Research on the Application of Intelligent Decision-Making and Machine Learning in Auditing

Xing Zhang*, Hong Liu

Finance and Auditing Department, Engineering University of Joint Logistics Support Force, Chongqing, China

*zhangxingaudit@163.com, professoraudit@163.com

Abstract. Amid the wave of digital transformation, auditing faces multiple challenges including surging data volumes, increasingly complex operational scenarios, and heightened oversight requirements. Traditional audit models reliant on manual expertise struggle to meet demands for high-quality development. Intelligent decision-making centers on data-driven approaches, integrating multidimensional information to generate optimal solutions; Machine learning, as a core branch of artificial intelligence, enables algorithmic models to uncover data patterns and achieve autonomous learning. The synergy between these two technologies provides critical support for overcoming bottlenecks in auditing. This paper systematically outlines the technical foundations and implementation details of intelligent decision-making and machine learning, analyzes their application pathways across the entire audit process—"pre-audit preparation, on-site execution, report generation, and corrective action tracking"—and offers references for audit digital transformation that combine theoretical depth with practical value.

Keywords: Intelligent Decision-Making; Machine Learning; Audit Digitalization; Risk Identification; Full-Process Application

1 Introduction

As the "immune system" of national governance and a critical component of corporate risk management, auditing delivers core value by ensuring economic activity compliance and enhancing resource allocation efficiency through independent review, risk identification, and issue rectification. With the rapid development of the digital economy, audit subjects have transitioned from "offline paper-based" to "online digital" formats, resulting in audit data characterized by "massive volume, heterogeneous nature, and real-time generation." The integrated application of intelligent decision-making and machine learning can overcome three major challenges in traditional auditing: low information processing efficiency, insufficient risk identification accuracy, and supervisory lag. Machine learning employs supervised learning, unsupervised learning, and deep learning algorithms to extract risk patterns and identify anomalies from massive heterogeneous datasets [1], addressing the challenges of

H. Dai et al. (eds.), *Proceedings of the 2025 5th International Conference on Business Administration and Data Science (BADS 2025)*, Advances in Computer Science Research 125,
https://doi.org/10.2991/978-94-6463-980-3_25

"difficult data processing and difficult risk detection." Intelligent decision-making builds upon machine learning analytics by integrating audit rules, industry standards, and business logic to construct a dynamic decision framework. This enables automated audit plan generation, real-time progress optimization, and precise effectiveness evaluation.

2 Technical Foundations, Collaborative Logic, and Implementation Details of Intelligent Decision-Making and Machine Learning

2.1 Machine Learning: The Core Engine for Audit Data Processing and Risk Identification

Machine learning achieves in-depth analysis of audit data through the "data input—model training—prediction output" process. Technical selection, implementation steps, and parameter configuration vary across different algorithms, as detailed below.

(a) Supervised Learning: Audit Risk Classification and Quantitative Assessment. Its core application scenarios involve audit risk classification and quantitative assessment, suitable for scenarios with clear classification standards and sufficient historical data. Implementation details include a technical stack utilizing Python (Scikit-learn library) and SparkMLlib; data inputs encompass financial indicators (debt-to-asset ratio, revenue growth rate, etc.), operational indicators (transaction frequency, contract value, etc.), and regulatory indicators (penalty records, etc.), totaling 28 feature variables; Model parameters are set to 100 trees, maximum depth 15, and minimum sample split 5. The training process utilizes 70% of historical audit cases (over 1,000 projects) as the training set and 30% as the test set, with model optimization achieved through 5-fold cross-validation. In audit adaptability validation, the model achieved 92.3% accuracy in risk level prediction and 89.7% accuracy in regulatory clause matching on the test set, both surpassing the 82.1% accuracy of traditional manual methods.

