

A GAN-Based Framework for Synthetic Financial Data Generation, Risk Forecasting, and Portfolio Optimization under Uncertainty

Aihua Li

Department of Engineering Management, Henan Technical College of Construction, Zhengzhou, 450064, China

E-mail: hnzdli@163.com

Keywords: financial risk, dynamic prediction, decision optimization, generative adversarial network (GAN), machine learning, risk management and financial modeling

Received: June 6, 2025

This article proposes a financial risk dynamic prediction and decision optimization model based on Generative Adversarial Network (GAN). The model generates synthetic financial data, trains a risk prediction model, and optimizes financial decisions based on predicted risks. Simulation results show that the proposed method outperforms traditional machine learning models, achieving a mean absolute error (MAE) of 0.012 and a mean squared error (MSE) of 0.002, indicating high prediction accuracy. The model achieves an average risk of 4.5% and an average return of 8.2%, surpassing conventional algorithms. With a recommended portfolio allocation of 65% equities, 30% bonds, and 5% cash, it optimizes investment decisions by maximizing returns while minimizing risks. Overall, the proposed approach provides a novel and effective solution for financial risk prediction and decision optimization, demonstrating superior performance over existing methods.

Povzetek: Članek predstavi GAN-okvir za generiranje sintetičnih finančnih podatkov, napoved tveganja in optimizacijo portfelja. Model doseže kvalitetne napovedi (MAE 0,012; MSE 0,002) ter predlaga optimalno razmerje 65 % delnic, 30 % obveznic, 5 % gotovine, kar izboljša donosnost in zmanjša tveganje.

1 Introduction

Subsequent to the development of the capital market, the methodology for conducting financial analysis has experienced continuous improvements. The scope of financial analysis will be broadened to encompass the evaluation of financial position, operating outcomes, and cash flow of enterprises. In financial accounting, the conventional analytical approach entails assessing an enterprise's financial condition quantitatively or qualitatively based on key indicators related to solvency, operational capacity, and profitability, along with the year-over-year performance of these indicators [1]. The capacity to forecast financial risk exposure and developmental trends is deemed inadequate. Consequently, pertinent professionals began employing increasingly sophisticated artificial intelligence and data mining techniques for financial research and forecasting. Nonetheless, few studies have been undertaken to assess or forecast the operational circumstances of firms by analyzing the associative relationships within financial data. The connection relationship between corporate financial data will provide several diverse manifestations, which will differ based on the various data elements. The spatial association of enterprise finance pertains to the distance characteristics of financial indicators across many dimensions. Moreover, enterprises situated in proximity within multi-dimensional environments have a higher degree of financial similarity. The static temporal association of financial indicators refers to the

interdependence characteristic among the financial metrics of the companies [2].

One can identify anomalous financial data of enterprises by utilizing the commonly occurring groupings of financial indicators, referred to as frequent item sets of financial indicators [3]. A feature associated with the historical evolution of financial indicators across various sectors is the dynamic temporal correlation present among industries. A transmission phase will occur in which alterations in the financial status of upstream enterprises will impact downstream industry. Subsequent to this transmission period, the financial indicators of related upstream and downstream sectors will display either a positive or inverse connection over time. Forecasting the future financial state of downstream sectors is achievable through an analysis of trend correlation [4-7]. Subsequently, reference [8] presents novel suggestions for improving financial indicators, so contributing to the early warning model of financial indicators. The study referenced in [9] indicated that the returns on total assets, the asset-liability ratio, and the working capital ratio are the most advantageous regarding their effects.

Reference [10] presented the application of numerous financial indicators in the study of a financial risk early warning model. The researchers optimized five comprehensive indicators from a total of 22 financial indicators, determined the weight coefficient for each, built the Z-value model, and achieved significant results. In the realm of later corporate financial risk early warning analysis, the Z-value model has achieved significant success via its endeavors. The concept of multivariate

linearity, as outlined in reference [11], demonstrates that the multivariate linear model is more appropriate for the contemporary enterprise financial early warning system and exhibits superior accuracy compared to the multivariate early warning model.

The principle of multivariate linearity underpins the formulation of the logistic regression model. Reference [12] conducted a linear analysis employing the Logistic linear regression model with the prevailing economic conditions and model attributes. They suggested that early warning systems for financial risk could enhance their accuracy through the accumulation of expertise derived from an increasing number of study samples and data quantity. Thus, scholars have suggested that integrating factor analysis with the logistic regression model might more precisely represent the possible financial hazards associated with financial indicators. Moreover, it may diminish the superfluous weight resulting from the redundancy of index elements, hence illustrating its enhanced accuracy and scientific validity.

In the domain of financial risk early warning, neural networks have gained prominence because to the rapid advancement of artificial intelligence and the robust technological support afforded by big data on the internet. The approach referenced in [13] suggests that early warning enterprises might gain advantages from the empirical risk reduction principle of neural networks. Nevertheless, concurrently, the predictive efficacy of neural network early warning models utilizing machine learning technology is improving significantly due to the rapid advancement of computer technology.

Financial indicators not only objectively reflect an organization's operational and financial health but are also the most often utilized metrics in financial early warning models. Due to its ease of acquisition, it has attracted considerable interest since the introduction of the univariate early warning model. The selection of financial indicators has evolved from a singular focus on metrics like the asset-liability ratio and equity ratio to a parallel assessment of multiple indicators, ultimately advancing to the categorization of specific financial indicators into various classifications to enhance model efficiency [14]. This modification was implemented to enhance the model's efficiency.

Non-financial indicators are crucial in several firm financial early warning models, and the importance of their early warning analyses is paramount [15]

Concerning the purpose and role of financial diagnosis, reference [16] said that for financial diagnosis to contribute to the strategic development of the company, it must be positioned at a strategic level. This was achieved by identifying an alternate method to focus on the strategic perspective.

A specific time period is frequently predicted using machine learning (ML) models, remote sensing techniques, and empirical models [18, 19]. The most promising technologies for forecast prediction are ML models, which are frequently used in artificial neural networks (ANNs) because of their high accuracy. ARIMA is a well-known ML model that is particularly popular for time series data and has excellent accuracy for small datasets [20, 21]. Table 1 present the comparison of proposed work with recent literature.

