Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on Audit Data Anomaly Detection and Risk Assessment Method Based on Data Mining Technology

Xia Li[1,] ✉

[1] *Yunnan Technology and Business University, Kunming, Yunnan, 650000, China*

ABSTRACT

Digital auditing has become the key to the transformation and upgrading of the auditing field. Financial audit data anomaly detection needs to combine multiple aspects of information, and it is of great practical significance to utilize the existing technical means to discover financial anomalies in the limited content. In this paper, based on the limitations of the weighted KNN deep neural network algorithm, a multi-branch deep neural network is proposed and a cost-sensitive loss function is designed. Combining the qualitative and quantitative methods of risk assessment, the enterprise audit risk assessment index system is constructed, the indexes are standardized, and the results of enterprise audit risk assessment are analyzed. The specific application effect of the assessment model is analyzed from the aspects of industry status and key financial performance, and the relevant strategies for corporate audit risk response are proposed. In the 1st risk assessment, 8 of the 20 enterprises are above higher risk, 6 are medium risk, and 6 are below lower risk. The results of the 2nd audit risk assessment have varying degrees of reduction between -0.3663 and -0.0119. From 2017, the overall net profit growth rate of enterprises is decreasing year by year, especially in the period from 2019 to 2020, and the net profit growth rate of the industry in 2020 is -24.87%, which predicts that the future development of the industry is not optimistic.

Keywords: Weighted KNN, Multi-branch deep neural network, Cost-sensitive loss function, Audit data anomaly detection, Audit risk assessment

1. Introduction

At present, strengthening enterprise supervision and financial auditing is becoming a new normal. In the face of increasing audit objectivity, scientific requirements and growing audit tasks, manual auditing based on specialized knowledge and experience gradually fails to meet the needs of audit work. The realization of manual audit based on the auditor's professional knowledge, expertise and experience, to data mining and artificial intelligence to assist the transformation of the wisdom of the

✉ Corresponding author.
E-mail address: yueyuemiguo@126.com (X. Li).
Received 25 October 2024; Revised 10 December 2024; Accepted 05 March 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-162
© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

audit, and constantly improve the efficiency and quality of the audit is becoming an imperative trend [1-4]. Financial voucher data is one of the most important data in the auditing process, the audited unit generally needs to provide financial voucher data in advance to carry out the audit, based on the financial voucher data combined with vouchers related to contracts, invoices and other information to carry out the audit [5-6]. The audit process involves more financial voucher data, and only through the auditor to verify all the financial voucher data is a large workload. Therefore, it is necessary to study a data anomaly detection method to help auditors narrow the scope of verification from a large number of financial vouchers and to get the financial vouchers with auditing doubtful expenditures [7-8]. On the other hand, in the face of the problem of multiple types of data (involving text data, category data, and numerical data) and a serious imbalance in the proportion of positive and negative samples in the financial vouchers, it is necessary to propose an effective detection method to narrow down the scope of the verification more effectively and to cover as many auditing issues as possible in the financial vouchers of the auditing doubtful expenditures, so as to realize the use of big data without the need of auditing expertise through the use of intelligent technology. The big data can be used without auditing expertise through intelligent technology [9-11]. Moreover, with the continuous enrichment of the negative sample data set, the anomaly detection method is optimized to better assist auditors in discovering potential errors, frauds and other auditing issues, and to improve the trust and reliability of the data [12-13].

With the continuous expansion of the enterprise scale and the increasing complexity of the operating environment, the role of auditing in enterprise management is becoming more and more prominent. Audit plays a vital role in helping enterprises to identify, assess, and manage risks, it can provide independent and objective assessment for enterprises, thus targeting the reduction of enterprise risks, helping to improve internal control, and ensuring that enterprises meet the established goals [14-17]. At the same time, risk assessment is a very important part of the audit process. By assessing potential risks, it can ensure that the auditor can identify risks that may lead to inaccurate financial reporting, thereby protecting the public interest, and the auditor can formulate appropriate strategies and measures to ensure the accuracy and effectiveness of the audit, thereby improving audit compliance [18-21].

In this paper, Word2vec is used to construct word vectors for financial “voucher summary” content segmentation, and a multi-branch deep neural network is proposed to transform this information into input data that can be recognized by the neural network. We choose the binary cross-entropy function as the loss function, design the cost-sensitive loss function, and test the classification performance of the proposed method through evaluation indexes. Combining qualitative and quantitative assessment methods, the enterprise audit risk assessment index system is constructed, and the indexes are standardized quantitatively and qualitatively respectively, and 20 enterprises are selected as research objects to carry out audit risk evaluation using the audit risk assessment model designed in this paper. Specific applications of the model assessment are made in terms of both industry conditions and key financial performance, and the future operating conditions of the enterprises are predicted. Based on the results of enterprise audit risk assessment analysis, corresponding countermeasures are proposed.

2. Overview

In recent years, audit data anomaly detection has been developed in technological advances.In 2019, literature [22] designed a prediction model for audit fraud detection to reduce cost, suppress fraud, and improve audit efficiency based on audit principal agent theory in the context of rule-based and classification. In 2020, literature [23] referred to generalized auditing software with controlled automated auditing techniques that can assist the auditor to quickly analyze the data and identify frauds in the audited data. In 2021, literature [24] states that unsupervised anomaly detection in auditing can not only detect anomalies in the data that can be expected by the auditor, but also detect audit process deviations and new frauds. However, this type of detection is mostly specific to a

particular dataset and lacks general applicability. In addition, literature [25] achieved data anomaly detection with Gauss-Beinoulli bidirectional restricted Cauchy machine neural network model for auditing with high accuracy, high speed, no dataset size limitation, real data, and heterogeneous associative memory function.2022, literature [26] detected auditing data anomalies with an algorithm based on the double-biased stochastic wandering property, and after optimally updating this algorithm, it can quickly, at 80% correctness rate, to detecting potential anomalies. Machine learning is a well-known tool for predictive analytics, and it is no exception in anomaly detection of audit data. In 2024, literature [27] designed an intelligent audit data anomaly detection method by combining the advantages and properties of bidirectional long and short-term memory, improved back-propagation neural network, attention mechanism, nonlinear prediction, long time series feature extraction, important information focusing ability, and correlation analysis algorithm, which still performs well in huge datasets. Literature [28] compared and analyzed supervised machine learning methods, and the K nearest neighbor algorithm performed the best in terms of audit data fraud detection, risk assessment, and efficiency. Literature [29] analyzed that artificial intelligence can assess audit datasets in order to identify anomalous data such as transactions, financial statement inconsistencies, and customer or supplier behavior in them, ensuring multifaceted and reliable data detection, while also predicting potential risks.