(b) Unsupervised Learning: Audit Anomaly Detection and Hidden Risk Mining. Primarily used for detecting abnormal transactions in audits, suitable for scenarios with unclear risk features or unknown issues requiring identification. The technology stack employs Flink (real-time computation) and Hadoop (data storage). Input data comprises corporate cash flow records encompassing four clustering dimensions: transaction amount, counterparty, transaction time, and account. Model parameters were determined using the elbow rule, setting k=8 for the number of clusters, with 100 iterations and Euclidean distance as the metric. Implementation involves real-time collection of T+1 cash flow data, followed by Z-score normalization for standardization before cluster training. Transactions deviating more than 3σ from the cluster center are flagged as anomalies.

(c) Deep Learning: Complex Unstructured Data Processing and Trend Forecasting.

Focuses on handling time-series prediction and image verification tasks in auditing, suitable for scenarios with high data dimensionality and reliance on long-term trend analysis. The technology stack includes TensorFlow (model training), OpenCV (image processing), and BERT (text extraction) [2]. LSTM model parameters are set to: 2 hidden layers, 64 units per layer, dropout coefficient 0.2, and time step 60 (corresponding to 5 years of financial data). CNN model parameters are: 3×3 convolution kernel size, max pooling for pooling layers, and 32 output channels. The implementation process comprises two stages: - Time series forecasting: Inputs five years of financial statement data, undergoes LSTM training to output 1-2 year risk probabilities. - Image verification: Performs CNN feature extraction on fixed asset photographs, identifies fraudulent images by comparing "pixel similarity" (threshold ≥95% indicates duplication).

2.2 Intelligent Decision-Making: Key Support for Audit Plan Optimization and Full-Process Control

Intelligent decision-making establishes a closed-loop system of "data-algorithm-decision." The system architecture adopts a cloud-native microservices design, decomposed into three core modules: the rule engine module, resource scheduling module, and real-time feedback module. The specific implementation steps are as follows [3].

(a) Automated Audit Plan Generation: Resource Allocation Based on Risk Priority.

Input data includes machine learning-derived risk priorities (e.g., procurement process risk score 85, sales process 60, finance process 45) and audit resource availability (10-person team comprising 3 procurement experts, 4 finance experts, 3 business experts; 15-day cycle). The rule engine incorporates built-in "Audit Standards" and "Industry Resource Allocation Guidelines," establishing core rules such as "High-risk areas must receive ≥60% of resources" and "Expert matching must reach ≥80%." The final plan allocates 6 personnel (3 procurement experts + 3 business experts) to the procurement process, focusing on tasks like "verifying supplier qualifications and payment approval records"; 3 personnel (2 finance experts + 1 business expert) to the sales process; and 1 personnel (1 finance expert) to the finance process, with clear deadlines for each phase.

(b) Dynamic Optimization During Audit Process: Decision Adjustments Based on Real-Time Data.

Triggered when auditors upload new field data via mobile devices—e.g., discovering "affiliated relationships among material suppliers" during an investment project audit. Upon receiving data, the system immediately invokes the association rule mining model (Apriori algorithm) to reanalyze supplier-project transaction records, identifying new risks like "affiliated suppliers quoting 15% above market average." Subsequently, the audit plan is updated to include "affiliated transaction penetration verification" and " Market Price Comparison Analysis" steps. One financial expert is reassigned from sales to procurement, extending the procurement review cycle by two days to ensure comprehensive risk coverage.

(c) Audit Result Validation and Iteration: Model Optimization Based on Effectiveness Retrospection.
Core evaluation metrics include: Risk Identification Accuracy (percentage of issues detected in recommended high-risk areas), Issue Omission Rate (percentage of issues detected in non-recommended areas), Resource Investment Return Rate (number of issues / resources invested) The iteration mechanism operates as follows: If a project exhibits "significant issues overlooked in low-risk areas," analyze causes (e.g., insufficient model feature variables, unreasonable risk weighting) and retrain the model after adding new feature variables or adjusting weights. If the return on resource investment falls below expectations, optimize resource allocation logic, such as increasing the resource allocation proportion for high-risk areas. For example, in one project, adding the "supplier change frequency" feature variable reduced the model's omission rate from 8% to 3%.

