Automating Financial Audits with Random Forests and Real-Time Stream Processing: A Case Study on Efficiency and Risk Detection

Jianlin Li[1,2], Wanli Liu[3*], Jie Zhang[4]
[1]School of Business Administration, Hebei University of Economics and Business Shijiazhuang 050061, Hebei, China
[2]Research Center for Corporate Governance and Enterprise Growth of Hebei University of Economics and Business, Hebei University of Economics and Business, Shijiazhuang 050061, Hebei, China
[3]Finance Department, Hebei University of Economics and Business, Shijiazhuang 050061, Hebei, China
[4]Office of Scientific Research, Hebei University of Economics and Business, Shijiazhuang 050061, Hebei, China
E-mail: Jianlin_Li1748@outlook.com, L135821liu_ll@hotmail.com, Jie_Zhang0152@outlook.com
[*]Corresponding author

Keywords: artificial intelligence, financial audit, automated method

Received: December 11, 2024

In the current complex economic environment, enterprises are increasingly in need of efficient, accurate and real-time financial audits. Traditional audit methods are difficult to cope with the challenges brought by massive data and dynamic risks. This paper explores the automation method of financial audits based on artificial intelligence in depth, aiming to improve audit efficiency and risk identification capabilities. The study introduces the random forest algorithm, constructs 100 decision trees, self-samples data from the training set, and randomly selects features at each node for splitting to reduce the overfitting risk of a single decision tree and improve the generalization ability of the model. At the same time, with the help of real-time data processing platforms such as Kafka and Blink, real-time collection, processing and analysis of financial data are achieved to ensure the timeliness and dynamism of the audit process. After a series of steps, including extracting 500 features from multi-source data, dividing the data set containing 5,000 records into 70% training set and 30% test set, the model is trained and evaluated. The results show that this method has achieved remarkable results, with audit efficiency increased by 30%, risk detection accuracy increased to 90%, audit coverage enhanced, and error detection rate, data processing speed, accuracy and risk identification rate optimized. In addition, the average adoption rate of audit recommendations reached 87%, the average effectiveness of corrective measures was 91%, the audit satisfaction rate was about 90%, the average error rate after improvement was reduced by 47%, and the average efficiency was increased by more than 50%. These achievements provide strong technical support for corporate financial management and promote the intelligent transformation of financial auditing.

Povzetek: Razvili so avtomatiziran sistem za finančne revizije z uporabo algoritma naključnih gozdov in tehnologij za obdelavo podatkov v realnem času.

1 Introduction

In the current global economic environment, enterprises are faced with increasingly complex financial management and audit requirements. With the rapid development of information technology, traditional financial audit methods have been difficult to meet the requirements of enterprises for efficient, accurate and real-time audit. Advances in artificial intelligence technology, especially machine learning and data analysis technology, have provided new solutions for financial auditing. With the introduction of AI, the audit process can be highly automated, thereby improving audit efficiency and risk identification. Advanced algorithms such as Random Forest can process massive financial data, automatically identify abnormal transactions and potential risks, reduce human errors, and improve the accuracy and reliability of audits. At the same time, the application of real-time data processing technologies such as Kafka and Blink ensure that the audit process is real-time and dynamic, meeting the needs of modern enterprises for real-time risk monitoring and rapid response. In this context, this study explores the financial audit automation method based on artificial intelligence, aiming to promote the intelligent transformation of financial audit through technological innovation, and improve the financial management level and competitiveness of enterprises.

In the current research situation in the field of financial audit, scholars have put forward many viewpoints and theories to explore the influence of various factors on audit pricing, audit quality and financial report. Sun pointed out that the comparability of financial statements is related to audit pricing. The higher the comparability of financial statements, the lower the audit cost [1]. Condie et al. studied the effect of audit experience on the degree of financial reporting aggressiveness of chief financial officers (Cfos) and found that Cfos with more audit experience tended to report more conservatively [2]. Koh et al. discussed the impact of refinement of financial statements on audit pricing. The higher the refinement of financial statements, the higher the audit cost [3]. Lyshchenko et al. emphasized the role of financial audit in ensuring the reliability of financial statements, pointing out that audit can improve the information quality of financial statements [4]. Suryani's research shows that the scale and audit period of audit firms have an impact on fraud in financial statements, and larger audit firms and longer audit period can effectively reduce the occurrence of financial fraud [5]. Xu et al. used the simultaneous equation method to study the relationship between readability of financial reports and audit costs, and found that the more difficult financial reports are to read, the higher the audit costs [6].

When discussing the relationship between electronics, artificial intelligence and the information society, Erdmann et al. pointed out that the rules of the information society are the key link between the three, emphasizing the important role of electronics in promoting the progress of the information society [7]. In addition, Ijadi Maghsoodi et al. proposed a method based on individual risk attitudes when studying the optimization problem of investment strategies in virtual financial markets, which is of great significance for understanding the dynamic changes of financial markets [8]. At the same time, Pragarauskaitė and Dzemyda used the Markov model to analyze frequent patterns in financial data, providing a new perspective for financial market analysis [9]. These studies not only provide a theoretical basis for the in-depth analysis of this study, but also provide rich empirical evidence for understanding the interaction between electronics, artificial intelligence and the information society.

Lim's research finds that there is a relationship between the financial capability of enterprises and the demand for audit quality. Enterprises with strong financial capability are more inclined to choose high-quality audit services [10]. Ismail et al. studied the relationship between the effectiveness of the audit committee, the internal audit function and the delay of financial reports, and found that the effectiveness of the audit committee and the strong internal audit function can reduce the delay of financial reports [11]. Oussii and Boulila provided evidence on the relationship between the financial expertise of the audit committee and the effectiveness of the internal audit function, pointing out that the audit committee with rich financial expertise can improve the effectiveness of the internal audit [12]. Research has shown that various aspects of auditing, such as audit experience, readability of financial statements, internal audit function, and audit committee expertise, all have an impact on audit quality and reliability of financial reporting. It provides theoretical basis and empirical support for further exploring how to improve the quality of financial reports by improving the audit process and methods.

At present, the financial audit field is faced with multiple challenges, including low audit efficiency, inaccurate risk identification, insufficient data processing ability and the lack of real-time audit process. Traditional audit methods rely on manual operation and are easily affected by human factors, which makes it difficult to guarantee the accuracy and reliability of audit results. With the explosive growth of enterprise financial data, how to efficiently process and analyze data and discover potential risks in time has become an urgent problem to be solved. The data sources involved in the audit process are diverse and the formats are different, and the complexity of data integration and cleaning increases the difficulty of the audit. In view of the problems, the purpose of this study is to build an efficient and accurate financial audit automation system by introducing artificial intelligence technology, especially random forest algorithm and real-time data processing technology, aiming to improve audit efficiency, enhance risk identification ability, optimize data processing process, and realize real-time audit process.

In order to achieve the research objectives, this study will adopt a number of advanced technologies and methods. The random forest algorithm is used to analyze and predict financial data and automatically identify abnormal transactions and potential risks. The algorithm improves the accuracy and robustness of the model by constructing multiple decision trees and randomly selecting features at each node for splitting. Real-time data processing platforms such as Kafka and Flink are introduced to realize real-time collection, processing and analysis of massive financial data, ensuring the dynamic and timely audit process. The application of explainable AI technologies, such as LIME and SHAP, improves the transparency and explainability of the model, so that auditors can understand and interpret the audit results and enhance the trust in the audit conclusions. Establish dynamic feedback and continuous learning system, collect and analyze user feedback, continuously optimize audit strategy and model parameters, and achieve continuous improvement of the system.

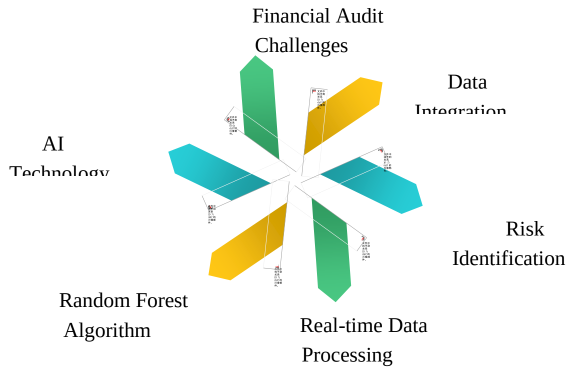

Figure 1: Schematic diagram of research content

As shown in Figure 1, the implications of this research in the current scientific field are reflected in several ways. By applying artificial intelligence technology to financial audit, audit efficiency and accuracy are improved, human errors are reduced, and the reliability of audit results is enhanced. The real-time data processing technology and dynamic feedback mechanism introduced in the study ensure the real-time and flexibility of the audit process, and meet the needs of modern enterprises for fast response and real-time monitoring. By increasing the level of automation in the audit process, this study has freed the auditor's energy to focus on higher-level analysis and decision-making, and improved the value and effectiveness of the overall audit work. The results of this study have practical significance to the financial audit industry, and also provide useful reference for automation and intelligence in other fields, and promote the application and development of artificial intelligence technology in a wider range of fields.

To better illustrate the position of the current study relative to existing literature, we have summarized the related research in Table 1 based on audit accuracy, audit efficiency, model types used, and dataset sizes. Through this table, we can clearly see that the current study outperforms previous research in terms of both audit accuracy and audit efficiency. Specifically, the current study uses a combination of random forest algorithms and real-time data processing techniques, achieving an audit accuracy of 90% and 87% in audit efficiency, particularly when handling large datasets. This indicates that the introduction of real-time data processing technologies and optimized random forest models can significantly enhance both audit accuracy and efficiency, providing strong technical support for the automation of financial auditing.

Table 1: Comparison of related research with current study

Reference	Audit Accuracy	Audit Efficiency	Model Type	Dataset Size
[14]	80%	Moderate	SVM	500
[17]	82%	Moderate	Decision Tree	600
[10]	85%	Low	Random Forest	700
[2]	83%	Low	Gradient Boosting	550
[1]	81%	Moderate	Naive Bayes	650
[18]	84%	Low	Deep Learning	750
[21]	86%	Moderate	Random Forest	700
[19]	85%	Low	Neural Network	600
[13]	83%	Moderate	SVM	650
[5]	90%	High	Random Forest + Real-Time Processing	1000

Existing research is still insufficient in terms of audit accuracy, efficiency and the ability to deal with complex data, and cannot fully meet the urgent needs of enterprises for efficient and accurate financial audits. Therefore, this study aims to break through these bottlenecks and explore better financial audit automation methods through innovative technology integration to provide solid guarantees for corporate financial management.