In terms of audit risk assessment, literature [30] first categorizes the key risks of public sector auditing, then develops a list of issues in the risks for auditors and departmental insiders, and designs a calculation of audit risk components to form a standard audit risk assessment methodology. Literature [31] used fuzzy theory to experiment with the inconsistency of the experience level of the audit team leader on the correctness of the prediction and assessment of risks in auditing. In order to avoid the audit efficiency and accuracy gap caused by manual auditing, seeking technological help is the way to go. Literature [32] describes risk assessment techniques based on risk matrices and questionnaires to analyze and assess risks by considering the probability of occurrence, importance, and degree of risk factors for a comprehensive audit understanding. Literature [33] intends to systematize the auditing behavior by using that inductive, deductive, cluster analysis and synthesis methods to detect the audit risk, and designed an audit risk assessment model based on analogy, fuzzy theory, and expert survey to solve the uncertainty brought by the auditor. Literature [34] optimizes the audit risk assessment model constructed by regression algorithm through back propagation neural network-machine learning technique, which not only guarantees the compliance of the audit process, but also solves the problems of resource allocation and financial discrepancies in the audit process. The difference is that literature [35] uses hierarchical analysis to design an audit risk assessment model, which analyzes the audit risk at multiple levels in the enterprise's business model, information system, and financial management, while the article also proposes a standardized audit process and risk prevention measures.

3. Audit data anomaly detection

3.1. Deep neural network algorithm based on weighted KNN

1) Divide the content of the financial "voucher summary" and use Word2vec to construct the word vector $\{v_1, v_2, ..., v_n\}$, and at the same time, in order to facilitate the calculation, and the voucher described by some "keywords" is more likely to have audit problems according to experience, so the weight w_i is different, so the weighted sentence vector of Word2vec based on prior knowledge is constructed to represent the content of "voucher summary", and the calculation formula is [36]:

$$s_{en\_vec} \frac{\sum_{i=1}^{m} v_i \times e^{wi}}{m} \tag{1}$$

where: m denotes the number of words in each voucher. The weight w_i of each word vector is:

$$w_i = \frac{n_{v_i}}{|v_i|} \times \frac{n_{abnormal-v_i}}{|a_{abnormal} - v_i|} \tag{2}$$

where: $|v_i|$ represents the total number of word vectors containing duplicate counts obtained after word segmentation of the "Voucher Summary" content of all vouchers, n_{v_i} represents the total number of occurrences of the specific word vectors containing duplicate counts v_i, $|a_{abnormal} - v_i|$ represents the total number of word vectors containing duplicate counts obtained after the segmentation of the "Voucher Summary" content of all historical accumulated abnormal vouchers, and $n_{abnormal-v_i}$ represents the total number of occurrences of the specific word vectors v_i of the obtained abnormal vouchers containing duplicate counts.

2) Sentence vector s_{en_vec}, which is the result of subwording the content of “Voucher Summary”, and the vectors obtained by One-Hot coding of “Audited Unit, Debit Issuance Cost Category, and Expense Detail” constitute the feature vector X of the vouchers.

3) In each subcluster C_i, calculate the weighted distance between the feature vector X of each voucher and the feature vectors Y of its k nearest neighbors:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n} \frac{w_i}{W}(x_i - y_i)^2} \tag{3}$$

where: w_i is the weights, $w_i = e^{-|x_i - y_i|\sigma}$, W are the normalization factors, $W = \sum w_i$ is the moderating factor, which is taken as $\sigma = 1$ in this paper.

3.2. Token-sensitive multi-branch deep neural network algorithm

3.2.1. Limitations of unsupervised algorithms based on weighted KNN and others. The idea of unsupervised algorithms based on weighted KNN and others is easy to understand and can effectively reduce the workload of financial data verification within a certain range, but in this experiment, it is found that if the unsupervised algorithm wants to include audit problem data records in the audit suspicious data records as much as possible, the amount of audit suspicious data records is at least 1/10 of the total amount of original audit data records, and the suspicious points sometimes cannot completely cover the real anomalies [37].

3.2.2. Multi-branch deep neural network structure. Audit data belongs to multiple types of data, that is, it contains textual information, numerical information, and category information. In order to effectively find audit doubts, this paper proposes a multi-branch deep neural network, and the network structure is shown in Figure 1. The text representation information corresponding to the "Slip Summary" in a financial voucher is recorded as x, then $x = [x_1, x_2, ..., x_n]$, where x_i represents the i th word in the text message, the "Account Category" information in the financial voucher is recorded as c, and the "Amount" numerical information in the financial voucher is recorded as n. Translate this information into input data that can be recognized by the neural network.

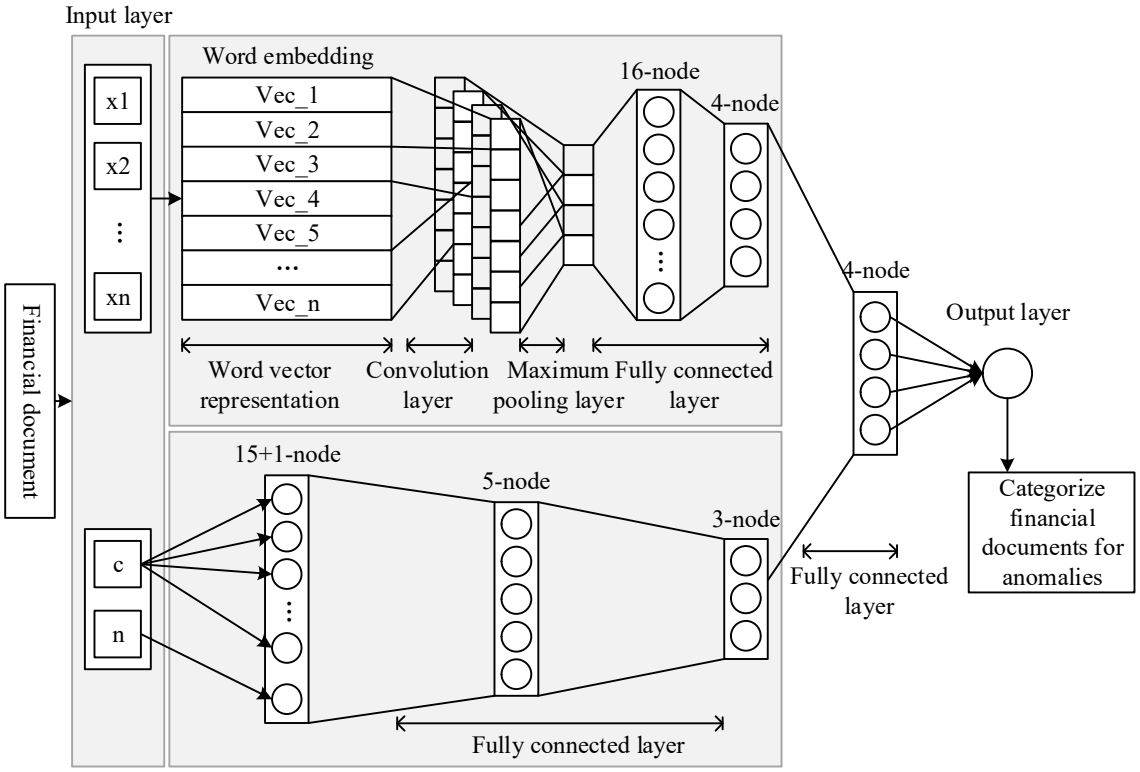


Fig. 1. Multi-branch deep neural network structure

Financial documents belong to various types of data, among which the "summary information" in the financial documents belongs to text-based data, "account name" belongs to category data, and "debit amount" belongs to numerical data. This raw data needs to go through some preprocessing operations before it can be fed into the neural network.

