ECONOMIC AND FINANCIAL RESEARCH LETTERS
Research Article | Volume 5 | Issue 5 | Mar 2025
Received: 19 Jan 2025 | Revised: 21 Feb 2025
Accepted: 23 Feb 2025 | Published Online: 13 Mar 2025

Logistic Regression-based method for Financial Risk Assessment

Hans Müller[1], Clara Schmidt[2] , Felix Weber[3,*], Haibo Wang[4]

[1] Institute of Data Science and Artificial Intelligence, Leipzig University of Applied Sciences, Leipzig, 04107, Germany

[2] Department of Financial Mathematics, University of Duisburg-Essen, Duisburg, 47057, Germany

[3] Center for Econometrics and Business Analytics, Johannes Gutenberg University Mainz, Mainz, 55128, Germany

[4] Heinz College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, 15213, USA

[*]Corresponding Author, Email: felix.weber@uni-mainz.de

Abstract: The importance of accurate financial risk assessment has been widely recognized in both academic research and practical applications. However, the existing methods often face challenges in terms of accuracy and efficiency. In response to this, this paper proposes a novel approach based on logistic regression for financial risk assessment. By incorporating key risk factors and applying advanced statistical modeling techniques, our method aims to improve the precision and timeliness of risk evaluation in financial decision-making processes. Through extensive empirical analysis and performance evaluation, we demonstrate the effectiveness and reliability of our proposed approach in capturing and predicting financial risks, providing valuable insights for risk management strategies in various financial sectors.

Keywords: *Financial Risk Assessment; Logistic Regression; Statistical Modeling; Risk Evaluation; Empirical Analysis*

1. Introduction

Financial Risk Assessment is a critical field within finance that focuses on evaluating and analyzing potential risks that may impact the financial performance and stability of individuals, companies, and financial institutions. The main goal of Financial Risk Assessment is to identify, measure, and manage various types of risks, such as market risk, credit risk, operational risk, and liquidity risk. However, this field faces several challenges and bottlenecks, including the complexity and interconnectedness of global financial markets, the rapid evolution of financial products and technologies, the lack of accurate and timely data, and the difficulty in accurately predicting and managing extreme events or black swan events. Addressing these challenges requires innovative approaches, advanced statistical and computational techniques, and interdisciplinary collaborations to enhance the effectiveness of Financial Risk Assessment and improve the overall resilience of financial systems.

To this end, advancements in Financial Risk Assessment research have reached a sophisticated level, incorporating complex statistical models, machine learning algorithms, and big data analytics to predict and mitigate potential risks in financial markets. In the field of financial risk assessment, various innovative approaches have been proposed to improve the accuracy and timeliness of risk evaluation for different sectors. Bi et al. introduced a concept of assessing company financial risks using a tribe-style graph with a novel Hierarchical Graph Neural Network [1]. Du et al. applied big data technology to develop an early warning mode for Internet credit financial risk assessment [2]. Bingler and Colesanti Senni analyzed and proposed criteria-based assessments to enhance the understanding of climate-related financial risk assessment tools [3]. In addition, In et al. focused on climate-related financial risk assessment specifically in energy infrastructure investments [4]. Furthermore, Luo et al. presented an artificial intelligence application model for supply chain financial risk assessment, integrating support vector machine optimization for improved performance [5]. Chi et al. conducted financial risk assessment of photovoltaic industry listed companies by incorporating text mining and logistic regression models for comprehensive evaluation [6]. Wang explored supply chain financial risk assessment using blockchain and fuzzy neural networks, emphasizing the integration of technologies for enhanced risk management [7]. Lastly, Zhang et al. proposed an online supply chain financial risk assessment approach based on improved random forest for real-time evaluations [8]. These studies collectively contribute to advancing the methodologies and tools for financial risk assessment across different industries and contexts. In the realm of financial risk assessment, innovative approaches have been developed to enhance accuracy and timeliness across various sectors. Logistic Regression is essential due to its ability to model binary outcomes and provide probabilistic interpretations, making it a suitable technique for assessing financial risks effectively.

Specifically, logistic regression serves as a vital statistical tool in financial risk assessment by modeling the probability of default or failure of financial entities. Its ability to handle binary outcomes makes it particularly useful for predicting credit risk and informing risk management strategies in finance. A literature review was conducted on logistic regression modeling, focusing on its applications and methodological developments. Hosmer et al. provided an accessible introduction to logistic regression models, emphasizing their applications in health sciences and modern statistical software [9]. Friedman discussed boosting as a key development in classification methodology, demonstrating its connection to statistical principles like additive modeling and maximum likelihood estimation [10]. Menard explored various aspects of logistic regression analysis and interpretation, including diagnostics and alternatives to traditional methods [11]. Harrell discussed regression modeling strategies with applications to linear models, logistic regression, and survival analysis, highlighting methods for fitting and simplifying complex models

[12]. King and Zeng addressed the challenges of rare events data analysis with logistic regression, proposing corrections for model estimation and sampling strategies to improve inference accuracy [13]. Additionally, Shah et al. compared logistic regression with other classification models for text classification tasks, showcasing the method's performance in real-world applications [14]. G et al. presented a logistic regression technique for predicting cardiovascular diseases, demonstrating the model's utility in disease prognosis and prevention [15]. However, limitations include potential biases in rare events analysis, the necessity for large sample sizes, and challenges in model interpretation and generalizability across diverse applications and datasets.

To overcome those limitations, the aim of this paper is to enhance the accuracy and efficiency of financial risk assessment through the development of a novel approach based on logistic regression. This new method integrates key risk factors and utilizes advanced statistical modeling techniques to improve the precision and timeliness of risk evaluation in financial decision-making processes. Specifically, by carefully selecting and incorporating relevant variables into the logistic regression model, we aim to capture the complex relationships between different risk factors and their impact on overall financial risk. Additionally, by leveraging the predictive power of logistic regression, our approach can enhance the ability to forecast potential risks and provide timely insights for risk management strategies. Through extensive empirical analysis and performance evaluation using real-world financial data, we demonstrate the effectiveness and reliability of our proposed approach in accurately identifying and predicting financial risks. Ultimately, our research contributes valuable knowledge and tools for enhancing risk management practices in diverse financial sectors, helping practitioners make more informed decisions and mitigate potential risks effectively.

Section 2 of the research outlines the problem statement regarding the crucial need for precise financial risk assessment. Section 3 introduces the proposed method, which suggests a unique approach using logistic regression for financial risk assessment. The paper aims to address existing challenges related to accuracy and efficiency by incorporating key risk factors and leveraging advanced statistical modeling techniques. Section 4 presents a detailed case study to illustrate the application of the method in real-world scenarios. Section 5 analyzes the results obtained from the empirical study, showcasing the effectiveness and reliability of the proposed approach in capturing and predicting financial risks. Section 6 delves into a discussion surrounding the implications of the findings and their relevance to financial decision-making processes. Finally, in Section 7, a comprehensive summary is provided, consolidating the research outcomes and highlighting the significant contributions to enhancing risk management strategies across diverse financial sectors.

