Available Online: 13 July 2026 DOI: 10.54254/3049-5768/2026.35342

Generative adversarial network-based learning for multi-source heterogeneous data

***Bo Xu*[1,2*], *Lifei Lu*[1,2], *Yulin Lu*[1,2], *Sitong Li*[1,2]**

[1]School of Big Data and Artificial Intelligence, Guangdong University of Finance & Economics, Guangzhou, China

[2]Key Laboratory of Collaborative Innovation in Digital Economy, Guangzhou, China

*Corresponding Author. Email: 20171080@gdufe.edu.cn

Abstract. Owing to the low occurrence frequencies of extreme financial risk events, risk prediction models are prone to being dominated by the majority of the normal samples encountered during the pretraining process when using financial risk data. This weakens the ability to effectively learn abnormal characteristics and thus reduces the sensitivity and stability of the developed predictive model. To overcome these limitations, a data class-balancing model, Financial Wasserstein Generative Adversarial Network with Gradient Penalty (FinWGAN-GP), is constructed on the basis of the Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) framework. Experimental results reveal that the FinWGAN-GP model achieves high-fidelity generation and expansion for high-risk samples, and a balanced dataset is conducive to improving the training effect of the early warning model. Through the integration of multisource heterogeneous data and Artificial Intelligence (AI) methods, the proposed model can identify potential financial risks earlier and more accurately, helping enhance the forward-looking and scientific nature of financial regulation.

Keywords: digital economy, financial technology, artificial intelligence, extreme financial risk

1. Introduction

The rapid development of the digital economy and financial technology has led to a new data acquisition and analysis paradigm. From financial market transaction data to unstructured information sources, such as news, public opinions, and social media, vast and diverse data provide a more comprehensive information foundation for conducting risk modelling, but they also impose higher demands on data processing, feature extraction, and modelling schemes for cases involving multisource, heterogeneous data [1]. More importantly, extreme financial risk events are characterized by low frequencies and high losses, resulting in a very small proportion of risk samples in the overall financial data and a serious class imbalance problem [2]. Faced with these challenges, deep learning and other Artificial Intelligence (AI) techniques have exhibited significant advantages in terms of processing large-scale, multidimensional, and complex-structured data. Unlike the traditional statistical methods that make assumptions or limit specific models to the relationships between data and variables, these methods learn directly through algorithms, set objective functions to explore the statistical relationships between variables and the overall characteristics of the input data, and address the issues encountered by the traditional methods, such as model mis-specifications, omitted variables, and the "curse of dimensionality". Thus, they provide powerful tools for integrating information from multiple sources, modelling nonlinear dynamic relationships, and addressing sample imbalance issues [3].

2. Status of research in China and other countries

In the field of finance, relatively few data related to specific risk events, such as financial crises, fraud risks, and financial risks, exist. This makes the traditional machine learning models prone to classification biases, i.e., overfitting the majority class while ignoring the key minority classes. Researchers in China and abroad have conducted studies on these issues. For example, Lin et al. [4] used the Shanghai Stock Exchange Composite Index and the Shenzhen Stock Exchange Component Index as subjects and combined Random Undersampling (RU) with the Synthetic Minority Over-sampling Technique (SMOTE) to improve a Support Vector Machine (SVM), and the experimental results revealed that the developed improved model significantly enhanced the accuracy achieved when predicting extreme risks in financial markets. Liu et al. [5] employed random sampling, the SMOTE, and bagging methods for implementing class balance adjustments when identifying internet finance risks. In response to the imbalanced nature of credit card transaction data, Ju et al. [6] integrated the SMOTE model with Long Short-Term Memory (LSTM) and the K-Nearest Neighbours (KNN) method, and this integrated model exhibited significantly improved performance in terms of addressing misclassification issues concerning the majority class. Wen [7] proposed an extreme financial risk data identification model by combining three different algorithms, i.e., the SMOTE, Particle Swarm Optimization (PSO) and a Least Squares Support Vector Machine (LSSVM). Xiao and Xiang [8] proposed a new model for predicting extreme risks in China's financial markets based on the XGBoost ensemble learning framework; their approach overcomes the biases of SVMs when predicting unbalanced extreme financial samples. Yang et al. [9] used a balanced loss function to construct a corrected cross-entropy imbalanced sample model.

