Scholar Publishing Group
Socio-Economic Statistics Research
https://doi.org/10.38007/SESR.2025.060206
ISSN 2790-2722 Vol. 6, Issue 2: 55-70

Design of Financial Data Sharing Algorithm Based on Differential Privacy and GAN Heterogeneous Optimization under the Federated Learning Framework

Yantao Tao [1,a*]
[1] *School of Artificial Intelligence, Jingchu University of Technology, Jingmen 448000, Hubei, China*
[a]*Email:* taoyantao@jcut.edu.cn
[*]*Corresponding author*

Keywords: Data sharing, Federated Learning optimization, Differential privacy protection, Generative Adversarial network, Heterogeneous communication scheduling

Abstract: Against the backdrop of financial data sharing, although federated learning can prevent the direct exposure of raw data, it still faces challenges such as device performance differences, data distribution imbalance, and privacy leakage risks. This paper proposes an optimization framework that integrates differential privacy and generative adversarial networks. During local training, it transmits effective information by generating alternative samples and feature distributions that comply with privacy constraints. In the communication process, it uses hierarchical scheduling and dynamic detection mechanisms to reduce congestion and invalid waiting. In the model aggregation stage, it dynamically adjusts weights through multi-dimensional contribution metrics to alleviate convergence shift. The experimental results show that this method significantly reduces the communication burden and improves robustness while ensuring accuracy. The final constructed prototype system realizes the complete functions from client management to global scheduling and security control, providing a feasible solution for achieving efficient sharing and collaborative modeling in the financial field under privacy protection.

1. Introduction

During the digital and intelligent transformation of the financial industry, data has gradually become the core driving force for risk control, business innovation and decision optimization. However, financial data must maintain an extremely cautious balance between sharing and protection due to the highly sensitive information it contains, such as transaction behaviors, account characteristics and asset flows. Although traditional technical means, such as encrypted transmission mechanisms, access control policies, and distributed ledger methods, have achieved certain results in reducing the risk of plaintext data exposure and enhancing information security, with the continuous expansion of the scale of collaboration, These methods are often subject to multiple constraints when dealing with cross-agency modeling tasks, such as computational and communication overheads, the complexity of permission coordination, and insufficient system scalability. For this reason, federated learning, a distributed training model that retains the original data locally and only exchanges model parameters, is gradually being regarded as a potential approach to solving the problem of cross-organizational data collaboration.

Although federated learning offers an effective way to avoid centralized data storage, its actual deployment in financial scenarios still faces many challenges: when the computing power of terminal devices varies, the training efficiency is often severely affected; When business characteristics and regional attributes lead to significant differences in data distribution, the model is prone to offset and unstable convergence. When model updates are frequently transmitted over the network, the potential risk of information leakage may further weaken its security and credibility. Based on this series of issues, this paper proposes an optimization framework that deeply combines the differential privacy mechanism with the generative adversarial network during the research process. It also introduces hierarchical scheduling and multi-dimensional weighting strategies for heterogeneous conditions in the specific mechanisms of communication and aggregation. Thus, while improving the model's accuracy and convergence speed, it achieves enhanced privacy protection and improved system robustness.

In the proposed framework, the client generates synthetic samples that can replace real data by introducing a generative model under differential constraints, thereby completing feature representation and model update without disclosing the original privacy. At the communication level, the system rationally allocates bandwidth and computing resources through a hierarchical scheduling mechanism and dynamic detection strategies, thereby avoiding resource waste while enhancing overall collaborative efficiency. In the aggregation stage, the framework dynamically adjusts the aggregation weights of each client by constructing a multi-dimensional contribution measurement system, enabling the global model to maintain faster convergence and reduce offsets under non-independent and iden-distributed conditions.

Based on this framework, this paper further designs and implements a prototype system for financial business scenarios. The system has completed functional integration in multiple links such as client management, task scheduling and security control. Through a series of experiments, the significant effects of the method in maintaining model accuracy, reducing communication load and alleviating privacy risks have been verified. This research provides a feasible path for the sharing and joint modeling of financial data under compliance constraints and lays the foundation for the exploration of the application of distributed intelligent algorithms in sensitive industries.

2. Relevant research

2.1 Data Security Sharing

As data gradually becomes a core production factor in various industries, researchers have proposed multiple technical solutions for the secure sharing of sensitive information. However, there are still limitations in scenarios with extremely high privacy and compliance requirements, such as finance. The traditional permission-based access control has gradually evolved into an attribute-driven fine-grained authorization mechanism [1]-[2], and combined with cryptographic means to enhance security and tamper-resistance. Lattice-based encryption and homomorphic computing have strengthened resistance to potential quantum threats, but it is difficult to balance efficiency and computational overhead [3]. The solution in the cloud storage environment relies on outsourcing computing and revocation mechanisms to alleviate the pressure on terminal computing power, but it is prone to increase communication costs [4]; Blockchain, due to its decentralized and immutable characteristics, is used to achieve access auditing and shared traceability. Some systems combine attribute-based encryption with smart contracts to enhance flexibility [5], but they still mainly rely on direct exchange of plaintext or ciphertext, making it difficult to meet the demands of financial applications where competition and cooperation coexist.

Against this backdrop, distributed modeling has emerged as a new research approach. It achieves cross-agent collaboration by completing training locally and uploading parameters or gradients, and reduces the risk of information leakage by integrating differential privacy, secure multi-party computation, and homomorphic encryption. It has already demonstrated feasibility in fields such as healthcare, energy, and the Internet of Things [6]-[9] However, it still faces challenges in terms of communication efficiency, computing power differences, uneven data distribution and model bias. To alleviate these issues, existing studies have attempted to enhance the robustness of global models through hierarchical communication protocols, task allocation, weight adjustment, and continuous learning mechanisms. At the same time, an alternative data strategy combining generative models and differential privacy has been introduced to improve the training effect while protecting privacy [10].

Existing research has provided ideas from multiple directions such as access control, cloud outsourcing, blockchain, and federated learning. However, a single technology is difficult to balance privacy, efficiency, and compliance requirements. Therefore, in the future, it is urgent to explore collaborative strategies such as multi-mechanism integration, heterogeneous optimization, and fair aggregation to build practical and sustainable solutions for the secure sharing of financial data.