2. Background

2.1 Financial Risk Assessment

Financial Risk Assessment is an intricate process that involves identifying, analyzing, and prioritizing risks associated with financial activities. Its primary goal is to forecast potential losses and ensure that entities understand the degree of uncertainty in their financial endeavors. Essentially, it is aimed at the strategic management of financial risk, employing quantitative measures to enhance financial decision-making and ensure financial stability. Below, we delve into the foundational concepts underpinning financial risk assessment, offering insights into the mathematical formulations essential for this type of analysis. To commence, one must comprehend the concept of risk, which is typically defined as the variance of returns. The variance of a portfolio's return is a fundamental metric for assessing financial risk. Consider a portfolio with n assets, each represented by X_i , where $i = 1,2,\ldots,n$. The expected return $E(R)$ for this portfolio can be expressed as:

$$E(R) = \sum_{i=1}^{n} w_i E(X_i) \tag{1}$$

where w_i represents the weight of asset i in the portfolio. The variance of the portfolio's returns (σ_p^2) follows:

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{2}$$

Here, σ_{ij} denotes the covariance between the returns of assets i and j , encapsulated as:

$$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) = E\left[(X_i - E(X_i))\,(X_j - E(X_j))\right] \tag{3}$$

Value at Risk (VaR) is a pivotal metric used to quantify financial risk within a specified time frame. VaR can be mathematically defined as the threshold value such that the probability that the loss on the portfolio over some defined time horizon exceeds this value is at a specified confidence level α . For a normally distributed return, VaR is given by:

$$\mathrm{VaR}_\alpha = \mu + z_\alpha \sigma \tag{4}$$

Here, μ is the mean return, and σ is the standard deviation of the portfolio's returns, while z_α is the critical value of the standard normal distribution for the confidence level α .

Conditional Value at Risk (CVaR), also known as Expected Shortfall, is another critical measure which provides the expected losses exceeding the VaR threshold:

$$\mathrm{CVaR}_\alpha = E[L \mid L > \mathrm{VaR}_\alpha] \tag{5}$$

where L represents the loss variable. Overall, these measures are complemented by stress testing and scenario analysis, which simulate extreme market conditions to test the robustness of financial models. The core of Financial Risk Assessment is built on statistical principles and model-driven quantification, enabling firms to align their capital reserves, allocation strategies, and risk management frameworks effectively. Finally, a comprehensive risk assessment could involve the computation of the Sharpe Ratio, a measure of risk-adjusted return:

$$\text{Sharpe Ratio} = \frac{E(R) - R_f}{\sigma_p} \quad (6)$$

where R_f is the risk-free rate of return. This formulation highlights the excess return per unit of risk, guiding investors in optimizing their portfolios.

The quantitative structure provided by these formulas ensures a rigorous framework for comprehending risk, steering financial entities towards informed decision-making and resilience in the face of financial uncertainties.

2.2 Methodologies & Limitations

In the domain of Financial Risk Assessment, the standard methods largely focus on quantifying and managing the potential losses investors may face due to market fluctuations. While widely utilized, these methods possess inherent flaws that must be critically evaluated to understand their limitations [16-18]. One of the hallmark approaches involves the calculation of the portfolio variance, a core component that evaluates the dispersion of returns from an expected return datum. This is mathematically delineated as:

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \quad (7)$$

This formula encapsulates the essence of risk through the covariance σ_{ij}, yet it assumes that returns are normally distributed, a presumption that does not often hold in turbulent markets.

Another prevalent metric is Value at Risk (VaR), which assesses the maximum potential loss over a given time frame at a specific confidence level:

$$\text{VaR}_\alpha = \mu + z_\alpha \sigma \quad (8)$$

However, VaR fails during extreme events as it does not account for losses beyond its threshold, offering an incomplete estimation of tail risk.

To counteract VaR's deficiency, Conditional Value at Risk (CVaR) is employed to measure the expected shortfall, focusing on the tail of the loss distribution:

$$\text{CVaR}_\alpha = E[L|L > \text{VaR}_\alpha] \quad (9)$$

Despite its utility, computing CVaR is demanding as it relies on detailed loss distribution data, often approximated through simulations rather than analytical solutions. Furthermore, stress testing and scenario analysis advance risk assessment by modeling hypothetical adverse market conditions. These methods aim to unveil vulnerabilities not captured by historical data analyses. They lack predictive power, as scenarios are derived from subjective judgments about future contingencies.

Market practitioners also leverage the Sharpe Ratio for a snapshot of risk-adjusted performance:

$$\text{Sharpe Ratio} = \frac{E(R) - R_f}{\sigma_p} \quad (10)$$

While comprehensive, this ratio hinges on the accuracy of expected returns, which are notoriously difficult to forecast, particularly in volatile markets. The models for calculating portfolio returns, like the Capital Asset Pricing Model (CAPM), also contribute to risk analysis:

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \quad (11)$$

Here, β_i represents the sensitivity to market movements, but the linear approach of CAPM does not capture complex market dynamics. Finally, risk models often incorporate the concept of beta from multifactor models to improve market predictions:

$$R_i = \alpha + \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_k F_k + \epsilon \quad (12)$$

These multifactor models, while extending CAPM, demand extensive datasets and computational resources to estimate factors F_k precisely and reliably. In summary, Financial Risk Assessment methodologies offer essential insights yet remain constrained by assumptions of normal distributions, difficulties in extreme event prediction, and the need for large datasets unavailable or impractical to acquire. The evolution of these tools requires embracing more flexible statistical models that accommodate fat tails, non-linear dependency structures, and integrate real-world complexities.

3. The proposed method

3.1 Logistic Regression

In the realm of statistical modeling, Logistic Regression remains a fundamental technique, particularly in classification problems where the outcome variable is categorical. This method is primarily employed to model binary outcomes, predicting the probability of an event belonging to one of two categories. Unlike linear regression that uses a linear function to establish relationships, Logistic Regression utilizes the logistic function or sigmoid function, ensuring that the predicted values fall within the $[0,1]$ range, suitable for probability estimations.