With the development of deep learning, researchers have applied Generative Adversarial Networks (GANs) to the processing of imbalanced samples in the financial field. Naritomi [10] proposed a data generation method based on a GAN-based artificial market simulation, and the experimental results revealed that the probability distribution of the synthetic order events generated by the GAN was close to reality. Zou [11] proposed an innovative Multivariate Empirical Mode Decomposition-Bidirectional Generative Adversarial Network (MEMD-BiGAN) risk prediction method, which combines multiscale analyses with GANs; in this approach, the BiGAN is used to generate an augmented dataset with a sufficient number of observations, thereby making the final risk prediction more accurate and reliable. Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) [12] exhibits strong stability in data generation tasks. The introduction of the Wasserstein distance and Gradient Penalty (GP) mechanism clearly mitigates gradient disappearance and modal collapse issues. However, even if WGAN-GP is used for minority class sample generation purposes, the core optimization objective is still to approximate the actual data distribution at the overall level without placing explicit weight constraints on the extreme risk region at the tail. This feature has great limitations in scenarios involving highly imbalanced financial risks [13]. Therefore, it is necessary to improve the WGAN-GP structure and optimize its mechanism to strengthen the ability of the associated model to characterize risk distributions. In this paper, a Temporal Convolutional Network (TCN) and the wavelet transform are combined to construct a dual-stream generator architecture, and the focal loss function is introduced to perform directional enhancement of the tail risk weights. This design not only preserves the superior distribution fitting ability of WGAN-GP but also overcomes the limitations of the traditional models with respect to imbalanced and nonlinear data by enhancing the ability to express time series structures and weighted losses, thus providing more balanced and high-quality training data for financial risk prediction models.

3. Financial time series generation model: FinWGAN-GP

In this paper, an improved financial time series generation model, i.e., Financial Wasserstein GAN with GP (FinWGAN-GP), is proposed. This model is aimed at capturing the multiscale spatiotemporal characteristics of the financial market through deep learning algorithms and generating synthetic risk data with high fidelity and statistical consistency to address the scarcity of extreme risk samples. Through a binary LSTM network, real financial time series data samples are divided into risk data and nonrisk samples, forming the conditional information C. The noise Z and the conditional information C are simultaneously input into the generator to generate data, and the synthetic risk data and real-time sample data are simultaneously input into the discriminator for adversarial learning. After the model training process is completed, the known abnormal samples are used for model evaluation and performance testing purposes, the reconstruction error and anomaly fraction of the samples are calculated, and a comparative experiment is conducted to evaluate the accuracy and performance of the model in terms of handling imbalanced samples. The architecture of the FinWGAN-GP model is shown in Figure 1.

The FinWGAN-GP model includes three key improvements based on the classical WGAN-GP framework.

First, an encoder-decoder structure is introduced. By introducing LSTM as the core module for time series modelling, the original high-dimensional data are mapped to a low-dimensional latent space, thereby effectively extracting the features of financial data and lowering the learning difficulty faced by the generator.

Second, a frequency-domain dual-stream generator is designed. The time-domain branch uses a TCN to capture local fluctuations and long memory dependencies. The frequency-domain branch uses the wavelet transform to decouple data for accurately modelling flows at different frequencies. In addition, to address the problem regarding the small number of abnormal risk samples contained in the input data, Featurewise Linear Modulation (FiLM) is used to add a sequence of risk labels generated by the Isolation Forest (IF) algorithm to the input of the generator, enabling the generator to produce a time series for a specified risk category.

Third, the loss function of the discriminator is improved, and the focal loss is introduced to replace the traditional cross-entropy loss so that the weights of hard and easy samples are dynamically adjusted, thus enhancing the attention paid by the model to the high-risk samples of the minority class.