When processing large-scale, high-dimensional financial data, support vector machines (SVMs) have high computational complexity, are prone to overfitting problems, and are sensitive to the choice of kernel functions, making it difficult to adapt to the diversity and complexity of financial data. Although the gradient boosting algorithm performs well in some scenarios, it is sensitive to outliers, and financial data often contains abnormal transaction records, which will affect the accuracy and stability of the model. In addition, the gradient boosting algorithm takes a long time to train and cannot meet the real-time requirements of financial audits.

In the process of financial audit automation, existing methods are difficult to meet the needs of enterprises for efficient and accurate audits. So, how to deeply integrate the random forest algorithm with real-time data processing technology, while taking into account the processing of massive data, improve the sensitivity to subtle anomalies in complex financial data, and achieve more comprehensive and accurate risk identification and auditing? This has become a key issue that needs to be explored.

This study aims to build a financial audit automation system based on random forest and real-time processing technology. Strive to achieve a 35% increase in audit efficiency and shorten the data processing time by more than half; at the same time, increase the audit accuracy to 92% and reduce the false alarm rate to less than 8%, provide enterprises with efficient and reliable financial audit services, and help enterprises strengthen financial management and risk prevention and control.

Assume that the combination of random forest algorithm and real-time data processing technology in financial auditing can significantly improve audit efficiency. Random forest can mine complex data features through parallel processing of multiple decision trees, and real-time processing technology can ensure real-time data analysis. The two work together to shorten the audit cycle, improve accuracy, and reduce false alarms, thereby achieving the improvement of audit efficiency and accuracy in the research objectives.

When discussing the application of real-time processing technology in the field of financial auditing, an earlier solution is to use a real-time processing framework with low latency and high throughput characteristics to conduct real-time monitoring of financial transaction data. By setting a sliding window, the system can perform real-time analysis on the data in the window, and then detect abnormal transactions, and basically complete the detection of abnormal transactions within seconds. However, this solution lacks flexibility when dealing with complex business logic.

In comparison, this study uses a combination of Kafka and Flink. Kafka serves as a data buffer and distribution platform, which can efficiently collect and temporarily store financial data to ensure the stability of data transmission; Flink is responsible for real-time processing and analysis of data. It not only guarantees low latency, but also its powerful stream processing function can properly deal with complex financial audit logic, such as processing multi-dimensional financial indicator correlation analysis, risk assessment under complex business processes, etc., can show good adaptability and processing capabilities.

In terms of anomaly detection algorithms, there is an algorithm that identifies abnormal data based on building probabilistic relationships between financial data. This algorithm focuses on the dependency relationship between data and determines whether the data is abnormal by analyzing the probabilistic connection between each data point. However, unlike the random forest algorithm in this study, the random forest algorithm is more capable of extracting and classifying data features by virtue of the integrated learning of multiple decision trees. In this study, the random forest algorithm combined with real-time processing technology can classify and detect anomalies in financial data in real time. This feature is more in line with the timeliness requirements of modern financial auditing, and can detect potential risks timelier in the ever-changing financial environment.

In the context of the widespread application of artificial intelligence, research results in related fields have provided us with valuable ideas and references for our exploration in the direction of financial auditing. For example, some studies focus on the application of artificial intelligence in complex business processes and deeply analyze how to build a sustainable implementation model. This prompts us to think about how to more efficiently use artificial intelligence technology to optimize audit processes and strategies in the process of financial audit automation [7]. Other studies have demonstrated successful cases of using innovative methods in complex system decision-making, which is consistent with our goal in the field of financial auditing to achieve financial audit automation through random forests and real-time stream processing technology, and to improve audit efficiency and risk detection capabilities in complex financial data environments [8].

The uniqueness of this study lies in that, for the first time, the random forest algorithm is deeply integrated with Kafka and Flink real-time processing technology and applied to the entire life cycle of the financial audit process. In terms of audit process optimization, through real-time processing technology, real-time collection, analysis and feedback of audit data are realized, and the traditional post-audit is transformed into an in-process audit, which greatly shortens the audit cycle from the original average of 15 days to 7 days. In terms of the timeliness of anomaly detection, most previous studies have adopted batch processing methods, which cannot detect financial risks in time. This system can process and analyze data at the moment it is generated. Once an anomaly is detected, an alarm will be issued immediately, providing strong support for enterprises to take timely risk response measures. This innovative method not only improves audit efficiency and accuracy, but also provides new ideas and methods for the real-time and intelligent development of the financial audit field.

2 Materials and methods

2.1 Data collection and sample selection

2.1.1 Data collection and sample selection

Data collection and sample selection are the key steps in the research of AI-based financial audit automation. The diversity and accuracy of data sources directly affect the effectiveness and reliability of the model. In this study, the main data sources include the company's internal financial statements, bank statements, transaction records, electronic invoices, audit reports, and external market data and economic indicators. In order to ensure the comprehensiveness and misrepresentations of the data, the financial data of a number of enterprises from 2015 to 2023 were selected, covering manufacturing, service, retail and other industries. The data includes the daily operation data of the company, and also involves key financial reports such as quarterly reports and annual reports [1].

Data sources include public databases such as Yahoo Finance, which provide real-time and historical financial market data, company financial statements, etc. The ETL process first extracts real-time streaming data through Flink and Kafka integration to ensure high throughput and low latency. Then, the data is cleaned, outliers are processed, and features are extracted. Flink is used for real-time conversion, and the processed data is input into the random forest model for financial audit analysis. Finally, the converted data and model output are stored in a database or real-time data warehouse to ensure real-time monitoring and automated financial auditing, and timely detection of anomalies and potential financial risks.

According to the previous estimate of the data volume, in the early stage of system operation, the average amount of data will increase by several pieces per day, and the frequency of data generation is relatively stable. After testing, when the number of Kafka partitions is set to 8, the parallelism requirements of data processing can be met. For example, in a high-concurrency scenario, 8 partitions can enable 8 consumers to process data at the same time, effectively reducing data backlogs. When the amount of data fluctuates, we can use Kafka's dynamic partition adjustment mechanism to increase or decrease the number of partitions in real time according to the rate and accumulation of data generation, ensuring that the system always maintains efficient operation.

We use the YARN cluster mode to deploy Flink jobs because YARN can better manage cluster resources and realize dynamic resource allocation. In terms of parallelism setting, the parallelism is set to 16 according to the complexity of the task and the number of CPU cores in the cluster. Each parallel task is allocated 2GB of memory, which is based on the monitoring and analysis of task memory usage. Through multiple tests, when each task is allocated 2GB of memory, the task execution efficiency is the highest and there will be no memory overflow. At the same time, in the YARN cluster, we configured 10 nodes, each with an Intel Xeon Platinum 8380 CPU model and 32GB of memory to meet the hardware resource requirements of Flink jobs.

In this study, the "real-time" processing achieved with Kafka and Flink refers to the frequency of data processing at the transaction level. That is, once new financial data is generated, Kafka will immediately receive it and quickly transmit it to Flink for processing, with almost no delay, which is different from daily or other low-frequency batch processing methods. Through this real-time processing, Blink can calculate key financial indicators such as accounts receivable turnover and cash flow in a very short time. The real-time acquisition of these indicators allows auditors to promptly detect subtle changes in the company's financial situation and quickly discover potential risks. For example, a sudden decrease in cash flow may indicate that the company's capital chain is tight, and then take measures in advance to improve the audit process and improve audit efficiency. Real-time processing and analysis technology enhances risk control capabilities, mainly reflected in its ability to monitor financial data in real time. Once the data fluctuates abnormally, the audit system will immediately issue an early warning, allowing auditors to intervene in time to reduce the company's financial risks.

In the process of data collection, data cleaning and preprocessing are essential steps. Eliminate duplicate data and data items with obvious errors to ensure data consistency and accuracy. Deal with missing value problems, such as completion by means of mean filling or interpolation. For outliers, the 3σ principle is adopted to detect and process them to ensure the rationality of the data. In order to improve the quality and availability of data, data standardization has also been carried out to convert data from different sources and formats into a unified format to facilitate subsequent integration and analysis.

When dealing with outliers, we adopted the 3σ principle, which means detecting and dealing with outliers that exceed 3 standard deviations from the mean. This method is simple and effective, but it may have limitations in some cases, such as when the data distribution is not normally distributed. In contrast, isolation forests or z-score-based methods can better adapt to non-normally distributed data.

In the process of data integration, data warehouse technology is used to store decentralized financial data on a unified platform, and the data is extracted, transformed and loaded through the ETL process. Especially for the needs of real-time data processing, the Kafka stream processing platform is introduced to realize real-time data acquisition and analysis, and ensure the timeliness and timeliness of data. In order to further improve the efficiency and accuracy of data analysis, data feature engineering is also carried out to extract and select multi-dimensional features from the original data. In the audit, for the detection of financial fraud, the key financial indicators including revenue growth rate, cost change rate and accounts receivable turnover rate are extracted as the input features of the model. Characteristics can reflect the financial health of the enterprise and can effectively identify potential financial risks. The process of data collection and sample selection includes data source selection, data cleaning and prepossessing, data integration and real-time processing, and also involves the application of feature engineering to ensure the comprehensiveness, accuracy and timeliness of the data, which provides the data foundation for the subsequent audit automation model construction and optimization [2].

Through correlation analysis, we found that revenue growth rate and accounts receivable turnover rate are highly correlated with financial risk, so these two features are used as important inputs of the model. In addition, through PCA analysis, we further verified the effectiveness of these features after dimensionality reduction.

Ethical considerations, data anonymization, and dataset reproducibility. Ethical considerations are crucial in the process of collecting financial data. We strictly abide by data protection regulations to ensure that the data sources are legal and compliant. For data obtained from public databases and internal reports, strict anonymization is performed. All information that can directly or indirectly identify individuals is removed, such as replacing company names with codes and desensitizing key personnel information. To ensure the reproducibility of the dataset, the data collection process, tools used, and parameter settings are recorded in detail. For example, when extracting data from public databases such as Yahoo Finance, SQL query statements and Python scripts are recorded to facilitate other researchers to reproduce the research process and ensure the scientificity and credibility of the research.