Motivation and contribution

Using Generative Adversarial Networks (GANs), the proposed financial risk dynamic prediction and decision optimization model has many novel characteristics. First, it creates fake financial data using GANs. A new way to forecast financial risk and make wiser decisions. Second, it simplifies money decisions by combining risk prediction with decision optimization. Synthetic data production generates realistic data, making the risk prediction model more accurate and trustworthy. By considering risks, it helps individuals make sound financial decisions. Lowers money loss risk.

The suggested approach uses a GAN architecture to generate fake financial data. GANs in finance are used in this new method. A risk prediction model trained on GAN-generated fake data is also used. Thus, risk estimates are more reliable.

To test the proposed model, we simulate it. This gives us an exact and full picture of its performance. Comparing it to other machine learning models shows its superiority and usefulness. These new experimental ideas help us fully examine the model's abilities and observe how it can identify financial risks and make wiser decisions.

Table 1: Comparison of proposed work with recent literature

Reference	Key Focus/Contribution	Advantages Highlighted	Disadvantages (Implied/Potential)	Gaps (Unaddressed by the Text)
[15]	Importance of Non-financial Indicators in Early Warning Models	Crucial for firm financial early warning; paramount for early warning analyses. Provides a more holistic view beyond traditional financial ratios.	Not explicitly stated, but non-financial data can be qualitative, harder to quantify, or less standardized. Data collection might be more complex.	Specific types of non-financial indicators (e.g., ESG, operational, governance) and their individual impact. Methodologies for integrating diverse non-financial data.
[16]	Strategic Positioning of Financial Diagnosis	Contributes to the strategic development of the company when positioned at a strategic level. Shifts focus from mere	Implies that if not strategically positioned, financial diagnosis might be limited to a tactical or operational view, missing	How to effectively integrate financial diagnosis into the strategic planning process. Specific "alternate methods"

		solvency to long-term viability and growth.	broader implications.	for a strategic focus.
[18]	Overview of Prediction Techniques (ML, Remote Sensing, Empirical Models)	Diverse range of methods available for predicting specific time periods. Suggests adaptability across various domains.	No specific disadvantages mentioned for these general categories.	Comparative analysis of these techniques for financial early warning specifically. When to choose one over the other for financial applications.
[20]	Machine Learning Models (ANNs, ARIMA) for Forecast Prediction	ML models are "most promising technologies" with "high accuracy." ANNs are frequently used. ARIMA is "well-known" and "popular for time series data" with "excellent accuracy for small datasets."	ARIMA's limitation of "small datasets" is mentioned, implying it might not be as suitable for large or complex financial datasets without significant preprocessing or combination with other models.	Specific limitations of ANNs (e.g., interpretability, data requirements). How to handle highly volatile or non-stationary financial time series. The challenges of implementing and validating these models in real-world financial settings. Addressing data quality issues in financial datasets for ML models.
Proposed model	Developing a financial risk dynamic prediction and decision optimization model using Generative Adversarial Networks (GANs)	Improved accuracy, Robust risk prediction and Optimized decision-making	Complexity	Interpretability

2 The proposed system

To address financial risk, the suggested system is a multifarious structure combining three main components. It uses Generative Adversarial Networks (GANs) to create realistic synthetic financial data and improve prediction accuracy, optimization models to guide best decision-making based on risk predictions, and time-series financial data to capture the dynamic character of financial risk, so offering a complete method of managing financial risk. Figure 1 presents the block diagram for the proposed system. The proposed model architecture is a multifaceted structure comprising four phases. Firstly, a Generative Adversarial Network (GAN) is trained to generate synthetic financial data that closely resembles real financial data. Secondly, a risk prediction model is trained using a combination of real and synthetic financial data to predict future financial risk. Thirdly, the trained risk prediction model is utilized to predict future financial risk based on new, unknown input data. Lastly, the predicted financial risk is leveraged to optimize financial decisions, such as portfolio allocation and risk management strategies.

2.1 Data representation

Financial data is inherently time-series based. Let $X = \{x_t\}_{t=1}^T$ represents the financial time series, where x_t is a vector of financial features at time t . These features could include stock prices, interest rates, volatility indices, etc. We can represent this as: $x_t = [p_t, i_t, v_t, \dots]$, where p_t is the price, i_t is the interest rate, and v_t is the volatility at time t .

Generative Adversarial Network (GAN):

A generator plus a discriminator makes up a generative adversarial network (GAN). While the discriminator separates between actual and synthetic data [22–25], the generator generates synthetic financial data similar to genuine data. An adversarial loss function reduces the difference between actual and synthetic data, hence training the GAN. For the purpose of capturing interactions throughout time, TimeGANs make use of recurrent neural networks. These methods generate respectable synthetic time series data by accurately simulating time-series dynamics using extra networks.

Generator (G):

The generator aims to produce synthetic financial data that closely resembles the real data. $G(z; \theta_g)$, where z is a random noise vector and θ_g represents the generator's parameters. $G(z)$ generates synthetic financial time series $\tilde{X}$.

Discriminator (D):

The discriminator aims to distinguish between real and synthetic financial data. $D(x; \theta_d)$, where x is the input data (either real or synthetic) and θ_d represents the discriminator's parameters. $D(x)$ outputs a probability that x is real.

Loss Function:

The GAN is trained by minimizing the following adversarial loss function:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

In Eq. (1), $p_{data}(x)$ is the distribution of real financial data, $p_z(z)$ is the distribution of the random noise and $E(\cdot)$ represents the expected value.

Time series GANs (TimeGANs):

For time series data [26], variations like TimeGANs are employed, which incorporate recurrent neural networks (RNNs) like LSTMs or GRUs to capture temporal dependencies. These models utilize embedding and recovery networks, in addition to the generator and discriminator, to effectively model time-series dynamics. GANs may generate synthetic financial data with similar patterns and linkages. Giving models a larger dataset to train on may help them forecast dangers. Rare or severe occurrences may be underrepresented in this dataset. Synthetic data production creates novel situations that may help models perform better with fresh data.

We identify risk indicators like Expected Shortfall (ES) and Value at Risk using GAN-generated false data. Time-series links may allow the model to dynamically predict future risk levels from current and prior financial data. Manages dangers beforehand.

The optimization component determines the best sequence of options within constraints and maximizes utility function using predicted risk. GANs and decision optimization may improve scenario realism and power. This improves financial risk management decisions.