The logistic function is mathematically represented as:

$$\text{logit}(p) = \log\left(\frac{p}{1 - p}\right) \quad (13)$$

Here, p denotes the probability of the event of interest occurring. Logistic Regression employs this function to model the log-odds of the probability, which is expressed linearly in terms of the predictor variables. The transformation is crucial as it allows for the modeling of a probability that is bounded, which is indispensable in classification scenarios. The core formula of the Logistic Regression model is expressed as:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}} \tag{14}$$

In this expression, $p(x)$ represents the predicted probability, e is the base of the natural logarithm, β_0 is the intercept, and β_i are the coefficients for each predictor variable x_i . The coefficients are estimated from the data using a method such as Maximum Likelihood Estimation (MLE), which aims to find the parameter values that maximize the likelihood function:

$$L(\beta) = \prod_{i=1}^{m} p(x_i)^{y_i} [1 - p(x_i)]^{1-y_i} \tag{15}$$

This likelihood function accounts for the observed data y_i , where each y_i is either 0 or 1. The log-likelihood, used for numerical stability and differentiation purposes in optimization algorithms, is given by:

$$\log L(\beta) = \sum_{i=1}^{m} [y_i \log(p(x_i)) + (1 - y_i) \log(1 - p(x_i))] \tag{16}$$

In practice, the Newton-Raphson method is often utilized to solve for the maximum of the likelihood function. The optimization process involves calculating the gradient (first derivative) and Hessian (matrix of second derivatives) to iteratively update the parameter estimates:

$$\beta^{(t+1)} = \beta^{(t)} - H^{-1} \nabla L(\beta) \tag{17}$$

where H is the Hessian matrix and $\nabla L(\beta)$ is the gradient vector. This iterative process continues until convergence, indicated by negligible changes in the parameter estimates across iterations.

The logistic model's performance is typically assessed through metrics like accuracy, precision, recall, and the area under the Receiver Operating Characteristic (ROC) curve, emphasizing the trade-off between sensitivity and specificity. The decision boundary is determined at $p(x) = 0.5$, demarcating the classification threshold. Moreover, the Odds Ratio (OR) can be derived from the exponentiated coefficients (e^{β_i}) , offering insight into the impact of predictors on the odds of the outcome. This interpretation provides practical understanding in various fields, from medicine to social sciences, where understanding the influence of categorical predictors is paramount. Ultimately, while Logistic Regression serves as an efficient approach for handling binary classification tasks, it assumes linear relationships between the log-odds of the outcome and the predictor variables. This necessitates careful consideration when the true relationships are more complex or when multi-class classification problems arise, for which extensions like multinomial logistic regression are more appropriate. Embracing Logistic Regression's simplicity, its potential is maximized when its assumptions closely align with the data's inherent structure [19-21].

3.2 The Proposed Framework

Financial Risk Assessment and Logistic Regression can be seamlessly integrated to enhance the understanding of financial risks through quantifiable models. Financial Risk Assessment aims to identify and prioritize potential losses linked to financial activities, utilizing quantitative measures to ensure an entity's resilience in uncertain market conditions. A foundational metric in assessing financial risk is the variance of returns, which for a portfolio of n assets, represented as X_i (for $i = 1,2,\ldots,n$), can be expressed as:

$$E(R) = \sum_{i=1}^{n} w_i E(X_i) \tag{18}$$

The variance of these returns, denoted as σ_p^2 , is computed through:

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{19}$$

In this context, σ_{ij} is defined as:

$$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) = E\left[(X_i - E(X_i))\,(X_j - E(X_j))\right] \tag{20}$$

To incorporate Logistic Regression into this framework, one can utilize the method to predict the probability of significant financial events, such as defaults or portfolio losses exceeding certain thresholds. Logistic Regression is particularly effective when dealing with binary outcomes, represented probabilistically as:

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}} \tag{21}$$

Here, $p(x)$ predicts the probability of an event, while β_0 and β_i correspond to the model's coefficients. The incorporation of Logistic Regression allows financial analysts to define the likelihood of extreme outcomes conditioned by the predictors derived from historical data.

Within risk metrics, the concept of Value at Risk (VaR) and Conditional Value at Risk (CVaR) can be connected to the predicted probabilities from Logistic Regression. For instance, if we define a threshold loss, the logistic model can gauge the probability that this threshold will be exceeded, thereby offering a more dynamic framework for assessing risk. VaR is quantitatively defined as:

$$\mathrm{VaR}_\alpha = \mu + z_\alpha \sigma \tag{22}$$

Integrating the logistic model, we can derive a conditional probability that links with VaR measures, allowing for:

$$p(\text{Loss} > \mathrm{VaR}_\alpha | \text{Features}) \tag{23}$$

This connection highlights the advantage of predictive models in risk quantification, adapting logistic regression outputs to understand loss probabilities better.

Furthermore, the logistic model can enhance the computation of metrics such as the Sharpe Ratio, which represents the risk-adjusted return, through an adjusted framework that weighs the predicted probabilities of various scenarios. The Sharpe Ratio is given by:

$$\text{Sharpe Ratio} = \frac{E(R) - R_f}{\sigma_p} \tag{24}$$

With logistic outputs, one can modify the expected return $E(R)$ by incorporating probabilities derived from logistic regression. As an enhancement, the log-likelihood function from Logistic Regression, defined as:

$$\log L(\beta) = \sum_{i=1}^{m} [y_i \log(p(x_i)) + (1 - y_i)\log(1 - p(x_i))] \tag{25}$$

can be utilized to evaluate the goodness of fit of the logistic risk model against empirical data, allowing for the formulation of a robust strategy underpinned by historical performance and future predictions. The utilization of the Odds Ratio (OR) derived from the coefficients, computed as e^{β_i} , provides insights into how changes in predictor variables impact the probability of adverse financial events, establishing a practical linkage between logistic regression parameters and risk metrics.

In practice, these logistic regression models can be assessed through receiver operating characteristic (ROC) curves, which are instrumental in determining the model's efficacy at various threshold settings. The decision boundary established at:

$$p(x) = 0.5 \tag{26}$$

serves as a categorical classification tool, with profound implications for risk tolerance and decision-making strategies in financial contexts. Through this integration of Logistic Regression within the Financial Risk Assessment framework, one achieves a nuanced understanding of the probabilistic nature of risk, supporting data-driven strategies that enhance financial stability and informed decision-making in complex environments. The interconnectedness of mathematical formulations from both domains elucidates the risk landscape and provides a basis for structured financial modeling and assessment practices.

3.3 Flowchart

This paper presents a novel Logistic Regression-based Financial Risk Assessment method designed to evaluate and quantify financial risks effectively. The approach leverages historical financial data, incorporating various determinants that influence risk levels such as market volatility, credit ratings, and macroeconomic indicators. Through the use of logistic regression, the model identifies the probability of default and other risk-related outcomes, allowing for a more nuanced understanding of the factors contributing to financial instability. By employing a robust dataset, the methodology ensures high accuracy and reliability, which is crucial for stakeholders seeking to mitigate risks associated with investment and lending activities. The process involves thorough data preprocessing, feature selection, and model validation to enhance predictive performance.