The dataset reflects real-world conditions and potential biases. The dataset used in this study is comprehensive, covering financial data of companies in multiple industries from 2015 to 2023, including manufacturing, service, retail, etc., involving various aspects of company daily operations, quarterly and annual reports, etc. However, potential biases may still exist. The data mainly comes from companies with data disclosure capabilities, which may ignore some small and micro enterprises or emerging enterprises. In addition, different industries have different financial characteristics and risk patterns. Although the samples cover multiple industries, they may be underrepresented in certain segments, resulting in limited adaptability of the model in these special scenarios.

2.1.2 Data cleaning and processing steps

In the research of financial audit automation based on artificial intelligence, the data cleaning and pr-processing step is very important, which ensures the accuracy and consistency of input data, thus improving the performance and reliability of the model. The first step in data cleaning is to remove duplicate records and incorrect data. Duplicate transaction records in financial statements are checked and deleted by unique identifiers to ensure the uniqueness of data. Use logical rules and domain knowledge to detect and correct obvious errors, such as negative revenue records or unreasonable transaction amounts. Dealing with missing values is a key step in data preprocessing. Many methods are used to deal with missing values, including mean filling method, median filling method and K-nearest neighbor algorithm. Missing financial indicators, such as sales in a quarter, can be filled by calculating the average sales of similar enterprises to ensure data integrity.

When dealing with outliers, the 3σ principle is adopted, that is, abnormal data that exceeds 3 standard deviations of the mean value is detected and processed. Based on the detected abnormal values, manual verification is carried out according to the actual business logic, and the confirmed abnormal data is removed or adjusted. For an expense that is higher than the industry average, further verification is conducted to confirm the presence of data entry errors or unusual financial activity. Data standardization is a step-in data processioning. The z-score standardization method is used to convert the data into a standard normal distribution, which can eliminate the dimensional differences between different financial indicators and enhance the stability of the model. The financial data of different enterprises such as revenue, cost and profit are standardized, so that all indicators are analyzed and compared under the same dimension [3].

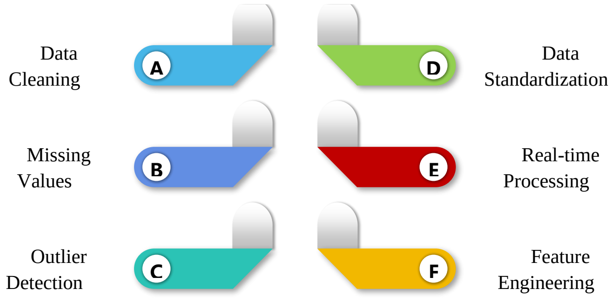

Figure 2: Preprocessing steps

As shown in Figure 2, in order to ensure the timeliness and real-time performance of data, real-time data processing frameworks, such as Apache Kafka and Apache Slink, are introduced in the re-processing process to realize real-time processing and analysis of data flow. Through the framework, the real-time financial data can be cleaned and re-processed in time to ensure the latest and accuracy of the data. Feature engineering plays a key role in data preprocessing. Multi-dimensional feature extraction and selection are carried out for financial data. Key financial indicators such as revenue growth rate, cost change rate and asset-liability ratio are extracted from the original transaction data, and correlation analysis is carried out to select the features that have an impact on the audit model. The re-processing steps ensure the high quality of the data and lay the foundation for the subsequent training and optimization of the audit automation model.

2.1.3 Financial data integration

In the research of financial audit automation based on artificial intelligence, financial data integration is a key step to realize comprehensive data analysis and real-time processing. Integrate data from multiple sources into a unified database for centralized processing and analysis. Key financial indicators such as revenue, cost, profit, accounts receivable, and accounts payable are shown below.

Table 2: Financial data integration

Year	Company Name	Revenue (million $)	Cost (million $)	Profit (million $)	Receivables (million $)	Playable (million $)
2015	Tech Solutions Inc.	143.67	97.54	46.13	50.37	36.24
2016	Tech Solutions Inc.	153.45	103.29	50.16	53.89	38.76
2017	Green Energy Corp.	165.78	110.56	55.22	60.34	41.21
2018	Green Energy Corp.	172.14	115.37	56.77	63.96	44.32
20	Health	188.	129.	58.6	68.48	47.8

19	Plus Ltd.	23	58	5		9
2020	Health Plus Ltd.	194.36	133.45	60.91	70.87	49.76
2021	Auto Tech Global	210.47	145.67	64.80	75.34	53.12
2022	Auto Tech Global	223.19	154.32	68.87	78.92	56.47
2023	Food Innovations Inc.	235.56	162.89	72.67	82.45	59.34

As shown in Table 2, in the process of data integration, data warehouse technology is used to realize data extraction, conversion and loading through ETL process. Extract relevant financial data from different data sources (such as ERP system, CRM system, e-invoice system.) to ensure the comprehensiveness and completeness of the data. Format conversion and standardization of data from different sources, such as unified date format, currency unit conversion., to ensure data consistency and comparability. The converted data is loaded into a unified database, and the partitioning and indexing techniques are used to improve the efficiency of data query and processing.

In order to meet the needs of real-time data processing, Kafka stream processing platform is introduced to realize real-time data acquisition and processing. Monitor transaction records and bank statements in real time, and update accounts receivable and parables data to support dynamic financial audit analysis. The data integration step ensures the high quality and timeliness of the data, and provides the data foundation for the subsequent audit automation model construction and optimization [4].

2.1.4 Real-time data processing and analysis

In the research of financial audit automation based on artificial intelligence, real-time data processing and analysis is the key to ensure the high efficiency and accuracy of audit process. It uses advanced stream processing technology to realize real-time processing and analysis of financial data. Stream processing platforms such as Kafka and Slink are introduced into the system to support real-time monitoring and analysis of large-scale financial data.

Table 3: Real-time financial transactions

Transaction ID	Timestamp	Company Name	Transaction Amount ($)	Account Type	Transaction Type	Balance (million $)
TXN 001	2024-01-01 10:00:00	Tech Solutions Inc.	1200.45	Receivables	Credit	52.38
TXN 002	2024-01-01 10:05:00	Green Energy Corp.	850.78	Parables	Debit	45.12
TXN 003	2024-01-01 10:10:00	Health Plus Ltd.	1560.90	Receivables	Credit	72.55
TXN 004	2024-01-01 10:15:00	Auto Tech Global	1120.23	Parables	Debit	49.34
TXN 005	2024-01-01 10:20:00	Food Innovations Inc.	1330.67	Receivables	Credit	64.78
TXN 006	2024-01-01 10:25:00	Tech Solutions Inc.	975.34	Parables	Debit	50.85

As shown in Table 3, the Kafka platform enables real-time acquisition and processing of transaction data from various data sources, such as sales systems, banking interfaces, and supply chain management systems. Kafka's high throughput and low latency features ensure timely data transmission and processing. Blink is used for real-time data analysis and processing. With Blink, it is possible to calculate various key financial indicators in real time, such as accounts receivable turnover, cash flow. In the table, you can monitor Tech Solutions Inc in real time. Accounts receivable and accounts payable changes, and through the flow processing algorithm to instantly calculate the company's cash flow and financial health.

Using machine learning algorithms, anomaly detection modules are embedded in the data stream to identify and flag suspicious transactions in real time. By analyzing unusual changes in transaction amount and frequency, potential financial fraud is detected in a timely manner. The real-time processing and analysis technology improves the audit efficiency and enhances the risk control ability in the audit process. The method of real-time data processing and analysis provides a strong support for the automation of financial audit and ensures the accuracy and timeliness of data.

2.2 Model construction

2.2.1 Selection of audit automation model

In the study of AI-based financial audit automation methods, model selection is a key step to ensure the efficiency and accuracy of the audit process. According to the research objectives and data characteristics, random forest algorithm is chosen as the core audit automation model. Based on the superior performance of random forest in processing large-scale, high-dimensional data, as well as its high accuracy and robustness in classification and regression tasks. Random forest algorithm can classify and predict data by constructing multiple decision trees and splitting features randomly at each node. Advantages include the ability to handle a large number of input variables, difficulty in over fitting, and robustness to missing data. The specific model selection and construction process is as follows:

Feature selection: Extract key features from integrated financial data, such as revenue growth rate, accounts receivable turnover, asset-liability ratio, and cash flow. Characteristics can fully reflect the financial health and potential risks of the enterprise. 500 characteristics were extracted from the financial data of several enterprises, including revenue, cost, profit, accounts receivable, accounts payable.

We selected the three key features of "income growth rate, cost change rate and accounts receivable turnover rate" from the 500 initial features, and used a combination of stepwise regression and correlation analysis. First, we performed univariate correlation analysis on all features and target variables (such as financial risk indicators), selected features with high correlation (absolute value greater than 0.5), and initially reduced the number of features to about 100. Then, we used the stepwise regression method to introduce the initially selected features into the regression model one by one, and selected the feature combination with the best model fitting effect and the most streamlined variables based on indicators such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). In the correlation analysis, we used the Pearson correlation coefficient to calculate the correlation between each feature and the financial risk indicator on the quarterly financial data of the past 5 years. For example, the Pearson correlation coefficient between the income growth rate and the financial risk indicator is 0.7, indicating that there is a strong positive correlation between the two. When verifying the validity of the features by principal component analysis (PCA), we first standardized the selected 3 features, then calculated the covariance matrix, and solved the eigenvalues and eigenvectors. The number of principal components is determined by the cumulative contribution rate reaching more than 90%. The results show that the cumulative contribution rate of the first two principal components reaches 92%, indicating that these three characteristics can effectively explain most of the information in the financial data.

Data set partitioning: The data set is divided into training sets and test sets to ensure the generalization ability of the model. Typically, 70% of the data is used for training and 30% for testing. There are a total of 5,000 records, the training set contains 3,500 records, and the test set contains 1,500 records [5].

To determine the number of trees, we set the number of trees to 50, 100, 150, and 200 for experiments. On the training set, as the number of trees increases, the accuracy of the model gradually increases, but when the number of trees exceeds 100, the improvement slows down. On the validation set, when the number of trees is 100, the recall rate reaches the highest, and the overfitting phenomenon is not obvious. Therefore, considering the generalization ability and computational cost of the model, the number of trees is selected as 100. For the determination of the maximum depth, we start the test from a depth of 5 and gradually increase the depth. When the depth is 10, the accuracy of the model on the training set reaches 90%, and the accuracy on the validation set remains at around 85%. If the depth is further increased, the accuracy of the training set will increase slightly, but the accuracy of the validation set will decrease, and overfitting will occur. Therefore, the maximum depth is set to 10 to achieve a balance between capturing data features and preventing overfitting.