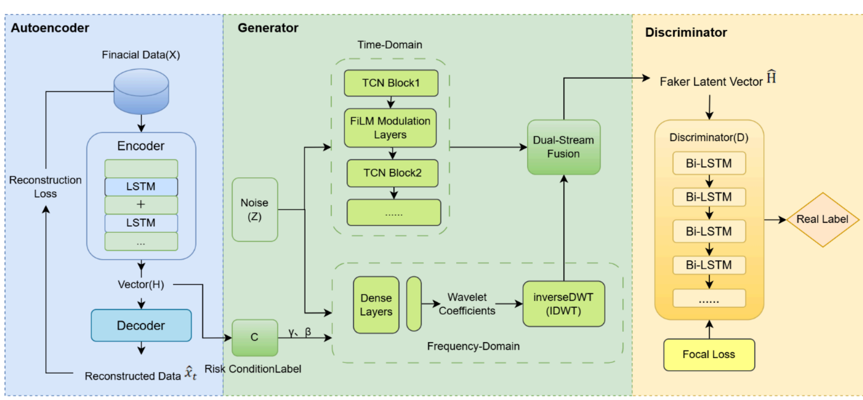

Figure 1. Structure of the FinWGAN-GP model

3.1. Encoder-decoder network

To solve the problems of high-dimensionality and sparseness of financial time series in the original data space, FinWGAN-GP constructs an autoencoder module composed of embedding and reconstruction submodules, which is designed to learn efficient latent data representations and nonlinear manifold features of the data through a joint training mechanism. The specific structure of the encoder-decoder network is shown in Figure 2.

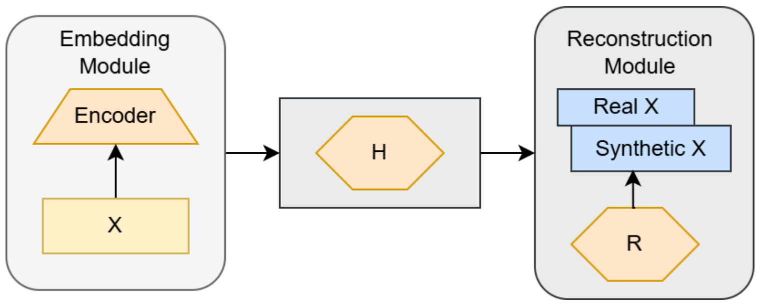

Figure 2. Encoder-decoder module

For the market state and core dynamic features that are implied in the input financial data, the embedding module uses a multilayer stacked LSTM network as the backbone architecture. After the original financial time series $X = \{x_1, x_2, \ldots, x_T\}$ is input into the autoencoder network, the LSTM network captures the long-range dependencies of the sequence through its gating mechanism and maps it to the low-dimensional latent vector space H. The mathematical expression of this dimensionality reduction process is $h_t = E(x_t)$, where h_t represents the latent variable at time t after undergoing feature compression, which effectively filters out the high-frequency micro-noise contained in the original data. Regarding the reverse process of the encoder network, the reconstruction module performs the task of reconstructing the original data from the latent space. This network is designed to learn the latent representations and temporal dynamics in the sequence and accurately decode the original time series $\widehat{x}_t$ from the low-dimensional latent vector h_t, i.e., $\widehat{x}_t = R(h_t)$. During the training stage, the model is optimized by minimizing the reconstruction error loss function between the original sequence and the reconstructed sequence, thereby forcing the latent space to maximally retain the complete information that is needed to generate the real data. This reversible mapping mechanism lays a solid feature foundation for the generator to subsequently perform high-fidelity adversarial learning in the latent space.

3.2. Generator network structure

The generator is the core component of FinWGAN-GP. In the early warning scenario concerning systemic financial risk, the sample distribution is significantly unbalanced. It is assumed that the actual data distribution is $P(x) = \mathit{\Pi}_n P_n(x) + \mathit{\Pi}_r P_r(x)$, where $\mathit{\Pi}_n \ll \mathit{\Pi}_r$, $P_n(x)$ is the distribution of the normal state, and $P_r(x)$ is the risk status distribution. The parameter update process of the traditional WGAN-GP model is dominated mainly by the majority-class distribution, the generator is inclined to converge to the optimal solution that is dominated by the master distribution, and the contribution of the risk distribution $P_r(x)$ to the overall gradient is significantly reduced.

Therefore, a generator network is designed for the frequency domain-time domain dual-stream architecture in this study. In the time-domain branch, a TCN is introduced to replace the traditional Recurrent Neural Network (RNN), leveraging its causal and dilated convolution properties to capture long-term dependencies and local fluctuations in risk data. In the frequency-domain branch, the multiresolution feature of the wavelet transform is used to decouple time series data to hierarchically model and accurately reconstruct market trends and high-frequency noise, and the FiLM mechanism is introduced to strengthen the conditional constraints imposed on the risk categories. The specific network structure of the generator is shown in Figure 3.