For categorical data, such as “account name” in financial vouchers, One Hot coding is performed. For continuous data, such as the “debit amount” in the financial voucher, it is generally necessary to carry out standardization before inputting into the neural network, such as min-max standardization, Z-score standardization and so on. In order to avoid the influence of extreme value to the greatest extent, Z-score method is chosen for standardization, and its formula is shown in Equation (4), where: $\bar{x}$ is the mean value of the original data, and σ is the standard deviation of the original data.

$$x^* = \frac{x - \bar{x}}{\sigma} \tag{4}$$

3.2.3. Branch Network Connections. For the category information and numerical information, a simple fully connected neural network is established for feature extraction. On this basis, the branching network for text-based data is connected to the fully connected branching network, and a fully connected layer is added, taking the number of nodes as 4.

Since the recognition of abnormal credentials is a binary classification problem, the number of nodes in the output layer is set to 1 and the activation function is selected as sigmoid function. Also considering that Adam optimization algorithm has the advantages of high computational efficiency and low memory requirement, the model in this paper is trained using Adam algorithm [38].

3.2.4. Cost-sensitive loss function design. Determine whether there is an audit suspicion of financial documents belongs to the binary classification, the binary classification problem usually choose the binary cross entropy function as the loss function, as shown in equation (5):

$$L(\hat{y}_i, y_i) = -[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{5}$$

where: y_i represents the true category of sample i, and $\hat{y}_i$ represents the probability that sample i is predicted to be an abnormal financial voucher.

The cost of mispredicting abnormal financial vouchers as normal financial vouchers is much greater than the cost of judging normal financial vouchers as abnormal financial vouchers. This is similar to the application scenario of fraud detection, where the number of fraudulent transactions is much smaller than the number of normal transactions, but the cost of misjudging each fraudulent transaction is extremely severe. The two-classified cross-entropy function does not have the above cost sensitivity, can not well support the unbalanced data, so this paper has improved it, the improved loss function is as follows:

$$L(\hat{y}_i, y_i) = -[(1 + t_{neg}) \times y_i \log(\hat{y}_i) + (1 + t_{pos}) \times (1 - y_i) \log(1 - \hat{y}_i)] \tag{6}$$

where: t_{neg} represents the cost of misclassifying an abnormal financial voucher as a normal financial voucher, and t_{pos} represents the cost of misclassifying a normal financial voucher as an abnormal financial voucher. t_{ncg} and t_{pos} are specifically calculated as:

$$t_{neg} = \frac{T}{d_{neg}} \tag{7}$$

$$t_{pos} = \frac{T}{d_{pos}} \tag{8}$$

where: T is the hyperparameter, d_{neg} is the total number of abnormal financial vouchers in the training set, and d_{pos} is the total number of normal financial vouchers in the training set. It can be seen that the misclassification cost is inversely proportional to the total number of financial vouchers under the corresponding category. On the dataset with unbalanced data distribution, the difference between the number of samples of majority and minority categories is large, and through cost-sensitive learning, the influence of minority categories on model parameters can be strengthened, so that the neural network can be more sensitive to the minority categories, and thus obtain better classification results.

3.3. Experimental design and analysis

3.3.1. Data sets. For a company, an experimental dataset is extracted from an SAP ERP instance containing all journal entries for a single fiscal year. Based on strict data privacy regulations, all the attributes of the journal entries are anonymized using an irreversible one-way hash function during the extraction process. The extracted data is also checked against the trial balance of the SAP ERP system to ensure the integrity of the dataset. A subset of the six most distinguishable attributes was extracted from the BKPF and BSEG forms.

Most of the attributes in the ERP system, for example, correspond to categorical (discrete) variables, such as bookkeeping date, account, bookkeeping type, currency, etc. When training the KNN neural network, the category entry attributes are preprocessed to obtain a binary representation of each entry encoded in solo heat, resulting in an experimental dataset with a coding dimension of 385 dimensions. Referring to the realistic auditing scenario, a small number of global and local anomalies are injected into the dataset to obtain highly unbalanced class distribution data. In the injected data, global anomalies contain attribute values that are not similar to the original dataset, while local anomalies contain combinations of attribute values that do not appear in the original data. True labels were added to the dataset to label each bookkeeping entry as a global anomaly, local anomaly, or normal entry. This resulted in an experimental dataset of 312,281 journal line items for six attribute

categories, which contained 90 abnormal entries, with 50 and 40 global and local abnormal entries, respectively. The experimental dataset is divided into training and test sets by 8:2.

3.3.2. KNN network training. In the training of each architecture, the learning rate is set to 10^{-4}, the number of small batches is 128 transcripts, adaptive momentum estimation is used, and each network layer weight is initialized. Standard backpropagation is used in training and the maximum number of training generations is 2000. Fig. 2 shows the training performance of different depth architectures with 100% recall is given for different KNN architectures. It can be found that increasing the number of hidden units can speed up the error convergence and reduce the number of detected anomalies. The results show that network depth is positively correlated with anomaly detection performance. Under the premise of ensuring 100% recall, the KNN5 architecture containing 17 hidden layers detects the lowest number of anomalous entry details in the total number of line items, which accounts for less than 0.02%, i.e., minimizing the false alarm rate.

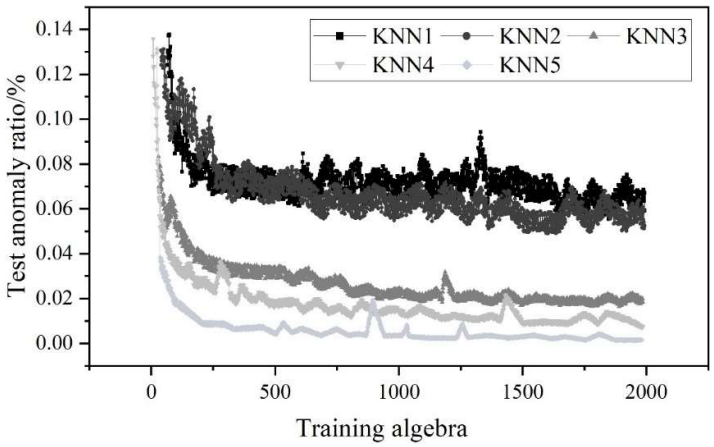
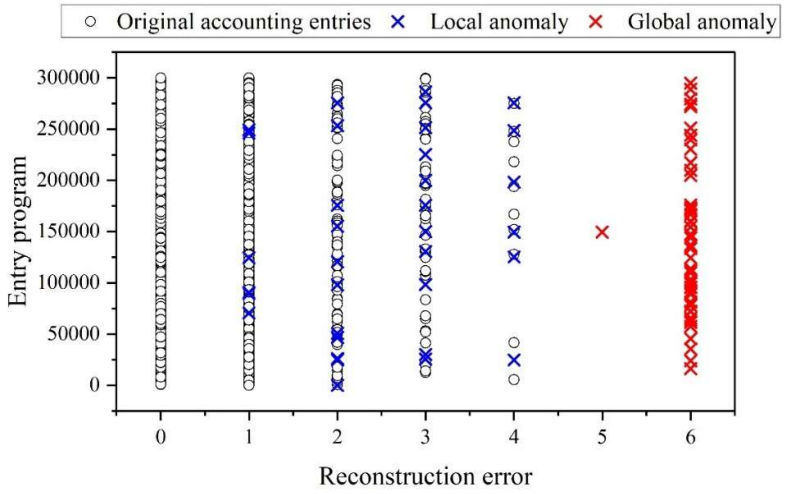
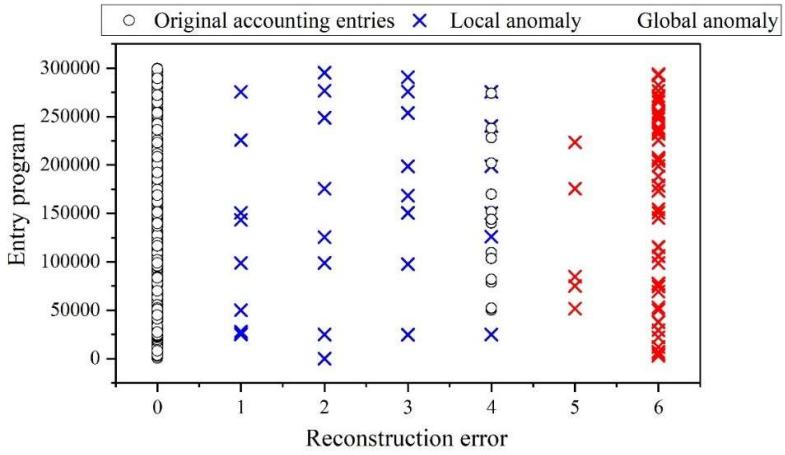


