ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 12 (2025) No. 12, pp. 122-137
DOI:10.33168/JLISS.2025.1208

ANFIS-Based Financial Data Quality Control: A Hybrid Genetic Algorithm Approach for Error Detection and Correction

Min Song

College of Financial Management, Chengdu Ginkgo Hotel Management College, Chengdu 611743, Sichuan, China

min.song@gingkoc.edu.cn

Abstract. The purpose of this study is to enhance the quality control and improvement effect of financial data through the adaptive intelligent algorithm. Traditional manual auditing and rule-based validation are costly, inefficient, and hard to deal with complex anomalies. The study formulates a unified adaptive data quality control framework based on ANFIS, clarifies its structural optimization logic, and explains its role in improving accuracy, consistency, and robustness of financial data processing in dynamic environments. In order to adapt the change of data characteristic, this article is based on the Adaptive Neuro Fuzzy Inference System (ANFIS) to automatically adjust the parameters and structures in uncertain and complex environment. This article carried out data cleaning, filling missing values, handling outliers on financial data collected from a large enterprise to make the data completer and more accurate. The experimental results showed that the mean squared error (MSE) and root mean square error (RMSE) of optimized ANFIS in data analysis stage were 0.025 and 0.158, respectively. In data quality improvement stage, MSE and RMSE decreased to 0.019 and 0.138, respectively. This shows that adaptive intelligent algorithms can improve the accuracy and efficiency of financial data, and shows the superiority of this method in financial data quality control and improvement.

Keyword: Financial Data Quality; Adaptive Intelligent Algorithms; Data Quality Control; Data Improvement Techniques; Intelligent Data Management

1. Introduction

Financial data quality control is an important link to ensure the effectiveness of enterprise financial control. High quality financial data (Cohn et al., 2022) plays an important role in strategic decision-making, risk management (So et al., 2022; Taylor, 2023), and regulation, providing a reliable basis for decision-making and help enterprise to avoid risks (FADJAR et al., 2024) and improve the efficiency. In traditional financial data management (Mosteanu and Alessio, 2020; Byrapu, 2023), statistical analysis (Wagenmakers et al., 2022), and other methods are commonly used to control and improve the quality of financial data. Manual audit method is adopted by financial experts to carefully check the data, find out the errors in the data and correct them. This method spends much time on checking data and is influenced by human factors, which leads to low efficiency and more errors. Existing financial audit and data validation systems typically exhibit error detection rates ranging from moderate to low when facing heterogeneous transaction patterns, and industry reports consistently indicate that poor financial data quality leads to substantial operational losses, delayed decision-making, and increased compliance risks, thereby revealing a clear gap between current practices and the requirements of large-scale, dynamic financial systems. Rule-based automation method performs data validation by rules and logic which have been set before. The shortcoming of rule-based automation method is that the rules are fixed and limited, which can't adapt to complex and changeable financial environment and are hard to deal with outliers and new type of errors. Risk avoidance is very important in risk management, and traditional methods have shortcoming in statistical analysis and outlier (Boukerche et al., 2020) mining, which leads to the fact that it is hard to realize accurate control of financial data when facing complex regression analysis (Shrestha, 2020).

In order to address the limitations of traditional methods, this article adopts an adaptive intelligent algorithm to control and improve the quality of financial data. Adaptive intelligent algorithm adapts to changes in data features by automatically adjusting its parameters and structure in uncertain and complex environments. This algorithm is based on machine learning and artificial intelligence technology. Adaptive intelligent algorithms are more efficient than traditional methods in processing nonlinear and high-dimensional data. This article controls the quality of financial data by optimizing the algorithm parameters and structure of ANFIS, which greatly improves the accuracy of data processing. The optimized ANFIS compensates for the shortcomings of manual auditing by automatically identifying and correcting errors in financial data. Based on changes in historical and real-time data, the optimized ANFIS can dynamically adjust control strategies to ensure continuous improvement in data quality.

2. Related Work

How to effectively control and improve data quality has always been an important research direction in academia and finance. Existing studies on financial data quality control can be systematically categorized according to accuracy assurance, completeness maintenance, consistency verification, and timeliness preservation, which jointly constitute the core dimensions of financial data quality. Ni, FuTao (Ni et al., 2020) et al. proposed a data compression and reconstruction framework based on deep learning (Shorten et al., 2021; Janiesch et al., 2021), which utilized one-dimensional convolutional neural networks (Dhillon and Gyanendra, 2020; Bharadiya, 2023) to detect abnormal data. He achieved high-precision data reconstruction at low compression ratios using an autoencoder (Chen et al., 2024) structure to address the data challenges in health monitoring of large-scale infrastructure structures. Persson, Rebecca (Persson et al., 2020), and others used the method of evaluating the correctness and integrity of diagnostic codes to study the diagnostic quality of type 2 diabetes, hyperlipidemia and anemia in the Clinical Practice Research Datalink (CPRD) Aurum database. From the perspective of financial data validation, machine learning based approaches have increasingly been applied to anomaly detection, consistency verification, and integrity assessment. These studies emphasize model-driven identification of data inconsistencies by learning implicit patterns from historical financial records, thereby reducing reliance on manually defined validation rules. However, many existing approaches focus on single-stage detection tasks and lack adaptive mechanisms to respond to evolving financial data distributions. Qin, Siyue (Qin, 2022) studied the positive impact of blockchain technology (Udeh et al., 2024; Gad et al., 2022) on accounting bookkeeping, accounting information quality, and digital financial information through comparative analysis methods, and explored its future development prospects. These studies provide various technical means for financial data quality control, but there are problems such as low efficiency and slow response speed when processing real-time and large-scale data, so further optimization and improvement are needed.