2.2 Federated Learning

In the research field of data security and sharing control, scholars have long explored how to achieve cross-subject data utilization while ensuring privacy. From the early access control to the attribute-based authorization framework and then to the multi-dimensional protection mechanism combined with cryptographic operations [11]-[12], research has continuously driven technological evolution. To address complex requirements such as access cancellation, policy adjustment, and encrypted retrieval, researchers have introduced lattice-based security models, homomorphic computing methods, and structured indexing mechanisms. These schemes theoretically enhance security protection capabilities, but in practical applications, they are often constrained by issues such as excessive computational load, excessive communication consumption, and increased management complexity [13]. Meanwhile, in the combination of cloud and edge scenarios, researchers attempt to share the computing overhead through trusted proxy nodes and maintain the correctness under dynamic policy changes through forward and backward security guarantees [14]; In the application direction of distributed ledgers, access behaviors are mapped to verifiable records, and transparency and collaboration efficiency are enhanced through automated execution rules. However, when interactions are frequent or concurrency is too high, the system often faces challenges in terms of latency and throughput.

With the proposal of distributed intelligent training frameworks, the data utilization approach of cross-domain collaboration has gradually shifted to "parameter aggregation" rather than "data centralization". This approach not only reduces the risk of direct data exposure but also retains the value of multi-source information. Based on the different situations of various participants, the research has constructed diverse collaboration models to adapt to the complex scenarios of sample size and feature difference distribution [15]; Meanwhile, improvements in algorithms and systems are emerging one after another. Scholars have proposed methods such as compressed update, asynchronous transfer, hierarchical scheduling, and cluster screening to alleviate the problems caused by communication pressure and heterogeneity, and to reduce the bias caused by non-independent data distribution by introducing dynamic weight distribution and continuous optimization strategies [16]. In terms of privacy protection, the academic community has proposed multiple protection paths around differential constraints, encryption operations, and secure collaborative computing, etc., to resist inference attacks, gradient restoration, and model contamination [17]; Some studies have also attempted to utilize methods such as generative models and knowledge distillation to provide the information support needed for training without disclosing the original data [18].

The current trend is moving towards integrating multiple protection mechanisms and optimization strategies, that is, reducing transmission and computing overheads through synthetic data and compression methods under limited privacy budgets [19]-[20], mitigating performance degradation caused by statistical differences through contribution-aware aggregation methods, and improving overall coordination among different computing power nodes with flexible scheduling strategies. This enables the system to demonstrate higher robustness and practicality in complex environments.

3. Research on Personalized Federated Intelligent Modeling Methods in Financial Scenarios

3.1 Data Representation and Knowledge Distillation Methods Oriented to Privacy Constraints

With the in-depth penetration of inclusive financial services, financial terminal devices have been widely applied at all levels of society [21-30]. While this trend improves business processing efficiency, it also brings multi-dimensional risks. Once local data is tampered with, it may become a tool for fraud. Once private information is leaked due to improper transmission or poor management, it will directly threaten users' funds and identity security. When the terminal is subject to external intrusion or internal abuse, it is more likely to cause large-scale leakage of sensitive information and induce systemic risks. Against this backdrop, federated learning is regarded as an important path to coordinate modeling efficiency and information protection by replacing the original data transmission with parameter sharing. However, in financial scenarios, it is still difficult to completely avoid potential risks. That is, if the terminal is maliciously attacked, the locally stored user data may still be lost. And if the attacker uses the model update process to implement reverse reasoning, The original records may also be partially reconstructed in a high-dimensional parameter space, thereby weakening the expected effect of privacy protection. Given that financial terminals often undertake the dual tasks of user interaction and institutional business processing simultaneously, the content they store and transmit involves identity information, transaction details and fund flows. Once lost, it will have a direct impact on customer rights and interests and the compliant operation of institutions. Therefore, in the design of federated learning, How to balance the security of end-side data and the stability of communication links has become a core problem that needs to be urgently solved.

In response to the above problems, this paper proposes a privacy-distilled federated modeling method for financial applications. The core idea lies in introducing a latent feature extraction mechanism based on variational autoencoders on the client side, enabling the local to learn the implicit distributed representation of the data, while the server builds a global representator relying on the distributed parameters uploaded by all parties. Thus, unified feature learning can be achieved under the premise of decentralized data storage.

The client generates optimized distillation samples based on the dual constraints of local distribution and global distribution, and uses them for subsequent training on the server side. This mechanism not only significantly reduces the communication frequency and transmission scale, but also effectively reduces the risk of restoring the original information through parameter reverse inference. Because the variational autoencoder can introduce probabilistic modeling in the latent space, thereby enhancing the diversity and stability of the generated samples, the global representator can capture the statistical patterns across clients more comprehensively after aggregation. Meanwhile, the distilled data maintains a balance between local semantic accuracy and overall pattern universality, thereby avoiding severe model shifts under non-independent homogeneous distribution conditions. As a result, this scheme reduces communication consumption and prevents privacy leakage while maintaining modeling performance, providing a path that is both secure and feasible for financial institutions to carry out multi-party collaboration under the premise of meeting regulatory requirements.

This study proposes a sample refinement method that combines differential privacy constraints with adversarial generation strategies for financial applications. The method starts with edge-side latent representation learning, characterizes the implicit distribution of sensitive records by locally training encoders with probabilistic modeling capabilities in each participant, and performs dense-aware clustering locally to select representative candidate samples. The study then performs bidirectional distribution comparison on these candidate samples and iteratively optimizes them with the help of the generator - discriminator structure to produce refined samples that retain local semantics while being compatible with overall statistical characteristics. The algorithm allows controlled perturbations to be injected into latent variables or synthetic outputs during the generation process to meet differential privacy requirements, and uses the discriminative network as a quality control unit to enhance the diversity and discriminability of synthetic samples, thereby significantly reducing the risk of restoring the original record through parameter reverse inference while ensuring usability. The server side adopts a weighted aggregation strategy to construct the training set after only receiving refined samples that have been privacy-protected and representationally compressed along with necessary lightweight characterization parameters, and completes the model training on centralized computing power. This study significantly reduces communication rounds and data transmission volume by synchronously replacing large-scale models with a limited number of sample and representation reports. This reduces the attack surface and enhances the robustness and energy efficiency of the overall system in a deployment environment with heterogeneous resources.