Fig. 2. Training performance of different depth architectures

After training convergence, the reconstruction error for each journal entry is obtained using the trained KNN5 model. Fig. 3 shows the reconstruction error of journal entry details with 100 generations of training in Fig. (a) and 400 generations of training in Fig. (b). The reconstruction error of 312,281 journal entries details for the experimental dataset, where black color indicates the original accounting entries, and red and blue colors indicate the global and local anomalies detected by KNN5, respectively. Set the anomaly detection threshold to β =0.01 and if the entry attribute is not reconstructed correctly or is extremely uncommon, it will be flagged as an anomaly. The number of local anomalies with a reconstruction error of 1-4 was around 6 to 12 for the 100-generation training condition.



(a)100 generation training



(b)400 generation training

Fig. 3. Journal entry reconstruction error

3.3.3. Assessment of indicators. To evaluate the classification performance of the proposed method, precision (P), F1 score (F_1) and top-k accuracy are used as performance metrics and precision is calculated as:

$$P = \frac{TP}{TP + FP} \tag{9}$$

The F1 score is the reconciled mean of recall R and precision P, approximately close to 1 representing better performance, and is calculated as:

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

where TP denotes the number of samples that are actually abnormal entries and are correctly subdivided into abnormal entries, FP denotes the number of samples that are actually normal entries but are incorrectly subdivided into abnormal entries, and FN is the number of samples that actually belong to the category of being an abnormal entry but are incorrectly subdivided into normal entries.

Due to the large size of the data in the accounting entries, the top-k precision index is invoked, and the k labels with the largest probability are selected in the abnormality detection, and the prediction is regarded as correct if there is a true label value among them. In this paper, the top-k accuracy is calculated by setting k=90, which corresponds to the number of manually injected anomalous entries in the experimental dataset.

3.3.4. Analysis of test results. After determining the optimal parameters of the model through training, the proposed method is quantitatively evaluated through experiments and compared with other methods. First, the quantitative test is conducted to analyze whether the proposed method is able to detect abnormal terms by training with normal entry data. Accuracy, F1 score and top-k accuracy are used as evaluation metrics in the quantitative evaluation, and the percentage of the number of detected anomalous transcription line items is given.

Table 1 shows the experimental results of different depth KNN structures, which are consistent with the training performance, and the increase in the number of hidden layers can effectively reduce the number of detected anomalies. Ensuring that the recall rate is 100%, the anomaly percentage of KNN1 is 5.9425, which proves that the shallow KNN generates a large number of false alarms, which in practice will result in an excessive amount of in-depth analysis data and reduce the efficiency of financial auditing.The anomaly percentage of KNN5 is only 0.5436%, which proves that the depth of the KNN has a significant impact on the performance of the anomalous data capturing.

Table 1. Experimental results of KNN structures at different depths

Model	Accuracy/%	F1/%	Top-k/%	Abnormal/%
KNN1	0.4896	1.1345	0.5169	5.9425
KNN2	0.5299	1.2978	0.6248	4.8163
KNN3	0.9946	2.0348	56.3485	3.5796
KNN4	3.1649	5.8978	71.3646	1.1348
KNN5	5.9765	13.4973	55.1348	0.5436

4. Enterprise audit risk assessment model design

4.1. Risk assessment methods and techniques

Risk assessment occupies a central place in the audit process and involves the use of a variety of methods and techniques to identify and assess risks that may affect the outcome of the audit. These methods and techniques not only require auditors to have specialized knowledge, but also require them to have access to the latest technological tools to improve the accuracy and efficiency of risk assessment.

4.1.1. Qualitative assessment. Qualitative assessment focuses on understanding and analyzing the essential characteristics of risks. This includes an assessment of the likelihood of the risk occurring and its potential impact on the audit outcome. The auditor needs to analyze the client's business environment in depth to identify those key factors that may affect the accuracy of the financial statements. For example, the auditor may consider the industry in which the business operates, the competitive market conditions, the business strategy, and even the experience and competence of management. All of these factors may increase the risk of material errors in the financial statements.

Qualitative assessment also involves the evaluation of a business's internal control system. By analyzing the design and implementation of internal controls, the auditor can identify risk points that could lead to financial errors or fraud. In addition, the auditor considers historical audit findings, prior errors and adjustments, and management's treatment of these issues and corrective actions.

4.1.2. Quantitative assessment. Unlike qualitative assessments, quantitative assessments are more concerned with the use of data and statistical methods to quantify risk. This includes calculating the probability of a particular risk occurring and the expected impact. An auditor may analyze historical financial data for patterns of anomalies or inconsistencies that may signal potential risks. For example, by comparing financial data from consecutive fiscal years, an auditor may identify fluctuations in profits or costs that are inconsistent with growth in the size of the business.

Quantitative assessments also include estimates of the impact of potential risks, such as the probable margin of error for financial statement items. Using statistical models and algorithms, auditors can predict financial results under different scenarios, resulting in a more accurate assessment of audit risk.

4.2. Construction of enterprise audit risk assessment index system

Since the impact of different risk assessment indicators on the risk assessment results varies, and even some indicators not only do not contribute to the assessment results, but also have a negative impact on the risk assessment results. In order to select the indicators that contribute more to the risk assessment results, it is necessary to compare the assessment indicators through the weight value to obtain the impact on the enterprise audit risk assessment. Hierarchical analysis is used to construct the judgment matrix B of risk assessment indicators, as in equation (12):

$$B=\begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & \cdots & x_{2,m} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & \cdots & x_{3,m} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & x_{n,3} & x_{n,4} & \cdots & x_{n,m} \end{bmatrix} \tag{12}$$

where $x_{i,j}$ is the degree of criticality of the two indicators x_j to x_i.