Adaptive intelligent algorithms (Salman et al., 2019) have shown great potential in processing complex data and dynamic systems. Wang, Gai-Ge et al. (Wang et al., 2022) used a multi-objective transformation to single objective optimization method based on fuzzy relative entropy (He et al., 2021), combined with the Hybrid Adaptive Differential Evolution (HADE) algorithm, to solve the job vehicle scheduling problem with fuzzy processing time and completion time, and verified that the HADE algorithm outperforms other advanced algorithms in performance. Cui, Hao et al. (Cui et al., 2020) used a two-point fitting method combined with Adaptive Genetic Algorithm (AGA) (Zhou et al., 2021; Gao et al., 2023) and cylindrical energy model to determine the Jones-Wilkins-Lee equation of state parameters for all ideal condensed explosives without the need for cylindrical experiments, and conducted relevant numerical simulations to verify their effectiveness. Sierra-Garcia, J. Enrique et al. (Sierra-Garcia and Matilde, 2022) proposed a hybrid system that combines fuzzy logic and deep learning (Zheng et al., 2021; Yuksel and Sedat, 2021) for controlling the pitch angle of wind turbines. Although the application domain differs, these studies demonstrate the effectiveness of fuzzy systems and adaptive optimization mechanisms in handling uncertainty, nonlinear relationships, and dynamic parameter variation, which are also fundamental characteristics of financial data quality control tasks. Through deep learning, the current wind speed is estimated, and future wind speed is predicted to improve the performance of intelligent controllers under different disturbances. The impact of deep learning configuration parameters on the training of the hybrid control system is also analyzed, achieving efficient processing of uncertain data. Although the above studies demonstrate the effectiveness of intelligent algorithms in handling complex and uncertain data, they do not establish a unified adaptive framework explicitly tailored to financial data quality control across multiple quality dimensions.

3. Quality Control and Improvement of Financial Data

3.1. Data Collection and Preprocessing

This article collects financial data covering the past five years from the financial management system of a large enterprise. The dataset originates from an internal enterprise-level financial management system in the manufacturing and service sector, covering routine operational transactions recorded between January 2019 and December 2023, and all data are anonymized prior to analysis to remove any identifiable institutional or individual information. The dataset reflects heterogeneous transaction behaviors, accounting structures, and audit records over multiple financial cycles, ensuring that the data distribution captures both routine operations and irregular financial patterns. The types of data include daily transaction records, financial statements, internal audit reports, and budget forecasts. The dataset contains 500000 records of daily transactions that have occurred as a result of differing financial activities including income, expenses, and transfers. The selection of this data volume follows the requirement of stable fuzzy rule learning and parameter convergence in ANFIS, ensuring sufficient coverage of diverse transaction states without introducing data sparsity effects. The financial statements portion includes the company’s monthly and annual financial statements for the company, which has resulted in 60 statements. The internal audit report includes 20 records of internal audits and the results for the past five years. The budget forecast data section includes 36 budget forecasts and fund utilization forecasts for three years. Each of these data sections provides rich data for the purposes of this research and also for training and testing algorithms to ensure the research results are reliable and practical.

To eliminate duplicate records and any data that is not needed, this article completes the data cleaning and simplifies the data structure. Multiple imputation methods are used to complete missing values, such as mean imputation, interpolation, and multiple imputations, depending on the data type and distribution. Outlier detection and processing are also very important. Outliers can be identified through box plot and standard deviation methods, and minor outliers can be corrected. Severe outliers can be processed using quantile method or substitute value method based on historical data. Data standardization processing is the final step, which converts data of different dimensions into the same scale range to improve the effectiveness of algorithm training. These preprocessing steps ensure the quality of the data and lay a solid foundation for subsequent algorithm applications and analysis. The characteristics of the collected dataset before and after preprocessing are summarized in Table 1.

Table 1 Data Preprocessing Results

Data Type	Original Data Volume	Missing Value Ratio (%)	Outlier Ratio (%)	Post-Processing Data Volume
Daily Transactions	500,000	5	3	485,000
Financial Statements	60	10	2	58
Internal Audit Reports	20	0	0	20
Budget Forecasts	36	8	1	35

In Table 1, the proportion of missing values in financial statements is the highest, reaching 10%; the proportion of outlier values in daily transaction records is the highest, at 3%; there are no missing or outlier values in internal audit reports. The reported data volumes reflect aggregated records after anonymization and consolidation procedures, and all data are processed in compliance with internal data governance and confidentiality requirements.

3.2. Adaptive Intelligent Algorithm

This article chooses an intelligent algorithm ANFIS (Armaghani and Panagiotis, 2021; Janardhana et al., 2021) based on adaptive reinforcement learning as the core method for financial data quality control. ANFIS combines the advantages of neural networks and fuzzy logic, capable of handling complex nonlinear relationships and performing well in the face of uncertain and fuzzy data. Its structure is shown in Figure 1.

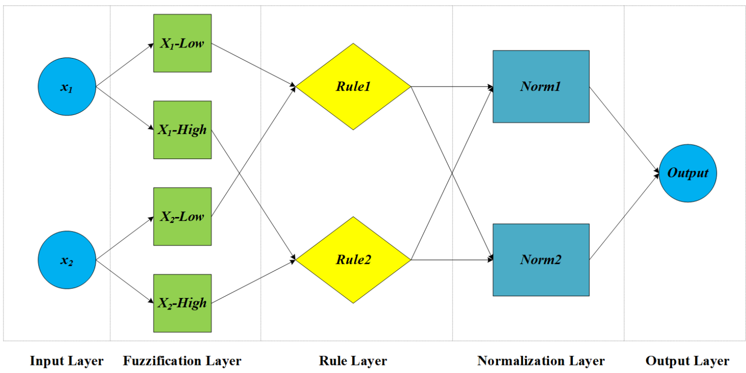

Fig.1: ANFIS Structure Diagram.