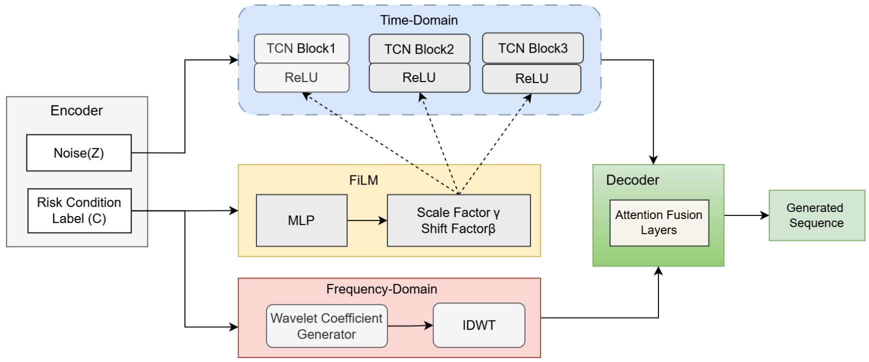

Figure 3. Generator network structure

The generator network adopts a parallel dual-stream structure for modelling the decoupling process. The time-domain generator branch uses a TCN as its backbone architecture and is mainly responsible for capturing the autocorrelations and long memory features of time series. Compared with traditional RNNs, TCNs possess parallel computing efficiency and gradient propagation stability advantages that are derived from the dilated causal convolution structure, so a TCN can efficiently extract long-range dependency features from series through an exponentially expanded receptive field while ensuring that no future information leakage occurs. Thus, it can effectively capture long-term dependencies spanning hundreds of time steps and accurately output time-domain representations that reflect macro market trends. The frequency-domain generator branch focuses on the microscopic fluctuations and multiresolution frequency-domain features of the given modelling data. The input noise and conditional features pass through a multilayer fully connected network or convolution layer, and the original financial data distribution $P(x)$ is mapped to a set of wavelet coefficients that approximate the energy distribution of the actual data in the frequency domain (Equation (1)):

$$x(t) = \sum_j A_j(t) + \sum_j D_j(t) \tag{1}$$

where $A_j(t)$ represents the low-frequency trend and $D_j(t)$ represents the high-frequency impact.

After generating the wavelet coefficients, the Inverse Discrete Wavelet Transform (IDWT) decomposes the complex distribution into multiple subdistributions for modelling purposes and reduces the difficulty of the distribution approximation step; that is (Equation (2)),

$$P(x) = T^{-1}(P(A_j), P(D_j)) \tag{2}$$

Compared with the direct generation of time series values, the approximation of sparse wavelet coefficients can more effectively decouple the trend term and the noise term in the given signal, thereby improving the accuracy of the model in terms of fitting the medium- and high-frequency burst risks observed in financial markets. The generator learns the distribution characteristics of different frequency bands in this branch, captures the market trend by modelling the low-frequency part, and simulates random fluctuations by modelling the high-frequency part. The output of the dual-stream branch is dynamically weighted and concatenated through an attention fusion layer. This layer adaptively adjusts the weights of the time-domain features and the frequency-domain features according to the currently generated contextual information and finally outputs the generated latent vector sequence $\widehat{H}$.

In particular, unlike the traditional vector concatenation strategy that is applied at the input end, the conditional input based on the FiLM mechanism dynamically modulates the network features. The goal is to map the discrete risk state labels identified by the IF algorithm to continuous affine transformation parameters to achieve precise control over the internal feature distribution of the generator network.

3.3. Discriminator network structurencoder-decoder network

The task of the Discriminator (D) is to distinguish the generated latent vector sequence from the real latent vector sequence, and the output is a probability value between 0 and 1. The discriminator of FinWGAN-GP adopts a Bi-LSTM network structure to simultaneously capture the forward and reverse dependencies of time series. The specific network structure is shown in Figure 4.