The product L_i of the elements of each row of the indicator judgment matrix B is calculated as in equation (13):

$$L_i=\prod_{j=1}^{n} b_{ij} \tag{13}$$

where b_{ij} represents the i rd element of row j. Determining the n th power root of L_i can be done using equation (14):

$$\overline{w_i}=\sqrt[n]{L_i} \tag{14}$$

Normalize to process Eq. (14) and obtain Eq. (15):

$$w_i=\frac{\overline{w}_i}{\sum_{i=1}^{n}\overline{w}_i} \tag{15}$$

Calculation of the maximum eigenvalue of the judgment matrix of risk assessment indicators B can be done using equation (16):

$$\partial_{\max}=\sum_{i=1}^{n}\frac{[Bw]_i}{nw_i} \tag{16}$$

Finally obtain the weights of the indicators in layer i: $C_i=(c_{i1},c_{i2},c_{i3})$, where c_i is the weight of the i th indicator.

Set the indicator weight threshold, and select the indicators whose weights are greater than the threshold to construct the enterprise audit risk assessment indicator system, as shown in Table 2.

Table 2. Enterprise audit risk assessment index system

Goal	CODE	Primary index	Code	Secondary index	Code
Enterprise audit risk assessment index system	M	Material error Risk	M1	The integrity of relevant laws and regulations	M11
				Macroeconomic environment	M12
				Relevant policy	M13
				Partner selection risk	M14
				Management collaboration risk	M15
				Key information and core technology outsourcing association	M16
				Moral hazard	M17
		Auditor risk	M2	Auditor business quality	M21
				Audit personnel are received to be tempted to influence audit	M22
				The audit personnel operate unregulated or misperformed	M23
		Audit procedure risk	M3	The completeness and science of the audit plan	M31
				The accuracy of obtaining data in the audit process	M32
				The comprehensive and professional results of the design	M33

4.3. Standardization of risk assessment indicators

Since there are not only qualitative but also quantitative indicators in the enterprise audit risk assessment indicator system, the standardization process of different risk assessment indicators can make different risk assessment indicators comparable.

4.3.1. Standardization of quantitative indicators. For the quantitative indicators within the enterprise audit risk assessment indicator system, the standardization process is as follows.

When the target value is in positive proportion to the assessment result, there is equation (17):

$$D_i = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}} \tag{17}$$

When the target value is inversely proportional to the assessment result, there is equation (18):

$$D_i = 1 - \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}} \tag{18}$$

In Eqs. (17) and (18), $D_i, x_{i\min}$ and $x_{i\max}$ are the standardized value of the target value x_i, the minimum value of the ith indicator set in advance, and the maximum value of the ith indicator set in advance, respectively.

4.3.2. Standardization of qualitative indicators. For the qualitative indicators within the enterprise audit risk assessment indicator system, the standardization process is as follows. The quantitative processing of the indicators is carried out using the expert scoring method, and the quantitative indicators are standardized using the quantitative indicator standardization process to ensure that there is comparability between the qualitative and quantitative indicators.

4.4. Example analysis

4.4.1. Data sources. Corporate auditing has a crucial impact on the development of enterprises, and ensuring the safety of corporate assets through audit risk assessment not only improves the internal management level of enterprises, but also greatly enhances the ability of enterprises to withstand huge market risks. Select 20 enterprises as the research object, using the audit risk assessment model designed in this paper for audit risk evaluation. With the help of the platform is MATLAB software, in which the input layer and intermediate layer connection function of the model is Tansig function, the intermediate layer and output layer connection function is Logsig function, and the training error of the audit risk assessment model is 10^{-4} [39].

4.4.2. Results of enterprise audit risk assessment. MATLAB software is used to prepare a program to train and learn the audit risk assessment model, and the relationship between the number of training times and training error is obtained, and the results are shown in Figure 4. When the number of model training in 160 times, its training error to meet the set requirements using the trained model to implement the assessment of the selected 20 enterprises audit risk.

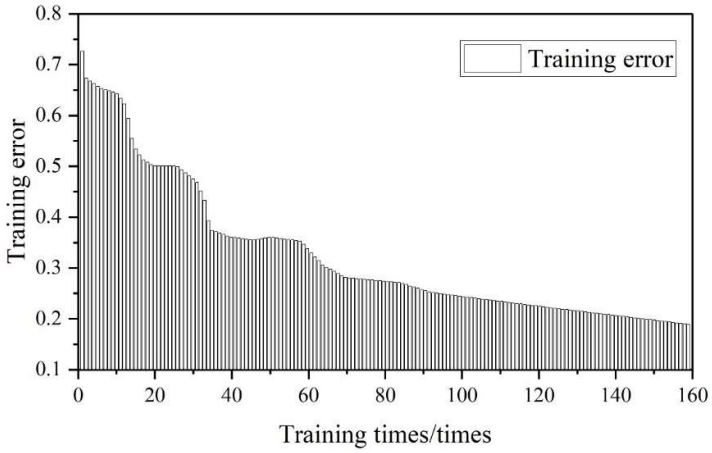


Fig. 4. Relationship between training frequency and training error

Table 3 shows the results of the corporate audit risk assessment, with eight of the 20 companies falling into the higher risk category or above, six in the medium risk category, and six in the lower risk category or below.

Table 3. Enterprise audit risk assessment results

Enterprise number	Network output	Audit risk level
1	0.5066	Medium risk
2	0.3198	Lower risk
3	0.4636	Medium risk
4	0.1679	Low risk
5	0.2961	Lower risk
6	0.6615	Higher risk
7	0.8364	High risk
8	0.5263	Medium risk
9	0.7034	Higher risk
10	0.4965	Medium risk
11	0.5064	Medium risk
12	0.6158	Higher risk
13	0.7168	Higher risk
14	0.8138	High risk
15	0.8134	High risk
16	0.3952	Lower risk
17	0.4348	Medium risk
18	0.6425	Higher risk
19	0.2498	Lower risk
20	0.3487	Lower risk

4.4.3. Comparison of audit risk assessment results. Implementing the assessment of enterprise audit risk, pinpointing the weaknesses in enterprise auditing, taking targeted measures to deal with and solve the weaknesses in the auditing process, and carrying out 2 audit risk assessments and comparing them with the results of the first audit risk assessment, the results are shown in Table 4.

For the 20 enterprises selected, except for enterprise 17, the results of the 2nd audit risk assessment have different degrees of reduction between -0.3663 and -0.0119. It can be seen that enterprises take more targeted measures to deal with and control audit weaknesses to help reduce audit risk, which is of practical value to ensure the financial security of enterprises and enhance their market competitiveness.