ANFIS converts the input financial data into fuzzy set by fuzzification process. Each financial input variable is mapped to continuous membership degrees through predefined Gaussian and Hyperbolic

Tangent membership functions, whose parameters are initialized based on the statistical distribution of the corresponding input features and iteratively optimized during training to minimize prediction error and rule inconsistency. For each input variable, it is assigned to several fuzzy subsets, and each subset has a fuzzy membership function. Through these fuzzy membership functions, ANFIS can reflect the fuzziness and uncertainty in financial data and establish more accurate and flexible data model. After fuzzification, ANFIS uses the learning mechanism of neural network to optimize fuzzy rules and constructs a series of "if then" fuzzy rules, which are adjusted and optimized by training process of neural networks. Neural networks utilize the relationship between input and output data to continuously update the parameters of fuzzy rules by backpropagation algorithms. Therefore, the system can adjust its structure and parameters by training process of neural networks to minimize the prediction errors.

ANFIS tunes to the data changes by learning and giving feedback constantly, and adjusting its model parameters online. This is the self-adjustment ability of ANFIS. Adaptive ability of ANFIS makes ANFIS to obtain satisfactory and precise prediction results in various application situations. When dealing with sudden financial data exception, ANFIS can identify and correct the wrong financial data effectively by rapidly adjusting rules and parameters, and then maintains high data quality. Interpretability ability of ANFIS combines fuzzy rules and neural network to make transparent and interpretable prediction results available to financial managers for understanding the process and result of financial data processing. Not only can improve the acceptable of algorithm, but also can enhance user's acceptance of algorithm and understand algorithm result. ANFIS combines the advantages of fuzzy logic and neural network to solve complex nonlinear relationship and uncertain data. By adjusting itself and optimizing itself, ANFIS makes use of self-adjustment and optimization to realize the financial data quality control in enterprise. It can achieve efficient, accurate, and interpretable predictive results, and provide solid data support for enterprise financial management by transparently, showing the process and results of financial data processing.

3.3. Model Training

This article employs a backpropagation algorithm for optimizing the parameters pertaining to fuzzy rules and membership functions of ANFIS throughout the entire model training stage in the training process, which is structured as illustrated in Figure 2.

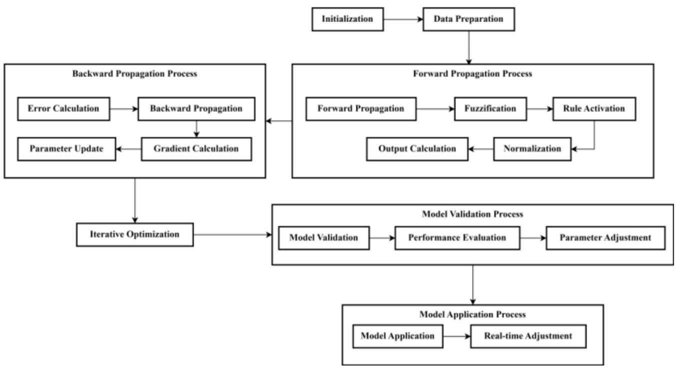

Fig.2: Training Process of the Model.

To train a model, the first stage is to establish or initialize the model itself and fuzzy rules by designating and distinguishing the input variables into several fuzzy subsets according to fuzzy rules, as each subset corresponds with a fuzzy membership function. Unknown initial conditions are set for the parameters in the membership functions and the fuzzy rules. The condition of model initialization is data preparation, which includes collecting and preprocessing data related to the financial markets to create training and testing datasets. The data is partitioned into training, validation, and testing sets that consist of the known data to train the model and subsequently evaluate its performance. The method utilized to divide the dataset into training and testing data is based on a mix of time series sampling and random purposive sampling in order to have sufficiently representative and temporally relevant training and testing datasets in the timeframe. The datasets represent transaction records on a daily basis as the training data consists of the last four years of all transaction records; the validation data includes most recent year of all transactions, and the testing dataset includes the last six months of transactions. For the financial statement dataset, the training set, validation set, and testing set are the annual reports of the past four years, the monthly report of the most recent year, and the annual report of the most recent year, respectively. In the internal audit report, the training set includes audit records from the past four years, while the validation set and test set each contain records from the most recent year. Budget forecast data. The training set includes the first two years of budget planning for the next three years, and the validation set and test set include the 25th to 30th data and the last 6 data, respectively.

In the forward propagation process, the input data is converted into a fuzzy set through a membership function. The activation degree of each rule is calculated based on fuzzy rules, and the activation degree is normalized to obtain the normalized activation degree. Then, the output of the model is calculated based on the normalized activation degree and fuzzy rules. In the error calculation stage, the MSE between the model output and the actual value is calculated, and the error is fed back to the model for parameter adjustment. In the backpropagation stage, the gradient of the error relative to the membership function parameters and fuzzy rule parameters can be calculated, and these parameters can be adjusted according to the gradient descent algorithm to reduce the error. The parameter update process follows gradient descent optimization, where gradients of the loss function with respect to membership function parameters and rule weights are computed and iteratively adjusted to minimize the overall prediction error. The forward propagation, error calculation, and backpropagation are continuously repeated until a predetermined number of training iterations are reached or the error converges to a predetermined threshold. During this process, the model performance can also be evaluated on the validation set to ensure that the model is not overfitting or underfitting, and the model parameters and structure can be further adjusted based on the validation results. In practical applications, the model can adjust and optimize itself in real-time based on new data to improve the accuracy and efficiency of financial data processing.

