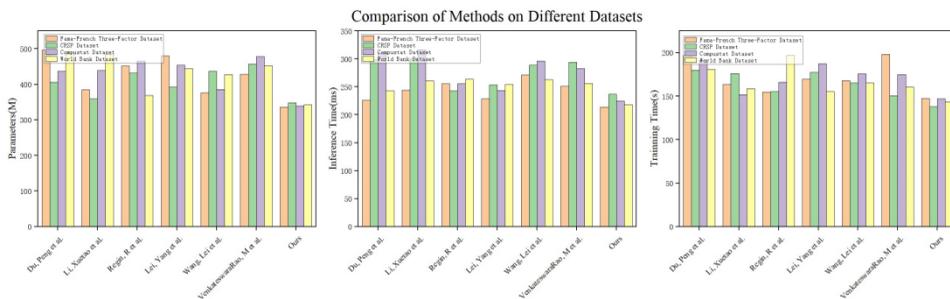


Table 4. In ablation experiments on Fama-French Three-Factor Dataset and CRSP Dataset, precision, recall, and F1-score metrics are selected for evaluation

Model	Dataset					
	Fama-French Three-Factor Dataset			CRSP Dataset		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
baseline	76.82	74.92	75.86	74.94	72.42	73.66
+LSTM	81.73	80.25	80.98	82.64	81.79	82.21
+Transformer	87.92	86.92	87.42	89.42	84.57	86.93
+LSTM Transformer	93.21	91.65	92.42	94.62	91.27	92.91

in improving model accuracy and recall. Additionally, we have visualized the results from Table 4 for comparison, as shown in Figure 9.

In Table 5, we conducted a comparative analysis of experimental results on the Compustat Dataset and World Bank Dataset. On the Compustat Dataset, the baseline model’s performance exhibited a precision of 75.25%, recall of 73.42%, and an F1-Score of 74.32%. When we introduced an LSTM layer, the performance significantly improved, with precision increasing to 82.93%, recall to 79.64%, and F1-Score to 81.25%. With the further addition of the Transformer, the performance improved once again, with precision reaching 86.27%, recall at 83.35%, and an F1-Score of 84.78%. Finally, the model that combined LSTM and Transformer achieved the best performance on the Compustat Dataset, with precision at 93.42%, recall at 91.49%, and an F1-Score of 92.44%. On the World Bank Dataset, a similar trend was observed. The baseline model’s performance was relatively low, with precision at 72.83%, recall at 75.28%, and an F1-Score of 74.03%. Adding LSTM and Transformer individually improved performance, but the best performance was still achieved when combining LSTM and Transformer, with precision at 92.54%, recall at 92.31%, and an F1-Score of 92.42%. This series of results emphasizes the outstanding performance of the model that combines LSTM and Transformer in financial risk monitoring, significantly enhancing model accuracy and recall, especially across different datasets. Finally, I have visualized the results from Table 5 for comparison, as shown in Figure 10.

Figure 9. Visual comparison of precision, recall, and F1 score metrics in ablation experiments on Fama-French Three-Factor Dataset and CRSP Dataset

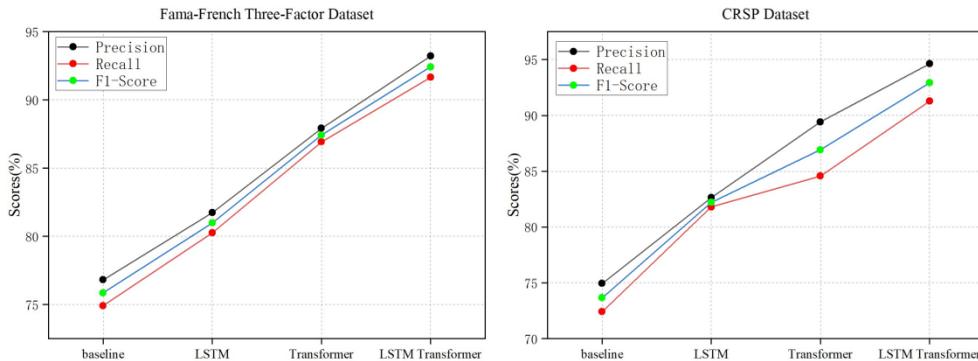
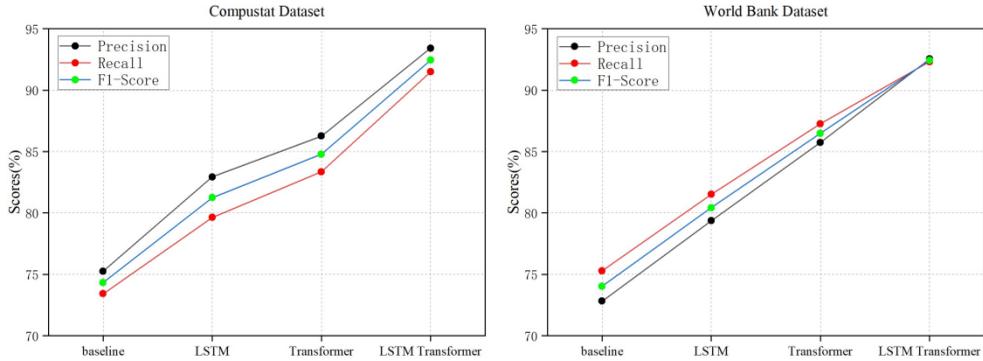