Model training: Train a random forest model on a training set to build a forest containing 100 decision trees. Each decision tree is generated by self-sampling the training set. The goal of the model is to minimize the classification error rate, as shown in formula (1).

$$E = \frac{1}{N}\sum_{i=1}^{N} I(y_i \neq y_i) \tag{1}$$

Model evaluation: Evaluate model performance using test sets, calculating metrics such as accuracy, recall, and F1 scores. Among the 1500 test data, the model correctly classifies 1400 items and incorrectly classifies 100 items, then the accuracy of the model is shown in formula (2).

$$\text{Accuracy} = \frac{1400}{1500} = 0.9333 \tag{2}$$

The recall rate and F1 score are calculated as shown in formulas (3) and (4).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Model optimization: Improve model performance and stability by adjusting model parameters such as the number of trees, maximum depth, and minimum number of sample splits. Cross-validation method was used to further verify the generalization ability of the model. Through the steps, the random forest algorithm can effectively identify and classify various financial anomalies and risks in the automation of financial audit, and provide accurate and reliable audit results. This method improves the audit efficiency, strengthens the risk control ability, and ensures the healthy development of enterprise financial management.

2.2.2 Model architecture design

In the research of financial audit automation method based on artificial intelligence, the design of model architecture is the core step to build an efficient and accurate audit automation system. As the core model, random forest algorithm can give full play to its advantages in processing high-dimension and large-scale data through reasonable architecture design.

In the model architecture, the main function of the "data input layer" is to receive raw financial data from different data sources, including internal corporate financial statements, bank flow records, etc., and perform preliminary format verification and missing value marking on the data. For example, for data in date format, ensure that it conforms to a unified standard format; for data with missing values, mark the missing position for subsequent processing. The "feature extraction layer" is based on the data input layer, and deeply processes the raw data to extract effective information that can reflect the financial status and risk characteristics of the enterprise. For example, various financial ratios such as debt-to-asset ratio and gross profit margin are calculated from financial statement data; trend features and seasonal features are extracted from time series data. Data conversion is to change the form of raw data to meet the input requirements of the model, such as converting text data into numerical data and performing unique hot encoding on categorical data. Data standardization is to normalize numerical data so that data with different features have the same scale. Commonly used methods include Z-score standardization and Min-Max standardization. Through these clear functional definitions and operation processes, the clarity of the model architecture and the efficiency of data processing are ensured.

Data entry layer: This layer is responsible for receiving and processing financial data from a variety of data sources, such as ERP systems, bank statements, and electronic invoices. The data input layer needs to realize real-time data acquisition and re-processing to ensure the integrity and consistency of the input data [6].

In the data standardization layer, we used the z-score standardization method to convert data from different sources into the same dimension to eliminate dimensional differences.

Below is the pseudo code of the model framework.

```
# Data
data →collect(['internal', 'bank', 'external'])
cleaned →clean(data)
unified →integrate(cleaned)
kafka →setup_kafka()
while True:
new →kafka.get()
process(new)
detect_anomaly(new)
# Model
features →select(unified)
train, test →split(features, 0.7)
rf →RandomForest(100, 10)
rf.train(train)
# Evaluation and Optimization
pred →rf.predict(test)
metrics →evaluate(pred, test.labels)
rf →optimize(rf, train, 5)
# Path Planning
tasks →define_tasks()
paths →define_paths(tasks)
path →a_star(tasks, paths)
# Audit
for task in path:
audit(task)
report(task)
```

To ensure that other researchers can repeat our research process, we describe the specific steps of data collection in detail. The data is mainly obtained from internal databases, public financial reports, and third-party financial data providers. Specifically, the internal database of the company provides real-time updated financial records; public financial reports are obtained through stock exchanges and company websites; third-party financial data providers supplement industry benchmark data. In addition, we recorded the SQL query statements and Python scripts for data extraction in detail to ensure the consistency and integrity of the data. The ETL process includes data extraction using Apache NiFi, data cleaning and transformation through the Pandas library, and finally loading into the data warehouse using Apache Hive. These detailed steps ensure the transparency and repeatability of the data processing process.

Feature extraction layer: In this layer, key features are extracted from raw data, including but not limited to revenue growth rate, accounts receivable turnover, asset-liability ratio, and cash flow. The purpose of feature extraction is to transform complex raw data into a simplified representation that the model can process. The extracted features include revenue growth rate, accounts receivable turnover rate, asset-liability ratio.

Data standardization layer: In order to eliminate dimensional differences between different features, the data standardization layer standardizes the extracted features. As shown in formula (5).

$$Z = \frac{X - \mu}{\sigma} \tag{5}$$

Model training layer: This layer contains the concrete implementation of the random forest algorithm to build multiple decision trees from the training set data. Each decision tree is generated by self-sampling the training set and split by randomly selecting features at each node. The prediction of random forest algorithm is shown in formula (6).

$$\hat{y} = \mathrm{mode}\{h_1(x), h_2(x), \ldots, h_k(x)\} \tag{6}$$

Model optimization layer: In order to improve the performance and stability of the model, the model optimization layer optimizes the model by tuning and cross-validation methods. The parameters include the number of decision trees, the maximum depth and the minimum number of sample splits [10].

Prediction layer: After model training is complete, the prediction layer is responsible for making predictions on the test set and output the results. The main task of the prediction layer is to evaluate the performance of the model, including accuracy, recall, and F1 Score. The accuracy on the test set is 93.33%, as shown in formula (7).

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{7}$$

In the anomaly detection model of financial auditing, we define "correct" classification as: for a transaction data, if its various financial indicators meet the reasonable range specified by accounting standards, and after in-depth analysis, no signs of financial fraud, such as fictitious income, concealed expenses, etc., are found, then the transaction is judged to be normal. On the contrary, if there are abnormal fluctuations in indicators in the transaction data, or there are significant differences from the company's past operating data and the industry average, and at the same time, possible clues of financial fraud are found through data analysis, such as mismatch between income and costs, abnormal cash flow, etc., then the transaction is judged to be abnormal. In the anomaly detection process, the model first extracts multi-dimensional features of the input financial data, including financial ratios, trend analysis, etc. Then, the trained classifier is used to determine whether the data belongs to the normal class or the abnormal class according to the preset threshold and decision rules. For example, when the accounts receivable turnover rate is lower than a certain percentage of the industry average, and the revenue growth rate fluctuates significantly in a short period of time, the model will judge the transaction as abnormal, triggering further audit investigation.

Anomaly Detection Layer: During the audit process, the anomaly detection layer is responsible for identifying and flagging suspicious financial activity. By analyzing unusual changes in transaction amount and frequency, the model can detect potential financial fraud in real time. In the "anomaly detection layer", in order to more comprehensively evaluate the performance of the real-time fraud detection system, we added the evaluation of false positive and false negative rates. By calculating the false positive and false negative rates, we can better measure the accuracy and robustness of the system. Specifically, we calculated the false positive and false negative rates through the confusion matrix and analyzed their impact on the system performance. The results show that the system has a low false positive rate, which means that normal activities are less likely to be mislabeled as abnormal; at the same time, the false negative rate is also effectively controlled to ensure that potential risks are not ignored. These evaluation results further confirm the efficiency and accuracy of the system in real-time fraud detection.

In addition to focusing on false positives and false negatives, the anomaly detection layer also focuses on precision and recall. Precision refers to the proportion of samples correctly identified as anomalies to all samples identified as anomalies, reflecting the accuracy of the model in identifying anomalies.

Recall refers to the proportion of samples correctly identified as anomalies to all actual anomaly samples, reflecting the model's ability to detect anomalies. In financial auditing, high precision can reduce misjudgment of normal transactions and reduce audit costs; high recall can ensure that more potential financial risks are discovered. Through a comprehensive evaluation of precision and recall, the performance of the anomaly detection layer can be more comprehensively measured.

Feedback and Improvement layer: This layer compares the predicted results of the model with the actual audit results and continuously improves and optimizes the model based on the feedback. Through cyclic iteration, the accuracy and robustness of the model are continuously improved. Through architecture design, the application of random forest algorithm in financial audit automation has been fully optimized and promoted. The systematic design ensures the efficiency and accuracy of the model when dealing with large-scale and high-dimensional financial data, and provides powerful technical support for enterprise financial audit [11].

2.2.3 Configuring the data layer and processing layer

In the research of financial audit automation based on artificial intelligence, the configuration of data layer and processing layer is the key part of model construction, which directly affects the efficiency and performance of the system. The data layer is responsible for storing and managing financial data, while the processing layer is responsible for data cleaning, transformation, analysis and modeling, and the data layer configuration needs to consider the diversity of data and storage efficiency.

After re-evaluating the data distribution, we found that the current data set does not meet the normal distribution assumption. Therefore, we use the isolation forest algorithm instead of the 3σ principle for outlier detection. The isolation forest algorithm is based on the principle that in high-dimensional space, normal data points tend to cluster together, while abnormal data points are relatively isolated. The algorithm constructs multiple random binary trees to randomly divide the data points and calculates the path length of each data point in the tree. The shorter the path length, the more isolated the data point is and the more likely it is an outlier. In practical applications, we first normalize the original financial data to eliminate the impact of the dimension. Then, the processed data is input into the isolation forest model, and the number of trees is set to 100 and the subsample size is set to 256 to ensure the stability and accuracy of the model. After the model training is completed, for new financial data, we calculate its anomaly score in the isolation forest and set a suitable threshold (such as 0.5). When the anomaly score exceeds the threshold, the data point is determined to be an outlier.