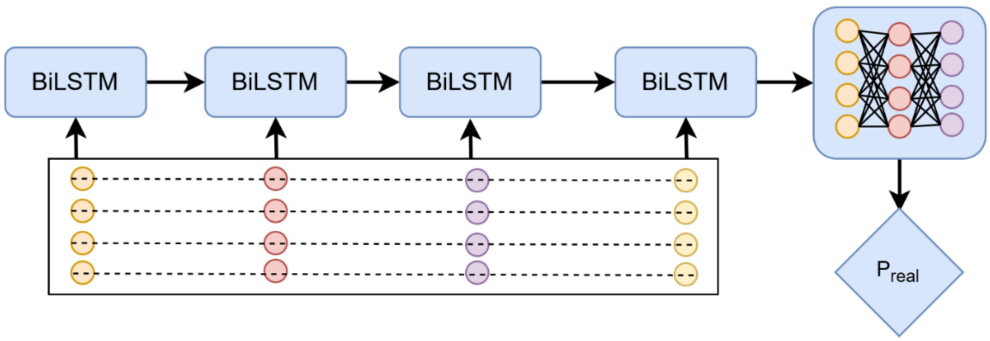

Figure 4. Structure of the discriminator network

The Bi-LSTM network can simultaneously process the characteristics of forward and reverse time flows and effectively capture the bidirectional dependencies that exist in the input sequence. In addition, the discriminator strictly follows the design specification of WGAN-GP, removes the sigmoid activation function from the last layer in the traditional GAN, and directly outputs the nonnormalized score, i.e., the Wasserstein distance-based estimation result. In addition, a GP term is integrated into the update process of the discriminator to force the discriminator to satisfy the 1-Lipschitz continuity constraint. This effectively solves the parameter distribution concentration and gradient explosion problems that are caused by weight pruning in the traditional WGAN-GP method and ensures the smooth convergence of the training process.

3.4. Focal loss function

In financial risk prediction scenarios, extreme risk events often belong to a minority of the available samples, whereas normal market fluctuations belong to the majority of the samples. The traditional loss functions are prone to being dominated by a large number of simple samples, making it difficult for the generator to learn extreme tail features. Therefore, in this study, the loss function of the discriminator is modified to the focal loss function. The core idea of the focal loss function is to reduce the weight of easy-to-classify samples and increase the weight of difficult-to-classify samples. The improved loss function of the discriminator L_D consists of the Wasserstein distance, the GP term, and the focal loss term (Equation (3)):

$$L_D = \underbrace{[D(\widetilde{x}) - D(x)]}_{\text{Wasserstein.Loss}} + \lambda_{gp}\underbrace{[(||\nabla_{\widehat{x}} D(\widehat{x})||_2 - 1)^2]}_{\text{Gradient Penalty}} + \lambda_{focal} L_{FL} \tag{3}$$

where the focal loss item L_{FL} is defined as follows (Equation (4)):

$$L_{FL} = -\alpha(1 - p_t)^{\gamma} log(p_t) \tag{4}$$

where p_t is the prediction probability yielded by the model for the sample class and γ is the focusing parameter. When $\gamma > 0$, for easy-to-class samples with high prediction probabilities ($p_t \to 1$), the weight of the loss $(1 - p_t)^\gamma$ sharply approaches 0; for hard-to-classify samples with low prediction probabilities, the weight is relatively large. By adjusting γ, FinWGAN-GP can focus its optimized attention on extreme-risk samples that are difficult to generate, thereby significantly improving the performance of the generated data across the long-tailed distribution.

4. Experimental results and analysis

4.1. Experimental dataset

To test the effectiveness of the proposed method with respect to processing imbalanced data in cases involving highly imbalanced data, the improved FinWGAN-GP model is empirically validated in this study. The dataset contains monthly data for a total of 39 indicators, including macroeconomics, monetary market, stock market, foreign exchange market, bond market, real estate market, news sentiment index and risk label indicators; the data range from January 1, 2005, to December 31, 2024, and fully reflect the characteristics of financial risk data. The valid dataset consists of 176 rows and 39 columns, containing a total of 6,864 time series data points. Among them, less than 5% of the original data is classified as high-risk data, indicating a significant class imbalance issue. Direct input of such data into the model can easily lead to a decrease in the gradient propagation efficiency rate and an imbalance among the contribution weights of different features in the loss function, thereby affecting the stability and convergence of the model training process.