Table 5. In ablation experiments on Compustat Dataset and World Bank Dataset, precision, recall, and F1-score metrics are selected for evaluation

Model	Dataset					
	Compustat Dataset			World Bank Dataset		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
baseline	75.25	73.42	74.32	72.83	75.28	74.03
+LSTM	82.93	79.64	81.25	79.37	81.52	80.43
+Transformer	86.27	83.35	84.78	85.74	87.26	86.49
+LSTM Transformer	93.42	91.49	92.44	92.54	92.31	92.42

Figure 10. Visual comparison of precision, recall, and F1 score metrics in ablation experiments on Compustat Dataset and World Bank Dataset



5. DISCUSSION

Our research aims to explore a novel deep learning model that combines the strengths of LSTM and Transformer for financial risk monitoring and warning. We conducted extensive experiments on multiple financial datasets to validate the exceptional performance of this model. The experimental results demonstrate that our deep learning model excels in key metrics such as accuracy, precision, recall, and F1-Score, affirming its significant advantages in the field of financial risk. However, it's important to acknowledge some limitations in our study. Firstly, the datasets we used still have room for improvement in terms of scale and diversity to better reflect the variability of real financial markets. Secondly, the model's computational resource requirements are relatively high, which may pose deployment challenges in practical scenarios. Further research is needed to optimize the computational efficiency of the model. Thirdly, the interpretability of deep learning models remains a subject that requires deeper investigation to help financial practitioners better understand the decision-making processes of the models. Additionally, our research did not consider the impact of external factors on risk, such as changes in the macroeconomic environment and policy adjustments. This is an important direction for future research. Finally, model performance may vary in different financial markets or industries, necessitating further experimental validation.

On the other hand, our experimental results suggest that the deep learning model combining LSTM and Transformer may offer new insights for financial risk monitoring. Beyond the improvement in accuracy and recall, our model can detect potential risk signals at an earlier stage, facilitating proactive risk management strategies. Furthermore, our research underscores the importance of data quality in model performance, emphasizing the critical role of data preprocessing and cleansing in financial risk monitoring. Future research directions include exploring the applicability of this model in different financial domains and further enhancing model interpretability. Our study validates the potential value of the proposed hybrid model in financial data analysis, but challenges related to interpretability and data quality need to be addressed further.

In the future, we will work on improving the efficiency of the model to accommodate larger-scale financial data and consider the impact of external factors. Additionally, we plan to apply the model to real financial markets to assess its practical utility in risk monitoring and decision-making. In summary, our research highlights several key advantages of the model we proposed, including enhanced predictive accuracy and earlier risk detection, providing new insights and opportunities for future research and development in the field of financial risk management.

6. CONCLUSION

In this study, we are committed to advancing innovation in the field of financial risk monitoring by introducing deep learning technologies, specifically Long Short-Term Memory (LSTM) and Transformer models. Our contributions are multifaceted. Firstly, the incorporation of the LSTM model successfully extends traditional monitoring methods, enabling our model to more effectively capture the long-term dependency relationships in time series data within financial markets. This is crucial for predicting risk events in financial markets, as traditional methods may struggle to accurately capture these complex temporal dynamics. Secondly, the introduction of the Transformer model further emphasizes our focus on nonlinear relationships and complex dynamics in financial markets. Compared to traditional statistical and linear models, the Transformer model, with its self-attention mechanism, more flexibly captures nonlinear relationships in the data, enhancing our model's adaptability to complex market changes. Most importantly, we have innovatively blended deep learning techniques with traditional statistical and machine learning approaches, creating a new methodological paradigm. This fusion allows our model to comprehensively understand the dynamic changes in financial markets, improving predictive accuracy and market sensitivity. In summary, our research brings a fresh perspective and methodology to the field of financial risk monitoring. Compared to existing work, our model possesses unique advantages in capturing nonlinear relationships, multidimensional data, and long-term dependencies in time series. This innovation not only provides new directions for theoretical research but also offers a critical tool for practical risk management in financial markets. In the future, we aim to further optimize the efficiency of the model, expand its application to larger-scale financial datasets, and consider the impact of external factors to validate its utility in real-world applications. Through continued research, we anticipate bringing more innovation and opportunities to the field of financial risk management.