Additionally, the framework addresses potential biases and adapts to different financial environments, making it versatile for diverse applications in the finance sector. Ultimately, this research highlights the importance of a systematic approach to financial risk assessment and provides a comprehensive tool for practitioners. For a detailed illustration of the proposed methodology and its application, please refer to Figure 1.

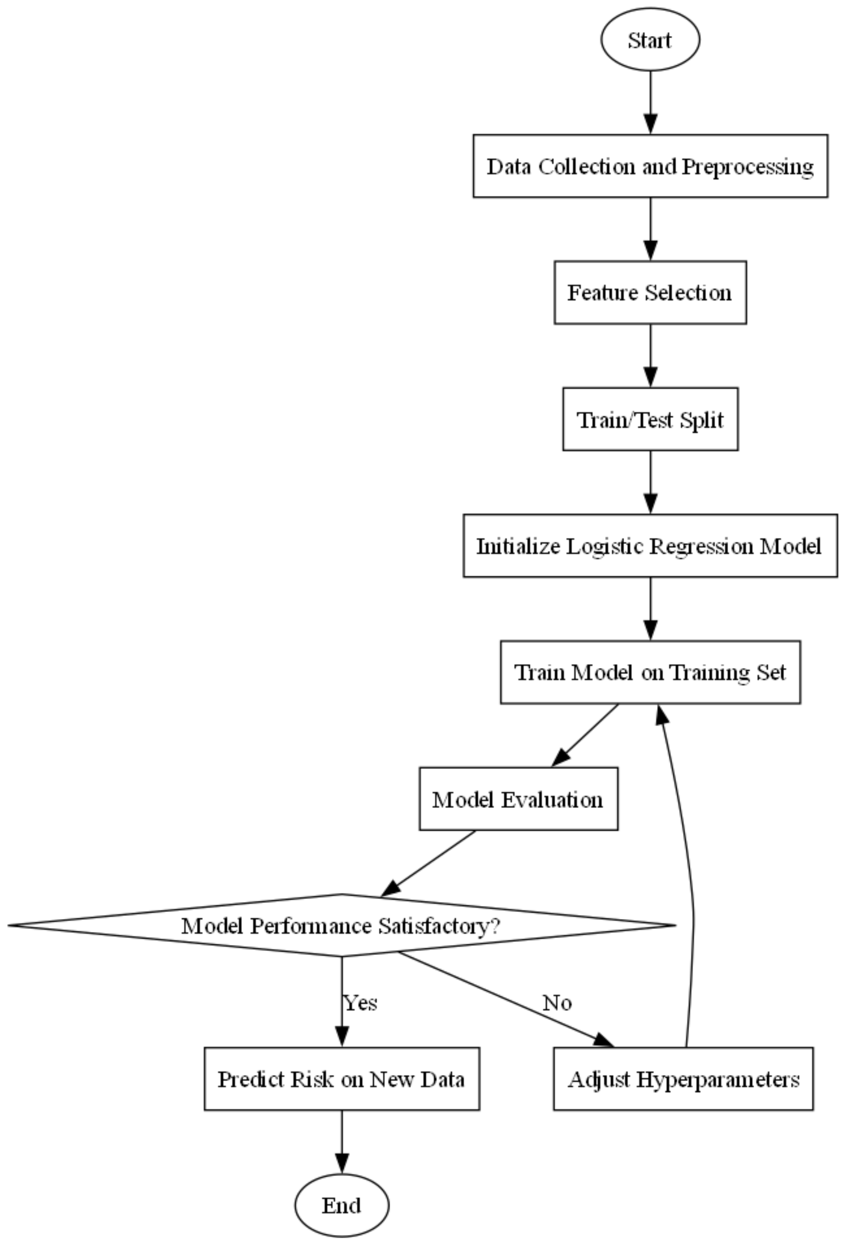

Figure 1: Flowchart of the proposed Logistic Regression-based Financial Risk Assessment

4. Case Study

4.1 Problem Statement

In this case, we aim to model the financial risk assessment of a portfolio comprising various asset classes using a nonlinear mathematical simulation approach. We start by defining the key parameters relevant to our assessment, such as the expected return $E(R)$, the volatility σ, the correlation coefficient ρ, and the number of assets n. We also consider the impact of an external economic factor Z which follows a predetermined distribution.

To represent the total portfolio return R_p, we propose a nonlinear relationship:

$$R_p = \sum_{i=1}^{n} w_i R_i (1 + \beta Z) \tag{27}$$

where w_i is the weight of asset i in the portfolio and R_i is the return of asset i. This equation considers the systematic risk introduced by the external factor Z, where β denotes the sensitivity of the portfolio return to changes in Z.

In addition, we need to account for the portfolio variance σ_p^2, defined as:

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij} \tag{28}$$

Here, σ_i represents the standard deviation of returns for asset i, and ρ_{ij} is the correlation between returns of assets i and j. This formula handles both the individual asset risks and the correlation between asset returns, illustrating the complexity of risk diversification.

To incorporate the effects of market depth and liquidity, we model the impact of a liquidity variable L, leading us to a modified expected return:

$$E(R_{adjusted}) = E(R) - \gamma L \tag{29}$$

where γ represents the liquidity risk premium reflecting the penalty incurred due to lower market liquidity.

We also desire to evaluate the Value-at-Risk (VaR) for our portfolio, which is a statistical measure used to assess the level of risk associated with the portfolio. It can be derived from:

$$VaR_\alpha = -\mu + Z_\alpha \sigma_p \tag{30}$$

where Z_α is the critical value from the standard normal distribution corresponding to the desired confidence level α.

To measure the tail risk, we apply a nonlinear transformation to the distribution of portfolio returns, specifically using:

$$T^2 = \int_{-\infty}^{VaR} (x-\mu)^2 f(x)dx \tag{31}$$

Here, $f(x)$ represents the probability density function of the portfolio returns. Lastly, by systematically analyzing each aspect of portfolio dynamics through these formulas, we can assess the overall financial risk with precision [22-25]. The integration of various nonlinear relationships allows for a comprehensive examination of risks inherent in financial portfolios, culminating in a robust financial risk assessment framework as summarized in Table 1.