2.3 Synergistic Logic: Establishing a Closed-Loop Audit Process of "Data-Driven → Intelligent Analysis → Dynamic Decision-Making"

Intelligent decision-making and machine learning achieve deep synergy through an ordered chain linking the "Data Input Layer—Intelligent Analysis Layer—Decision Output Layer—Practical Validation Layer—Self-Evolution Layer," forming a self-optimizing audit closed loop. The functional positioning and data flow of each segment are as follows.

(a) Data Input Layer (Foundation of the Collaborative Closed Loop).
The core objective is to provide high-quality standardized data sources for machine learning while establishing the foundation for data credibility in intelligent decision-making. First, data is integrated through a "multi-source data acquisition matrix": historical audit cases, regulatory databases, and personnel resource data are obtained from internal systems of audit agencies; financial ledgers, operational logs, and rectification progress data are collected from auditee systems; and industry benchmark data, regulatory policy data, and third-party evaluation data are extracted from external platforms. Second, standardized processing is achieved through a "three-tier data governance mechanism": The first tier uses ETL (Extract-Transform-Load) tools to convert unstructured and semi-structured data into structured formats [4], standardizing naming conventions and data types for core fields such as "Issue ID," "Rectification Deadline," and "Responsible Department." The second layer employs a "rules engine + manual review" approach to eliminate duplicate data, correct logical errors, and fill missing data (e.g., using multiple imputation methods for missing values); The third

layer establishes multi-source data associations through "audit issue IDs," forming a complete data chain linking "financial data – operational data – rectification data – external data" to provide comprehensive support for subsequent analysis.

(b) Intelligent Analysis Layer (Collaborative Closed-Loop Computing Core)".
Responsible for transforming standardized data into audit-ready risk leads, prioritization rankings, and trend prediction outcomes, providing data-driven foundations for intelligent decision-making. Different machine learning models are invoked based on audit scenario requirements: In risk identification scenarios, when classifying "known risk types," supervised learning models are trained using historical labeled samples to output "risk type + risk probability"; For uncovering "unknown hidden risks," it employs unsupervised learning models to analyze intrinsic data distribution patterns, flagging "transactions deviating from normal clusters" or "abnormal correlation relationships." In trend prediction scenarios, when forecasting remediation cycles or risk recurrence probabilities, it utilizes deep learning models with time-series data inputs, outputting "prediction results + error margins."

(c) Decision Output Layer (Core of Collaborative Closed-Loop Implementation).
Integrates machine learning analysis results with audit business rules to generate actionable audit plans, early warning mechanisms, and intervention strategies, with real-time adjustment capabilities [5]. During plan generation, a "multi-dimensional decision framework" is constructed: combining risk priority with resource constraints to formulate "resource allocation plans," and referencing audit regulations and industry standards to define "task lists and timelines." The dynamic adjustment phase features a "real-time feedback—model recalculation—solution update" mechanism. When auditors upload new field data, the system immediately triggers machine learning model reanalysis and adjusts decision plans based on the new results.

(d) Practice Validation Layer (Effectiveness Verification of Collaborative Closed-Loop).
Implement solutions generated by intelligent decision-making and collect practical data for iterative optimization, forming a "decision-execution-feedback" chain. During plan execution, auditors follow task lists: in pre-audit preparation, they prioritize high-risk subjects and formulate plans; during field implementation, they coordinate rectifications based on real-time alerts; in report generation, they use system-generated drafts supplemented with details; and in rectification tracking, they increase verification frequency based on recurrence risk predictions. Data feedback is collected through multiple channels: the audit system automatically records "task completion status" and "number of issues identified"; auditors complete "Practice Feedback Forms" to document "plan rationality issues" and "model analysis deviations"; audited entities report "difficulty in implementing corrections." This data serves as critical basis for subsequent optimization.

(e) Self-Evolution Layer (Core of Collaborative Closed-Loop Continuous Optimization).