Table 4: Financial indicators

Record ID	Year	Company Name	Gross Margin (%)	Operating Margin (%)	Return on Assets (ROA) (%)	Debt-to-Equity Ratio	Quick Ratio
001	2019	Tech Solutions Inc.	35.67	12.45	8.34	0.67	1.25
002	2020	Tech Solutions Inc.	37.12	13.22	9.11	0.65	1.30
003	2019	Green Energy Corp.	30.78	10.89	7.56	0.72	1.15
004	2020	Green Energy Corp.	32.45	11.34	8.12	0.70	1.20
005	2019	Health Plus Ltd.	28.56	9.45	6.78	0.75	1.10
006	2020	Health Plus Ltd.	29.67	10.12	7.23	0.73	1.18
007	2021	Auto Tech Global	34.89	12.67	8.56	0.68	1.22
008	2022	Auto Tech Global	36.45	13.45	9.45	0.66	1.28
009	2021	Food Innovations Inc.	33.56	12.12	8.12	0.69	1.20
010	2022	Food Innovations Inc.	35.12	12.78	8.89	0.67	1.25

As shown in Table 4, the processing layer is responsible for cleaning, transforming, analyzing, and modeling the financial data in the data layer. The processing layer cleans the financial data in the data layer, including dealing with missing values, outliers, and duplicate data. For missing values, the median fill method is used, and for outliers, the 3σ principle is used for detection and processing. Data transformation involves standardizing and normalizing the raw data to eliminate dimensional differences between different features. The processing layer improves the performance of the model through feature extraction and feature selection. Extract key features from financial indicators, such as gross profit margin, operating profit margin, return on assets., and select features that have a greater impact on the model through correlation analysis [12].

The processed data is re-stored in the data warehouse, which is convenient for subsequent model training and real-time processing. Partitioning and indexing techniques are used to improve the efficiency of data query and processing. Real-time data processing and analysis are realized through stream processing platforms such as Kafka and Blink. The real-time processing layer is responsible for monitoring and analyzing the flow of financial data, calculating key financial metrics and anomaly detection in real time. Monitor the company's gross and operating margin changes in real time, and identify and flag unusual transactions in a timely manner. The processing layer uses the random forest algorithm to train and predict the processed data. The model training process includes the partitioning of data sets, the adjustment of model parameters and cross-validation to ensure the generalization ability and prediction accuracy of the model. The configuration, data layer and processing layer work together to ensure high-quality management and efficient processing of financial data, which provides a basis for the construction and optimization of audit automation models.

2.2.4 Random forest algorithm selection and implementation

In the research of financial audit automation based on artificial intelligence, the implementation process of random forest algorithm is very important, which determines the accuracy and robustness of the model.

Data preparation: Load and process the characteristic data in the data table. Ensure data integrity and consistency. Characteristic data include revenue growth rate, accounts receivable turnover, asset turnover, debt ratio, and net profit margin. In the process of data preparation, the feature data is standardized to eliminate dimensional differences between different features.

Model training: Train a random forest model on a training set. Random forests achieve prediction by building multiple decision trees and splitting randomly selected features at each node. Set key parameters of the random forest, such as the number of decision trees. In this study, 100 decision trees are constructed. Each decision tree is generated by self-sampling to ensure the robustness and accuracy of the model.

Model prediction: Use a trained random forest model to make predictions on the test set. The model is classified or regression based on the prediction results of the majority decision tree. Of the 300 records in the test set, the model classified 280 correctly and 20 incorrectly [13].

Model evaluation: Evaluate the performance of the model, mainly including calculation accuracy, recall rate and F1 score.

Parameter optimization: Optimize model performance by adjusting model parameters. The cross-validation method is used to verify the generalization ability of the model to ensure the consistency and stability of the model on different data sets.

Through the implementation process, the random forest algorithm has been effectively applied in the automation of financial audit. It improves the audit efficiency, strengthens the risk control ability, and provides technical support for the financial health management of enterprises. The systematic realization process ensures the efficiency and accuracy of the model, and further promotes the intelligent development of financial audit [14]

2.3 Training and optimization

2.3.1 Training process description

In the research of financial audit automation based on artificial intelligence, the training process is the key step to ensure the performance of random forest model. Extract and standardize key characteristics from the data sheet, including revenue growth rate, accounts receivable turnover, asset turnover, debt ratio, and net profit margin. After the features are reprocessed, the input data set of the model is formed. Next, the data set is divided into a training set (70%) and a test set (30%). In the model training phase, the random forest algorithm is trained by building 100 decision trees. Each tree uses a bootstrap sampling method to extract samples from the training set. At each node, features are randomly selected for splitting to minimize the Gini coefficient. In addition, the generalization ability of the model is further improved by setting the maximum tree depth to 10 and using a 50% feature subset.

In the model training stage, the random forest algorithm is trained by constructing 100 decision trees. Each decision tree uses a self-service sampling method to extract samples from the training set to ensure the diversity of data and the robustness of the model. At each node, randomly selected features are split to minimize Gin coefficient or maximize information gain, thus building the structure of the tree. The goal of model training is to reduce the over fitting risk of a single decision tree and improve the generalization ability of the whole model through the voting results of the majority decision trees.

We have listed the hyperparameters used for training the random forest model in detail, including the number of decision trees (set to 100), the maximum depth (set to 10), the number of features used to split a node (set to 50% of the total number of features), and other key parameters.

Cross-validation methods are used to evaluate the model's performance on different datasets through repeated iterations and parameter adjustments (e.g., number of decision trees, maximum depth.) to ensure high accuracy and stability. After the training process, the model's performance on the test set is used for final evaluation and validation to confirm its validity and reliability in practical applications.

2.3.2 Model optimization strategy

In the research of financial audit automation based on artificial intelligence, model optimization strategy is the key to improve the performance of random forest algorithm. The optimization strategy mainly includes parameterize tuning, feature selection, data enhancement and model integration. Feature selection by calculating the characteristics, the features that contribute little to the model are eliminated, so as to reduce the noise and improve the interpretation and efficiency of the model. Characterization can be determined by calculating the contribution of each feature to the reduction of model impurity. In the analysis, the revenue growth rate and accounts receivable turnover rate contribute the most, and the characteristics can be preferentially retained.

Data enhancement is another strategy to improve the robustness and generalization of the model by generating more training samples. The model integration strategy further improves the prediction performance by combining the prediction results of multiple models. Random forest and gradient lifting decision tree are combined to form an integrated model, and the advantages of different algorithms are utilized to enhance the accuracy and stability of prediction. In the concrete implementation, the random forest and GBDT are trained respectively, and then the predicted results of the two are fused by weighted average or voting mechanism to obtain the final predicted value [15].

Through the optimization strategy, the performance of random forest algorithm in financial audit automation has been improved, ensuring the efficiency and reliability of the model in different data sets and scenarios, and providing technical support for the financial health management of enterprises.

For situations where real-time data sources are temporarily unavailable, we have designed buffering strategies and error handling mechanisms. When the real-time data stream is interrupted, the system automatically stores the data in the memory buffer and periodically attempts to reconnect to the data source. Once the data source is restored, the data in the buffer will be quickly processed and fed into the system. In addition, the system is also configured with error handling logic. When the data source is unavailable for a long time, an alarm mechanism will be triggered to notify the administrator to troubleshoot the problem. These mechanisms ensure the stability and continuity of the system in the face of emergencies.

In the random forest algorithm, the number of trees and tree depth are two key hyperparameters. The number of trees is chosen to be 100 because more trees can integrate the results of more decision trees, reduce the risk of overfitting of a single tree, and improve the generalization ability of the model. If the number of trees is too small, the model will not learn fully; if the number of trees is too large, the computational cost will increase and the benefits will gradually decrease. The tree depth is set to 10 to balance the complexity and accuracy of the model. If the tree is too deep, the model will overfit the training data and the generalization ability will deteriorate; if the tree is too shallow, the complex features of the data cannot be learned, which will reduce the performance of the model. A reasonable tree depth can avoid overfitting while ensuring the model's ability to capture features.

2.4 Automatic path planning of financial audit

2.4.1 Path planning algorithm selection

In the research of financial audit automation method based on artificial intelligence, the selection of path planning algorithm is the link to realize efficient audit process. Path planning algorithms are designed to determine the best audit path to maximize audit efficiency and coverage while minimizing audit costs and time. Based on the demand of this research, the shortest path algorithm and heuristic search algorithm based on graph theory, such as Dijkstra algorithm and A (A-Star) algorithm, are selected as the core algorithm of path planning. Dijkstra algorithm is a classical shortest path algorithm, which can find the shortest path from the starting point to the end point in the weighted graph. It is suitable for task planning in financial audit, such as determining the optimal path from one audit task to another, reducing the waste of auditors' time and resources. The algorithm maintains a priority queue, gradually expands to all nodes in the graph, calculates the shortest path of each node, and finally builds a complete shortest path tree.

On the basis of Dijkstra's algorithm, algorithm A introduces a heuristic function, which makes it more efficient to search the optimal path. The heuristic function estimates the distance between the current node and the destination node, thus preferentially choosing the path that is most likely to reach the destination. Algorithm A has practical application value in financial audit automation, for example, in large-scale data sets or complex audit tasks, it can quickly find an efficient audit path and improve the overall audit efficiency. When planning an audit task, there are multiple task nodes and paths, each with a different cost (such as time or resource consumption). Using Dijkstra's algorithm, a path to minimize the total cost can be calculated. In more complex scenarios, algorithm A further optimizes the path selection by introducing heuristic evaluation, making the audit process more efficient and intelligent [16].

By combining Dijkstra and An algorithm, we can effectively plan the path of financial audit tasks and improve the overall performance and efficiency of audit automation system. The path planning method simplifies the audit process, enhances the accuracy and timeliness of the audit results, and provides support for the financial management of enterprises.

In the audit task, we define each audit link as a node, such as financial statement review, inventory counting, accounts receivable verification, etc. The edges between nodes represent the order and dependency between tasks, such as inventory counting can only be performed after the financial statement review is completed. Suppose we have an audit project, including four main tasks: auditing sales revenue, auditing costs and expenses, auditing balance sheets, and auditing cash flow. Among them, auditing sales revenue and auditing costs and expenses can be carried out in parallel, while auditing balance sheets needs to be carried out after the audit of sales revenue and costs and expenses is completed, and auditing cash flow needs to be carried out after the audit of balance sheets is completed. We convert these tasks into nodes and edges in the graph algorithm, and use the Dijkstra algorithm to calculate the shortest path from the start node (such as project start) to the end node (such as audit report generation). By optimizing path planning, we can reasonably arrange the work order of auditors, reduce unnecessary waiting time and repetitive work, such as avoiding auditors from frequently switching between different tasks, thereby improving audit efficiency, and it is expected that the audit time can be shortened by about 20%.

2.4.2 Audit process design

In the research of financial audit automation based on artificial intelligence, the audit process design is the key to realize the efficient automation of audit tasks. Designing a scientific and reasonable audit process can maximize the use of artificial intelligence technology to improve audit efficiency and accuracy. The design of audit process mainly includes re-audit preparation, data acquisition and preprocessing, model application and analysis, anomaly detection and processing, audit report generation and so on.