REFERENCES

- Adebayo, T. S., Akadiri, S. S., & Rjoub, H. (2022). On the relationship between economic policy uncertainty, geopolitical risk and stock market returns in South Korea: A quantile causality analysis. *Annals of Financial Economics*, 17(01), 2250008. doi:10.1142/S2010495222500087
- Al Janabi, M. A. (2022). Optimization algorithms and investment portfolio analytics with machine learning techniques under time-varying liquidity constraints. *Journal of Modelling in Management*, 17(3), 864–895. doi:10.1108/JM2-10-2020-0259
- Ali, M., Khan, D. M., Alshambari, H. M., & El-Bagoury, A. A. A. H. (2023). Prediction of complex stock market data using an improved hybrid emd-lstm model. *Applied Sciences (Basel, Switzerland)*, 13(3), 1429. doi:10.3390/app13031429
- Alizamir, M., Shiri, J., Fard, A. F., Kim, S., Gorgij, A. D., Heddami, S., & Singh, V. P. (2023). Improving the accuracy of daily solar radiation prediction by climatic data using an efficient hybrid deep learning model: Long short-term memory (LSTM) network coupled with wavelet transform. *Engineering Applications of Artificial Intelligence*, 123, 106199. doi:10.1016/j.engappai.2023.106199
- Ashtiani, M. N., & Raahmei, B. (2023). News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review. *Expert Systems with Applications*, 217, 119509. doi:10.1016/j.eswa.2023.119509
- Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. *Engineering Applications of Artificial Intelligence*, 120, 105843. doi:10.1016/j.engappai.2023.105843
- Chen, Q., Huang, Z., & Liang, F. (2023). Measuring systemic risk with high-frequency data: A realized GARCH approach. *Finance Research Letters*, 54, 103753. doi:10.1016/j.frl.2023.103753
- Chen, X., Peng, H., Wang, D., Lu, H., & Hu, H. (2023). SeqTrack: Sequence to Sequence Learning for Visual Object Tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14572–14581). IEEE. doi:10.1109/CVPR52729.2023.01400
- Cheng, L., van Dongen, B. F., & van der Aalst, W. M. (2019). Scalable discovery of hybrid process models in a cloud computing environment. *IEEE Transactions on Services Computing*, 13(2), 368–380. doi:10.1109/TSC.2019.2906203
- Dai, W. (2022). Application of improved convolution neural network in financial forecasting. [JOEUC]. *Journal of Organizational and End User Computing*, 34(3), 1–16. doi:10.4018/JOEUC.289222
- FuchsF.HorvathB. (2023). A Hybrid Quantum Wasserstein GAN with Applications to Option Pricing. Available at SSRN 4514510. 10.2139/ssrn.4514510
- Gianfreda, A., Maranzano, P., Parisio, L., & Pelagatti, M. (2023). Testing for integration and cointegration when time series are observed with noise. *Economic Modelling*, 125, 106352. doi:10.1016/j.econmod.2023.106352
- Hassan, S. G., Kieuvan, T. T., Liu, S., Garg, H., Hassan, M., & Iqbal, S. (2023). A Novel First-Order Fuzzy Rules-Based Forecasting System Using Distance Measures Approach for Financial Market Forecasting. *Journal of Mathematics*, 2023, 2023. doi:10.1155/2023/8027664
- Hong, J., Zhang, H., & Xu, X. (2023). Thermal fault prognosis of lithium-ion batteries in real-world electric vehicles using self-attention mechanism networks. *Applied Thermal Engineering*, 226, 120304. doi:10.1016/j.applthermaleng.2023.120304
- Hu, J., Chang, Q., & Yan, S. (2023). A GRU-based hybrid global stock price index forecasting model with group decision-making. *International Journal on Computer Science and Engineering*, 26(1), 12–19.
- Kaur, J., Parmar, K. S., & Singh, S. (2023). Autoregressive models in environmental forecasting time series: A theoretical and application review. *Environmental Science and Pollution Research International*, 30(8), 19617–19641. doi:10.1007/s11356-023-25148-9 PMID:36648728
- Korthikanti, V. A., Casper, J., Lym, S., McAfee, L., Andersch, M., Shoeybi, M., & Catanzaro, B. (2023). Reducing activation recomputation in large transformer models. *Proceedings of Machine Learning and Systems*, 5.