To alleviate these problems, the original features are preprocessed using the max-min normalization method, in which the dimensional variables are linearly mapped to $[0,1]$. This process effectively eliminates the dimensional differences, allows different features to receive relatively balanced attention weights during the model training process, improves the convergence speed of the parameter optimization procedure, and enhances the ability of the prediction model to analyse the internal pattern of the acquired features. The normalization formula is as follows (Equation (5)):

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{5}$$

where x' is the normalized value, x is the original data, and $min(x)$ and $max(x)$ are the minimum and maximum values in the training dataset, respectively. To visually characterize the difference between the generated data and the real data in terms of their distribution characteristics, two typical dimensionality reduction methods, i.e., t-distributed Stochastic Neighbourhood Embedding (t-SNE) and Principal Component Analysis (PCA), are selected for visually analysing high-dimensional time series data.

4.2. Comparative experiments and results

To verify the effectiveness of the improved GAN in financial data generation tasks, PCA and t-SNE are used to compare and analyse the distribution characteristics of the original and generated samples in the low-dimensional space. The results are shown below.

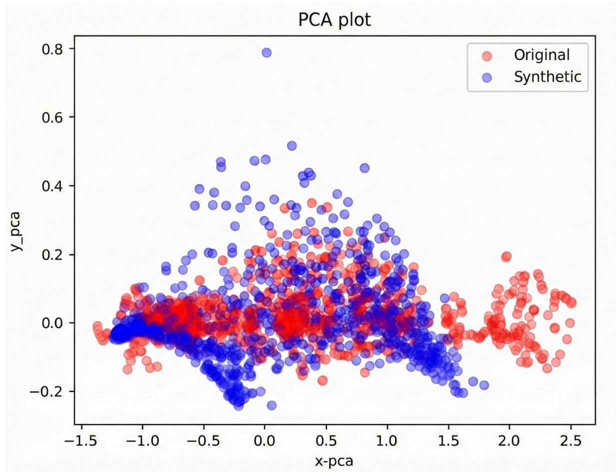

Figure 5. Visualization results of PCA-based dimensionality reduction

The distributions of the actual high-financial-risk data (red dots) and the data generated by FinWGAN-GP (blue dots) in the first two principal component spaces are shown in Figure 5. The PCA plot shows that the generated samples and the original samples exhibit a high degree of overlap in the principal component space; especially in the sample-dense areas, the distribution centres, expansion directions, and overall shapes of the two sample sets are relatively consistent. These results reveal that FinWGAN-GP can effectively learn the main statistical characteristics of the original financial data and reproduce their low-dimensional structural characteristics. In addition, the generative model is not limited to fitting mode samples but successfully captures the outlier characteristics of the distribution edge. While the blue scatters cover the main distribution of the red scatters, they maintain the same variance spread as that of the original data, which indicates that the generated samples do not exhibit significant clustering or collapse effects. This suggests that the model does not experience severe collapse during training and possesses a certain ability to preserve the diversity of data. In terms of the direction of the first principal component, the generated samples essentially cover the main distribution range of the original samples, reflecting that FinWGAN-GP has a strong ability to fit the overall trend and dominant fluctuation structure of the data.

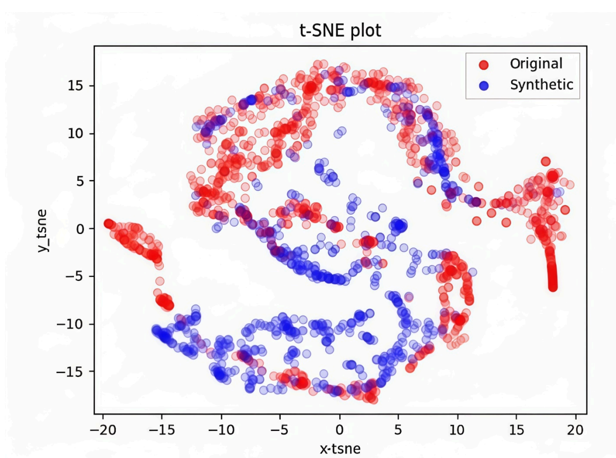

Figure 6. Visualization results of t-SNE-based dimensionality reduction

The original samples are not simply aggregated into several isolated clusters in the low-dimensional embedding space but rather are distributed along a continuous, curved low-dimensional nonlinear manifold as shown in Figure 6, reflecting the relatively smooth evolution trends of financial markets across different risk states. In addition, even if a small number of sparse clusters are present at the edge, the generated samples can be distributed along the manifold and can achieve high degrees of overlap with the original samples in many regions, indicating that the improved FinWGAN-GP model can successfully capture the nonlinear geometric structure and local similarity relationships in the data to some extent. There are a small number of outlier samples; however, the generated data are generally distributed within and at the edges of the original feature clusters, with no significant deviations from the observed category structure. The overall distribution is relatively stable, indicating that the data are feasible for practical applications.