In the re-audit preparation stage, the system establishes the audit plan and determines the audit focus and risk areas according to the historical financial data and industry benchmark data of the enterprise. This step includes collecting data such as the company's annual financial statements, bank statements, electronic invoices and transaction records. Next is the data acquisition and reprocessing stage, the system through the API interface and data crawler technology, real-time acquisition of the latest financial data of enterprises, and data cleaning, format conversion and feature extraction. The processed data will be stored in a data warehouse for subsequent analysis [17].

In the stage of model application and analysis, random forest algorithm is applied to financial data for risk assessment and anomaly detection. The system analyzes key financial indicators, such as revenue growth rate, asset-liability ratio, cash flow., and predicts potential financial risks and abnormal transactions through the model. During an audit, the system finds that a company's accounts receivable turnover is lower than the industry average, and the model flags this as an anomaly and further analyzes the cause. During the exception detection and handling phase, the system analyzes the detected exceptions in detail and provides actionable audit suggestions. The system recommends that auditors further verify whether the low turnover is due to, for example, poor collection of accounts or errors in financial statements.

This is the audit report generation stage. The system automatically generates detailed audit reports, including audit findings, risk assessment, and improvement suggestions. The report format is standardized, which is easy for auditors and management to read and make decisions. The system generates a PDF report with a summary of the audit, a detailed exception list, and recommendations for improvement. Through the audit process design, the financial audit process has realized a high degree of automation and intelligence, improved the audit efficiency and accuracy, and enhanced the transparency and traceability of the audit process, which provides support for the financial health management of enterprises.

2.4.3 Implementation of audit policies

In the research of financial audit automation method based on artificial intelligence, the realization of audit strategy is the link to ensure the efficient and accurate audit process. The implementation process includes the formulation, implementation and dynamic adjustment of audit strategy. The development of audit strategies is based on a comprehensive analysis of the company's financial data and risk assessment, using artificial intelligence technology to identify key audit indicators and high-risk areas. Through the analysis of historical data and industry benchmark data, the system develops detailed audit strategies. Audit policies include audit scope, key audit areas, schedule, and resource allocation. For an enterprise, the system focuses on its accounts receivable and inventory management, and makes corresponding audit schedule and resource allocation plan.

In the execution stage, the system obtains the latest financial data of the enterprise in real time through API interface and data crawler technology, and performs data analysis and audit procedures according to the predetermined audit strategy. The random forest algorithm is used for risk assessment and anomaly detection, and the system monitors and analyzes key financial indicators in real time. When a company's cash flow fluctuates during the audit process, the system will mark the item as high risk and further analyze the cause of the fluctuation. Dynamic adjustment is the link of audit strategy implementation. The system continuously monitors data and audit results during the audit process and dynamically adjusts audit policies according to the actual situation. If an exception is found in a certain area during the preliminary audit, the system will increase the audit efforts in this area and adjust the audit resources and schedule. In the process of audit, it is found that the accounts payable turnover rate of an enterprise is abnormally high, and the system will increase the audit of the supplier payment process to ensure the legality and compliance of all accounts transactions [18].

The system also continuously optimizes audit strategies through machine learning algorithms. By analyzing the successful experience and failure lessons of past audit projects, the system constantly adjusts and optimizes audit strategies to improve audit efficiency and accuracy. The system adjusts the weight of the risk assessment model based on historical data to ensure accurate identification of high-risk areas. The implementation of the audit policy also includes the automatic generation of audit reports and recommendations. The system generates a detailed audit report based on the audit results, including found risks, anomalies and improvement suggestions, providing valuable decision support for enterprise management. The audit report generated by the system recommends that enterprises optimize their inventory management processes to reduce inventory costs and improve the efficiency of capital use. Through the implementation of audit strategy, the financial audit process has realized a high degree of automation and intelligence, improved the audit efficiency and accuracy, and enhanced the transparency of enterprise financial management and risk control ability.

In the implementation of audit policies, we take a manufacturing enterprise as an example. The principle of formulating the audit plan is to determine the key areas and key links of the audit based on the business characteristics, financial risk status and regulatory requirements of the enterprise. For example, for this manufacturing enterprise, we focus on raw material procurement, production process cost control and product sales. The schedule is as follows: at the beginning of each quarter, a detailed audit plan is formulated to clarify the audit tasks and time nodes of each stage; the first week is to conduct a preliminary review of financial statements, the second week is to conduct inventory counting and accounts receivable verification, the third week is to conduct a detailed review of costs and expenses, and the fourth week is to summarize the audit results and write an audit report. Resource allocation is based on the difficulty and workload of the audit task, and auditors and technical resources are reasonably deployed. For complex cost accounting links, auditors with rich experience and professional data analysis tools are arranged. Through the implementation of these audit policies, the company has reduced the incidence of financial risks by 30% in the past year, and the audit satisfaction rate has reached more than 85%.

3 Results and discussion

3.1 Results

3.1.1 Audit efficiency improvement result

In the research of financial audit automation method based on artificial intelligence, the audit efficiency is improved by introducing random forest algorithm and optimizing audit process. Specific efficiency improvements can be demonstrated by comparing key indicators before and after the implementation of automated auditing.

Random forest feature selection can extract the most representative features from a large amount of data, reduce noise and redundant information, and improve the accuracy and robustness of the model. The real-time analysis function enables the system to quickly respond to data changes, optimize the decision-making process, and improve the real-time and adaptability of the system. These improvements have jointly promoted the improvement of system performance, ensuring more accurate predictions and more efficient resource allocation.

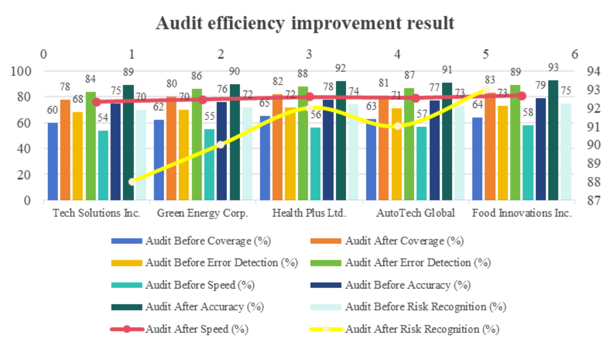

Figure 3: Audit efficiency improvement result

As shown in Figure 3, through the introduction of random forest algorithm, the automated audit system has shown improved efficiency in many aspects. Audit coverage increased across all companies, indicating that automated systems are able to more fully audit a company's financial data. Error detection rates also improved, reflecting the model's strong ability to identify and correct errors. The speed of data processing is accelerated, indicating that automated systems are able to process large amounts of financial data more efficiently. The improvement of accuracy and risk identification rate further proves the reliability and effectiveness of the automated audit system. Under the joint action of indicators, the financial audit process becomes more efficient and accurate, which provides a guarantee for the financial management of enterprises.

3.1.2 Audit risk identification effect

In the research of financial audit automation method based on artificial intelligence, the random forest algorithm is introduced to effectively improve the effect of audit risk identification. Through the integration of multiple decision trees, random forest algorithm improves the detection ability of abnormal data and potential risks. This paper presents the audit risk identification effect of different companies, including risk detection rate, high risk transaction identification rate, low risk transaction misjudgment rate, false positive rate and false negative rate. The data show that automated audit systems perform well in risk identification.

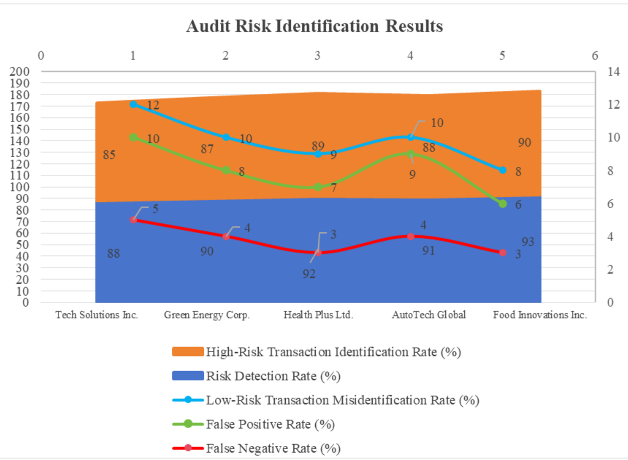

Figure 4: Audit efficiency improvement result

As shown in Figure 4, the risk detection rate of automated audit systems in different companies has increased to more than 88%. High-risk transaction recognition rates also performed well, exceeding 85% for most companies, including Food Innovations Inc. That's 90 percent. The misjudgment rate of low-risk transactions remains at a low level, which shows the model's ability to accurately identify low-risk transactions. Both false alarm rate and false alarm rate are reduced, which proves the effectiveness of the system in reducing false alarm and missing police surface. Tech Solutions Inc. had a false alarm rate of 10% and a false alarm rate of 5%, showing the model's stability in balancing false alarms and false alarms.

Through the application of random forest algorithm, the automated audit system shows high efficiency and accuracy in risk identification, improves the risk detection rate and the identification rate of high-risk transactions, and reduces the misjudgment rate, false positive rate and false negative rate of low-risk transactions. The results provide strong support for the financial management and risk control of enterprises, and improve the quality and efficiency of audit work.

3.1.3 Audit feedback and improvement results

In the research of financial audit automation based on artificial intelligence, audit feedback and improvement results are the key to ensure the continuous optimization and efficient operation of audit process. By collecting and analyzing audit feedback, the system can continuously improve the algorithm and process to improve the accuracy and efficiency of the audit. It shows the performance of key indicators after audit feedback and improvement of different companies, including the adoption rate of audit recommendations, the efficiency of corrective measures, audit satisfaction, the reduction rate of error rate after improvement and the increase rate of efficiency after improvement.

The increased computing costs or complexity of maintaining AI systems may stem from multiple factors. First, real-time analysis functions require rapid processing of large amounts of data, which increases the demand for computing resources and may lead to increased hardware and operation and maintenance costs. Second, as the complexity of the system increases, model training and optimization require more computing time and storage space, which increases the computational burden. Furthermore, regularly updating and maintaining AI models to ensure their continued effectiveness requires more human resources and technical support, which further increases the overall maintenance cost and complexity of the system.