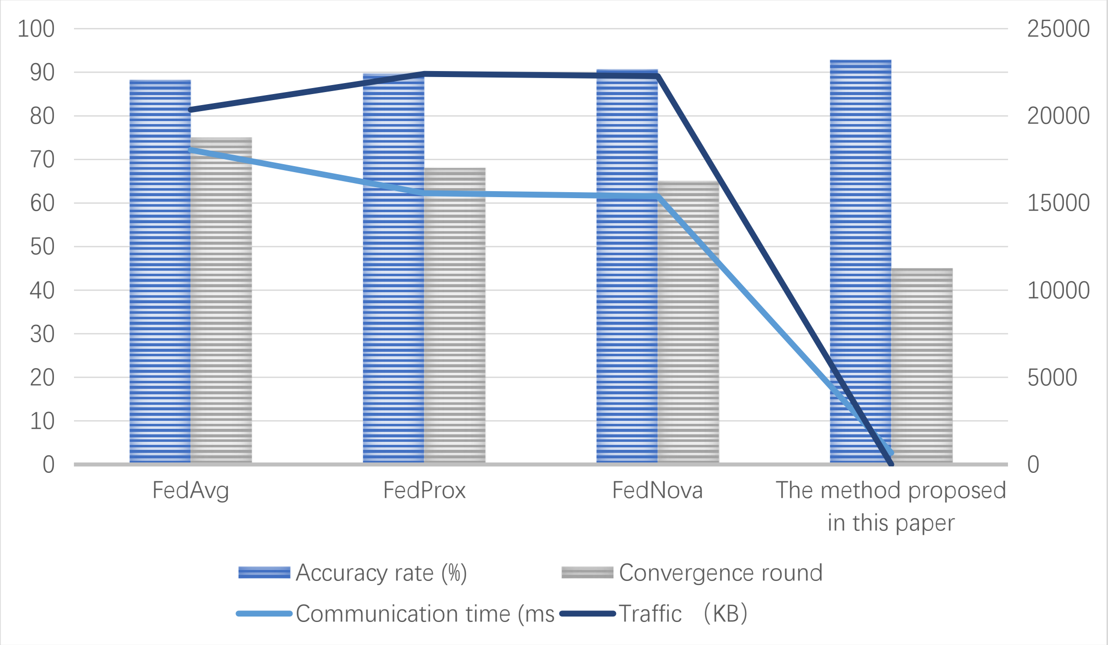

Figure 1. Comparison of credit data model performance and communication costs

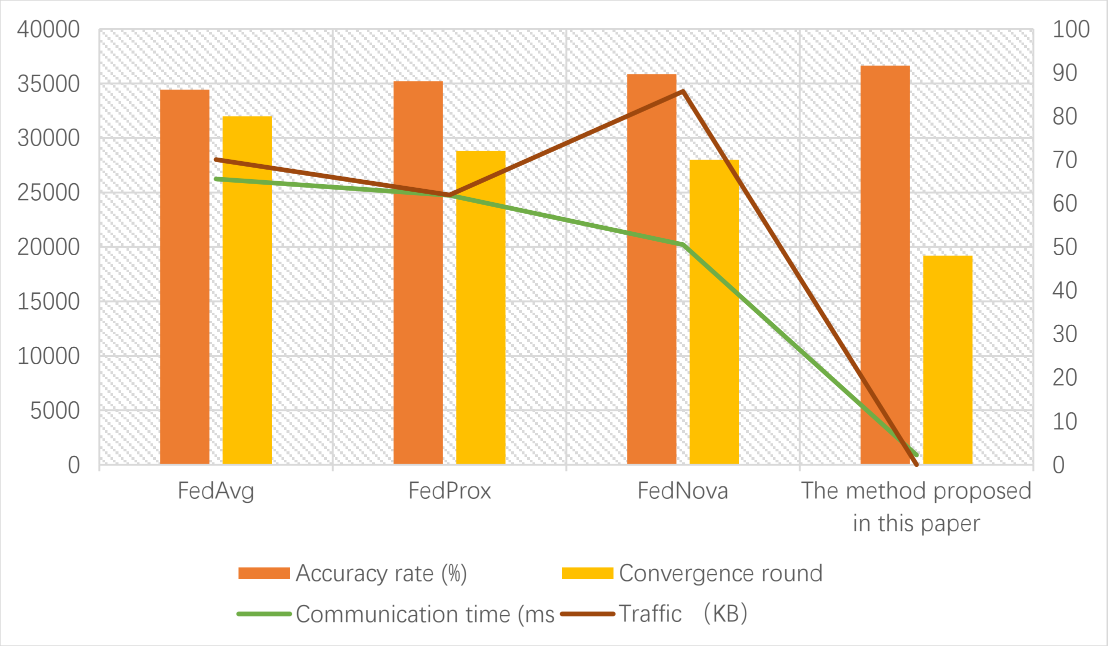

Figure 2. Comparison of the performance of the insurance claims image data model and communication costs

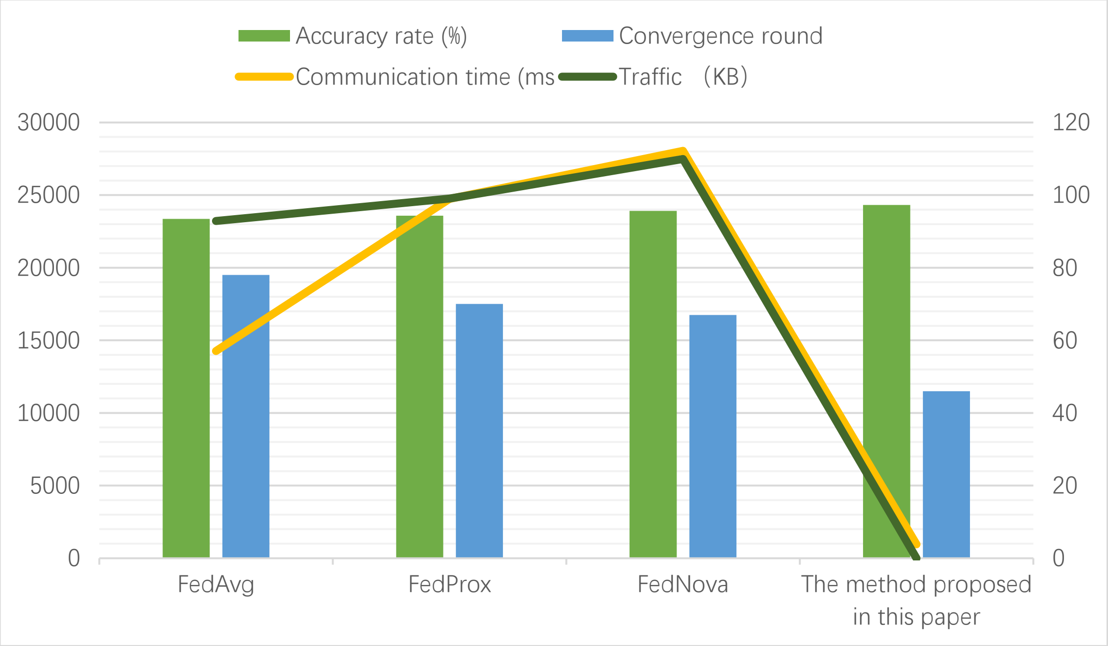

Figure 3. Comparison of the performance and communication cost of the enterprise bankruptcy data model