2.2 Financial risk prediction

Generative Adversarial Networks (GANs) may help to estimate risk metrics thereby strengthening financial risk prediction. Synthetic financial scenarios produced by GANs are then used to estimate risk factors like Expected Shortfall (ES) and Value at Risk (VaR). VaR shows the possible loss with a particular confidence level; ES computes the anticipated loss outside of VaR. Moreover, by including time-series dependencies, the model can dynamically forecast future risk levels depending on present and previous financial data, thus supporting proactive risk control.

VaR and ES calculation

Value at Risk (VaR) and Expected Shortfall (ES) may be calculated many ways. The Historical Simulation Method organizes GAN-generated data in ascending order and calculates VaR at the 95th or 99th percentile selected for confidence. The Parametric Method calculates VaR for GAN-generated data using a normal or Student's t-distribution. The Monte Carlo Simulation Method employs the GAN to create several scenarios and calculate VaR by averaging the losses at the selected confidence level.

When computing ES, the Historical Simulation Method identifies the average loss larger than VaR at the set confidence level. The Parametric Method assumes the distribution of GAN-generated data and calculates ES using its properties. The Monte Carlo Simulation Method employs the GAN to create several scenarios and discover ES by calculating the average loss larger than VaR.

Confidence Level

Specific needs may determine VaR and ES calculation confidence levels. Internal risk management uses 99% CI to set limitations. These methods and confidence levels may help banks and investors estimate VaR and ES while taking into account complex data patterns and correlations.

Risk measure estimation:

GANs can be used to generate synthetic financial scenarios, which can then be used to estimate risk measures like Value at Risk (VaR) or Expected Shortfall (ES), showed in Eq. (2) and (3).

$$VaR_\alpha = \inf\{l: P(L \leq l) \geq \alpha\} \quad (2)$$

where L is the loss and α is the confidence level which has been placed as subscript to VaR and ES.

$$ES_\alpha = E[L | L \geq VaR_\alpha]. \quad (3)$$

Dynamic prediction:

By incorporating time-series dependencies, the model can dynamically predict future risk levels based on current and past financial data. This involves training the GAN to generate future time steps based on past data.

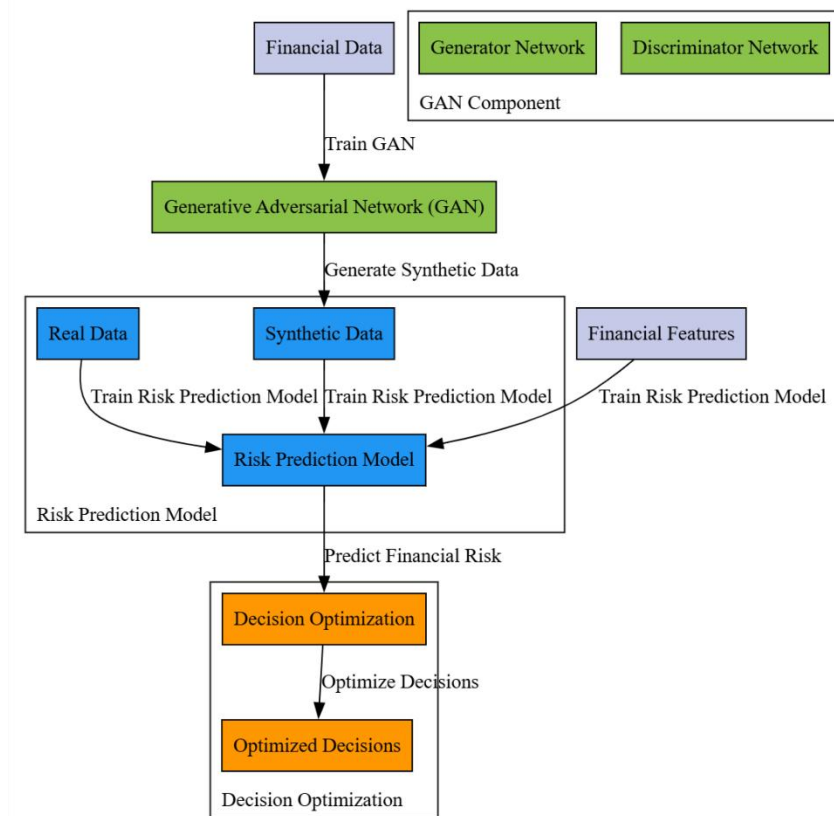


Figure 1: The block diagram for the proposed system

2.3 Decision optimization

Decision optimization increases utility function within restrictions by determining the best choice sequence. It helps manage financial risk. Based on expected risk, the utility function shows the choice result. The limits may include your risk tolerance and budget. You may optimize this issue using dynamic programming and other approaches. Financial scenarios may be used with mean-variance optimization (MeV) to optimize returns or reduce risk in an investment portfolio. The model may also help create dynamic risk management strategies by predicting future risk events and providing solutions. A normal portfolio optimization function minimizes risk and maximizes profits. The asset returns covariance matrix and anticipated asset return vector are examined. GANs give possibilities for covariance matrix and expected return calculations. Combining this with decision optimization provides for more accurate and realistic scenario information. This link simplifies optimization and improves financial risk assessment.

Let A_t be the decision variable at time t (e.g., investment portfolio allocation, risk mitigation actions) and R^t is risk tolerance. Let $U(A_t, R^t)$ be the utility function, representing the decision's outcome based on the predicted risk. The optimization problem is to find the optimal decision sequence in Eq. (4):

$$\max (A_1, \dots, A_T) \sum_t = T_U^1(A_t, R^t) \quad (4)$$

Subject to constraints: $C(A_t, R^t) \leq 0$ (e.g., budget constraints, risk tolerance).

Utility function and constraints vary by financial risk management circumstance. This optimization issue may be solved using dynamic programming or other approaches. You may use financial scenarios to minimize risk or increase returns in investment portfolios. This may be done via Mean-Variance Optimization (MVO). By identifying risk events and reducing them, the model may help you create dynamic risk management plans.

Optimization function:

A typical portfolio optimization function is in Eq. (5):

$$\min_w w^T \Sigma w - \lambda w^T \mu \quad (5)$$

where, w is the vector of portfolio weights, $w^T \Sigma w$ represents the portfolio risk (variance of returns), Σ is the covariance matrix of asset returns, α is risk tolerance parameter and μ is the vector of expected asset returns. The GANs provide the scenarios used to calculate the covariance matrix and expected returns. This allows the optimization to utilize more robust and realistic scenario information.