Machine learning model parameters and intelligent decision rules are adjusted based on practice feedback data to enhance application accuracy and rationality. Model optimization follows a "performance evaluation—parameter adjustment—retraining" mechanism: core metrics (e.g., accuracy, error rate) are calculated first. If metrics deviate, causes (e.g., insufficient samples, unreasonable parameters) are analyzed. Subsequently, labeled samples are supplemented, parameters are adjusted, or algorithms are replaced for retraining. Finally, optimization effectiveness is validated using a test set. Intelligent decision rule optimization adjusts parameters like "resource allocation logic" and "task time limit settings" based on practical feedback. For instance, in actual industry audits, feedback may indicate increasing the resource allocation for high-risk areas from 60% to 65% to ensure decision schemes align with industry characteristics. Optimized model parameters and decision rules then reinvigorate subsequent operations in the data input layer, forming a complete closed-loop system. (see Table 1).

Table 1. Key action of Each Link in the Collaborative Closed-Loop

Closed-Loop Link	Key Actions
Data Input Layer: Integration and Standardization of Audit Data	1. Multi-source data collection (internal systems + audited entities + external platforms)2. Three-layer data governance (format conversion → error correction → association modeling)3. Build a data chain based on "audit issue ID"
Intelligent Analysis Layer: Machine Learning Model Analysis and Result Output	1. Risk identification: supervised learning (classification of known risks), unsupervised learning (mining of unknown risks)2. Trend prediction: deep learning (prediction of rectification cycle/risk recurrence probability)3. Generate analysis reports supported by data
Decision Output Layer: Intelligent Decision-Making System Scheme Generation and Dynamic Adjustment	1. Scheme generation: formulate resource allocation and task time limits based on risk priority and resource constraints2. Dynamic adjustment: receive new data → trigger model recalculation → update the scheme3. Align with audit regulations and industry standards
Practice Validation Layer: Audit Practice Execution and Data Feedback	1. Scheme execution: promote pre-audit/on-site/report/rectification work in accordance with the task list2. Data collection: automatically record task completion status + manually fill in feedback forms + feedback from audited entities
Self-Evolution Layer: Model and Decision Parameter Iterative Optimization	1. Model optimization: evaluate indicators (accuracy rate/error rate) → supplement samples/adjust parameters/replace algorithms → retrain and verify2. Rule optimization: adjust resource allocation ratio, task time limit buffer, and add industry-specific rules

3 Application Pathways of Intelligent Decision-Making and Machine Learning Across the Audit Lifecycle

Taking the "Annual Financial Audit of a Major Manufacturing Enterprise" (Data Volume: 10TB financial data, 500 million transaction records, 20,000 contract texts) as an example, the end-to-end application pathways and implementation details are as follows.

3.1 Pre-audit Preparation Phase: Intelligent Planning and Risk Focus (Completed in 3 days, vs. 15 days traditionally)

First, process 500 million transaction records using the Flink real-time computing tool to identify over 300 suspicious transactions meeting the criteria "transaction amount > RMB 5 million, transaction frequency < 1 per quarter." Subsequently, apply a random forest model to assess risks across all business segments, generating risk priority scores (Procurement: 85 points, Inventory Management: 72 points, Sales: 60 points). Finally, the intelligent decision system allocated resources based on the audit team's expertise (4 procurement specialists, 2 inventory specialists, 1 sales specialist, 3 finance specialists): 4 members for procurement, 2 for inventory management, 1 for sales, and 3 for consolidated financial verification. Specific tasks and timelines were defined: "Procurement Contract Compliance Review (3 days), Inventory Discrepancy Analysis (2 days), and Sales Receivables Tracking (2 days)" with defined timelines, ensuring resources are prioritized toward high-risk areas.

3.2 On-site Implementation Phase: Real-time Monitoring and Intelligent Intervention (Completed in 12 days, vs. traditional 30 days)

Real-time Alerting: The K-Means model continuously monitors inventory data. Alerts trigger immediately upon detecting anomalies like "a warehouse's inventory turnover rate falling below the industry average of 60%." Dynamic Adjustment: The system employs association rule mining to analyze inventory and procurement data. Upon identifying anomalies such as "a supplier changing its name three times within a short period," it automatically adds a "supplier qualification penetration verification" task. This requires auditors to cross-check the supplier's business registration changes with actual controlling shareholder information. Automated Verification Phase: A CNN model performs feature extraction and comparison on 2,000 fixed asset photos, identifying 5 reused fraudulent images. This replaces manual repetitive image verification, reducing verification time by 80%.