The comparative evaluation carried out on a unified experimental platform covers three types of tasks: credit default discrimination, auto insurance claim identification, and enterprise operation risk prediction. The study selects a multi-layer perception structure for tabular tasks and uses a convolutional backbone for image-based tasks to verify the applicability and stability of the method. The experimental results in Figures 1, 2, and 3 show that the method proposed in this paper can maintain performance similar to that of centralized training in most scenarios while outperforming several common federated baseline algorithms. In terms of communication overhead, it achieves an order of magnitude reduction. This improvement is mainly attributed to the replacement of frequent model parameter exchanges with a limited number of refined sample transmissions and one-time reporting of representation parameters. Researchers also found that by adjusting the noise intensity of differential privacy and the sample compression strategy, a controllable trade-off can be achieved between model accuracy and privacy protection. Moreover, empirical evidence shows that this scheme can accelerate convergence and reduce the computational burden on the terminal, thereby providing a practical and feasible technical path for regulated financial institutions to promote cross-organizational collaborative modeling while maintaining compliance.

3.2 Hierarchical Scheduling and Communication Optimization Strategies for Heterogeneous Terminals

In an environment where financial terminals are widely distributed and computing capabilities vary significantly, if the traditional communication mechanism that assumes all clients have consistent performance and remain online at all times is still adopted, it will lead to a significant delay in the overall training progress of the central server while waiting for a few low-performance or temporarily failed nodes to complete model updates. Therefore, this paper proposes a communication strategy based on hierarchical transmission channels and reservation application mechanisms. In this strategy, the central end is responsible for setting transmission Windows with different priorities and resource limitations. Meanwhile, the client side dynamically generates upload applications based on its own computing power, storage status and network conditions, and completes the transmission of the local model at the agreed time. So as to achieve flexible scheduling and full utilization of various performance terminals. During each round of aggregation, the central server not only screens, sorts and allocates all submitted reservation applications, but also dynamically activates standby nodes to replace failed nodes based on their active status and historical performance, to prevent the training process from being blocked by individual nodes. At the same time, through the dispersion of upload time and hierarchical channel design, the bandwidth occupation is effectively balanced. It has reduced the risk of communication congestion and significantly improved the efficiency and stability of global model aggregation.

To further reduce the risk of training interruption caused by sudden terminal offline, network fluctuations or hardware anomalies, this paper introduces multi-stage dynamic state detection in the communication mechanism. The central server will conduct real-time detection of node status before task initialization, appointment application reception, and channel allocation at all levels. To ensure that all clients participating in this round of aggregation are available and responsive, and to adjust the scheduling plan and aggregation sequence in a timely manner based on the detection results, thereby enhancing the accuracy of node selection and the reliability of the communication process. During the scheduling process, the central server comprehensively considers the computing power potential of each node, historical training contributions, and the current network throughput capacity. It allocates more training rounds and data samples for high-performance devices to enhance the accuracy and representativeness of the global model, while retaining reasonable participation opportunities for low-performance or restricted devices. To enable it to complete the upload task of model updates under limited resources, thereby achieving efficient hierarchical scheduling and stable communication management of heterogeneous financial terminals while taking into account efficiency, data integrity and security, and providing an implementer technical path for cross-institutional financial data sharing and collaborative modeling.

In the federated learning scenario of financial data sharing, there are significant differences in computing performance and communication performance among different client terminals. This heterogeneity will directly affect the training efficiency and accuracy of the global model. To solve this problem, this paper designs a hierarchical scheduling strategy, dividing the terminals into three levels - high, medium and low - based on computing power and communication capacity, and dynamically adjusting the local training rounds and communication time Windows according to the levels. In actual experiments, the resource usage of each client is restricted through the Python environment. Low-performance devices are limited by CPU and memory, and communication-restricted devices introduce delays during the model upload process to simulate the heterogeneous characteristics of real financial terminals. Meanwhile, the central server collects the status information of each terminal through dynamic reservation and multiple heartbeat check mechanisms, achieving dynamic allocation of communication time Windows. This effectively reduces server waiting time and prevents low-performance terminals from blocking the training process.

The experimental results in Table 1 and Figure 4 show that the DHFMFLM mechanism proposed in this paper demonstrates excellent model performance and communication efficiency in scenarios where communication performance is heterogeneous, computing performance is heterogeneous, or both computing and communication are heterogeneous. Specifically, it is manifested as follows: The accuracy of DHFMFLM on different datasets is higher than that of FedAvg, q-FedAvg and FedAsync mechanisms, while the communication time and communication volume have been significantly optimized. Even in the most complex heterogeneous environment, its aggregation success rate remains above 96%, significantly outperforming other communication mechanisms. This indicates that by integrating hierarchical scheduling, dynamic training rounds, and the scheduled heartbeat mechanism, DHFMFLM can effectively enhance model accuracy, reduce communication latency, and increase task completion rates when dealing with heterogeneous terminals, providing an efficient and stable solution for financial data sharing.