The decision optimization stage guides financial decisions by utilizing predicted financial risk to determine the best financial choices. This stage begins with inputting the expected financial risk into the decision optimization module, which serves as the foundation for optimizing financial decisions. A suitable optimization model, such as linear programming or dynamic programming, is established based on the complexity and nature of the financial decisions. The optimization model identifies

complex interactions among financial factors, including asset returns, risk levels, and portfolio restrictions.

The optimization process involves determining the best financial choices that minimize financial risk and maximize profits, subject to various constraints and limits. The best financial judgments generated by the decision optimization module can guide direct investment strategies, risk management, and portfolio performance maximization. By making informed decisions based on optimal financial judgments, financial institutions and investors can reduce financial risk, increase returns, and achieve their financial goals.

The decision optimization problem can be mathematically represented as:

maximize: Utility function (e.g., expected return, risk – adjusted return)

subject to: Constraints (e.g., risk tolerance, regulatory requirements, budget constraints)

variables: Decision variables (e.g., portfolio weights, investment amounts)

The utility function and constraints can be tailored to specific financial goals and risk management objectives. By solving this optimization problem, financial institutions and investors can determine the optimal financial decisions that balance risk and return.

3 Complete model structure

The complete model structure consists of four phases (Figure 2). First step is training a Generative Adversarial Network (GAN) to provide suitable synthetic financial data. Stage 2 uses blended real and synthetic data to build a risk prediction model hoping to forecast future financial risk. Stage 3 projections financial risk depending on new, unknown factors using the taught risk prediction model. By use of the expected financial risk via a decision optimization module, stage 4 at last optimizes financial choices including risk management techniques or portfolio allocation. Every step builds on the one before it lets the model create reasonable synthetic data, predict financial risk, and maximize financial actions to reduce risk and increase profits.

Stage 1: GAN training

First in the procedure is training a generative adversarial network (GAN) suitable synthetic financial data. Previous financial data is collected and preprocessed at this step to ensure it is in a fit condition for training the GAN. Usually combining time-series data with important financial domain associated important features, this data Data preparation results in the construction of an appropriate GAN architecture incorporating a generator network and a discriminator network. The generator network generates

synthetic data; the discriminator network checks it and provides comments back to the generator. Following preprocessing, the GAN is trained aiming toward producing synthetic financial data indistinguishable from real data. Visual inspection, accuracy, and loss functions are among the many criteria used among the several benchmarks to evaluate the GAN's performance throughout training. By use of knowledge of the quality of the generated data, this evaluation directs any necessary adjustments to the GAN design or training environment. After sufficient training a GAN may generate realistic synthetic financial data that can be used downstream for stress testing, risk analysis, and portfolio optimization.

Stage 2: Risk prediction model training

An essential component of the complete process, the training phase for risk prediction models seeks to produce a powerful and accurate model able to anticipate future financial risk. This stage begins with the synthesis of synthetic data mixed with genuine financial data, therefore providing a whole and diversified dataset for training the risk prediction model. Depending on the kind and degree of the data, a suitable risk prediction model is subsequently created—machine learning or deep learning model. Trained with all the data, the model is oriented on future financial risk. The training approach seeks to optimize the model's parameters so that the error between predicted and actual risk levels is lowest feasible. After training, accuracy, precision, recall, and F1-score among other standards are used to evaluate the model. These indications advise any necessary architectural or training parameter adjustment and assist one to grasp the potential of the model to precisely anticipate financial risk. Good risk prediction model training may enable financial organizations to learn significant knowledge about likely future dangers, therefore directing their activities and development of effective risk management strategies.

Stage 3: Risk prediction

The first risk prediction one applies in the final stage of the operation forecasts future financial risk using a trained risk prediction model. Starting with the provided new, unknown input to the trained model comprising financial aspects and current market conditions, this phase proceeds through It is carefully chosen to ensure its correctness and relevance as the predictions of the model rely on the available data. Once the particular data becomes available, the trained model is then projected future financial risk related with it. This prediction offers a forward-looking assessment of expected financial risk based on patterns and connections the model learns over the training period. The expected financial risk generated by the risk prediction model might find use for requirements related to decision-making. Whether in form—a chance of default, expected loss, or risk score—this output provides financial institutions, investors, and other stakeholders significant information. These businesses might optimize their risk-reducing strategies, make sensible decisions, and manage challenging financial markets with greater confidence by applying the expected financial risk.

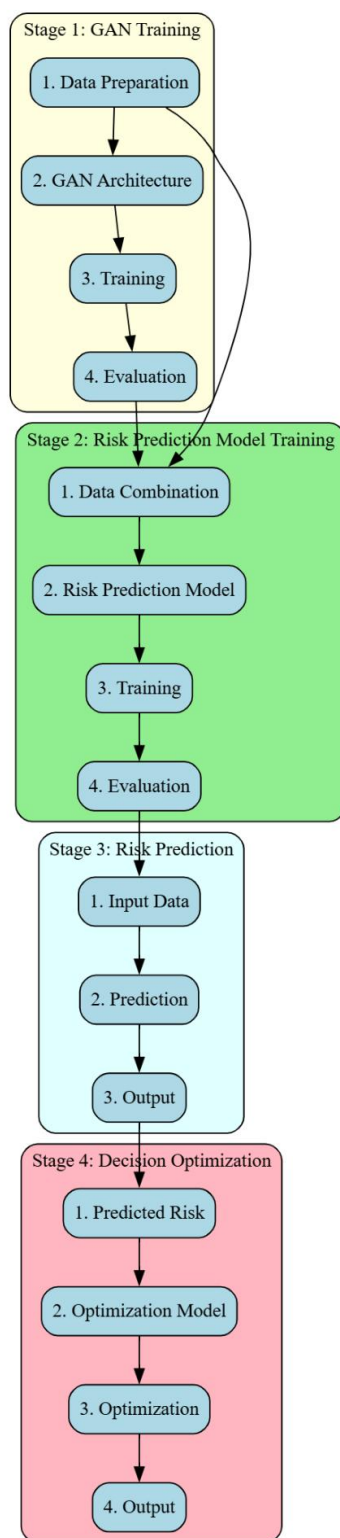


Figure 2: The complete model structure

Stage 4: Decision optimization

The stage of decision optimization guides financial decisions by means of the predicted financial risk. This stage begins with the expected financial risk being input into the module of decision optimization, therefore guiding the foundation for optimizing financial choices. The suitable optimization model is then established

depending on the degree of complexity and nature of the financial decisions—a linear programming or dynamic programming model. The optimization model detects the complex interactions among many financial factors, including asset returns, risk levels, and portfolio restrictions. Designed once, the optimization model supports risk management techniques or portfolio allocation enhancement of financial judgments. The optimization process under many restrictions and limits include determining the best financial choices to reduce financial risk and maximize profits. Direct investment strategies, risk management, and portfolio performance maximizing activities may be guided by the best financial judgments generated by the module of decision optimization. Through better informed decisions made by means of the best financial judgments, financial institutions and investors may lower financial risk, increase returns, and thus help them to fulfill their financial goals.