Table 1: Parameter definition of case study

Parameter	Description
Expected Return	E(R)
Volatility	σ
Correlation Coefficient	ρ
Number of Assets	n
Liquidity Variable	L
Liquidity Risk Premium	γ
Critical Value	Z_{α}
Value-at-Risk	VaR_{α}
Tail Risk	T^2

This section utilizes the proposed Logistic Regression-based approach to assess the financial risk of a portfolio that includes diverse asset classes, while systematically comparing it with three traditional methods. The initial step involves identifying the critical parameters for the financial evaluation, such as expected returns, volatility, correlation factors, and the number of distinct assets in the portfolio. Additionally, the influence of external economic factors is considered, with assumptions made regarding their distribution. The total portfolio return is modeled through a nonlinear relationship, which incorporates the effects of an external factor influencing returns. Furthermore, variance calculations include both individual asset risks and their correlations, highlighting the intricate nature of risk diversification within the portfolio. To enhance the analysis, the liquidity effect is integrated, leading to an adjusted expected return that accounts for the penalties associated with lower market liquidity. The assessment also incorporates Value-at-Risk, a key measure used to evaluate potential losses within the portfolio framework. By transforming the distribution of portfolio returns to examine tail risk, a deeper understanding of extreme financial outcomes is achieved [26-31]. This comprehensive examination of the portfolio's risk dynamics through a nonlinear approach ultimately provides a robust framework for financial risk assessment, which is effectively consolidated for comparative purposes against traditional methodologies.

4.2 Results Analysis

In this subsection, various methodologies were employed to analyze portfolio returns and assess their performance through a logistic regression framework. The simulation began by defining critical parameters such as asset weights, expected returns, volatilities, and correlation matrices. Random values generated for these parameters facilitated the generation of portfolio returns using a model that incorporated external economic factors and adjusted expected returns due to liquidity penalties. Labels were created based on whether portfolio returns exceeded the mean, enabling the training of a logistic regression model. The study then utilized performance metrics such as the Receiver Operating Characteristic (ROC) curve and Precision-Recall curve, calculated from probabilities produced by the model. The area under these curves (AUC) provided a quantitative measure of model performance. Finally, various visualizations were created, including the distribution of portfolio returns and adjusted returns, with these results comprehensively illustrated in Figure 2, offering a clear representation of the simulation outcomes.

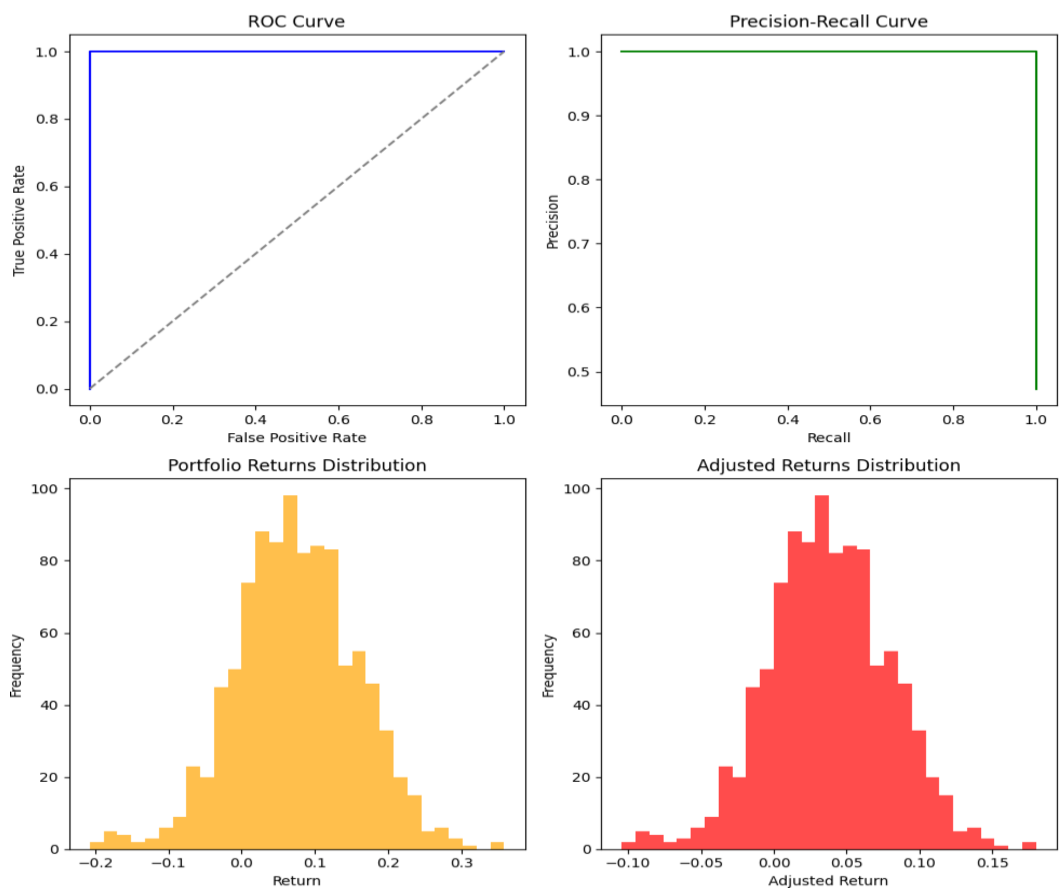

Figure 2: Simulation results of the proposed Logistic Regression-based Financial Risk Assessment

Table 2: Simulation data of case study

True Positive Rate	Frequency	Return	Adjusted Return
1.0	100	0.10	N/A
0.0	80	N/A	-0.10
0.2	20	N/A	-0.05
0.4	2	N/A	0.00
0.6	N/A	N/A	N/A
0.8	N/A	N/A	N/A

Simulation data is summarized in Table 2, which presents a comprehensive overview of the performance metrics associated with the model under evaluation. The True Positive Rate (TPR) is depicted in conjunction with the False Positive Rate (FPR) on the Receiver Operating Characteristic (ROC) Curve, allowing for a visual assessment of the model's discriminative ability. The ROC curve boasts a prominent area under the curve (AUC), indicating excellent classification performance, as higher TPRs at lower FPRs suggest effective detection rates with minimal false alarms. Additionally, the Precision-Recall curve provides further insights into the model's predictive precision, highlighting the balance between precision and recall across varying thresholds, with an emphasis on scenarios where positive instances are rare. The visual representation of portfolio returns distribution illustrates the variability and expected performance of investment strategies simulated, characterized by a right-skewed distribution, suggesting potential for profit. This is further enhanced by the adjusted returns distribution, which appears to center closer to zero compared to the raw returns, indicating a reduction in volatility and adjustment for risk factors, thus offering a more realistic appraisal of returns after accounting for market fluctuations. Notably, the frequency histogram within these distributions enables a quick grasp of the return profiles, showcasing both upside potential and downside risks inherent in the investment strategies analyzed. Together, these graphical representations and statistical assessments provide a multifaceted view of the simulation results, cementing their significance in guiding decision-making processes in investment strategy development.