Table 1 Comparison of Model Accuracy in Different Datasets

Metric / Dataset	Lending Club	Vehicle Insurance Fraud	Taiwanese Bankruptcy Prediction
FedAvg Accuracy (%)	87.95	89.02	95.48
q-FedAvg Accuracy (%)	91.00	90.05	96.36
FedAsync Accuracy (%)	94.34	92.22	97.38
DHFMFLM Accuracy (%)	96.15	92.79	96.97

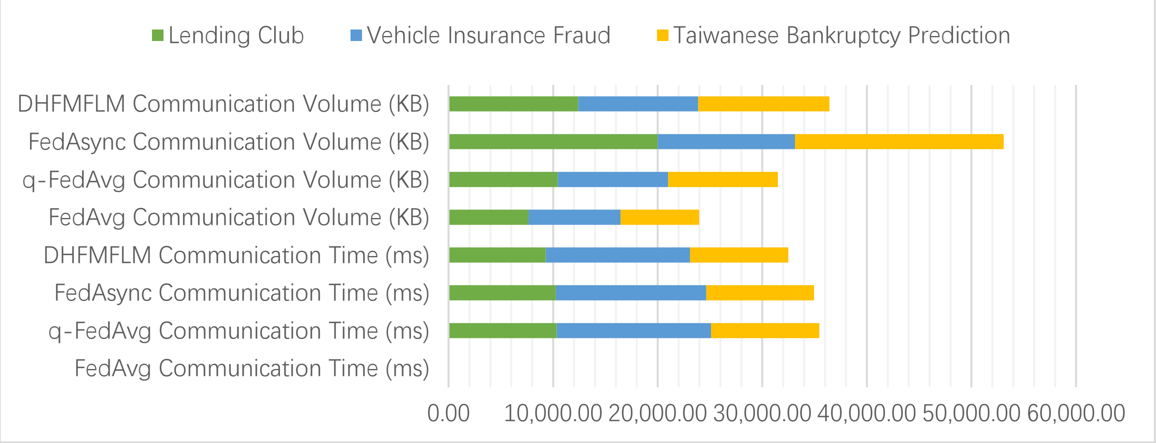

Figure 4 Comparison of different datasets and metrics

As can be seen from the comprehensive analysis in Figure 5, the DHFMFLM mechanism can provide high-precision models and low communication overhead in all heterogeneous scenarios.

Especially when both communication and computing are restricted, its aggregation success rate still remains above 96.94%, fully demonstrating the effectiveness of the dynamic hierarchical scheduling and communication optimization strategy in heterogeneous terminal environments.

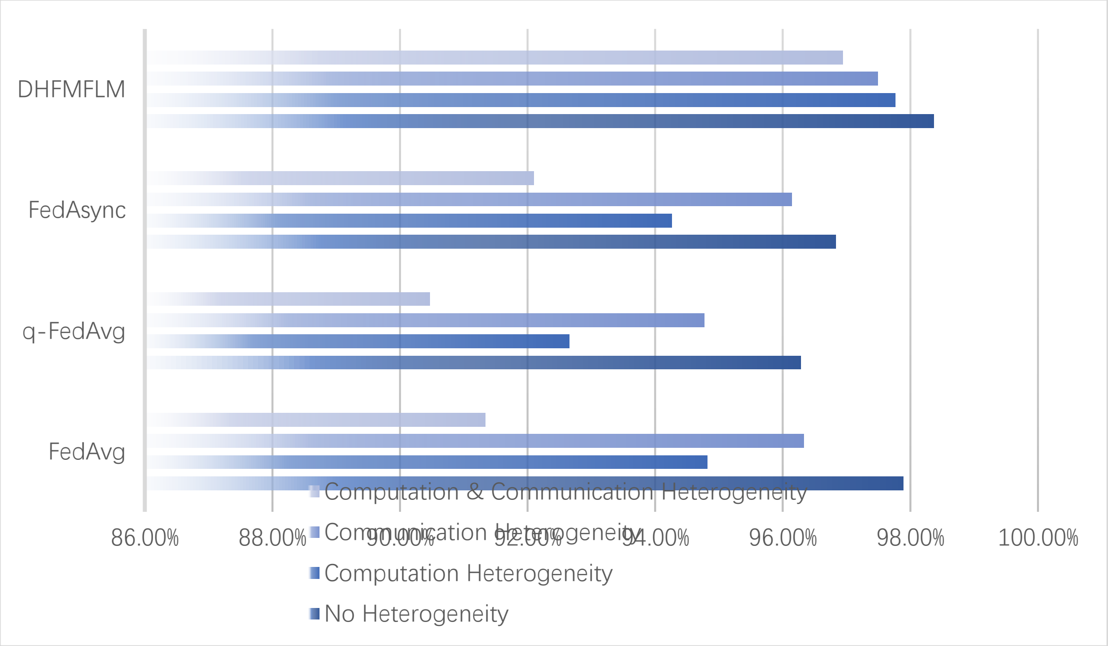

Figure 5 shows the success rate of global model aggregation under heterogeneous conditions of different devices

Through this strategy, not only has the stability and efficiency of federated learning in financial data sharing been enhanced, but also a foundation has been laid for subsequent heterogeneous optimization schemes that combine differential privacy and generative adversarial networks. It provides a practical and feasible method for financial institutions to achieve efficient model training while ensuring data security and privacy.

3.3 Multi-dimensional Evaluation and aggregation Modeling Mechanism for Distribution Differences

In the scenario of financial data sharing, federated learning frameworks are confronted with significant client heterogeneity issues. This heterogeneity is not only reflected in the differences in the data volume of each client, but also in the imbalance of feature distribution and the differences in the proportion of label categories. These factors jointly lead to the possibility that each client may form a biased model based on its own data during the local training process. This thereby affects the generalization ability and stability of the entire global model. To address this challenge, this study proposes a multi-dimensional contribution evaluation and intelligent aggregation mechanism. This mechanism achieves the robust construction of a global model in a highly heterogeneous environment by integrating local continuous learning strategies, personalized weight distribution methods, and generative model enhancement techniques, while ensuring the secure sharing of data among various clients.

During the local training stage, each client not only updates the model based on its own data, but also introduces a historical global model as a training constraint to prevent the model from overly relying on local data during the training process. At the same time, through a continuous learning strategy, the client can retain the memory of previous global information while constantly accepting new data This effectively reduces the risk of local models developing biases during training. During the training process, the client can not only fully utilize the unique features of local data for model optimization through this continuous learning approach, but also take into account the integration of global knowledge, making the model parameters provided by each client in the subsequent aggregation process more stable and reliable.

During the aggregation stage, the server established a multi-dimensional evaluation system for each client's contribution. This system comprehensively considered the volume of data, label balance, model performance, and the rationality of parameter update amplitude, thereby being able to dynamically adjust the aggregation weight of each client to effectively suppress the negative impact of a small amount of data anomalies or potential malicious models on the global model. For clients with relatively small amounts of data or sparse label distribution, the system enhances their contribution to global training by introducing generative models to generate auxiliary samples, thereby ensuring that the performance of the global model on diverse financial data is both balanced and robust.