- Lazcano, A., Herrera, P. J., & Monge, M. (2023). A Combined Model Based on Recurrent Neural Networks and Graph Convolutional Networks for Financial Time Series Forecasting. *Mathematics*, *11*(1), 224. doi:10.3390/math11010224
- Liu, C., Zeng, Q., Cheng, L., Duan, H., & Cheng, J. (2021). Measuring similarity for data-aware business processes. *IEEE Transactions on Automation Science and Engineering*, *19*(2), 1070–1082. doi:10.1109/TASE.2021.3049772
- Liu, L., Cai, L., Zhang, C., Zhao, X., Gao, J., Wang, W., & Li, Q. (2023, July). Linrec: Linear attention mechanism for long-term sequential recommender systems. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 289-299). ACM. doi:10.1145/3539618.3591717
- Mba, J. C., & Mai, M. M. (2022). A particle swarm optimization copula-based approach with application to cryptocurrency portfolio optimisation. *Journal of Risk and Financial Management*, *15*(7), 285. doi:10.3390/jrfm15070285
- Megdad, M. M., Abu-Naser, S. S., & Abu-Nasser, B. S. (2022). *Fraudulent financial transactions detection using machine learning*.
- Mgammal, M. H., Al-Matari, E. M., & Alruwaili, T. F. (2023). Value-added-tax rate increases: A comparative study using difference-in-difference with an ARIMA modeling approach. *Humanities & Social Sciences Communications*, *10*(1), 1–17. doi:10.1057/s41599-023-01608-y PMID:36969312
- Mousapour Mamoudan, M., Ostadi, A., Pourkhodabakhsh, N., Fathollahi-Fard, A. M., & Soleimani, F. (2023). Hybrid neural network-based metaheuristics for prediction of financial markets: A case study on global gold market. *Journal of Computational Design and Engineering*, *10*(3), 1110–1125. doi:10.1093/jcde/qwad039
- Rapoo, M. I., Chanza, M. M., & Motlhwe, G. (2023). Inflation Rate Modelling Through a Hybrid Model of Seasonal Autoregressive Moving Average and Multilayer Perceptron Neural Network. In *Research Anthology on Macroeconomics and the Achievement of Global Stability* (pp. 551–567). IGI Global.
- Vuletić M. Prenzel F. Cucuringu M. (2023). Fin-gan: Forecasting and classifying financial time series via generative adversarial networks. Available at SSRN 4328302. 10.2139/ssrn.4328302
- Wang, P., & Han, W. (2021). Construction of a New Financial E-Commerce Model for Small and Medium-Sized Enterprise Financing Based on Multiple Linear Logistic Regression. [JOEUC]. *Journal of Organizational and End User Computing*, *33*(6), 1–18. doi:10.4018/JOEUC.286808
- Wang, X., Kang, Y., Hyndman, R. J., & Li, F. (2023). Distributed ARIMA models for ultra-long time series. *International Journal of Forecasting*, *39*(3), 1163–1184. doi:10.1016/j.ijforecast.2022.05.001
- Wang, Y., Yang, G., Li, S., Li, Y., He, L., & Liu, D. (2023). Arrhythmia classification algorithm based on multi-head self-attention mechanism. *Biomedical Signal Processing and Control*, *79*, 104206. doi:10.1016/j.bspc.2022.104206
- Wang, Z., Zhao, Q., & Qiu, L. (2022). Multi-Dimensional Factor Correlation, Multiple Interbank Network Contagion, and Conditional VaR of Banks. *Frontiers in Physics (Lausanne)*, *10*, 895603. doi:10.3389/fphy.2022.895603
- Xu, X., Jin, X., Xiao, D., Ma, C., & Wong, S. C. (2023). A hybrid autoregressive fractionally integrated moving average and nonlinear autoregressive neural network model for short-term traffic flow prediction. *Journal of Intelligent Transport Systems*, *27*(1), 1–18. doi:10.1080/15472450.2021.1977639