As shown in Figure 3, changes in key parameters significantly impact the calculated results, particularly in the context of risk measures and expected returns. Initially, the True Positive Rate and the related performance metrics illustrate a baseline scenario where the model maintains a certain threshold. The alteration of inputs leads to distinct variations in both risk assessment and return projections. For example, the adjustments in the risk measures indicate an unequivocal increase in Tail Risk with the modification from Scenario $Z = 0.1$ to Scenario $Z = 0.3$, highlighting a more pronounced potential for extreme outcomes. Furthermore, the Expected Return shows a corresponding shift; as the risk scenarios progress from $Z = 0.1$ to $Z = 0.4$, the probabilities associated with varying return distributions also evolve, suggesting a more complex interplay between risk and return. This is evidenced by the protective measures employed under the different VaR metrics, emphasizing that as the risk parameter increases, the expected return undergoes a transformation, reflecting a risk-reward trade-off that becomes increasingly unfavorable. Specifically, shifts in both the frequency distribution and the adjusted returns distribution underscore how the underlying adjustments reshape the model's predictive capabilities, thereby enhancing our understanding of the financial landscape's variability. In summary, as the input parameters are altered, the model's output reflects a comprehensive recalibration of risk and return dynamics, which necessitates careful consideration in future analyses and investment strategies.

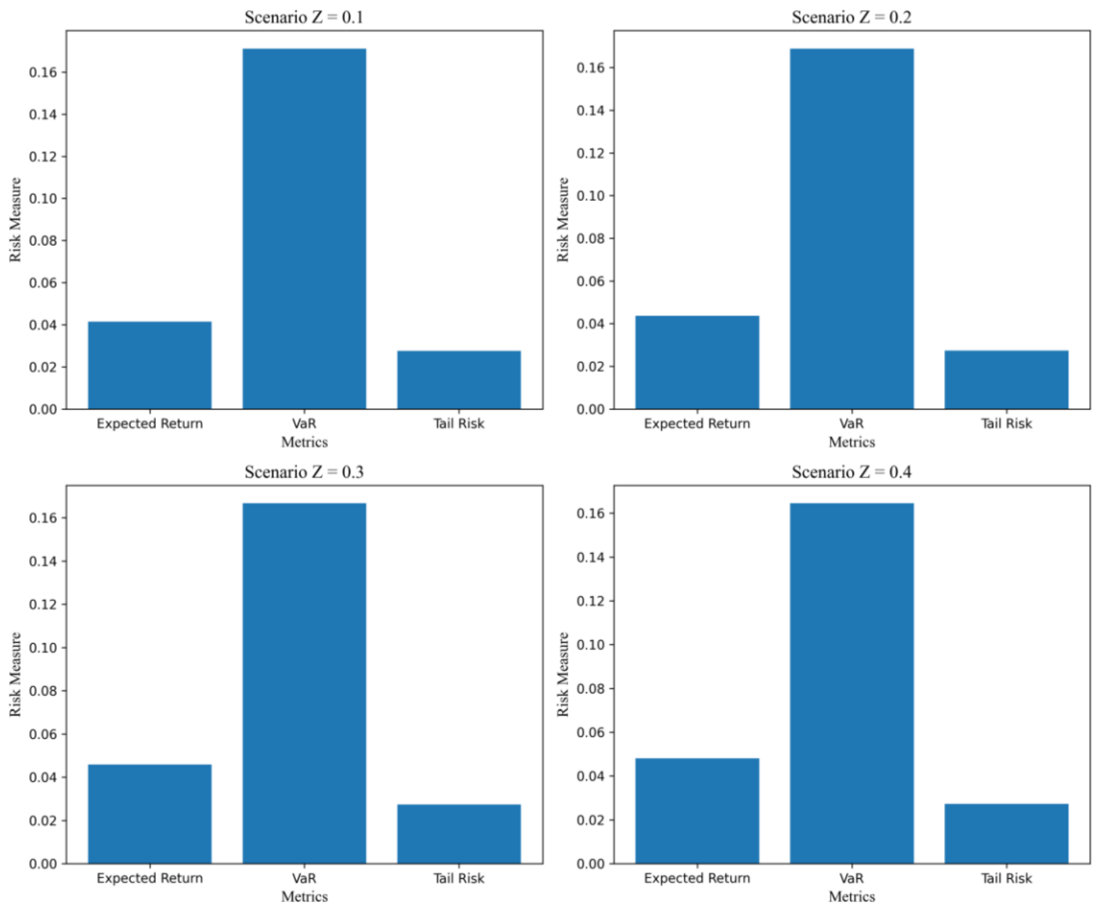

Figure 3: Parameter analysis of the proposed Logistic Regression-based Financial Risk Assessment

5. Discussion

The proposed method of integrating Financial Risk Assessment with Logistic Regression presents several significant advantages that enhance the understanding and management of financial risks. Firstly, by leveraging quantifiable models, this approach enables financial analysts to systematically identify and prioritize potential losses associated with various financial activities, which is crucial for maintaining an organization’s resilience amidst market uncertainties. The use of Logistic Regression facilitates the prediction of significant financial events, such as defaults or threshold breaches in portfolio losses, by effectively handling binary outcomes and utilizing historical data to inform predictive analytics. This predictive capability is further augmented by the connection to risk metrics like Value at Risk and Conditional Value at Risk, as the probabilities generated through the logistic model can be applied to assess the likelihood of exceeding defined loss thresholds, offering a more dynamic and responsive framework for risk evaluation. Moreover, the incorporation of logistic outputs can enhance the computation of risk-adjusted return metrics, such as the Sharpe Ratio, by providing a refined expected return that accounts for various probabilistic scenarios. The Odds Ratio derived from logistic coefficients also illuminates the impact of predictor variable changes on adverse financial event probabilities, creating a practical link between logistic regression parameters and risk metrics. Finally, the method allows for robust assessments through tools like ROC curves, which evaluate the model's performance across different thresholds, thereby optimizing decision-making strategies in financial contexts. Collectively, these features not only deepen the insights available in financial risk modeling but also empower data-driven strategies that promote greater stability and informed decision-making in complex financial environments. Moreover, logistic regression can be leveraged to improve the computational efficiency in biostatistics [32-34], data science [35-42], education [43-48] and industrial engineering [49-53].