The implementation of this mechanism has brought about significant advantages: While alleviating the deviation caused by heterogeneous data on the global model, the system enables the model to demonstrate stronger adaptability and generalization ability in different business scenarios. Moreover, through multi-dimensional contribution evaluation and generative model enhancement methods, it optimizes the performance and fairness of the global model while protecting the privacy of client data. This thus provides a practical and feasible technical solution for secure and efficient data collaboration among financial institutions.

In the scenario of cross-institutional sharing of financial data, the data itself exhibits obvious heterogeneity, especially in terms of user behavior characteristics, credit records, and transaction patterns, showing non-independent and co-distributed features. This difference poses a severe challenge to the model training and global aggregation performance of traditional federated learning methods. To address this issue, this study proposes a heterogeneous optimization strategy that integrates differential privacy protection with generative adversarial networks (Gans) for sample generation. This strategy achieves adaptive weight distribution in the global aggregation stage by conducting multi-dimensional quantitative evaluations of the model contributions of each client, thereby enhancing the convergence speed and prediction accuracy of the model under distribution difference conditions.

In the experimental design, this study constructed a data partitioning scheme based on the number of different clients, covering both independent and identic distribution as well as various non-independent and identic distribution scenarios. The division method of non-independent homogeneous distribution takes into account factors such as label distribution skew and uneven sample size, and uses the Dirichlet distribution to generate multiple data distribution combinations to simulate the data heterogeneity existing in real financial business scenarios. During the local training process, each client updates the local model parameters through a continuous training strategy. At the same time, differential privacy constraints are combined to ensure the confidentiality of the original data. The GAN generator is used to supplement the samples of sparse categories, enabling the local model to achieve better generalization ability on categories with fewer samples, thereby improving the overall training effect.

In the global aggregation stage, this study constructed a multi-dimensional contribution evaluation system to quantify the actual contribution of each client model to the global performance improvement. The evaluation metrics include the marginal impact of each client on the decline of the global loss function, the number and diversity indicators of local training samples on the client, as well as the measurement of the consistency between the update direction of local model parameters and the global model. Based on these indicators, this study assigns adaptive aggregation weights to each client, enabling the global model to dynamically adjust according to data distribution and client model characteristics, thereby alleviating the problem of model aggregation performance degradation caused by data distribution skew and sample size differences.

Figure 6 shows the performance comparison results of this method with the traditional FedAvg, FedProx and FedNova algorithms under different client numbers and data distribution conditions. The experimental results show that the method proposed in this study achieves higher accuracy in the initial aggregation round. Meanwhile, the performance advantage of the aggregated model is the most significant under non-independent homogeneous distribution conditions. This fully validates the effectiveness of the multi-dimensional evaluation mechanism and heterogeneous optimization strategy in accelerating model convergence and improving prediction accuracy.

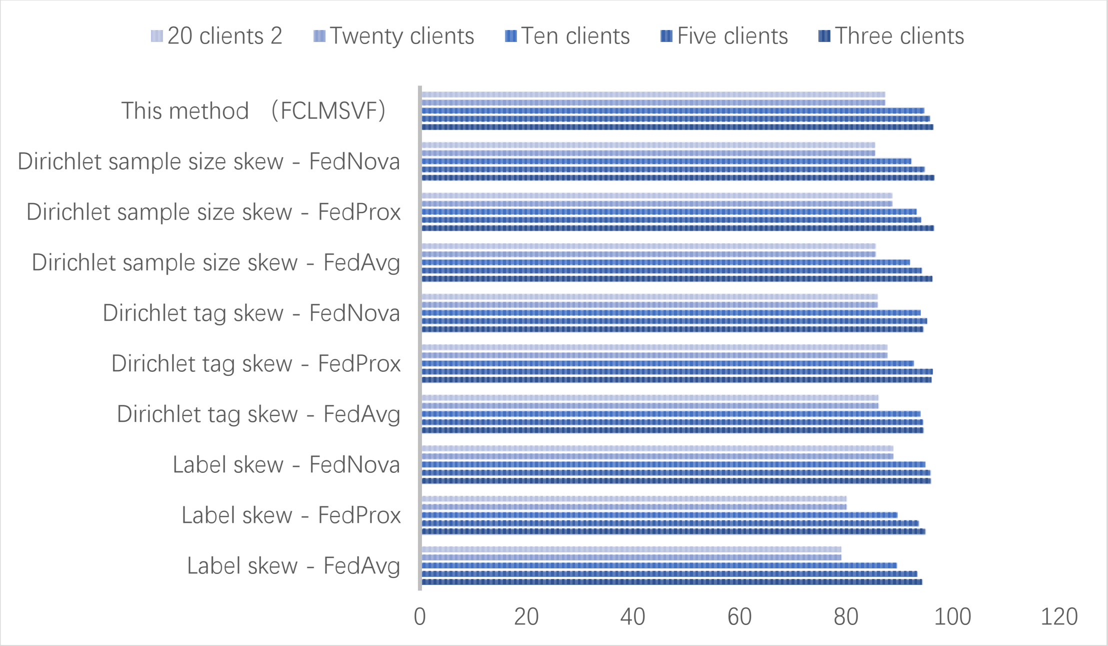

Figure 6 Performance comparison of aggregation algorithms under multi-client and multi-distribution conditions

4. Design and Implementation of Personalized Federated Learning System for Financial Data

As a data collaborative intelligent training platform for the financial industry, the core objective of the financial data personalized federated learning system is to achieve cross-institutional data collaboration and model optimization through distributed learning technology. At the same time, under the premise of strictly ensuring information privacy and security, it aims to break the data isolation among various financial institutions. This significantly enhances the adaptability and predictive ability of artificial intelligence models in complex financial scenarios. This system performs local data processing and model training within each institution, and sends the generated model parameters or gradient information to the central server for unified aggregation, so as to form a global model and continuously optimize algorithm performance. During this process, it effectively prevents the leakage of sensitive information, enabling institutions to achieve joint learning and collaborative innovation without sharing the original data. The central server is responsible for communication management, task scheduling, global control, and model fusion strategy formulation for participating institutions. It can flexibly adjust the training process and aggregation strategy according to different business requirements, thereby ensuring the efficient operation and stability of the entire system. The client is responsible for local data storage, model training and update, computing resource allocation, and secure communication verification, etc. It can independently configure training strategies and parameters based on the computing power and business characteristics of the institution, achieving personalized and efficient model training.