3.3 Report Generation Phase: Intelligent Drafting and Quality Control (2 days, vs. traditional 7 days)

Information extraction: A BERT pre-trained model automatically extracts key details—such as "problem descriptions, responsible departments, violation amounts, and rectification requirements"—from 20,000 contract texts, generating standardized audit ledgers to prevent manual omissions and errors. Report generation: An intelligent decision system automatically drafts reports based on the fixed format of the "Audit Report Standards," integrating ledger information to include core elements like risk point lists, audit evidence links, and rectification recommendations [6]. Quality review phase: Machine learning models cross-reference report content against the audit regulations database, correcting 2 instances of "incorrectly cited regulatory provisions" (e.g., misattributing Accounting Standard for Business Enterprises No. 14 as No. 15). Manual intervention is limited to supplementing explanations for special circumstances identified during on-site verification, significantly shortening the report drafting cycle.

3.4 Rectification Tracking Phase: Long-Term Oversight and Effectiveness Evaluation (6 months)

Progress Monitoring: An LSTM model predicts the completion timeline for each rectification task (e.g., 45 days for procurement compliance, 30 days for inventory variance) based on historical data (duration of past rectifications, resource allocation). The audit system tracks progress in real-time, triggering alerts when any task falls more than 10% behind schedule. Effect Evaluation Phase: A BP neural network model is constructed. Inputting business data from 3 months post-rectification (e.g., procurement approval compliance rate, inventory turnover rate) outputs the probability of risk recurrence (5% for procurement, 3% for inventory). Subsequent Verification Phase: The intelligent decision-making system generates differentiated verification plans based on recurrence probabilities, recommending "verification every 2 months" for procurement and "verification every 3 months" for inventory to prevent risk resurgence.

4 Conclusion

In conclusion, the integration of intelligent decision-making and machine learning is pivotal for auditing's digital transformation. This synergy creates a self-improving closed-loop system—spanning data input, intelligent analysis, decision output, practical validation, and self-evolution—that comprehensively enhances audit efficiency and effectiveness. Machine learning acts as the analytical core, automating risk identification and prediction through advanced algorithms, while intelligent decision-making translates these insights into optimized plans and dynamic adjustments. Practical applications, as demonstrated in large-scale audits, confirm significant reductions in process timelines and substantial improvements in risk detection accuracy, from planning to rectification tracking. Looking forward, the evolution of intelligent auditing hinges on advancing model interpretability, strengthening data security, and integrating emerging AI technologies. Ultimately, this technology-driven paradigm is set to become a cornerstone of robust economic governance and high-quality development.

References

1. Gustavo Fleury Soares; Induraj Pudhupattu Ramamurthy, "A Comparison of Machine Learning and Deep Learning Models with Advanced Word Embeddings: The Case of Internal Audit Reports," in Optimization and Machine Learning: Optimization for Machine Learning and Machine Learning for Optimization, Wiley, 2022, pp.151-167.
2. Z. Xu, Y. Sheng, Q. Bao, X. Du, X. Guo and Z. Liu, "BERT-Based Automatic Audit Report Generation and Compliance Analysis," 2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA), Xi'an, China, 2025, pp. 1233-1237.
3. G. V. Krishnan, J. J. Lee, and A. K. Srivastava, "Artificial Intelligence and the Future of Audit Evidence: An Analysis of PCAOB Inspection Data," Contemporary Accounting Research, vol. 39, (4) 2022, pp. 2526-2556.
4. M. Balta and V. Felea, "Using Shannon Entropy in ETL Processes," Ninth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2007), Timisoara, Romania, 2007, pp. 151-156.
5. F. Caputo, B. Keller, M. Möhring, L. Carrubbo, R. Schmidt, Advancing beyond technicism when managing big data in companies' decision-making, J. Knowl. Manag., 27 (10) (2023), pp. 2797-2809.
6. M. Alles, "Drivers of the adoption and integration of AI technologies in auditing: A focus on the audit future", Journal of Information Systems, vol. 35, (3) 2021, pp. 5–22.