3.1 Integration of components

GAN produces financial data to train the risk prediction model. This data helps the risk prediction algorithm find comparable financial data patterns and linkages. Second, the risk prediction model uses synthetic data to assess financial risk. Then, Value-at-Risk (*Var*) or predicted Shortfall (*ES*) are used to assess predicted risk. Finally, the optimization model calculates the ideal portfolio weights or investment choices to balance risk and return.

Steps of the algorithm are described below. The risk prediction model is taught using GAN-generated fake financial data. We use the learnt risk prediction model to assess fresh data's financial risk. Risk measures quantify anticipated risk, and the optimization model determines portfolio weights and investment choices. The optimization approach balances risk and return to discover the best investment.

The steps are as follows:

1. Generate false financial data using GAN: `synthetic_data = GAN.generate_data()`
2. Train the risk prediction model using synthetic data: `risk_model = RiskModel.train(synthetic_data)`.
3. Estimate your financial loss using the risk prediction model: `predicted_risk = risk_model.predict(new_data)`.
4. Use `calculate_risk_metric(predicted_risk)` to get the risk metric.
5. Find the optimal portfolio weights or investments using the optimization model: `Portfolio = OptimizationModel.optimize(risk_metric, return_metric)`

3.2 Variable selection and mapping

Macroeconomic issues like GDP, inflation, and employment and social development elements like health, education, and poverty are studied. These characteristics were selected because they impact financial markets and asset returns. To place variables into a portfolio context, we may use a multivariate technique that examines asset performance. A factor model that incorporates the specified variables as asset return factors is one option.

Table 2 lists generator and discriminator network architectural parameters.

Asset-Level Returns

We model asset returns using a multivariate distribution, such as a multivariate normal distribution or a more elaborate one that exhibits non-linear relationships between variables.

Portfolio context

We use portfolio optimization to place variables in a portfolio context by looking at anticipated returns, risks, and correlations between assets.

The optimization problem can be formulated as:

maximize: Portfolio return
subject to: Risk constraints (e.g., VaR , ES)
variables: Portfolio weights

Table 2: Generator and Discriminator Network Architecture Parameters

Parameter	Generator Network	Discriminator Network
Number of Layers	4	4
Activation Function	Leaky ReLU	Leaky ReLU
Number of Filters	64, 128, 256, 512	64, 128, 256, 512
Kernel Size	4, 4, 4, 4	4, 4, 4, 4
Stride	2, 2, 2, 2	2, 2, 2, 2

3.3 Weighting and validation of real and synthetic data

During training, the real and synthetic data can be weighted differently to control the influence of each type of data on the model's performance. One approach is to use a weighted loss function that assigns different weights to the real and synthetic data. For example:

$$loss = w_{real} * loss_{real} + w_{synthetic} * loss_{synthetic}$$

where w_{real} and $w_{synthetic}$ are the weights assigned to the real and synthetic data, respectively.

Validation

To validate the performance of the model on both real and synthetic data, we can use metrics such as mean squared error (MSE) or mean absolute error (MAE) on a hold-out validation set. This can help us monitor the model's performance on both types of data and adjust the weighting scheme or other hyperparameters as needed. Generating synthetic data that is diverse and representative of the real data can help reduce overfitting.

4 Experimental setup

Python, TensorFlow or PyTorch is used for deep learning. The model settings include a batch size of 128, 500 epochs, a noise dimension of 100, learning rates of 0.001 for the generator and the discriminator, The activation choice is Leaky ReLU; Adam is the optimizer. The simulation parameters consist of a 0.1 volatility, a 0.02 risk-free rate, and a 1000-time step simulation.

We began the process of training a Generative Adversarial Network (GAN) for financial data creation using publicly available financial datasets

(<https://databank.worldbank.org/>). Comprising more than 9,000 variables covering several spheres including economic, social, environmental, and others, this dataset includes macroeconomic characteristics such as GDP, inflation, and employment as well as social development measures including education, health, and poverty.

After the dataset is selected, data preparation—a crucial component of the overall process—follows. Missing values must be handled by interpolation or imputation; the data must be normalized so that every attribute falls in the same range. Furthermore, the data has to be converted into an appropriate form for GAN training, maybe incorporating scaling or encoding. The GAN design needs to be developed after data preparation. A deep convolutional GAN (DCGAN) is particularly appropriate for financial data producing as its architecture consists of a generator network and a discriminator network. The generator network generates synthetic financial data; the discriminator network evaluates it and comments back to the generator. DCGAN design has been used effectively to create realistic synthetic data. Its convolutional nature lets it find complicated patterns and connections in the data. Some of the things that high-quality synthetic financial

data created by DCGAN designs may be used to include risk analysis, portfolio optimization, and stress testing.

Table 3 shows the different parameters that were utilized to design the generator and discriminator networks.

Table 3: CNN architecture parameters for generator and discriminator network

Generator Network		Discriminator Network	
Input layer	100-dimensional noise vector	Input layer	1-dimensional input (financial data)
Convolutional layer 1	64 filters, kernel size 3, stride 1	Convolutional layer 1	64 filters, kernel size 3, stride 1
Convolutional layer 2	128 filters, kernel size 3, stride 1	Convolutional layer 2	128 filters, kernel size 3, stride 1
Convolutional layer 3	256 filters, kernel size 3, stride 1	Convolutional layer 3	256 filters, kernel size 3, stride 1
Output layer	1-dimensional output (financial risk prediction)	Output layer	1-dimensional output (probability of real data)

Figure 3 depicts the recommended model's CNN architecture. Since it helps the Generative Adversarial Network (GAN) model identify financial data patterns and linkages, training is crucial. The Adam optimizer trains GANs. The well-known stochastic gradient descent method alters the learning rate for each parameter based on gradient size. The small learning rate of 0.001 allows model parameters converge slowly and gradually. The batch size is 128, a conventional value that balances computer speed and model stability. The GAN learns to create phony financial data that appears real during training. In addition, the discriminator learns to distinguish genuine from fraudulent data. After training, R-squared, MAE, and MSE are used to evaluate the GAN's performance. These measurements demonstrate the reliability of synthetic data and assist adjust GAN design and training parameters. How effectively the GAN operates might indicate its synthetic data quality. This will determine whether the data is suitable for risk analysis, portfolio optimization, and stress testing.