3.4. Algorithm Optimization

For the purpose of improving the performance of ANFIS and make it perform better in financial data quality control, this article carries out a series of optimizations on ANFIS. The fuzzy membership functions of traditional ANFIS usually use triangular or trapezoidal membership functions, but these simple functions cannot accurately capture the characteristics of complex data. For this purpose, this article introduces Gaussian Membership Function (GMF) and Hyperbolic Tangent Membership Function (HTMF), whose calculation formulas are shown in formulas (1) and (2) respectively:

$$\mathrm{G(x)} = \exp\left({}^{-(\mathrm{x}-\mathrm{c})^2}/_{2\sigma^2}\right) \quad (1)$$

$$\mathrm{H(x)} = {}^{(1+\tanh\,({}^{\mathrm{x}-\mathrm{c}}/_{\mathrm{a}}))}/_{2} \quad (2)$$

Among them, a is the width scale parameter used to control the functions, which is the center of the two functions, and σ is the standard deviation of the Gaussian function used to control the width of the functions. GMF provides continuous differentiability and stable gradient behavior during parameter optimization, while HTMF enhances sensitivity in boundary regions of financial data distributions, jointly improving rule smoothness and inference stability. Smoother and more flexible make the model more accurate in describing the ambiguity and uncertainty in financial data. During the training process, the parameters of these membership functions are continuously adjusted through optimization algorithms to better adapt to the characteristics of the input data. The selection of Gaussian and Hyperbolic Tangent membership functions is driven by their smooth differentiability and stable gradient behavior, which supports efficient parameter optimization and avoids abrupt rule transitions that commonly degrade inference stability in financial data quality control tasks.

This article combines Backpropagation Algorithm (BP) with Genetic Algorithm (GA) to solve the problem of getting stuck in local optima. The integration follows a sequential optimization strategy in which GA is first applied to perform global parameter search and determine high-quality initial solutions, followed by BP-based gradient optimization to refine parameters until convergence criteria are satisfied. GA simulates the processes of natural selection and genetic variation to globally search for the optimal solution, avoiding getting stuck in local optima. In the early stages of training, GA conducts a global search to find optimal initial parameters, which are then fine tuned through BP to further optimize the model parameters. This hybrid training algorithm not only ensures the global optimization ability of the model, but also improves the training speed and accuracy. In ANFIS, the fuzzy rule base is the core of model performance. The size and quality of the rule library directly affect the predictive ability of the model. This article adopts a rule extraction method based on fuzzy clustering to improve the efficiency of the rule base. By using the fuzzy C-means clustering algorithm, similar data samples are clustered together, with each cluster center corresponding to a fuzzy rule. This method reduces the number of rules, simplifies the complexity of the model, and improves the representativeness and effectiveness of the rules. This article introduces a rule simplification mechanism to optimize the rule base, removing redundant and inefficient rules by calculating the contribution of each rule.

3.5. Quality Control of Financial Data

The article inputs the preprocessed data into the model and fuzzifies the data using GMF and HTMF, and preliminary evaluation of data is achieved by applying a trained fuzzy rule library. When a new transaction data is input into the system, ANFIS can immediately analyze it to detect any outlier and input error. If there is a problem detected in the transaction, the system alarms and notifies the related operator to check and correct it so as to minimize the human effort error.

The system dynamically adjusts the changing financial data using the adaptive characteristic of optimized ANFIS during data transmission and data storage. When a financial statement data is transmitted between different department, ANFIS can apply fuzzy clustering method to detect any abnormal change in transmitted financial statement data. When a problem detected that the financial statement data has been tampered or lost during transmission, the system can automatically trigger the error correction function to repair abnormal financial statement data based on historical data so as to ensure the integrity of transmitted financial statement data.

The system applies global search using GA, combined GMF and HTMF to perform intensive analysis of data and automatically screen all financial record during year end financial audit, detects all the errors and abnormal data and marks it accordingly. The abnormal data generated from the financial audit system can be reviewed by financial auditor so as to ensure the accuracy of the financial report generated by the system. The adaptive characteristic of ANFIS used by the system plays an important role in the real time data monitoring. The system continuously optimizes its model parameters based on the real time feedback achieved during the execution so as to adapt to the dynamic changes of financial data. When preparing quarterly financial report, ANFIS monitors the changes in various financial indicator in real time, adjusts the fuzzy rule and parameter setting based on the dynamic changes achieved during the execution, and ensures that the financial data in each quarterly financial report generated by the system passed strict quality control. Through the specific application above, the optimized ANFIS achieves high accuracy and efficiency in financial data quality control. The system can achieve real time monitoring and automatic error correction at various stages of data processing so as to improve the accuracy and consistency of financial data and provided solid technical support for enterprise financial management.

The adaptive structure of ANFIS allows continuous rule adjustment in response to evolving financial data distributions, mitigating performance degradation caused by concept drift. By updating membership parameters and rule weights based on incoming data streams, the system maintains long-term consistency without requiring manual retraining.

3.6. Improvement of Financial Data Quality

When improving the quality of financial data, this article uses GA to optimize the parameter of center and width of membership function so that the membership function is closer to the actual data distribution so as to improve the accuracy and adaptability of the model. During the training process, GA searches the optimal parameter configuration globally so as to avoid the situation that the parameter configuration is trapped in the local optimum. BP uses gradient descent to locally optimize the initial parameter configuration obtained from GA, and refines the model parameter to further improve the speed of convergence and training accuracy.

In article, statistical analysis on historical financial data, extraction of representative rules and every rule can cover different financial data patterns. Anomaly detection technology can apply to the whole rules to detect redundant or invalid rules and remove them. The number of rules can decrease and the speed and accuracy of inference also can improve. Optimized rule library can reduce computational complexity and improve system efficiency. Meanwhile, it can guarantee comprehensive coverage. Through above optimization, the effectiveness and accuracy of ANFIS algorithm in financial data quality control has been improved greatly. It can identify and process outlier in financial data more accurately and improve accuracy and reliability of data verification. By timing detecting financial risk through optimized ANFIS, the accuracy and efficiency of risk management can be improved. Optimized ANFIS has higher accuracy and stability in handling regression analysis and statistical analysis.

4. Financial Data Quality Control and Improvement Experiment

4.1. Experimental Environment

The experimental environment configuration of this article is shown in Table 2.