Table 4. Results and comparison of two audit evaluations of enterprises

Enterprise number	Before	After	Audit risk level	Result variation
1	0.5066	0.4864	Medium risk	-0.0202
2	0.3198	0.1864	Low risk	-0.1334
3	0.4636	0.3915	Lower risk	-0.0721
4	0.1679	0.1526	Low risk	-0.0153
5	0.2961	0.2348	Lower risk	-0.0613
6	0.6615	0.5158	Medium risk	-0.1457
7	0.8364	0.8245	High risk	-0.0119
8	0.5263	0.4597	Medium risk	-0.0666
9	0.7034	0.5245	Medium risk	-0.1789
10	0.4965	0.4364	Medium risk	-0.0601
11	0.5064	0.3485	Lower risk	-0.1579
12	0.6158	0.2495	Lower risk	-0.3663
13	0.7168	0.5482	Medium risk	-0.1686
14	0.8138	0.5421	Medium risk	-0.2717
15	0.8134	0.6415	Higher risk	-0.1719
16	0.3952	0.2485	Lower risk	-0.1467
17	**0.4348**	**0.5431**	**Medium risk**	**0.1083**
18	0.6425	0.5154	Medium risk	-0.1271
19	0.2498	0.1648	Low risk	-0.085
20	0.3487	0.2484	Lower risk	-0.1003

4.5. Analysis of specific applications in the model risk assessment procedure

In the implementation of risk assessment procedures, the model provides auditors with a module for analyzing industry conditions and key financial indicators, which assists auditors in understanding the situation of the audited entity and enabling them to identify unusual transactions or events in a timely manner, thus improving audit efficiency. This section introduces the convenience for auditors in implementing risk assessment from the perspectives of industry condition analysis and key financial performance analysis in the company's implementation of risk assessment procedures.

4.5.1. Analysis of the state of the industry. First of all, from the point of view of the revenue growth rate of the industry in which the audited unit is located, the auditor's model provides an analysis of the industry's revenue growth rate chart, which can clearly understand the revenue growth rate of the industry in which the company is located in comparison with the overall level of the country's enterprises above the national scale for the same period, as shown in Figure 5.

During the period from 2019 to 2020, the revenue situation of the industry in which the subject company is located is not optimistic, and the industry's revenue growth rate in both years is lower than the overall level of enterprises above the national scale in the same period, and the revenue growth rate in these two years is 18.79% and 16.48%, respectively. Further analysis, the industry in the past two years, the decline in revenue growth rate, mainly due to the national policy of regulating product prices. As a result, it can also be predicted that the company's operating income will be adversely affected.

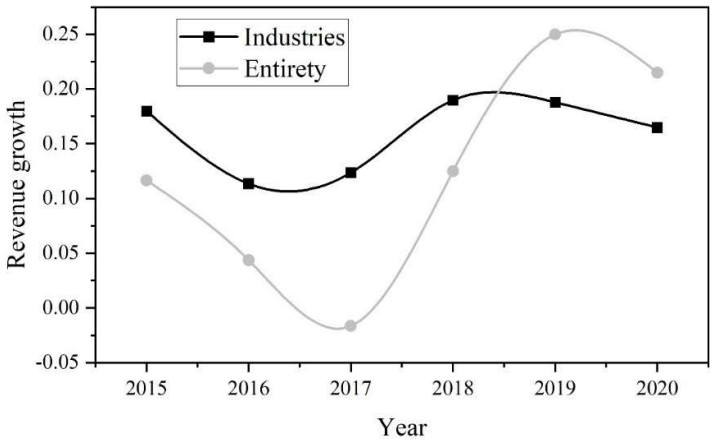


Fig. 5. Comparative analysis of industry revenue growth rate

Then, from the point of view of the changes in the net profit of the industry in which the audited unit is located, the auditor can understand the net profit growth rate of the industry in which the company under test is located in comparison with the overall level of enterprises above the national scale in the same period through the industry net profit growth rate analysis chart provided by the model, as shown in Figure 6.

From 2017, the overall net profit growth rate of enterprises is decreasing year by year. Through the industry revenue growth rate and the industry net profit growth rate, it can be learned that from 2017 the industry development trend is slowing down. Especially in the period from 2019 to 2020, the net profit growth rate of the industry in 2020 is negative (-24.87%), and the industry situation is not optimistic.

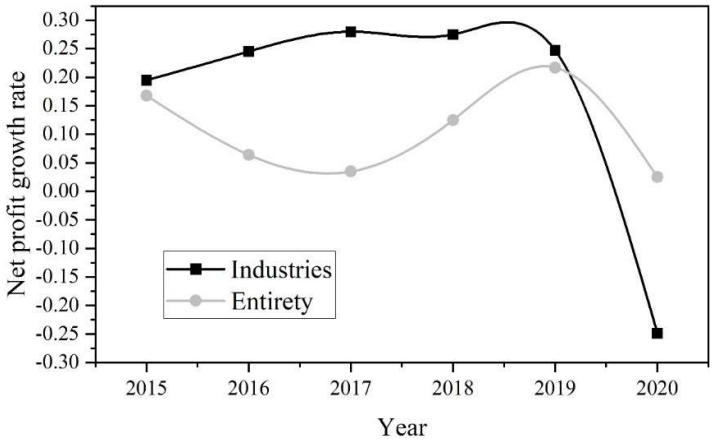


Fig. 6. Comparative analysis of industry net profit growth rate

4.5.2. Analysis of key financial performance. Figure 7 shows a comparative analysis of the net sales margin of the subject company, which ranges from 15% to 16.5% during the period from 2017 to 2020, which is much higher than that of its competitors in the same industry and higher than the industry average. The net sales margin of the subject company has been high and is five to six times higher than that of its industry competitors. The chart shows that the test company's ability to generate sales revenue is strong.

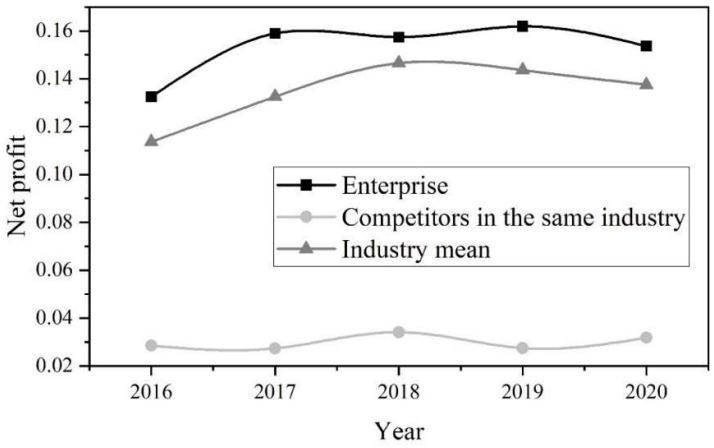


Fig. 7. Comparative analysis of corporate sales net profit rate

Figure 8 shows a comparative analysis of the company's gross sales margin, from the perspective of profit, the gross sales margin of the subject company from 2016 onwards is much higher than that of its competitors in the same industry, and higher than the average level of the industry in which it operates, in particular, between 2016 and 2017, the gross sales margin has achieved an increase of 31.06%. The level of gross profit margin on sales is not in line with the conventional value of companies in this industry in general, and is two to three times higher than that of competitors in the same industry. The high gross sales margin resulted in a surprisingly high net sales margin for the subject company as well. The gross sales margin of the test company appeared to be clearly abnormal, and therefore, the auditors believed that there might be the possibility of inflating operating income and underestimating operating costs in order to maintain profitability levels. This affects the determination of the occurrence of operating income and the determination of the completeness of operating costs of the test company. The auditors should implement further audit procedures to obtain sufficient audit evidence.