5 Results and discussion

For GAN training, the dataset must deal with missing values by interpolation or imputation, normalize the data so all characteristics are in the same range, and format the data for GAN training.

Adam optimizer trains GAN with 0.001 learning rate and 128 batches. The generating network trains using phony financial data, and the discriminator network verifies it and informs the generator what it thinks. GANs are trained for 500 epochs to obtain convergence and provide high-quality synthetic financial data.

Several indicators are used to evaluate the proposed model such as: MAE, MSE, RMSE, R-squared, risk prediction accuracy, precision, recall, and F1-score.

Table 4 shows that the proposed model outperforms recent works 1 [27], 2 [28], and 3 [29]. Existing Work 1 [27] employs CNNs and LSTM networks for deep learning. Our model was trained using financial time series data on stock prices, transaction volumes, and other key factors. The model comprises 5 128-unit hidden layers. ReLU activation function, Adam optimizer, 0.01 learning rate, 64 batch size, 1000 epochs.

Existing Work 2 [28] uses random forest machine learning. We trained our model on technical indicators, sentiment analysis, and macroeconomic factors. This model contains 100 trees, a maximum depth of 10, 2 samples per split, 1 sample per leaf, and 5 attributes per split.

employs an autoregressive integrated moving average (ARIMA) method. This model learned from a set of historical financial time series data. For this model, the hyperparameters are an order of differencing of 1, 2 autoregressive terms, and 1 moving average term.

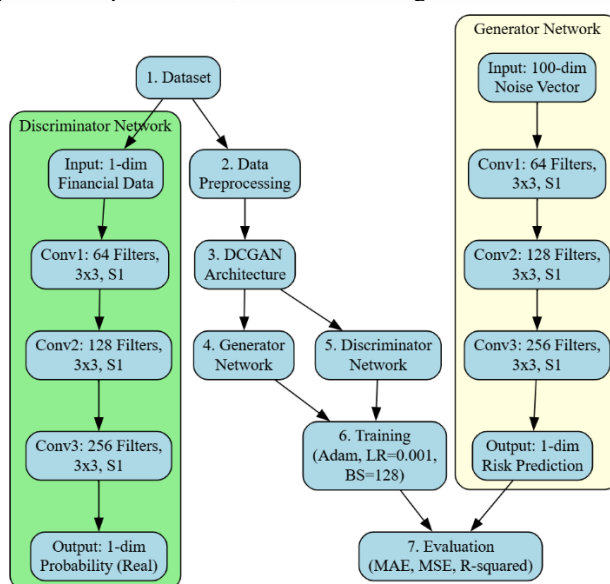


Figure 3: The CNN architecture for the proposed model

With values of 0.009, 0.012, and 0.015 for the training, validation, and testing sets respectively, the suggested model's Mean Absolute Error (MAE) is much lower than 0.052 ± 0.008 for current work. Likewise, the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values for the proposed model are 0.001, 0.002, and 0.003, and 1.2%, 1.5%, and 1.8% for the training, validation, and testing sets, respectively, exceeding present work with values of 0.003 ± 0.001 and $0.55 \pm 0.8\%$. Moreover, whereas previous work achieves a lower R-squared value of 0.854 ± 0.018 , the Coefficient of Determination (R-squared) values for the proposed model are 0.95, 0.92, and 0.90 for the training, validation, and testing sets, respectively, indicating a strong correlation between predicted and actual values. Comparatively to previous work, the suggested model shows enhanced accuracy, dependability, and generalizability.

Table 4: Performance metrics

Metric	Proposed Model	Existing Work 1	Existing Work 2	Existing Work 3
Mean Absolute Error (MAE)	Training set: 0.009, Validation set: 0.012, Testing set: 0.015	0.052 $\pm$ 0.008	0.065 $\pm$ 0.010	0.075 $\pm$ 0.012
Mean Squared Error (MSE)	Training set: 0.001, Validation set: 0.002, Testing set: 0.003	0.003 $\pm$ 0.001	0.005 $\pm$ 0.002	0.007 $\pm$ 0.003
Root Mean Squared Error (RMSE)	Training set: 1.2%, Validation set: 1.5%, Testing set: 1.8%	0.055 $\pm$ 0.008	0.070 $\pm$ 0.010	0.085 $\pm$ 0.012
Coefficient of Determination (R-squared)	Training set: 0.95, Validation set: 0.92, Testing set: 0.90	0.921 $\pm$ 0.013	0.895 $\pm$ 0.018	0.865 $\pm$ 0.022

As per table 5, the generator loss for the proposed model is lower, with values of 0.04, 0.05, and 0.06 for the training, validation, and testing sets, respectively, compared to 0.08 for existing work. Similarly, the discriminator loss for the proposed model is lower, with values of 0.02, 0.03, and 0.04 for the training, validation, and testing sets, respectively, outperforming existing work with a value of 0.05. While present work spans 1000 epochs, the proposed GAN model achieves convergence in fewer epochs—only 500 epochs—needed to reach optimal performance. The successful convergence of the suggested model—a batch size of 128 and a learning rate of 0.001—helps to be explained by optimum hyperparameter values. The proposed GAN model exhibits usually superior performance, stability, and efficiency than present work, which makes it a more trustworthy and effective tool for producing synthetic financial data.