Table 2 Experimental Environment Configuration Table

Component	Configuration	Details	Notes
Operating System	Windows 11	Latest version	For compatibility
CPU	Intel Core i5-13600KF	High performance CPU	Supports parallel processing
Memory	128GB DDR4	Large memory capacity	Handles large datasets
Storage	4TB SSD	High-speed storage	For fast read/write
Programming Language	Python	Popular language for ML	Versatile usage
Software Environment	Python 3.8	Stable release	Standard for ML tasks
Model Training and Optimization	TensorFlow 2.4, Keras 2.4	Popular deep learning frameworks	Efficient model building
Data Preprocessing and Analysis	Scikit-learn 0.24	Widely used library	For preprocessing tasks
Data Visualization	Matplotlib 3.3, Seaborn 0.11	Visualization libraries	For data analysis
Software and Library Management	Anaconda 2020.11	Comprehensive package manager	Manages dependencies
Network Environment	High bandwidth, low latency	Stable network connection	Ensures data transfer
Version Control System	Git	Version control	Tracks changes

All algorithms are implemented using Python, with numerical computation, machine learning, and optimization libraries maintained at fixed versions throughout the experiments to ensure consistency and reproducibility of the reported results.

4.2. Parameter Settings

The initial values for the learning rate, iteration times, and batch size of ANFIS in this experiment are 0.01, 200, and 64, respectively. The initial center and standard deviation of GMF are set to 1.5 times the data mean and standard deviation, respectively, while the center and scale parameters of HTMF are set to the data mean and standard deviation. The population size of GA is 50, with crossover and mutation rates set at 0.8 and 0.1, respectively. The maximum number of iterations for GA is 100. BP uses the Adam optimizer with an initial learning rate of 0.001, a learning rate decay factor of 0.96, and 10000 decay steps. The fuzzy C-means clustering algorithm has a clustering number of 5, a maximum iteration count of 300, and a fuzzy coefficient of 2.0. The fuzzy coefficient controls the degree of cluster overlap, and the value of 2.0 is selected to balance clustering stability and membership smoothness, avoiding excessive fuzziness or hard partitioning during rule extraction. The rule simplification mechanism adopts a contribution threshold of 0.05, which means deleting rules with a contribution level below 5%. The parameter configurations are determined to maintain training stability and avoid gradient oscillation, rather than to maximize performance under a single experimental setting. All parameters have been validated through multiple experiments to ensure the stability and optimization effectiveness of the model.

4.3. Experimental Evaluation Indicators

This article uses MSE, RMSE, and data processing efficiency to evaluate the effectiveness of ANFIS in financial data quality control. The calculation formulas for MSE and RMSE are shown in formulas (3) and (4):

$$MSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / n \quad (3)$$

$$RMSE = \sqrt{MSE} \quad (4)$$

In formula (3), y_i and $\hat{y}_i$ are the actual and predicted values, respectively, and n is the sample size.

This experiment uses Data Accuracy Improvement Rate (DAIR), Error Correction Rate (ECR), and Improved Data Consistency Index (CI) to evaluate the quality improvement effect. The calculation formulas for the three are shown in formulas (5) to (7):

$$DAIR = a_2 - a_1 / a_1 \times 100\% \quad (5)$$

$$ECR = N_c / N_d \times 100\% \quad (6)$$

$$CI = N_r / N_t \times 100\% \quad (7)$$

In formula (5), a_1 represents the accuracy before data quality improvement, and a_2 represents the accuracy after data quality improvement. In formula (6), N_d is the total number of detected errors, and N_c is the number of corrected errors. In formula (7), N_t represents the total number of records, and N_r represents the number of consistent records. Consistent records refer to financial data entries that satisfy predefined structural, logical, and cross-field validation constraints after quality control, indicating internal coherence within the dataset rather than comparison against an external reference.

In addition to these indicators, this article uses Training Time (TT) and Convergence Rate (CR) as auxiliary indicators for model training. TT refers to the time required from the start of model training to reaching the predetermined stopping condition, while CR is evaluated through the relationship between iteration times and error changes, reflecting the stability and efficiency of the model during the training process. These evaluation indicators comprehensively reflect the performance of ANFIS in financial data quality control and improvement, providing comprehensive quantitative basis for experimental results.

4.4. Cross-Validation and Statistical Validation

The reliability of experimental results depends not only on evaluation metrics but also on the stability of model performance under different data partitions. To address this issue, a k-fold cross-validation strategy is employed in the experimental evaluation. The complete dataset is partitioned into k subsets with consistent data distributions. In each round, one subset is used as the testing set while the remaining subsets are used for training and validation. This process is repeated until all subsets have served as the testing set once. Based on the results obtained from each fold, the mean and standard deviation of error metrics are calculated to quantify both average performance and variability. The cross-validation results and corresponding statistical significance analysis are summarized in Table 3.

Table 3 Cross-Validation and Statistical Significance Results

Algorithm	MSE Mean	RMSE Mean	MSE Std	RMSE Std	Significance vs ANFIS
ANFIS	0.020	0.141	0.003	0.005	—
RBVM	0.034	0.184	0.006	0.008	Significant
OD	0.028	0.168	0.005	0.007	Significant
CNN	0.026	0.159	0.004	0.006	Significant
LSTM	0.024	0.152	0.004	0.006	Significant

As shown in Table 3, ANFIS exhibits the lowest mean error values across all validation folds, while maintaining relatively small standard deviations, indicating stable performance under different data partitions. In contrast, baseline methods present higher error means and larger variability, reflecting weaker robustness. Statistical significance testing based on fold-wise error distributions further confirms that the performance differences between ANFIS and comparison methods are not attributable to random variation, thereby validating the effectiveness of the proposed approach from a statistical perspective.

5. Result Analysis

5.1. Performance of Adaptive Intelligent Algorithms

To assess the robustness and statistical reliability of the experimental results, repeated experiments and statistical analysis are conducted for all evaluation metrics. This article records the MSE and RMSE of ANFIS during the data analysis, data transmission, and data quality improvement stages in the experiment, in order to analyze the performance of ANFIS in financial data quality control. Multiple experiments are conducted at each stage to ensure the reliability and accuracy of the results. Figure 3 shows the MSE and RMSE changes of ANFIS at different stages of data processing.