In this paper, the improved FinWGAN-GP model, the WGAN-GP model [12], the WGAN model [14], the TimeGAN model [15], and the TransGAN model [12] are compared. The data generated by each model are input into the constructed systemic financial risk measurement models. The risk labels of the generated data and the actual risk labels obtained after the measurement process are compared and analysed. The experimental results are shown in Table 1.

Table 1. Evaluation indicator values achieved by different models

Model	Accuracy	AUC	F1
FinWGAN-GP	92.21%	0.9264	0.9215
WGAN	81.73%	0.8124	0.8284
WGAN-GP	82.18%	0.8127	0.8298
TransGAN	86.53%	0.8552	0.8547
TimeGAN	87.47%	0.8743	0.8762

The experimental results indicate that the proposed model has the highest risk label classification accuracy at 92.21%, which is the highest value among all the models. Compared with the basic WGAN model, the FinWGAN-GP model achieves a performance jump of more than 10%, which reveals the key role of the introduction of the GP mechanism in terms of stabilizing the training process and improving the generation quality level. Compared with WGAN-GP, which also uses GP, the proposed model still maintains an accuracy advantage of approximately 10%. This finding reveals that simply relying on GP or time series modelling is not sufficient for perfectly capturing complex financial risk characteristics. In addition, although TransGAN introduces an advanced transformer architecture, its performance is inferior to that of FinWGAN-GP. These results strongly verify the effectiveness of the frequency domain-time domain dual-stream architecture-based strategy that is designed in this paper. This targeted improvement successfully solves the problems regarding the scarcity of high-risk samples and long-tailed distribution, making the generated data more consistent with the real data in risk label classification tasks.

The Receiver Operating Characteristic (ROC) curves displayed in Figure 7 reveal that the ROC curve of FinWGAN-GP is always above all the other curves in the entire False Positive Rate (FPR) interval, indicating that this method attains optimal classification performance. The Area Under the Curve (AUC) reaches 0.9264, which is significantly greater than that of the second-ranked TimeGAN method. This means that under the same FPR, the proposed model can identify more genuine high-risk samples, indicating that the data generated by FinWGAN-GP can effectively enhance the discriminative capabilities of downstream classifiers.

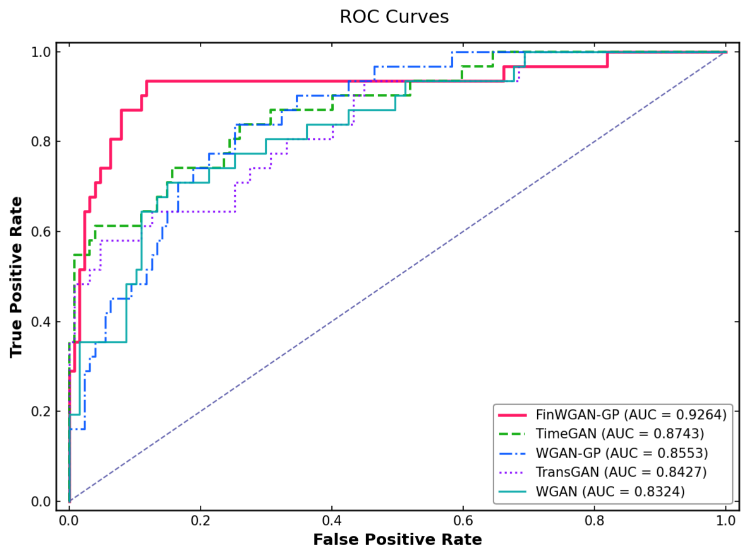

Figure 7. Comparison among the ROC curves of the models

These results indicate that the proposed improved GAN method offers improvements in areas such as handling low-sample-size datasets, generating time series data, and enhancing the attained classification performance. On the top of the advantage provided by the GP mechanism of WGAN-GP, FinWGAN-GP effectively alleviates the gradient disappearance and mode collapse problems that are encountered in the traditional GAN training procedure and exhibits improved training stability and sample distribution approximation capabilities. The added frequency domain-time domain dual-stream architecture can simultaneously characterize the local fluctuation characteristics and multiscale periodic structures of financial time series. Compared with models that rely only on time series modelling or pure transformer structures, the dual-stream architecture is more conducive to capturing time series fluctuations in the evolution trends of financial risks. Moreover, inputting risk condition information strengthens the targeted risk category feature learning process, which enhances the risk discrimination information of the generated samples while maintaining strong overall distribution consistency.