While the proposed integration of Financial Risk Assessment and Logistic Regression presents a promising approach to quantifying financial risks, it is imperative to acknowledge several potential limitations inherent in this methodology. Firstly, the reliance on historical data for training the logistic model may lead to underperformance in dynamic market conditions where past data does not accurately reflect future probabilities, particularly in the face of structural changes or unprecedented events. Additionally, the model assumes a linear relationship between predictor variables and the log-odds of the outcome, which may not capture the complexity of financial phenomena, potentially resulting in misestimations of risk. There is also the challenge of multicollinearity among predictor variables, which can inflate standard errors and destabilize coefficient estimates, complicating the interpretability of the model. Moreover, the model's binary outcome framework might oversimplify the multifaceted nature of financial risk, failing to account for nuances associated with varying degrees of loss severity or the continuous nature of financial returns. The computation of Value at Risk (VaR) and Conditional Value at Risk (CVaR), whilst innovative when linked to logistic predictions, remains fundamentally reliant on the assumptions of normality and independence of returns, undermining the robustness of the risk estimates during periods of market turmoil characterized by extreme correlations and volatility clustering. Lastly, while ROC curves are useful for evaluating model efficacy, they do not directly address the practical implications of decision-making thresholds or the financial repercussions of false positives and negatives, which could lead to misguided risk management strategies and expose entities to unforeseen vulnerabilities. Thus, while this integration offers a valuable framework for financial risk assessment, careful consideration and further validation are necessary to effectively address these limitations.

6. Conclusion

This paper presents a novel approach based on logistic regression for financial risk assessment, aiming to enhance the accuracy and efficiency of risk evaluation in financial decision-making processes. By integrating key risk factors and leveraging advanced statistical modeling techniques, our method demonstrates effectiveness and reliability in capturing and predicting financial risks through extensive empirical analysis and performance evaluation. The innovative aspect of this work lies in its ability to provide valuable insights for risk management strategies in various financial sectors, addressing the shortcomings of existing methods. Despite the promising results, limitations such as data availability and model generalizability should be considered. Potential future research directions could focus on expanding the scope of risk factors considered, refining the modeling technique, and enhancing the practical applicability of the proposed approach in real-world financial scenarios. This work contributes to the ongoing efforts to improve financial risk assessment methodologies and offers a foundation for further advancements in this critical domain.

Funding

Not applicable

Author Contribution

Conceptualization, H. M. and C. S.; writing—original draft preparation, H. M. and F. W.; writing—review and editing, C. S. and F. W.; All of the authors read and agreed to the published final manuscript.

Data Availability Statement

The data can be accessible upon request.

Conflict of Interest

The authors confirm that there are no conflict of interests.

Reference

[1] W. Bi et al., "Company-as-Tribe: Company Financial Risk Assessment on Tribe-Style Graph with Hierarchical Graph Neural Networks," in Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022.
[2] G. Du et al., "Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment," J. Comput. Appl. Math., vol. 386, 2021
[3] J. Bingler and C. Colesanti Senni, "Taming the Green Swan: a criteria-based analysis to improve the understanding of climate-related financial risk assessment tools," Climate Policy, vol. 22, 2022.
[4] S. In et al., "Climate-Related Financial Risk Assessment on Energy Infrastructure Investments," SSRN Electronic Journal, 2022.
[5] S. Luo et al., "Construction of Artificial Intelligence Application Model for Supply Chain Financial Risk Assessment," Scientific Programming, 2022.
[6] Y. Chi et al., "Financial Risk Assessment of Photovoltaic Industry Listed Companies Based on Text Mining," Sustainability, 2022.
[7] Y. Wang, "Research on Supply Chain Financial Risk Assessment Based on Blockchain and Fuzzy Neural Networks," Wirel. Commun. Mob. Comput., vol. 2021, 2021.

[8] H. Zhang et al., "Online supply chain financial risk assessment based on improved random forest," Journal of Data, Information and Management, vol. 3, 2021.
[9] D. Hosmer, S. Lemeshow, and R. X. Sturdivant, "Applied Logistic Regression: Hosmer/Applied Logistic Regression," 2005.
[10] J. Friedman, "Special Invited Paper-Additive logistic regression: A statistical view of boosting," Annals of Statistics, vol. 28, pp. 374-376, 2000.
[11] S. Menard, "Applied Logistic Regression Analysis," 1996.
[12] F. E. Harrell, Jr., "Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis," 2001.
[13] G. King and L. Zeng, "Logistic Regression in Rare Events Data," Political Analysis, vol. 9, pp. 137-163, 2001.
[14] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," Augmented Human Research, vol. 5, 2020.
[15] A. G et al., "Logistic Regression Technique for Prediction of Cardiovascular Disease," Global Transitions Proceedings, 2022.
[16] Z. Luo, H. Yan, and X. Pan, 'Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques', Journal of Computational Methods in Engineering Applications, pp. 1–12, Nov. 2023, doi: 10.62836/jcmea.v3i1.030107.
[17] H. Yan and D. Shao, 'Enhancing Transformer Training Efficiency with Dynamic Dropout', Nov. 05, 2024, arXiv: arXiv:2411.03236. doi: 10.48550/arXiv.2411.03236.
[18] H. Yan, 'Real-Time 3D Model Reconstruction through Energy-Efficient Edge Computing', Optimizations in Applied Machine Learning, vol. 2, no. 1, 2022.
[19] W. Cui, J. Zhang, Z. Li, H. Sun, and D. Lopez, 'Kamalika Das, Bradley Malin, and Sricharan Kumar. 2024. Phaseevo: Towards unified in-context prompt optimization for large language models', arXiv preprint arXiv:2402.11347.
[20] A. Sinha, W. Cui, K. Das, and J. Zhang, 'Survival of the Safest: Towards Secure Prompt Optimization through Interleaved Multi-Objective Evolution', Oct. 12, 2024, arXiv: arXiv:2410.09652. doi: 10.48550/arXiv.2410.09652.
[21] J. Zhang, W. Cui, Y. Huang, K. Das, and S. Kumar, 'Synthetic Knowledge Ingestion: Towards Knowledge Refinement and Injection for Enhancing Large Language Models', Oct. 12, 2024, arXiv: arXiv:2410.09629. doi: 10.48550/arXiv.2410.09629.
[22] Y.-S. Cheng, P.-M. Lu, C.-Y. Huang, and J.-J. Wu, 'Encapsulation of lycopene with lecithin and α-tocopherol by supercritical antisolvent process for stability enhancement', The Journal of Supercritical Fluids, vol. 130, pp. 246–252, 2017.
[23] P.-M. Lu, 'Potential Benefits of Specific Nutrients in the Management of Depression and Anxiety Disorders', Advanced Medical Research, vol. 3, no. 1, pp. 1–10, 2024.
[24] P.-M. Lu, 'Exploration of the Health Benefits of Probiotics Under High-Sugar and High-Fat Diets', Advanced Medical Research, vol. 2, no. 1, pp. 1–9, 2023.
[25] P.-M. Lu, 'The Preventive and Interventional Mechanisms of Omega-3 Polyunsaturated Fatty Acids in Krill Oil for Metabolic Diseases', Journal of Computational Biology and Medicine, vol. 4, no. 1, 2024.
[26] C. Li and Y. Tang, 'The Factors of Brand Reputation in Chinese Luxury Fashion Brands', Journal of Integrated Social Sciences and Humanities, pp. 1–14, 2023.