In terms of technical implementation, the system mainly uses Python and Java as development tools, combines the MVC architecture and multiple mature frameworks to support the construction of front-end and back-end as well as federated learning algorithms. At the same time, it adopts high-performance relational databases and in-memory storage solutions to meet the requirements of large-scale data processing and rapid system response. This ensures the stability and reliability of the system's operation in financial business. During the design stage, the system fully considers economic input and maintenance costs. It achieves efficient development by leveraging existing hardware and open-source tools, while ensuring flexible adaptability and sustainable development potential during business expansion. This enables financial institutions to rely on this platform to complete complex data analysis tasks, improve decision-making efficiency, and reduce operational risks. And throughout the entire process, comply with national information security regulations and financial data compliance requirements to ensure the legal use and management of data. The system functionally covers multiple links such as global communication coordination, task scheduling, model training and aggregation management, client management, and security auditing. It can adapt to the needs of different institutions and business scenarios through flexible parameter configuration and policy setting, ensuring that all participants share the model results under the premise of protecting information security At the non-functional level, it emphasizes operational safety, user-friendly interface, strong maintainability, module scalability and cross-platform compatibility, thereby providing stable, efficient and compliant technical support for financial data-driven intelligent applications.

The development of the personalized federated learning system for financial data follows a systematic design approach. The entire process from requirement analysis to system implementation is strictly managed and advanced in accordance with the principles of modularization and gradual refinement. The research is conducted through precise task division, scientific resource allocation, and reasonable arrangement of stage goals. Ensure that development risks can be effectively controlled and overall efficiency can be improved in every link, including system planning, requirement analysis, architecture design, functional implementation and operation and maintenance. The selection of the development environment fully considers cross-platform compatibility and technical scalability. Both front-end and back-end development are carried out in Visual Studio Code, and efficient coding and debugging are achieved with its multilingual support and plugin ecosystem. Tomcat is selected as the application server to undertake the task of running Web applications. At the same time, it ensures the flexibility of deployment and the optimal utilization of resources. The MySQL database not only provides high-performance data storage and management capabilities, but also is compatible with multiple operating system environments and supports complex query operations and transaction processing. The overall architecture of the system is constructed based on the MVC three-tier model. The back end relies on the Spring Boot framework to handle business logic and service interfaces. The front end uses the Vue.js framework to build the user interface. The client uses the NW.js framework to achieve the cross-platform operation experience on the desktop. Redis, as a high-speed in-memory storage system, provides support for data caching and real-time access. The federated learning algorithm part relies on the FEDLAB framework for model training and parameter updates, achieving the core function of distributed collaborative learning.

In terms of functional implementation, the central server is responsible for the overall scheduling and management of federated learning tasks. It can initialize the global training tasks, configure the model structure and training hyperparameters, and collect model updates from each client in each iteration to complete the synchronization of the global model. At the same time, it determines whether the training termination condition has been met to ensure the efficiency of the learning process. The client management module of the central server continuously monitors the status of all participating nodes, dynamically adjusts task allocation, conducts identity verification and authorization management, and collects model data uploaded by legitimate clients during the training process to ensure the stable operation of the system in a distributed environment. To integrate the model parameters distributed across various clients, the aggregation module of the central server standardizes the local model and generates a new global model version in combination with personalized strategies. At the same time, it records the iteration information of each version for subsequent backtracking and analysis. The communication management module plays a crucial role in ensuring the security and efficiency of data transmission. It can maintain communication protocols, monitor node status, regulate network load, and promptly identify and resolve potential issues through log recording and abnormal early warning mechanisms. The personalized federated learning management module supports personalized adjustments to task configuration, communication policies, and aggregation methods, enabling users to import custom algorithms and flexibly invoke, update, or deprecate policies to meet different application requirements. On the client side, the model management module is responsible for the storage of locally trained models, version control, and the execution of privacy protection policies. It also supports statistical analysis and visual presentation of the training process, enabling users to keep abreast of changes in model performance in real time. The personalized Settings module of the client enables users to freely configure the occupation of computing resources, training parameters and data processing methods while ensuring the stability of system operation, thereby optimizing the local training efficiency. The communication configuration module ensures that the interaction between the client and the server is both secure and efficient by managing IP addresses, ports, encryption keys and traffic priorities. The data management module offers functions such as data storage, backup, recovery, and visual analysis. Users can flexibly manage datasets and their privacy policies, while achieving traceability of operation records and policy adjustments. The system security module comprehensively guarantees the security of system operation through identity authentication, access permission control, log monitoring, and security audit early warning mechanisms, ensuring that all operations are carried out within a legal and controllable scope. The user management module provides the system with functions such as account registration, login, information maintenance, and permission allocation. It ensures the legitimacy of user identities through multiple verification methods. At the same time, it offers flexible management tools for role permissions to system administrators, supporting the long-term stable operation and security management of the system.

During the system testing phase, the research conducted a comprehensive and detailed functional verification of each module, which included task management, client monitoring, model aggregation, communication management, and personalized policy management, as well as client model management, personalized Settings, communication configuration, data management, system security, and user management modules. The test results show that the system can accurately complete operations such as global task initialization, model parameter update, local data management, communication transmission, policy invocation, and permission control. The entire system operates stably, has complete functions, and fully meets the multiple requirements of financial data personalized federated learning in practical application scenarios for data privacy protection, training efficiency, and model accuracy. It has laid a solid foundation for subsequent system optimization, functional expansion and reliability improvement.

5. Conclusions and Prospects

The distributed collaborative training platform for financial data takes cross-institutional modeling as its core idea. It completes data processing and model training locally for each participant and only passes parameters or gradients to the central aggregation end for unified fusion without touching the original information, thereby continuously optimizing the performance of the global model under the premise of ensuring data privacy and compliance. During this process, the central aggregation end is not only responsible for communication coordination, task allocation and model update, but also can flexibly adjust the aggregation strategy based on the different demands of financial business in risk prediction, customer analysis and market judgment to ensure the stability and adaptability of the system in complex environments. Meanwhile, the clients of each institution perform model iteration and security verification by invoking local computing power. And flexibly configure the training program based on business positioning and resource conditions, thereby achieving a balance between personalization and collaboration. The system adopts an extensible architecture and mainstream development languages, combines efficient databases and caching mechanisms to meet the demands of large-scale data processing and high concurrent access. Meanwhile, during development and operation and maintenance, it utilizes open-source tools and existing hardware to reduce costs and enhance efficiency, and reserves upgrade interfaces to strengthen future expansion capabilities. The platform functionally covers communication scheduling, model aggregation, client management, and security auditing, enabling various institutions to share modeling results while maintaining data independence and security. At the non-functional level, it emphasizes user-friendly interfaces, cross-platform compatibility, flexible modules, and feasible maintenance, thereby providing efficient and reliable technical support for financial institutions in risk management and decision support.