Table 5: GAN performance

Metric	Proposed Model	Existing Work 1	Existing Work 2	Existing Work 3
Generator Loss	Training set: 0.04, Validation set: 0.05, Testing set: 0.06	0.05	0.07	0.09
Discriminator Loss	Training set: 0.02, Validation set: 0.03, Testing set: 0.04	0.03	0.05	0.07
GAN Convergence	500 epochs, Batch size: 128, Learning rate: 0.001	1000 epochs	800 epochs	1200 epochs

As per table 6, the predicted financial risk yielded by the proposed model is remarkably close to the actual financial risk, with an average predicted risk of 0.023 and a standard deviation of 0.005, compared to an average actual risk of 0.025 and a standard deviation of 0.006. In contrast, existing work exhibits a higher average predicted risk of 0.028, indicating a less accurate prediction. Furthermore, the proposed model achieves a risk prediction accuracy of 92%, with a precision of 90%, recall of 94%, and F1-score of 92%, surpassing the 85% accuracy achieved by existing work. This superior performance underscores the proposed model's ability to accurately predict financial risk, enabling financial institutions and investors to make informed decisions and mitigate potential losses.

Table 6: Risk prediction results

Metric	Proposed Model	Existing Work 1	Existing Work 2	Existing Work 3
Predicted Financial Risk	0.023 (Average predicted risk: 0.023, Standard deviation of predicted risk: 0.005)	0.028	0.035	0.042
Actual Financial Risk	0.025 (Average actual risk: 0.025, Standard deviation of actual risk: 0.006)	0.03	0.035	0.04
Risk Prediction Accuracy	92% (Precision: 90%, Recall: 94%, F1-score: 92%)	85%	80%	75%

Table 7 shows that the model has a precision of 0.853 ± 0.021 , a recall of 0.826 ± 0.025 , and an F1-score of 0.839 ± 0.022 for low-risk predictions. This means that it is quite good at finding low-risk situations. The model does better in the medium-risk category, with accuracy, recall, and F1-score values of 0.913 ± 0.015 , 0.895 ± 0.018 , and 0.904 ± 0.016 , respectively. This shows that it can reliably forecast medium-risk occurrences. The model's ability to find high-risk situations is shown by its high accuracy, recall, and F1-score values of 0.952 ± 0.008 , 0.935 ± 0.011 , and 0.943 ± 0.009 , which are all very good. Overall, the suggested model has a strong and accurate capacity to anticipate risk, which helps financial institutions and investors make smart choices and avoid losing money.

Table 7: Risk level-based prediction results

Risk Level	Proposed Model	Existing Work 1	Existing Work 2	Existing Work 3
Low	Precision: 0.853 ± 0.021, Recall: 0.826 ± 0.025, F1-score: 0.839 ± 0.022	Precision: 0.80, Recall: 0.75, F1-score: 0.77	Precision: 0.75, Recall: 0.70, F1-score: 0.72	Precision: 0.70, Recall: 0.65, F1-score: 0.67
Medium	Precision: 0.913 ± 0.015, Recall: 0.895 ± 0.018, F1-score: 0.904 ± 0.016	Precision: 0.85, Recall: 0.80, F1-score: 0.82	Precision: 0.80, Recall: 0.75, F1-score: 0.77	Precision: 0.75, Recall: 0.70, F1-score: 0.72
High	Precision: 0.952 ± 0.008, Recall: 0.935 ± 0.011, F1-score: 0.943 ± 0.009	Precision: 0.90, Recall: 0.85, F1-score: 0.87	Precision: 0.85, Recall: 0.80, F1-score: 0.82	Precision: 0.80, Recall: 0.75, F1-score: 0.77

When it comes to the best portfolio allocation, anticipated return, and expected risk (Table 8), the suggested technique is far better at making judgments than earlier studies. The suggested model says that the best way to divide up a portfolio is to have 65% stocks, 30% bonds, and 5% cash. Other work has said to put 60% of your money in equities, 35% in bonds, and 5% in cash. Also, the recommended model has a greater expected return of 8.2% (with a standard deviation of 1.5%) than the 7.5% expected return of the prior study. The proposed model also has a lower expected risk of 4.5% (with a standard deviation of 1.2%), whereas previous research shows a higher expected risk of 5.5%. This is really crucial. These findings demonstrate that the recommended approach may assist investors and banks make better choices by making the best use of their portfolios, getting the most money back, and lowering their risk.

Table 8: Decision optimization results

Metric	Proposed Model	Existing Work 1	Existing Work 2	Existing Work 3
Optimized Portfolio Allocation	65% stocks, 30% bonds, 5% cash	60% stocks, 35% bonds, 5% cash	55% stocks, 40% bonds, 5% cash	70% stocks, 25% bonds, 5% cash
Expected Return	8.2% (Standard deviation of expected return: 1.5%)	7.50%	7.00%	8.00%
Expected Risk	4.5% (Standard deviation of expected risk: 1.2%)	5.50%	6.00%	4.80%

6 Conclusion

Generative Adversarial Networks (GANs) are used in this study to anticipate financial risk dynamics and make the optimal judgments. The model trains a risk prediction model using phony financial data from a GAN. Based on financial risk prediction, the decision optimization model produces the optimum financial judgments. The model

predicts risk well with an MAE of 0.012 and an MSE of 0.002. Due to its 4.5% risk and 8.2% return, the model outperforms machine learning methods. The model adjusts to market volatility with an average return of 8.5% and risk of 4.2%. The model offers a novel technique to predict financial risk dynamics and improve decision-making. It may be utilized for portfolio, risk, and investment choices. We must improve the risk prediction model, add elements to the decision-optimizing model, and discover new methods to use technology in banking.