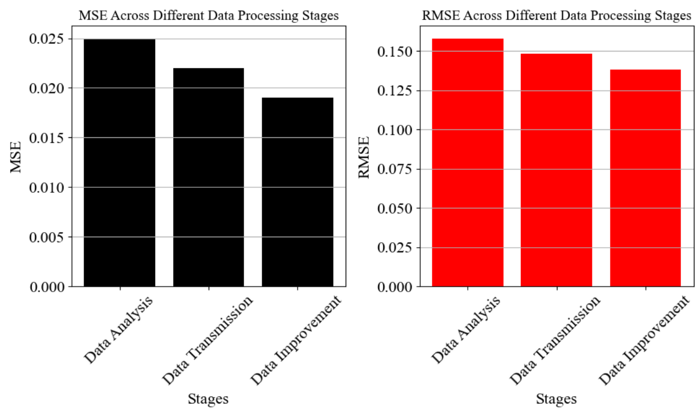

Fig.3: MSE and RMSE changes of ANFIS at different data processing stages.

Figure 3 shows significant changes in MSE and RMSE of ANFIS at different stages of data processing. The observed error reduction reflects the progressive alignment between fuzzy rules and underlying financial data patterns as the adaptive mechanism continuously refines model parameters. In the data analysis stage, the MSE is 0.025, and the RMSE is 0.158, indicating that the algorithm has been able to process the data well in this stage. During the data transmission phase, MSE and RMSE decrease to 0.022 and 0.148, respectively, indicating the stability of the data during transmission and the effectiveness of the algorithm. The MSE and RMSE of ANFIS during the data quality improvement phase are 0.019 and 0.138, respectively, indicating that after further optimization and adjustment, ANFIS has reached its optimal state in ensuring data quality and minimizing errors.

The statistical significance and uncertainty of the reported error metrics are further quantified through repeated experiments, as summarized in Table 4.

Table 4 Statistical Analysis of Error Metrics Across Repeated Experiments

Stage	Mean MSE	Std. Dev.	95% Confidence Interval
Data Analysis	0.025	0.002	[0.021, 0.029]
Data Transmission	0.022	0.002	[0.018, 0.026]
Data Improvement	0.019	0.001	[0.016, 0.022]

5.2. Model Training Effectiveness

This article uses fuzzy logic combined with neural networks to train ANFIS by using BP to optimize fuzzy rules and parameters of the membership functions. The changes in the losses during the learning progress show that the model improves in every learning step. As illustrated in Figure 4, the losses change over 1000 iterations learned by the model, with the x-axis showing the number of iterations, and the y-axis showing the losses.

Fig.4: Loss variation during ANFIS training process.

Since the model learns and improves quickly in the first iterations, the loss curve in Figure 4 indicates that in the first 200 iterations, the loss has dropped rapidly to 0.03, and as the iterations increase, the reduction rate of the loss has slowed down and converged to approximately 0.01. This shows that BP minimizes errors during the course of improving the model.

Table 5 presents both the original settings and the optimized changes of the training parameters. Modification of training parameters has an important impact on the optimization.

Table 5 Model Training Parameters and Performance

Training Parameter	Initial Setting	Optimized Setting	Improvement (%)
Learning Rate	0.01	0.005	50
Epochs	200	1000	40
Batch Size	64	128	100
Number of Fuzzy Rules	10	8	20

As demonstrated in Table 5, changing the learning rate from 0.01 to 0.005 results in more stable parameter updates. The number of training iterations increases from 200 to 1000. The batch size increases from 64 to 128, enabling improved training efficiency. Additionally, the number of fuzzy rules is truncated from 10 to 8, mitigating complexity in the model. These parameter adjustments jointly reduce training instability and model redundancy, thereby improving convergence behavior and inference efficiency.

5.3. Effectiveness of Financial Data Quality Control

The CNN and LSTM models are implemented in their one-dimensional sequential forms to accommodate financial transaction time-series characteristics, and their network depth, learning rates, and training iterations are adjusted to ensure stable convergence under the same data partitions and training conditions as ANFIS, thereby providing a consistent baseline for performance comparison. To assess ANFIS's advantages in managing financial data, this article compares the performance of ANFIS against four other data monitoring algorithms: rules-based verification methods (RBVM), outlier detection methods (OD), convolutional neural networks (CNN), and long short-term memory networks (LSTM). The performance of these data monitoring algorithms is presented in the experimental data in Figure 5.

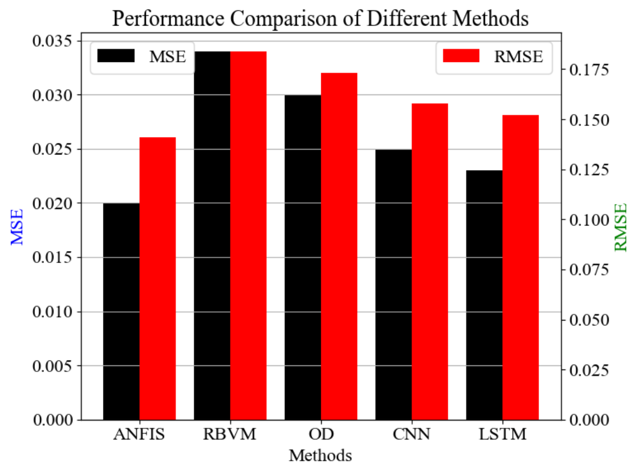

Fig.5: Performance Comparison of Different Methods.

Looking at the five algorithms demonstrated in figure 5, it is determined that the ANFIS MSE and RMSE are both 0.020 and 0.141, respectively, which are lower than the other algorithms presented in the research, and demonstrates superior performance in managing high-dimensional and nonlinear data. The RBVM, although a simple algorithm and easy to implement, produces excessively high errors in complex data environments, with MSE and RMSE reaching 0.034 and 0.184, respectively. The CNN has some improvement in processing data but performs at a lower level than ANFIS. The LSTM has some improvement in processing data but does not perform at levels higher than ANFIS.