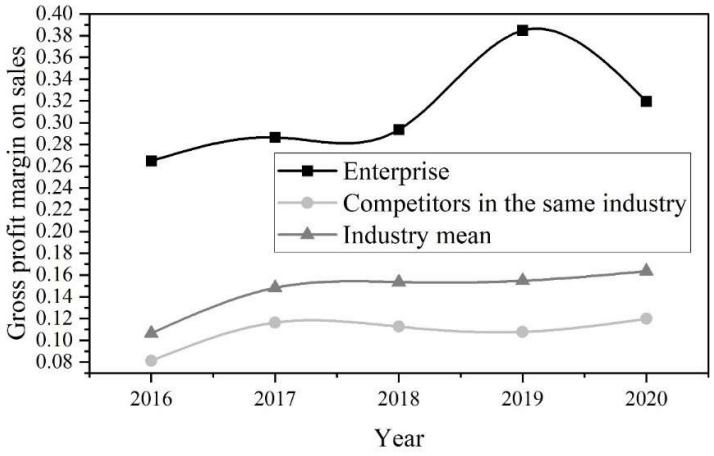


Fig. 8. Comparative analysis of company sales gross profit margin

5. Recommendations for addressing enterprise audit risks

5.1. Recommendations for addressing the risk of material misstatement

5.1.1. Recommendations for addressing the risk of material misstatement at the financial statement level. Auditors need to maintain a high degree of professional skepticism in the audit of such operations, to maintain prudence in obtaining audit information, and to reduce dependence on audit

evidence that has an important impact on the audit opinion, and to be able to issue objective and fair audit opinions. The person in charge of the audit should strengthen supervision in the whole audit process on the one hand, and on the other hand, should also carry out more process guidance to strengthen the scientific rationality of the audit process, thus reducing the inspection risk.

5.1.2. Recommendations for addressing the risk of material misstatement at the determination level. Matters involving revenue and cost recognition are high-risk matters, and therefore the overall audit risk should be addressed by setting up substantive procedures to match these high-risk items in particular when developing the audit plan.

5.2. Inspection Risk Response Recommendations

5.2.1. Strengthening management and risk control in the conduct of audits. The transactions of the company under test are characterized by the real-time nature of the process, so its data may be changing at any time, so in order to improve the efficiency of the auditor's work to reduce the process of the auditor to go to the field for data acquisition, as well as to increase the unpredictability of the audit and reduce the risk of tampering with the audit information, so that the auditor can access the audit information at any time, and the data acquisition is also synchronized with the enterprise to reduce the mismatch. Among them, there are four processes in the auditing process that can be unfolded through networked auditing:

1) Audit data collection. Auditors need to use the specific data of the audited unit in determining whether to undertake the audit case and in formulating the audit plan after undertaking the case, so auditors can use the networked audit system in data collection to collect the synchronized dynamic data of the enterprise instead of static data through the same network environment to supervise the real-time situation of the enterprise.

2) Audit data transmission. After the audit data collection is completed, the audit data need to be transmitted to the accounting firm through the transmission channel. Through the networked audit, the auditor can select the most suitable transmission method from the many transmission methods for the transmission of data.

3) Audit data storage. After the collected audit data is transmitted to the accounting firm's system, the storage of a large amount of audit information is not affordable by ordinary software, so the cloud data storage technology of networked auditing plays an important role.

4) Audit data analysis. When analyzing the received audit data, the financial information needs to be calculated, and the networked audit can analyze the data with the help of cloud computing and other tools, which not only improves the accuracy of calculation but also saves labor.

5.2.2. Increased investment in audit software development. Problems in the audit process include not only the lack of auditing standards and guidance, but also the lack of auditing software at the technical level, as most of the more mature auditing software available today is for auditing traditional industries. Therefore, the proposal to address the inspection risk suggests increasing the research and development costs for audit software. This initiative is not only an effective suggestion for one accounting firm, but also a progressive initiative for the entire auditing industry, so we can call for the Chinese Institute of Certified Public Accountants to take the lead and the accounting firms to implement it, and to work hand in hand to develop auditing software suitable for the emerging Internet enterprises. This can be used to reduce the risk of inspection due to mistakes made by inexperienced auditors.

6. Conclusion

This paper proposes a weighted KNN deep neural network algorithm and introduces a multi-branch deep neural network structure on this basis to improve the cost sensitivity of audit data anomaly detection. The audit risk assessment index system is designed to assess the risk assessment results of the test company from the industry status and key financial performance in various aspects.

1) The model KNN model is trained to converge, and the reconstruction error of each entry is obtained using the trained KNN5 model. Under 100 generations of training conditions, the reconstruction error of the model for the detection of data anomalies is 1-4 the number of local anomalies in the range of 6 to 12.

2) Selected 20 companies before and after two audit risk assessment, compared with the 1st risk assessment, the 2nd audit risk assessment results have different degrees of reduction between -0.3663~-0.0119. This shows that the intervention of weighted KNN deep neural network model helps to reduce the audit risk.

3) From the perspective of the revenue growth rate of the industry in which the audited unit is located, the revenue growth rate of the audited company is 18.79% and 16.48% during 2019-2020, and the industry revenue growth rate in both years is lower than the overall level of the national enterprises above the scale in the same period, so it is predicted that the future operating income of the audited company will be adversely affected.

About the Author

Xia Li was born in Dali, Yunnan, P.R. China, in 1988. She received the master's degree from Tianjin University of Technology, P.R. China. Now, she worked in the School of Economics and Management in Yunnan University of Technology and Business. Her research interest includes corporate governance, internal control, accounting and auditing.

References

[1] Zhong, H., Yang, D., Shi, S., Wei, L., & Wang, Y. (2024). From data to insights: the application and challenges of knowledge graphs in intelligent audit. Journal of Cloud Computing, 13(1), 114.

[2] Dande Alekya, D. P. M. (2021). Study on Manual Auditing for Web Application Vulnerability Detection. Annals of the Romanian Society for Cell Biology, 19612-19618.

[3] Gamayuni, R. R. (2018). The effect of internal auditor competence and objectivity, and management support on effectiveness of internal audit function and financial reporting quality implications at local government. International Journal of Economic Policy in Emerging Economies, 11(3), 248-261.

[4] Lee, K. K., & Levine, C. B. (2020). Audit partner identification and audit quality. Review of Accounting Studies, 25(2), 778-809.

[5] Rakipi, R., De Santis, F., & D'Onza, G. (2021). Correlates of the internal audit function's use of data analytics in the big data era: Global evidence. Journal of International Accounting, Auditing and Taxation, 42, 100357.

[6] No, W. G., Lee, K., Huang, F., & Li, Q. (2019). Multidimensional audit data selection (MADS): A framework for using data analytics in the audit data selection process. Accounting Horizons, 33(3), 127-140.

[7] Post, R., Beerepoot, I., Lu, X., Kas, S., Wiewel, S., Koopman, A., & Reijers, H. (2021, October). Active anomaly detection for key item selection in process auditing. In International Conference on Process Mining (pp. 167-179). Cham: Springer International Publishing.