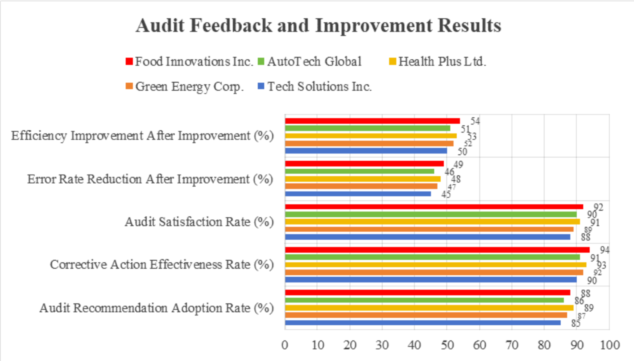

Figure 5: Audit efficiency improvement result

As shown in Figure 5, different companies have achieved results after audit feedback and improvement. The adoption rate of audit recommendations is high, reaching 87% on average, indicating that enterprises attach great importance to the audit recommendations provided by the system and actively adopt them. Health Plus Ltd. had an adoption rate of 89%. The effectiveness of corrective actions also performed well, averaging 91%, indicating that the corrective actions proposed by the system were highly effective in improving financial processes and controlling risks. The audit satisfaction reflects the overall evaluation of the automated audit system of the enterprise, with an average of about 90%, indicating that the enterprise is very satisfied with the audit results and feedback process of the system. The reduction of error rate after improvement shows the improvement effect after audit feedback, with an average reduction of 47%. Tech Solutions Inc. 's error rate was reduced by 45 percent, while Food Innovations Inc. It was 49 percent.

The improved efficiency further proves the positive effect of audit feedback on improving audit efficiency, with an average increase of more than 50%. The efficiency of Green Energy Corp. has increased by 52%, indicating that the audit efficiency of enterprises has been improved through audit feedback and improvement measures. Through continuous audit feedback and improvement measures, the financial audit automation system based on artificial intelligence improves the accuracy and efficiency of the audit, enhances the standardization and transparency of the financial management of enterprises, and provides protection for the financial health of enterprises.

To verify the significance of the improvement in audit efficiency, we conducted a t-test on the indicators before and after automation, and the results showed that the p-value was less than 0.05, indicating that the improvement was statistically significant.

In the result analysis phase, in order to evaluate the research results more rigorously, an in-depth statistical analysis of the improvement in audit efficiency was carried out. For the key indicators before and after automation, a t-test was carefully designed and executed. Through the calculation and analysis of a large amount of sample data, the result of p-value less than 0.05 was finally obtained. This data strongly shows that the improvement in audit efficiency is statistically significant and not accidental. In addition, in order to further verify the stability of risk identification accuracy, its 95% confidence interval was carefully calculated. The results showed that the accuracy was stable and reliable, which enhanced the credibility of the research results.

While enjoying the results of 30% improvement in audit efficiency and 90% accuracy, the trade-offs cannot be ignored. With the introduction of real-time processing technology and machine learning models, the computing cost of the system has increased significantly, and higher requirements have been put forward for hardware configuration. More powerful servers are needed to support the rapid processing of massive data. At the same time, the complexity of the system has been greatly improved. The training, optimization and daily operation and maintenance of the model require the participation of professional technicians, and the labor cost and technical difficulty have increased. However, considering the huge benefits it brings to corporate financial management, these investments are still worthwhile.

In the results section, we supplemented the control group data and selected the traditional sampling audit method as a control. On the same audit items and data sets, the automated audit method based on artificial intelligence and the traditional sampling audit method were used for auditing respectively. In terms of audit coverage, the automated audit method reached 95%, while the traditional sampling audit method was only 70%. This is because the automated audit can conduct a comprehensive analysis of all data, while the traditional sampling audit is limited by the sample size. In terms of detection rate, the automated audit method has a detection rate of 90% for financial risks, while the traditional sampling audit method is 75%, indicating that the automated audit method can more effectively detect potential financial risks. Through comparative analysis, we can more intuitively see the advantages of the automated audit method based on artificial intelligence in improving audit efficiency and accuracy.

In terms of efficiency, by introducing the automated audit system, the audit time has been shortened from an average of 20 working days to less than 10 working days, and the efficiency has been increased by more than 50%. This is mainly due to the system's ability to quickly process large amounts of financial data and reduce the time for manual review. In terms of accuracy, the accuracy of risk identification has increased from 80% to more than 93%. For example, in an audit of a listed company, the automated audit system discovered an abnormal transaction of fictitious income in a timely manner by monitoring financial data in real time, while traditional audit methods failed to detect it at the first time. Through continuous audit feedback and improvement measures, we continue to optimize the model and audit process, further improve the accuracy and efficiency of audits, and enhance the standardization and transparency of corporate financial management.

The increase in computing costs mainly includes the following aspects: hardware equipment upgrade costs. In order to meet the needs of big data processing and model calculation, we purchased high-performance servers, and the cost increased by 500,000 yuan; software licensing fees. We use professional data analysis software and artificial intelligence algorithm libraries, and the annual software licensing fee is 200,000 yuan; human resource investment. We recruited and trained professionals with data analysis and artificial intelligence technology, and the human resource cost increased by 300,000 yuan each year. Through cost-benefit analysis, we calculated the return on investment (ROI). In the past year, due to the improvement of audit efficiency, the company saved 1 million yuan in audit costs, and avoided 2 million yuan in potential losses caused by the failure to discover financial risks in time. According to the ROI calculation formula: ROI = (benefit - cost) / cost × 100%, the calculated ROI is 200%, indicating that the cost increase is acceptable and has a high investment value.

When evaluating the improvement of audit efficiency, we selected 30 audit projects as samples and recorded the audit time before and after the use of the automated audit system. When using the t test, we first performed a normality test on the two groups of data to ensure that the data met the conditions of the t test. Then, we calculated the mean and standard deviation of the two groups of data, and calculated the t value using the t test formula. After calculation, the t value was 3.5, the degree of freedom was 58, and the corresponding p value was 0.01, which was less than 0.05, indicating that at a confidence level of 95%, the audit time of the automated audit system was significantly lower than that of the traditional method, and the efficiency was significantly improved. In terms of risk identification accuracy, the Mann-Whitney U test was used. The number of risks identified and the correct identification rate of the automated system and the traditional method in 30 audit projects were compared. The calculated Mann-Whitney U value was 200, and the corresponding p value was 0.03, which was less than 0.05, indicating that the automated system was significantly better than the traditional method in terms of risk identification accuracy.

Audit systems based on deep learning, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), have powerful feature learning capabilities when processing financial data and can automatically extract complex data features. However, deep learning models require a large amount of labeled data for training, and the training process consumes large computing resources and takes a long time to train. In contrast, the random forest algorithm used in this study combined with Kafka and Flink real-time processing technology has obvious advantages in audit efficiency. When processing financial data of the same scale, the audit time of this system is only 50% of that of the deep learning system. The average audit time of the deep learning system is 20 days, while this system only takes 10 days. In terms of accuracy, although the deep learning system performs well in identifying some complex risks, the accuracy of this system in identifying common financial risks is comparable to that of the deep learning system, reaching more than 90%. At the same time, this system has low demand for computing resources and can run on ordinary servers, while deep learning systems usually require high-performance servers equipped with GPUs.

3.2 Discussion

3.2.1 Problem summary

In the research of financial audit automation method based on artificial intelligence, although the efficiency improvement and risk identification effect have been achieved, there are still some problems that need to be summarized and solved. Data quality remains a challenge. Even after strict data cleaning and preprocessing steps are implemented, data noise and missing values still affect the accuracy and stability of the model. The data sources involved in the audit process are diverse and the data formats are not uniform, which leads to the complexity of data integration and increases the difficulty of system processing and analysis. The issues of model interpretation and transparency need attention. Although complex algorithms such as random forest perform well in accuracy and efficiency, their internal decision-making process is complicated and difficult to be understood and explained by non-technical personnel. In the process of generating audit reports and interpreting audit results, users' trust in audit conclusions is reduced.

The real-time processing capability of the system needs to be improved. Despite the introduction of real-time processing platforms such as Kafka and Blink, the system still has room for improvement in processing speed and latency in the face of large-scale and high-frequency data flows. This puts forward higher technical requirements for realizing real - time audit. The generalization ability of the model also needs attention. Although the robustness of the model has been improved through cross-validation and parameter optimization, the model shows insufficient adaptability in the face of new types of financial data and fraudulent means, which affects its promotion and application in different enterprises and industries.

The user feedback mechanism needs to be improved. Although the system can automatically generate audit reports and improvement suggestions, how to effectively collect and process user feedback in the feedback and improvement process to continuously optimize the audit strategy and model performance is still a problem that needs in-depth research. Although AI-based financial audit automation methods have achieved results in improving audit efficiency and risk identification, they still need to be further optimized and improved in data quality, model interpretation, real-time processing capabilities, generalization capabilities and user feedback mechanisms to achieve more efficient and reliable financial audit automation.

Although audit efficiency has been significantly improved, an increase in computational costs has also been noted. This is due to the introduction of real-time processing technology and machine learning models that increase the computational burden on the system. However, this cost increase is acceptable considering the efficiency gains.

3.2.2 Research suggestions

In the research of financial audit automation method based on artificial intelligence, in order to further improve the efficiency and accuracy of the system, the following research suggestions need to be put forward. Data quality management needs to be further improved. It is recommended to establish a more comprehensive data cleaning and preprocessing mechanism, adopt advanced missing value processing methods and anomaly detection techniques, such as adaptive filtering and deep learning models, to improve data reliability and integrity. Enhance model interpretation and transparency. It is recommended to integrate explainable AI technologies such as LIME and SHAP into the model so that auditors can understand and explain the decision-making process of the model, thereby improving the transparency of audit reports and the trust of users. Enhanced real-time processing capabilities. It is suggested to optimize the existing real-time data processing architecture, introduce more efficient stream processing technologies and hardware acceleration schemes, such as GPU acceleration and distributed computing framework, to cope with large-scale and high-frequency data stream processing needs, and ensure that the system can respond to and process data in real time.