To further demonstrate the effectiveness of ANFIS in data processing efficiency, this experiment records the specific performance of five algorithms, as shown in Table 6.

Table 6 Specific Performance of Different Algorithms in Data Processing Efficiency

Algorithm	Data Processing Efficiency (%)	Average Processing Time (s)	DAIR (%)
ANFIS	95	0.5	10
RBVM	85	1	5
OD	90	0.8	7
CNN	90	0.7	8
LSTM	92	0.6	9

Table 6 shows that ANFIS has achieved a data processing efficiency of 95%, which is the highest among the five algorithms. The average processing time and DAIR of the data are 0.5s and 10%, respectively, both of which are optimal, reflecting the advantages of ANFIS in financial data processing.

This experiment also compares the processing effects of different types of financial data, as shown in Figure 6.

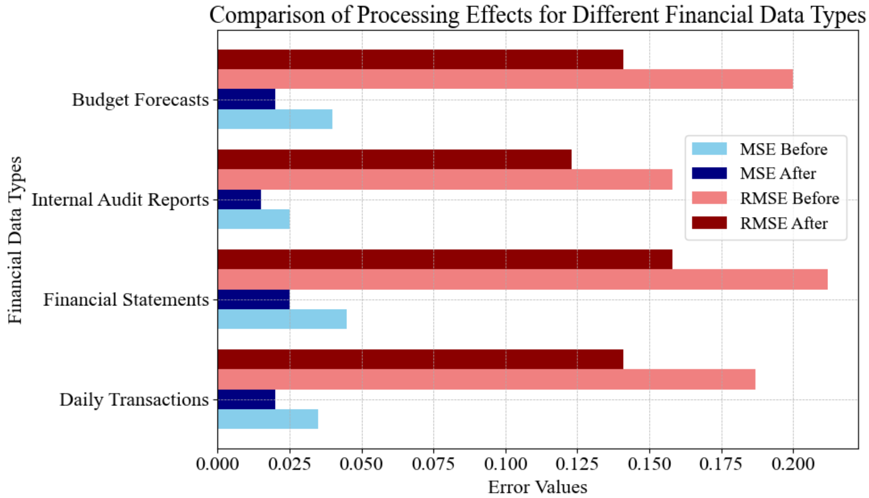

Fig. 6 Processing effect of various financial data

From the data in Figure 6, it can be seen that the MSE and RMSE of daily transaction records after ANFIS processing have decreased to 0.020 and 0.141, respectively. The MSE and RMSE values of financial statements and internal audit reports have also significantly decreased. The improvement effect of budget forecast data is most significant, with MSE reduced from 0.040 to 0.020 and RMSE reduced from 0.200 to 0.141. These data demonstrate that adaptive intelligent algorithms have good performance in processing high-dimensional and nonlinear data.

5.4. Improvement Effect of Financial Data Quality

This article uses three indicators, DAIR, ECR, and improved CI, to compare and analyze the financial data quality before and after improvement, which clearly demonstrates the effectiveness of adaptive intelligent algorithms in improving data quality. The radar chart in Figure 7 shows the performance of these three indicators before and after improvement.

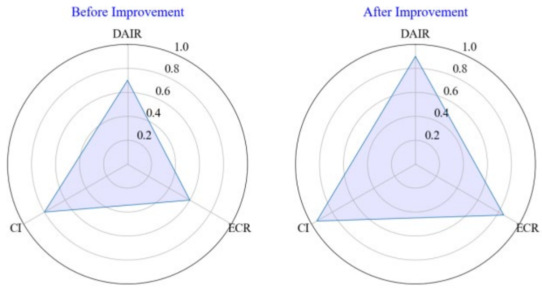

Fig.7 Improvement effect of financial data quality

The left side of Figure 7 presents the performance metrics prior to the enhancement of financial data quality, and the right side illustrates the performance metrics after the improvement. DAIR increases from 70% before improvement to 90% after improvement; ECR increases from 0.6 before improvement to 0.85 after improvement; CI increases from 0.8 to 0.95. This demonstrates that the improved ANFIS does enhance the quality of financial data.

6. Conclusions

This article discusses how do adaptiver intelligent algorithm enhance the quality control and effectiveness of financial data. Traditional method of financial data quality control based on manual auditing and rule verification have certain limitation, such as high cost, low efficiency, and difficult to handle complex data. This article uses ANFIS to adapt to the change of data feature and meet challenge of complex environment. It can adjust the parameters and structure of ANFIS dynamically. Data cleaning, filling missing value, and outlier handling in preprocessing can improve the integrity and accuracy of data. Optimized ANFIS has excellent performance in reducing MSE and RMSE. It can verify the effectiveness of improving data quality by algorithm. This study not only can improve the accuracy and efficiency of financial data processing, but also can provide new technological approach of financial data quality control. Through analysis of historical data and real-time data, ANFIS can adjust the control strategy in time and optimize the data processing flow. It can guarantee the high quality and consistency of financial data. It also provides interpretable fuzzy rules that support audit traceability and regulatory inspection requirements. In addition, this study also does multiple rounds of optimization on algorithm. It can improve the adaptability and prediction accuracy of system. It can further confirm the efficiency and reliability of adaptive intelligent algorithm in practical application. The computational complexity of the proposed framework is mainly influenced by the number of fuzzy rules and membership functions, and under a fixed rule base its training and inference costs scale linearly with the data size. The method is intended for structured financial data quality control scenarios and may exhibit degraded performance when data distributions shift abruptly without sufficient historical feedback or when data volumes are insufficient to support stable rule learning, in which case retraining or complementary validation mechanisms are required. This study not only can provide new view for research of financial data management, but also can provide the solution of data quality control which need to handle the similar challenge of financial data.

References

Armaghani, D. J. & Panagiotis, G. A. (2021). A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength. *Neural Computing and Applications*, 33 (9), 4501-4532.

Bharadiya, J. (2023). Convolutional neural networks for image classification. *International Journal of Innovative Science and Research Technology*, 8 (5), 673-677.