[8] Bernardino, D., Pedrosa, I., & Laureano, R. M. (2018, June). Analytical methods for auditing and anomaly/fraud detection. In 2018 13th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE.

[9] Peng, C., & Tian, G. (2023). Intelligent auditing techniques for enterprise finance. Journal of Intelligent Systems, 32(1), 20230011.

[10] Zhang, L. (2021, November). The Construction of Rural Poverty Alleviation Audit using Big Data Anomaly Detection. In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 879-882). IEEE.

[11] Shan, R., Xiao, X., Che, J., Du, J., & Li, Y. (2022). Data Mining Optimization Software and Its Application in Financial Audit Data Analysis. Mobile Information Systems, 2022(1), 6851616.

[12] Schumann, G., Kruse, F., & Nonnenmacher, J. (2020). A practice-oriented, control-flow-based anomaly detection approach for internal process audits. In Service-Oriented Computing: 18th International Conference, ICSOC 2020, Dubai, United Arab Emirates, December 14–17, 2020, Proceedings 18 (pp. 533-543). Springer International Publishing.

[13] Kiefer, S., & Pesch, G. (2021). Unsupervised anomaly detection for financial auditing with model-agnostic explanations. In KI 2021: Advances in Artificial Intelligence: 44th German Conference on AI, Virtual Event, September 27–October 1, 2021, Proceedings 44 (pp. 291-308). Springer International Publishing.

[14] Yousif, N. S., & Mohamed, S. A. (2022). The Role of Internal Audit in Assessing the Risks of Management Decisions regarding Strategic Operations) Acquisition). Journal of Economics and Administrative Sciences, 28(133), 172-186.

[15] Sudirman, S., Sasmita, H., Krisnanto, B., & Muchsidin, F. F. (2021). Effectiveness of internal audit in supporting internal control and prevention of fraud. Bongaya Journal of Research in Accounting (BJRA), 4(1), 8-15.

[16] Hassan, N. A., Zailani, S., & Rahman, M. K. (2021). Impact of integrated audit management effectiveness on business sustainability in manufacturing firms. Management Research Review, 44(12), 1599-1622.

[17] AL-QUDAH, L. A. (2021). The Impact of business risk-based audit approach on reducing unsystematic risks: Evidence from Jordanian banks. The Journal of Asian Finance, Economics and Business, 8(1), 343-352.

[18] Tuan, D. A., & Dung, N. N. K. (2023). Development of Audit Risk Model Applied in Public Investment Project Audit: The State Audit in Vietnam. Przestrzeń Społeczna (Social Space), 23(1), 193-221.

[19] Fakhfakh, I., & Jarboui, A. (2023). Research on the relationship between audit risk assessment and risk governance: evidence from Tunisia. Journal of African Business, 24(1), 95-110.

[20] El-Khaldi, N., Nashwan, I., & Kullab, A. (2018). Business Risk Assessment Using Client Strategy Analysis Approach in order to Increase the Efficiency and Effectiveness of the Audit Process. Journal of Economics and Administrative Sciences, 24(103), 43-43.

[21] Țîrcovnicu, G. I., & Hategan, C. D. (2023). The Audit Risk Assessment of European Small-and Mid-Size Enterprises. Journal of Risk and Financial Management, 16(3), 158.

[22] Singh, N., Lai, K. H., Vejvar, M., & Cheng, T. E. (2019). Data-driven auditing: A predictive modeling approach to fraud detection and classification. Journal of Corporate Accounting & Finance, 30(3), 64-82.

[23] Leo Handoko, B., Rosita, A., & Leonarda Warganegara, D. (2020, October). Forensic and Anomaly Detection Using Generalized Audit Software. In Proceedings of the 2020 4th International Conference on E-Business and Internet (pp. 47-51).

[24] Nonnenmacher, J., & Gómez, J. M. (2021). Unsupervised anomaly detection for internal auditing: Literature review and research agenda. International Journal of Digital Accounting Research, 21.

[25] Neskorodieva, T., Fedorov, E., Smirnov, O., Rudakov, K., & Neskorodieva, A. (2021). Method Detection Audit Data Anomalies on Basis Restricted Cauchy Machine. In CPITS II (1) (pp. 1-12).

[26] Kleboth, J. A., Kosorus, H., Rechberger, T., & Luning, P. A. (2022). Using data mining as a tool for anomaly detection in food safety audit data. Food Control, 138, 109004.

[27] Wang, S. (2024). Intelligent BiLSTM-Attention-IBPNN method for anomaly detection in financial auditing. IEEE Access.

[28] Özbaltan, N. (2024). Applying machine learning to audit data: Enhancing fraud detection, risk assessment and audit efficiency. EDPACS, 69(9), 70-86.

[29] Antwi, B. O., Adelakun, B. O., Fatogun, D. T., & Olaiya, O. P. (2024). Enhancing audit accuracy: The role of AI in detecting financial anomalies and fraud. Finance & Accounting Research Journal, 6(6), 1049-1068.

[30] Safina, A. R. (2018). Internal control and audit of public sector organizations: Risk assessment. Revista San Gregorio, (25), 151-157.

[31] Porcuna-Enguix, L., Bustos-Contell, E., Serrano-Madrid, J., & Labatut-Serer, G. (2021). Constructing the audit risk assessment by the audit team leader when planning: using fuzzy theory. Mathematics, 9(23), 3065.

[32] Demirović, L., Isaković-Kaplan, Š., & Proho, M. (2021). Internal audit risk assessment in the function of fraud detection. Journal of Forensic Accounting Profession, 1(1), 35-49.

[33] Sysoieva, I., Zagorodniy, A., Pylypenko, L., Tomilin, O., Balaziuk, O., & Pohrishchuk, O. (2021). Analysis of potential risks of audit of agricultural enterprises. Agricultural and Resource Economics: International Scientific E-Journal, 7(1), 164-191.

[34] Fang, L., & Sun, H. (2024). Research and Optimization of Audit Risk Assessment Model Based on Regression Algorithm. International Journal of High Speed Electronics and Systems, 2540140.

[35] Sun, G., & Guan, B. (2024). Enterprise Audit Risk Assessment and Prevention based on AHP Analysis. Scalable Computing: Practice and Experience, 25(3), 2013-2020.

[36] Wang Rongrong. (2024). Legal Privacy Protection Machine Learning Method Based on Word2Vec Algorithm. International Journal of Information Security and Privacy (IJISP)(1),1-19.

[37] Bohang Chen,Jun Ma,Lingfei Zhang,Zhuang Xiong,Jinyu Fan & Haiming Lan. (2023). An improved weighted KNN fingerprint positioning algorithm. Wireless Networks(6),6011-6022.

[38] Chen Sian,Zuo Yajuan & Wang Rui. (2024). RETRACTED: An optimal framework for natural English translate processing: On the application of artificial neural network and Adam optimization algorithm. Journal of Intelligent & Fuzzy Systems(5-6, Supplement 1),205-219.

[39] Zhenzheng Yan. (2024). Audit Risk Assessment Model Construction by Incorporating Deep Learning Techniques. Applied Mathematics and Nonlinear Sciences(1).