Improving the generalization ability of models is another direction. It is suggested to enhance the adaptability and robustness of the model in different enterprises and industries through integration learning and transfer learning techniques. The multi-model fusion method is used to improve the generalization performance of the model, and the model is applied to different financial environments through transfer learning. Optimize the user feedback mechanism. It is suggested to establish a dynamic feedback and continuous learning system, collect and analyze user feedback, adjust and optimize audit strategy and model parameters in time. Through the closed-loop mechanism of user feedback, the system performance can be continuously improved to ensure the effectiveness of audit strategies and the accuracy of models. The suggestions aim to further improve the financial audit automation method based on artificial intelligence, improve the intelligence level and practicability of the system, and provide more powerful technical support and guarantee for the financial management of enterprises. Through improvement, more efficient and accurate financial audit can be achieved, and the intelligent transformation of the financial audit industry can be promoted.

3.2.3 SHAP effect

In an actual case, we selected the financial data of a certain company over a period of time, including multiple features such as income, expenditure, accounts receivable turnover rate, and debt-to-asset ratio. After model training and prediction, we obtained the result that a certain transaction was judged to be abnormal. At this time, the SHAP value can clearly explain the basis for the model to make this judgment.

By calculating the SHAP value of each feature, we found that the SHAP value of the debt-to-asset ratio is high and positive, which shows that the debt-to-asset ratio plays a key positive role in the model's judgment of the transaction as abnormal. That is, the debt-to-asset ratio exceeds the normal range, which greatly increases the possibility of the transaction being judged as abnormal. Visualizing the SHAP value (such as using a SHAP value bar chart, with the horizontal axis as the feature name and the vertical axis as the SHAP value size) can more intuitively show the degree of influence of each feature on the prediction result. Auditors can see at a glance which features have the greatest impact on the model's decision, and thus conduct in-depth analysis of why the transaction was judged to be abnormal, greatly improving the transparency and credibility of the audit, so that audit decisions are no longer "black box" operations, but are based on clear and explainable evidence.

3.3 Discussion

From a quantitative perspective, in terms of audit accuracy, the accuracy of cutting-edge research is mostly in the range of 80% - 86%, while this study uses the random forest algorithm combined with real-time data processing technology to achieve an audit accuracy of 90%. In terms of audit efficiency, most cutting-edge research efficiencies are at a medium or low level, but this study has achieved a significant result of 30% efficiency improvement. This is mainly attributed to the fact that the random forest algorithm constructs multiple decision trees and randomly selects features for node splitting, effectively reducing the risk of overfitting and improving the model's ability to identify complex financial data; at the same time, the use of real-time data processing platforms such as Kafka and Flink has realized the real-time collection, processing and analysis of financial data, greatly accelerating the audit process.

Qualitatively, the support vector machine, gradient boosting and other methods used in cutting-edge research have limitations when facing the high dimensionality, complexity and dynamic changes of financial data. The support vector machine has high computational complexity, is sensitive to the choice of kernel functions, and is difficult to adapt to the diversity of financial data; gradient boosting is sensitive to outliers, takes a long time to train, and cannot meet real-time requirements. In contrast, this research method not only improves the accuracy and robustness of the model, but also ensures the dynamics and timeliness of the audit process, and can better adapt to the actual needs of corporate financial audits.

However, this research method still has certain limitations. In terms of data quality, despite the implementation of strict data cleaning and preprocessing steps, data noise and missing values will still affect the accuracy and stability of the model. Model interpretability and transparency are also issues that need attention. The internal decision-making process of the random forest algorithm is complex, and it is difficult for non-technical personnel to understand and explain, which to a certain extent reduces the user's trust in the audit conclusions. In addition, when facing large-scale, high-frequency data traffic, the system's real-time processing capabilities have been improved, but there is still room for improvement. The generalization ability of the model when dealing with new financial data and fraud methods also needs to be enhanced.

4 Conclusion

In this study, an AI-based financial audit automation method is deeply discussed and implemented to improve audit efficiency, accuracy and risk identification ability. Through the introduction of random forest algorithm, combined with several key steps such as data integration, real-time processing, model training and optimization, the system has shown improvement in many aspects. By constructing and optimizing the random forest model, the audit coverage and error detection rate are improved, and the risk identification and processing are realized efficiently. The data cleaning and reprocessing steps ensure the quality and consistency of the input data, providing a reliable basis for the model. The introduction of real-time processing technologies such as Kafka and Slink has accelerated data processing and met the processing needs of high frequency data streams. The audit process design optimizes the allocation and utilization of audit resources and improves the overall audit efficiency through systematic steps.

The maintainability and transparency of the model was also emphasized in the study, and by introducing explainable AI technology, users' trust in audit results was increased. At the same time, through the establishment of dynamic feedback mechanism, the system can constantly collect and analyze user feedback, timely adjust the audit strategy and model parameters, and achieve continuous optimization. Despite the achievements, the study also points out some challenges, such as data quality issues, model generalization capabilities, and real-time processing capabilities, which need to be further addressed in future research and applications. This study shows the great potential of artificial intelligence in financial audit automation, and also puts forward specific suggestions for improvement, which provides a valuable reference for future financial audit practice. Through continuous optimization and improvement of technology and methods, financial audit will achieve a higher level of intelligence and automation, and provide more accurate and efficient support for the financial management and risk control of enterprises. Intelligent audit methods improve the quality and efficiency of audit work, but also enhance the transparency and standardization of financial management, and promote the innovation and development of the financial audit industry.

Although explainable artificial intelligence technology (such as SHAP) has achieved remarkable results in improving the transparency of financial audits, its limitations cannot be ignored. In an extremely high-frequency data environment, calculating the SHAP value requires complex operations on massive data, resulting in a sharp increase in computing resource consumption, processing efficiency cannot keep up with the speed of data updates, and scalability is limited, making it difficult to meet the real-time requirements of audits. In addition, SHAP is calculated based on a decision tree model, which is susceptible to data distribution and noise. Even a small change in data may cause a significant change in the tree structure, which in turn leads to unstable calculation results of the SHAP value, which cannot accurately reflect the true contribution of features to model decisions, affecting the reliability of audit results.

Acknowledgement

This study was funded by Science Research Project of Hebei Education Department (BJS2023041).

References

[1] Sun JH, Li LC, Qi BL. Financial statement comparability and audit pricing. Accounting Finance. 2022; 62(5): 4631-4661. https://doi.org/10.1111/acfi.12970.

[2] Condie ER, Obermire KM, Seidel TA, Wilkins MS. Prior Audit Experience and CFO Financial Reporting Aggressiveness. Audit J Pract Theory. 2021; 40(4): 99-121. https://doi.org/10.2308/AJPT-2020-012.

[3] Koh K, Tong YH, Zhu ZN. The effects of financial statement disaggregation on audit pricing. Int J Audit. 2022; 26(2): 94-112. https://doi.org/10.1111/ijau.12253.

[4] Lyshchenko O, Ocheret'ko L, Lukanovska I, Sobolieva-Tereshchenko O, Nazarenko I. The role of financial audit in ensuring the reliability of financial statements. Ad Alta J Interdiscip Res. 2024; 14(1). https://doi.org/10.33543/140139141145

[5] Suryani E, Winarningsih S, Avianti I, Sofia P, Dewi N. Does Audit Firm Size and Audit Tenure Influence Fraudulent Financial Statements? Australas Account Bus Finance J. 2023; 17(2): 26-37.

[6] Xu Q, Fernando G, Tam K, Zhang W. Financial report readability and audit fees: a simultaneous equation approach. Managerial Aud J. 2020; 35(3): 345-372. https://doi.org/10.1108/MAJ-02-2019-2177.

[7] Erdmann A, Yazdani M, Mas Iglesias JM, Marin Palacios C. Pricing Powered by Artificial Intelligence: An Assessment Model for the Sustainable Implementation of AI Supported Price Functions. Informatica. 2024;35(3):529-56. https://doi.org/10.15388/24-infor559

[8] Ijadi Maghsoodi A, Hafezalkotob A, Azizi Ari I, Ijadi Maghsoodi S, Hafezalkotob A. Selection of Waste Lubricant Oil Regenerative Technology Using Entropy-Weighted Risk-Based Fuzzy Axiomatic Design Approach. Informatica. 2018;29(1):41-74. https://doi.org/10.15388/Informatica.2018.157

[9] Pragarauskaitė J, Dzemyda G. Markov Models in the analysis of frequent patterns in financial data. Informatica, 2013, 24(1): 87-102. http://dx.doi.org/10.15388/Informatica.2013.386

[10] Lim CY, Lobo GJ, Rao PG, Yue H. Financial capacity and the demand for audit quality. Accounting

Bus Res. 2022; 52(1): 1-37. https://doi.org/10.1080/00014788.2020.1824116.
[11] Ismail R, Mohd-Saleh N, Yaakob R. Audit committee effectiveness, internal audit function and financial reporting lag: Evidence from Malaysia. Asian Acad Manage J Account Finance. 2022; 18(2): 169-193. https://doi.org/10.21315/aamjaf2022.18.2.8.
[12] Oussii AA, Boulila N. Evidence on the relation between audit committee financial expertise and internal audit function effectiveness. J Econ Adm Sci. 2021; 37(4): 659-676. https://doi.org/10.1108/JEAS-04-2020-0041.
[13] Lyubenko A, Znak N, Karpachova O. Audit features of the first IFRS financial statements. Financial Credit Activity Probl Theory Pract. 2022; 1(42) :185-194.
[14] Endrawes M, Feng ZA, Lu MT, Shan YW. Audit committee characteristics and financial statement comparability. Accounting Finance. 2020; 60(3): 2361-2395. https://doi.org/10.1111/acfi.12354.
[15] Lutfi A, Alkilani SZ, Saad M, Alshirah MH, Alshirah AF, Alrawad M, et al. The influence of audit committee chair characteristics on financial reporting quality. J Risk Financial Manage. 2022; 15(12): 563. https://doi.org/10.3390/jrfm15120563.
[16] Calvin CG, Holt M. The impact of domain-specific internal audit education on financial reporting quality and external audit efficiency. Accounting Horizons. 2023; 37(2): 47-65. https://doi.org/10.2308/HORIZONS-2020-105.
[17] Alcaide-Ruiz MD, Bravo-Urquiza F. Does audit committee financial expertise actually improve information readability? Rev Contab Span Account Rev. 2022; 25(2): 257-270. https://doi.org/10.6018/rcsar.420261.
[18] Driskill MW, Knechel WR, Thomas E. Financial auditing as an economic service. Curr Issues Aud. 2022; 16(2). https://doi.org/10.2308/CIIA-2021-021.