Funding

Project: Fund Project: Design and Implementation of Smart Office System Based on WeChat official account, Project No.: HZ250096

References

[1] Li, W. (2025). Discussion on Using Blockchain Technology to Improve Audit Efficiency and Financial Transparency. Economics and Management Innovation, 2(4), 72-79.

[2] Hu, Q. (2025). Implementation and Management of a Cross-Border Tax System Oriented Towards Global Tax Administration Informatization. Economics and Management Innovation, 2(4), 94-101.

[3] Zhang M. Discussion on Using RNN Model to Optimize the Accuracy and Efficiency of Medical Image Recognition[J]. European Journal of AI, Computing & Informatics, 2025, 1(2): 66-72.

[4] Cui, N. (2025). The Practical Application of Traffic Flow Forecasting and Capacity Analysis. Journal of Computer, Signal, and System Research, 2(5), 65-71.

[5] Huang, J. (2025). Promoting Cross-field E-Commerce Development by Combining Educational Background and Technology. Economics and Management Innovation, 2(4), 26-32.

[6] Ye, J. (2025). Optimization and Application of Gesture Classification Algorithm Based on EMG. Journal of Computer, Signal, and System Research, 2(5), 41-47.

[7] Li, B. (2025). Research on Data-Driven Environmental Policy in Water Resource Management. European Journal of Public Health and Environmental Research, 1(1), 101-107.

[8] Zhu, Z. (2025). Application of Database Performance Optimization Technology in Large-Scale AI Infrastructure. European Journal of Engineering and Technologies, 1(1), 60-67.

[9] An, C. (2025). Exploration of Data-Driven Capital Market Investment Decision Support Model. European Journal of Business, Economics & Management, 1(3), 31-37.

[10] Wei, X. (2025). Practical Application of Data Analysis Technology in Startup Company Investment Evaluation. Economics and Management Innovation, 2(4), 33-38.

[11] Cai, Y. (2025). Research on Positioning Technology of Smart Home Devices Based on Internet of Things. European Journal of AI, Computing & Informatics, 1(2), 80-86.

[12] Wang, C. (2025). Exploration of Optimization Paths Based on Data Modeling in Financial Investment Decision-Making. European Journal of Business, Economics & Management, 1(3), 17-23.

[13] Han, Wenxi. "The Practice and Strategy of Capital Structure Optimization under the Background of the Financial Crisis." European Journal of Business, Economics & Management 1, no. 2 (2025): 8-14.

[14] Lu, C. (2025). The Application of Point Cloud Data Registration Algorithm Optimization in Smart City Infrastructure. European Journal of Engineering and Technologies, 1(1), 39-45.

[15] Jing, X. (2025). Research on the Application of Machine Learning in the Pricing of Cash Deposit Products. European Journal of Business, Economics & Management, 1(2), 150-157.

[16] Pan, H. (2025). Development and Optimization of Social Network Systems on Machine Learning. European Journal of AI, Computing & Informatics, 1(2), 73-79.

[17] Li W. Audit Automation Process and Realization Path Analysis Based on Financial Technology[J]. European Journal of Business, Economics & Management, 2025, 1(2): 69-75.

[18] Liu X. The Role of Generative AI in the Evolution of Digital Advertising Products[J]. Journal of Media, Journalism & Communication Studies, 2025, 1(1): 48-55.

[19] Wu X, Bao W. Research on the Design of a Blockchain Logistics Information Platform Based on Reputation Proof Consensus Algorithm[J]. Procedia Computer Science, 2025, 262: 973-981.

[20] Wu, H. (2025). The Commercialization Path of Large Language Models in Start-Ups. European Journal of Business, Economics & Management, 1(3), 38-44.

[21] Hao, L. (2025). Research on Perception and Control System of Small Autonomous Driving Vehicles. International Journal of Engineering Advances, 2(2), 48-54.

[22] Jing X. Real-Time Risk Assessment and Market Response Mechanism Driven by Financial Technology[J]. Economics and Management Innovation, 2025, 2(3): 14-20.

[23] Liu Y. The Impact of Financial Data Automation on the Improvement of Internal Control Quality in Enterprises[J]. European Journal of Business, Economics & Management, 2025, 1(2): 25-31.

[24] Xu Q. AI-Based Enterprise Notification Systems and Optimization Strategies for User Interaction[J]. European Journal of AI, Computing & Informatics, 2025, 1(2): 97-102.

[25] Ren B. Research Progress of Content Generation Model Based on EEG Signals[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 97-103.

[26] Liu Z. Research on the Application of Signal Integration Model in Real-Time Response to Social Events[J]. Journal of Computer, Signal, and System Research, 2025, 2(2): 102-106.

[27] Tang X, Wu X, Bao W. Intelligent Prediction-Inventory-Scheduling Closed-Loop Nearshore Supply Chain Decision System[J]. Advances in Management and Intelligent Technologies, 2025, 1(4).

[28] Xu Q. Design and Future Trends of Intelligent Notification Systems in Enterprise-Level Applications[J]. Economics and Management Innovation, 2025, 2(3): 88-94.

[29] Zhu P. Construction and Experimental Verification of Automatic Classification Process Based on K-Mer Frequency Statistics[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 391-400.

[30] Yang D, Liu X. Collaborative Algorithm for User Trust and Data Security Based on Blockchain and Machine Learning[J]. Procedia Computer Science, 2025, 262: 757-765.

[31] Zhang Y. Research on Optimization and Security Management of Database Access Technology in the Era of Big Data[J]. Academic Journal of Computing & Information Science, 2025, 8(1): 8-12

[32] Hua X. Optimizing Game Conversion Rates and Market Response Strategies Based on Data Analysis[J]. European Journal of AI, Computing & Informatics, 2025, 1(2): 37-43.

[33] Huang, J. (2025). Reuse and Functional Renewal of Historical Buildings in the Context of Cultural Heritage Protection. International Journal of Humanities and Social Science, 1(1), 42-50.