5. Conclusion

The FinWGAN-GP model proposed in this paper effectively expands the high-risk sample size, the statistical characteristics of the generated data are highly consistent with those of the real data, and the overfitting problem encountered by deep learning models in cases with sparse samples is solved; this is a key step towards improving the generalizability of early warning models. Future work should introduce AI algorithms and models to assist with decision-making processes and improve the intelligence levels of risk early warning schemes; therefore, the transformation from a "reactive" to a "proactive" regulatory model can be effectively realized, thus reducing the likelihood of occurrence of extreme risks and tail events.

Funding project

The Project of Philosophy and Social Science Planning of Guangdong (GD24CGL08)

References

[1] Wang, G. J., Chen, Y., Zhu, Y., & Xie, C. (2024). Systemic risk prediction using machine learning: Does network connectedness help prediction? *International Review of Financial Analysis, 93*, 103147. https: //doi.org/10.1016/j.irfa.2024.103147

[2] Li, H. Q., & Zhou, L. (2023). Systemic financial risk monitoring and early warning based on machine learning model. *Operations Research and Management Science, 32*(11), 212–219.

[3] Yang, H., Liu, L., Cui, J., & Zhang, Y. (2025). Research on dynamic measurement and early warning of systemic financial risk in China based on TVP-FAVAR and deep learning model. *Systems, 13*(8), 720. https: //doi.org/10.3390/systems13080720

[4] Lin, Y., Huang, X., & Xu, K. (2013). Research on extreme risk warning for financial market based on RU-SMOTE-SVM. *Forecasting, 32*(4), 15–20.

[5] Liu, H. L., Lin, B., Yun, W. J., & Wang, J. (2019). Comparison of balancing methods in internet finance overdue recognition: Taking PPDai.com as case. *Computer Science, 46*(S2), 595–598, 608.

[6] Ju, C. H., Chen, G. Y., & Bao, F. G. (2021). KNN-Smote-LSTM based consumer financial risk detection model: A case credit card fraud detection. *Journal of Systems Science and Mathematical Sciences, 41*(2), 481–498.

[7] Wen, T. X., & Kong, X. B. (2020). Research on extreme risk warning in financial market from imbalance distribution of samples. *Computer Engineering and Applications, 56*(8), 256–260.

[8] Xiao, Y. L., & Xiang, Y. T. (2021). Early warning of corporate bond default risk: Based on the GWO-XGBoost method. *Shanghai Finance*, (10), 44–54.

[9] Yang, L., Shi, B. F., & Dong, Y. Z. (2022). A credit risk evaluation model for imbalanced data classification based on class balanced loss modified cross entropy function. *Journal of Systems & Management, 31*(2), 255–269, 289.

[10] Naritomi, Y., & Adachi, T. (2020). Data augmentation of high frequency financial data using generative adversarial network. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 641–648). IEEE. https: //doi.org/10.1109/WIIAT50758.2020.00093

[11] Zou, Y., Yu, L., & He, K. (2023). Forecasting crude oil risk: A multiscale bidirectional generative adversarial network-based approach. *Expert Systems with Applications, 212*, 118743. https: //doi.org/10.1016/j.eswa.2022.118743

[12] Jiang, Y., Chang, S., & Wang, Z. (2021). TransGAN: Two pure transformers can make one strong GAN, and that can scale up. *Advances in Neural Information Processing Systems, 34*, 14745–14758.

[13] Hart, O., & Zingales, L. (2011). A new capital regulation for large financial institutions. *American Law and Economics Review, 13*(2), 453–490. https: //doi.org/10.1093/aler/ahr010

[14] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 214–223). PMLR.

[15] Yoon, J., Jarrett, D., & Van der Schaar, M. (2019). Time-series generative adversarial networks. *Advances in Neural Information Processing Systems, 32*, 5509–5519.