Boukerche, A. Zheng, L. N. and Omar, A. (2020). Outlier detection: methods, models, and classification. *ACM Computing Surveys (CSUR)*, 53 (3), 1-37.

Byrapu, S. R. (2023). Big data analysis in finance management. *Journal of Algebraic Statistics*, 14 (1), 142-149.

Chen, X. K. Ding, M. Y. Wang, X. D. Xin, Y. Mo, S. T. Wang, Y. H. Han, S. M. Luo, P. Zeng, G. and Wang, J. D. et al. (2024). Context autoencoder for self-supervised representation learning. *International Journal of Computer Vision*, 132 (1), 208-223.

Cohn, J. B. Zack, L. and Malcolm I. W. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, 146 (2), 529-551.

Cui, H. Song, P. Gu, X. H. Zhou, H. Xing, B. Y. Jiang, L. and Guo, R. et al. (2020). Determination of the JWL parameters of detonation products using adaptive genetic algorithm. *Propellants, Explosives, Pyrotechnics*, 45 (12), 1920-1930.

Dhillon, A. and Gyanendra, K. V. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9 (2), 85-112.

Fadjar, N. S. Tyas, D. H. and Risqi, N. H. P. (2024). Risk avoidance in intergenerational investment decisions. *International Journal of Environmental, Sustainability, and Social Science*, 5 (3), 599-605.

Gad, A. G. Diana, T. Mosa, L. A. and Amr, A. A. (2022). emerging trends in blockchain technology and applications: a review and outlook. *Journal of King Saud University-Computer and Information Sciences*, 34 (9), 6719-6742.

Gao, Y. G. Ma, C. L. and Sheng, A. (2023). Compound fault diagnosis for cooling dehumidifier based on RBF neural network improved by kernel principle component analysis and adaptive genetic algorithm. *Soft Computing*, 27 (3), 1599-1613.

He, L. J. Raymond, C. Li, W. F. Sandeep, D. Cao, Y. L. and Zhang, Y. et al. (2021). Multiobjective optimization of energy-efficient job-shop scheduling with dynamic reference point-based fuzzy relative entropy. *IEEE Transactions on Industrial Informatics*, 18 (1), 600-610.

Janardhana, K. Sridhar, S. Chandra, K. Dixit, M. Deivakani, S. T. Abdul, R. K. Asif, A. and Maughal, A. A. B. et al. (2021). ANFIS modeling of biodiesels' physical and engine characteristics: a review. *Heat Transfer*, 50 (8), 8052-8079.

Janiesch, C. Patrick, Z. and Kai, H. (2021). Machine learning and deep learning. *Electronic Markets*, 31 (3), 685-695.

Mosteanu, N. R. and Alessio, F. (2020). Digital systems and new challenges of financial management– FinTech, XBRL, blockchain and cryptocurrencies. *Quality–Access to Success*, 21 (174), 159-166.

Ni, F. T. Zhang, J. and Mohammad, N. N. (2020). Deep learning for data anomaly detection and data compression of a long-span suspension bridge. *Computer-Aided Civil and Infrastructure Engineering*, 35 (7), 685-700.

Persson, R. Catherine, V. S. Katrina, W. H. Todd, S. Tim, W. Puja, M. and Susan, S. J. et al. (2020). CPRD aurum database: assessment of data quality and completeness of three important comorbidities. *Pharmacoepidemiology and Drug Safety*, 29 (11), 1456-1464.

Qin, S. Y. (2022). A review of research on the impact of blockchain on financial reporting. *Accounting, Auditing and Finance*, 3 (1), 51-58.

Salman, A. D. Osamah, I. K. and Ghaida, M. A. (2019). An adaptive intelligent alarm system for wireless sensor network. *Indonesian Journal of Electrical Engineering and Computer Science*, 15 (1), 142-147.

Shorten, C. Taghi M. K. and Borko, F. (2021). Deep learning applications for COVID-19. *Journal of big Data*, 8 (1), 1-54.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8 (2): 39-42.

Sierra, G. Enrique, J. and Matilde, S. (2022). Deep learning and fuzzy logic to implement a hybrid wind turbine pitch control. *Neural Computing and Applications*, 34 (13), 10503-10517.

So, M. K. P. Thomas, W. C. C. and Amanda, M. Y. C. (2022). Efficient estimation of high-dimensional dynamic covariance by risk factor mapping: applications for financial risk management. *Journal of Econometrics*, 227 (1), 151-167.

Taylor, N. (2023). Making financial sense of the future: actuaries and the management of climate-related financial risk. *New Political Economy*, 28 (1), 57-75.

Udeh, E. O. Prisca, A. Kudirat, B. A. and Anwulika, O. S. (2024). The role of Blockchain technology in enhancing transparency and trust in green finance markets. *Finance & Accounting Research Journal*, 6 (6), 825-850.

Wagenmakers, E. J. Alexandra, S. and Balazs, A. (2022). One statistical analysis must not rule them all. *Nature*, 605 (7910), 423-425.

Wang, G. Ge. Gao, D. and Witold, P. (2022). Solving multiobjective fuzzy job-shop scheduling problem by a hybrid adaptive differential evolution algorithm. *IEEE Transactions on Industrial Informatics*, 18 (12), 8519-8528.

Yuksel, A. S. and Sedat, A. (2021). Driver's black box: a system for driver risk assessment using machine learning and fuzzy logic. *Journal of Intelligent Transportation Systems*, 25 (5), 482-500.

Zheng, Y. H. Xu, Z. H. and Wang. X. X. (2021). The fusion of deep learning and fuzzy systems: a state-of-the-art survey. *IEEE Transactions on Fuzzy Systems*, 30 (8), 2783-2799.

Zhou, M. Long, Y. X. Zhang, W. P. Pu, Q. L. Wang, Y. Nie, W. and He, W. et al. (2021). Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization. *IEEE Transactions on Evolutionary Computation*, 25 (5), 913-927.