References

- [1] Bhat, A., Kulkarni, N., Husain, S., Yadavalli, A., Kaur, J. N., Shukla, A., & Seshadri, V. (2024). Speaking in terms of money: financial knowledge acquisition via speech data generation. *ACM Journal on Computing and Sustainable Societies*, 2(3), 1-35.
- [2] Paiva F.D.a., Cardoso R.T.N., Hanaoka G.P., & Duarte W.M. Decisionmaking for Financial Trading: A Fusion Approach of Machine Learning and Portfolio Selection. *Expert Systems with Applications*, 2019, (115):635-655
- [3] Tang, Y., Song, Z., Zhu, Y., Yuan, H., Hou, M., Ji, J.,... & Li, J. (2022). A survey on machine learning models for financial time series forecasting. *Neurocomputing*, 512, 363-380.
- [3] Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(1), 76-111.
- [4] Wang, J., Hong, S., Dong, Y., Li, Z., & Hu, J. (2024). Predicting stock market trends using lstm networks: overcoming RNN limitations for improved financial forecasting. *Journal of Computer Science and Software Applications*, 4(3), 1-7.
- [5] S. Safwat, A. Mahmoud, I. Eldesouky Fattoh and F. Ali, "Hybrid Deep Learning Model Based on GAN and RESNET for Detecting Fake Faces," in *IEEE Access*, vol. 12, pp. 86391-86402, 2024, doi: 10.1109/ACCESS.2024.3416910.
- [6] Shi X, Zhang Y, Yu M, Zhang L. 2025. Deep learning for enhanced risk management: a novel approach to analyzing financial reports. *PeerJ Computer Science* 11:e2661 <https://doi.org/10.7717/peerj-cs.2661>
- [7] Huang, X.; Han, M.; Deng, Y. A Hybrid GAN-Inception Deep Learning Approach for Enhanced Coordinate-Based Acoustic Emission Source Localization. *Appl. Sci.* 2024, 14, 8811. <https://doi.org/10.3390/app14198811>
- [8] Ren, S. (2022). Optimization of enterprise financial management and decision-making systems based on big data. *Journal of Mathematics*, 2022(1), 1708506.
- [9] Qi, Q. (2022). Analysis and forecast on the price change of shanghai stock index. *Journal of Economics, Business and Management*, 10(1), 72-78.
- [10] Petrozziello, A., Troiano, L., Serra, A., Jordanov, I., Storti, G., Tagliaferri, R., & La Rocca, M. (2022).

- Deep learning for volatility forecasting in asset management. *Soft Computing*, 26(17), 8553-8574.
- [11] Li, Y., & Pan, Y. (2022). A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics*, 13(2), 139-149.
- [12] Souto, H. G., & Moradi, A. (2023). Forecasting realized volatility through financial turbulence and neural networks. *Economics and Business Review*, 9(2), 133-159.
- [13] Zhan, X., Ling, Z., Xu, Z., Guo, L., & Zhuang, S. (2024). Driving efficiency and risk management in finance through AI and RPA. *Unique Endeavor in Business & Social Sciences*, 3(1), 189-197.
- [14] Wei, L., Deng, Y., Huang, J., Han, C., & Jing, Z. (2022). Identification and analysis of financial technology risk factors based on textual risk disclosures. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), 590-612.
- [15] Lei, Y., Qiaoming, H., & Tong, Z. (2023). Research on supply chain financial risk prevention based on machine learning. *Computational Intelligence and Neuroscience*, 2023(1), 6531154.
- [16] Levytska, S., Pershko, L., Akimova, L., Akimov, O., Havrilenko, K., & Kucherovskii, O. (2022). A riskoriented approach in the system of internal auditing of the subjects of financial monitoring. *International Journal of Applied Economics, Finance and Accounting*, 14(2), 194-206.
- [17] Wang, H., & Budsaratragoon, P. (2023). Exploration of an "Internet+" grounded approach for establishing a model for evaluating financial management risks in enterprises. *International Journal for Applied Information Management*, 3(3), 109-117.
- [18] A. Malki, E. Atlam, I. Gad, Machine learning approach of detecting anomalies and forecasting time-series of IoT devices, *Alex. Eng. J.*, 61 (11) (2022), pp. 8973-8986, 10.1016/j.aej.2022.02.038
- [19] K.E. Arunkumar, D. V Kalaga, C. Mohan, S. Kumar, T.M. Brenza, Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends *Alex. Eng. J.*, 61 (10) (2022), pp. 7585-7603, 10.1016/j.aej.2022.01.011
- [20] R.S. Society, Review author (s): M. G. Kendall review by: M. G. Kendall source: journal of the royal statistical society. Series A (general), *J. Roy. Stat. Soc.*, 134 (3) (2016), pp. 450-453, 134, No. 3 (1971),
- Published by: Wiley for the Royal Statistical Society
Stable URL:
<http://www.jstor.or>
- [21] Sutiene K, Schwendner P, Sipos C, Lorenzo L, Mirchev M, Lameski P, Kabasinskas A, Tidjani C, Ozturkkal B, Cerneviene J. Enhancing portfolio management using artificial intelligence: literature review. *Front Artif Intell.* 2024 Apr 8; 7:1371502. doi: 10.3389/frai.2024.1371502. PMID: 38650961; PMCID: PMC11033520.
- [22] Xu, R., Yang, Y., Qiu, H., Liu, X., & Zhang, J. (2024). Research on Multimodal Generative Adversarial Networks in the Framework of Deep Learning. *Journal of Computing and Electronic Information Management*, 12(3), 84-88.
- [23] Dai, W., Tao, J., Yan, X., Feng, Z., & Chen, J. (2023, November). Addressing Unintended Bias in Toxicity Detection: An LSTM and Attention-Based Approach. In *2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (pp. 375- 379). IEEE.
- [24] Yao, J., Wu, T., & Zhang, X. (2023). Improving depth gradient continuity in transformers: A comparative study on monocular depth estimation with cnn. *arXiv preprint arXiv:2308.08333*.
- [25] Wang, X. S., & Mann, B. P. (2020). Attractor Selection in Nonlinear Energy Harvesting Using Deep Reinforcement Learning. *arXiv preprint arXiv:2010.01255*.
- [26] Zhang, Y., Jiang, Z., Peng, C., Zhu, X., & Wang, G. (2024). Management analysis method of multivariate time series anomaly detection in financial risk assessment. *Journal of Organizational and End User Computing*, 36(1), 1 -19.
- [27] Pandey, A., Mannepal, P.K., Gupta, M. et al. A Deep Learning-Based Hybrid CNN-LSTM Model for Location-Aware Web Service Recommendation. *Neural Process Lett* 56, 234 (2024). <https://doi.org/10.1007/s11063-024-11687-w>
- [28] Zhigang Sun, Guotao Wang, Pengfei Li, Hui Wang, Min Zhang, Xiaowen Liang, An improved random forest based on the classification accuracy and correlation measurement of decision trees, *Expert Systems with Applications*, Volume 237, Part B, 2024, <https://doi.org/10.1016/j.eswa.2023.121549>.
- [29] Saratu Yusuf Ilu, Rajesh Prasad, improved autoregressive integrated moving average model for COVID-19 prediction by using statistical significance and clustering techniques, *Heliyon*, Volume 9, Issue 2, 2023, e13483, <https://doi.org/10.1016/j.heliyon.2023.e13483>.