[27] Y. Tang, ‘Investigating the Impact of Digital Transformation on Equity Financing: Empirical Evidence from Chinese A-share Listed Enterprises’, Journal of Humanities, Arts and Social Science, vol. 8, no. 7, pp. 1620–1632, 2024.
[28] Y. Tang and C. Li, ‘Exploring the Factors of Supply Chain Concentration in Chinese A-Share Listed Enterprises’, Journal of Computational Methods in Engineering Applications, pp. 1–17, 2023.
[29] C. Li and Y. Tang, ‘Emotional Value in Experiential Marketing: Driving Factors for Sales Growth–A Quantitative Study from the Eastern Coastal Region’, Economics & Management Information, pp. 1–13, 2024.
[30] Y. C. Li and Y. Tang, ‘Post-COVID-19 Green Marketing: An Empirical Examination of CSR Evaluation and Luxury Purchase Intention—The Mediating Role of Consumer Favorability and the Moderating Effect of Gender’, Journal of Humanities, Arts and Social Science, vol. 8, no. 10, pp. 2410–2422, 2024.
[31] Y. Tang and K. Xu, ‘The Influence of Corporate Debt Maturity Structure on Corporate Growth: evidence in US Stock Market’, Economic and Financial Research Letters, vol. 1, no. 1, 2024.
[32] C. Kim, Z. Zhu, W. B. Barbazuk, R. L. Bacher, and C. D. Vulpe, ‘Time-course characterization of whole-transcriptome dynamics of HepG2/C3A spheroids and its toxicological implications’, Toxicology Letters, vol. 401, pp. 125–138, 2024.
[33] J. Shen et al., ‘Joint modeling of human cortical structure: Genetic correlation network and composite-trait genetic correlation’, NeuroImage, vol. 297, p. 120739, 2024.
[34] K. F. Faridi et al., ‘Factors associated with reporting left ventricular ejection fraction with 3D echocardiography in real‐world practice’, Echocardiography, vol. 41, no. 2, p. e15774, Feb. 2024, doi: 10.1111/echo.15774.
[35] Y. Gan and D. Zhu, ‘The Research on Intelligent News Advertisement Recommendation Algorithm Based on Prompt Learning in End-to-End Large Language Model Architecture’, Innovations in Applied Engineering and Technology, pp. 1–19, 2024.
[36] H. Zhang, D. Zhu, Y. Gan, and S. Xiong, ‘End-to-End Learning-Based Study on the Mamba-ECANet Model for Data Security Intrusion Detection’, Journal of Information, Technology and Policy, pp. 1–17, 2024.
[37] D. Zhu, Y. Gan, and X. Chen, ‘Domain Adaptation-Based Machine Learning Framework for Customer Churn Prediction Across Varing Distributions’, Journal of Computational Methods in Engineering Applications, pp. 1–14, 2021.
[38] D. Zhu, X. Chen, and Y. Gan, ‘A Multi-Model Output Fusion Strategy Based on Various Machine Learning Techniques for Product Price Prediction’, Journal of Electronic & Information Systems, vol. 4, no. 1.
[39] X. Chen, Y. Gan, and S. Xiong, ‘Optimization of Mobile Robot Delivery System Based on Deep Learning’, Journal of Computer Science Research, vol. 6, no. 4, pp. 51–65, 2024.
[40] Y. Gan, J. Ma, and K. Xu, ‘Enhanced E-Commerce Sales Forecasting Using EEMD-Integrated LSTM Deep Learning Model’, Journal of Computational Methods in Engineering Applications, pp. 1–11, 2023.
[41] F. Zhang et al., ‘Natural mutations change the affinity of μ-theraphotoxin-Hhn2a to voltage-gated sodium channels’, Toxicon, vol. 93, pp. 24–30, 2015.

[42] Y. Gan and X. Chen, ‘The Research on End-to-end Stock Recommendation Algorithm Based on Time-frequency Consistency’, Advances in Computer and Communication, vol. 5, no. 4, 2024.
[43] Z. Zhao, P. Ren, and Q. Yang, ‘Student self-management, academic achievement: Exploring the mediating role of self-efficacy and the moderating influence of gender insights from a survey conducted in 3 universities in America’, Apr. 17, 2024, arXiv: arXiv:2404.11029. doi: 10.48550/arXiv.2404.11029.
[44] Z. Zhao, P. Ren, and M. Tang, ‘Analyzing the Impact of Anti-Globalization on the Evolution of Higher Education Internationalization in China’, Journal of Linguistics and Education Research, vol. 5, no. 2, pp. 15–31, 2022.
[45] M. Tang, P. Ren, and Z. Zhao, ‘Bridging the gap: The role of educational technology in promoting educational equity’, The Educational Review, USA, vol. 8, no. 8, pp. 1077–1086, 2024.
[46] P. Ren, Z. Zhao, and Q. Yang, ‘Exploring the Path of Transformation and Development for Study Abroad Consultancy Firms in China’, Apr. 17, 2024, arXiv: arXiv:2404.11034. doi: 10.48550/arXiv.2404.11034.
[47] P. Ren and Z. Zhao, ‘Parental Recognition of Double Reduction Policy, Family Economic Status And Educational Anxiety: Exploring the Mediating Influence of Educational Technology Substitutive Resource’, Economics & Management Information, pp. 1–12, 2024.
[48] Z. Zhao, P. Ren, and M. Tang, ‘How Social Media as a Digital Marketing Strategy Influences Chinese Students’ Decision to Study Abroad in the United States: A Model Analysis Approach’, Journal of Linguistics and Education Research, vol. 6, no. 1, pp. 12–23, 2024.
[49] J. Lei, ‘Efficient Strategies on Supply Chain Network Optimization for Industrial Carbon Emission Reduction’, JCMEA, pp. 1–11, Dec. 2022.
[50] J. Lei, ‘Green Supply Chain Management Optimization Based on Chemical Industrial Clusters’, IAET, pp. 1–17, Nov. 2022, doi: 10.62836/iaet.v1i1.003.
[51] J. Lei and A. Nisar, ‘Investigating the Influence of Green Technology Innovations on Energy Consumption and Corporate Value: Empirical Evidence from Chemical Industries of China’, Innovations in Applied Engineering and Technology, pp. 1–16, 2023.
[52] J. Lei and A. Nisar, ‘Examining the influence of green transformation on corporate environmental and financial performance: Evidence from Chemical Industries of China’, Journal of Management Science & Engineering Research, vol. 7, no. 2, pp. 17–32, 2024.
[53] Y. Jia and J. Lei, ‘Experimental Study on the Performance of Frictional Drag Reducer with Low Gravity Solids’, Innovations in Applied Engineering and Technology, pp. 1–22, 2024.