World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3115

DIGITAL-INTELLIGENT PENETRATING SUPERVISION AND DATA ASSET AUDIT RISK EARLY WARNING: AN EMPIRICAL STUDY ON LISTED ELECTRONIC INFORMATION FIRMS IN SICHUAN PROVINCE

YongHui Li, HaiBo Zhang*
School of Digital Economy and Management, Sichuan University of Technology and Business, Meishan 620000, Sichuan, China.
**Corresponding Author: HaiBo Zhang*

Abstract: The digital economy has elevated data assets to a strategic corporate resource, yet their unique characteristics — non-rivalry, value volatility, and valuation ambiguity — introduce unprecedented audit risks. This study develops a novel theoretical framework integrating Digital-Intelligent Penetrating Supervision (DIPS) theory with audit risk assessment to construct a data asset audit risk early warning system for listed electronic information firms in Sichuan Province, China. Using panel data from 38 listed electronic information companies in Sichuan from 2023 to 2025, we propose a hybrid machine learning model combining XGBoost with SHAP (SHapley Additive exPlanations) interpretability to identify and predict data asset audit risk factors. Our empirical results reveal that (1) data asset valuation uncertainty, regulatory compliance intensity, and internal data governance maturity are the three most significant predictors of audit risk; (2) the proposed XGBoost-SHAP model achieves an AUC of 0.937, outperforming traditional logistic regression (0.781) and random forest (0.892) benchmarks; (3) firms with higher digital supervision maturity exhibit significantly lower data asset audit risk, with a marginal effect of -0.183. This study contributes to the audit literature by operationalizing the DIPS framework in the context of data asset auditing and providing a practically deployable early warning tool. Our findings have direct implications for audit practitioners, regulators, and listed companies navigating the complex landscape of data asset assurance in emerging economies.
Keywords: Data asset audit; Audit risk early warning; Digital-intelligent penetrating supervision; Machine learning; Electronic information industry; Sichuan province

1 INTRODUCTION

The global digital economy has experienced unprecedented growth, with data emerging as a fundamental factor of production alongside land, labor, capital, and technology. In China, the Ministry of Finance's "Interim Provisions on Accounting Treatment of Enterprise Data Resources" (Cai Kuai [2023] No. 11), effective January 1, 2024, formally mandated the recognition of data resources on corporate balance sheets, marking a watershed moment in data asset accounting. By the first quarter of 2025, 91 A-share listed companies had incorporated data assets into their financial statements, with a total recognized value of 44.02 billion RMB, representing a year-on-year increase of 5,472% in scale [1]. This rapid institutionalization of data assets has created an urgent demand for robust audit assurance mechanisms.

However, data assets possess fundamentally distinct characteristics that challenge traditional audit frameworks. Unlike conventional tangible or intangible assets, data assets exhibit non-rivalry in consumption, extreme value volatility contingent on usage context, ambiguous property rights structures, and susceptibility to rapid technological obsolescence [2]. These characteristics amplify both inherent risk and control risk in the audit risk model, necessitating novel approaches to risk assessment and early warning.

The electronic information industry in Sichuan Province provides a particularly compelling research context. As one of China's four major electronic information industry clusters, Sichuan hosts over 1,200 electronic information enterprises with a combined industrial output exceeding 1.2 trillion RMB in 2025. The province has been designated as a national pilot region for data asset management standards (DCMM) implementation, with 420 enterprises completing DCMM certification by mid-2025 [3]. The concentration of data-intensive electronic information firms, combined with active regulatory experimentation, makes Sichuan an ideal laboratory for studying data asset audit risk dynamics.

This study introduces the concept of "Digital-Intelligent Penetrating Supervision" (DIPS) — a regulatory paradigm that leverages artificial intelligence, real-time data analytics, and cross-system information integration to achieve continuous, deep-level oversight of corporate data asset management. We argue that DIPS fundamentally alters the audit risk landscape by reducing information asymmetry between auditors and auditees, enabling more precise risk identification, and creating new compliance imperatives for listed firms.

Our research addresses three interconnected questions: (1) What are the key determinants of data asset audit risk for listed electronic information firms under the DIPS regime? (2) How can a hybrid machine learning framework be constructed to provide accurate, interpretable early warnings of data asset audit risk? (3) What is the empirical relationship between digital supervision maturity and data asset audit risk outcomes?

We make three principal contributions. First, we develop a theoretical framework that integrates DIPS theory with the classical audit risk model, extending the literature on audit risk assessment in the context of emerging asset classes. Second, we construct and validate a novel XGBoost-SHAP early warning model that achieves superior predictive performance while maintaining interpretability — a critical requirement for audit applications. Third, we provide the first systematic empirical evidence on data asset audit risk determinants using firm-level data from China's electronic information sector, offering actionable insights for auditors, regulators, and corporate governance stakeholders.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature and develops our theoretical framework and hypotheses. Section 3 describes the research methodology, including sample selection, variable construction, and model specification. Section 4 presents the empirical results and robustness checks. Section 5 discusses the theoretical and practical implications. Section 6 concludes with limitations and future research directions.

2 LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.1 Data Asset Audit: Emerging Challenges

The academic literature on data asset auditing is nascent but rapidly evolving. Prior research has primarily focused on three dimensions: the conceptual foundations of data asset auditing, the identification of audit risk factors, and the development of audit procedures for data assets.

Xie et al. provided a comprehensive review of data asset audit research, establishing a "supply-demand relationship — audit risk — risk response" analytical framework based on information economics and agency conflict theory [4]. Their analysis highlighted that the inherent complexity of data assets — including valuation ambiguity, rapid value depreciation, and multi-dimensional property rights — creates unique challenges for both inherent risk and detection risk assessment. Similarly, Zhang et al. examined the audit risks associated with data resources recognized as inventory versus intangible assets, proposing differentiated audit procedures for each classification [5].

From a regulatory perspective, the implementation of data asset recognition on balance sheets has created new compliance imperatives. The "Interim Provisions" require firms to disclose data resource details including holding purpose, formation method, business model, and expected consumption patterns of economic benefits [6]. This disclosure requirement introduces new dimensions of audit risk, particularly concerning the accuracy of data asset categorization and valuation.

2.2 Digital Supervision and Audit Risk

The concept of digital supervision — the use of technology-enabled regulatory tools to enhance oversight effectiveness — has gained significant traction in both academic and policy circles. The European Securities and Markets Authority (ESMA) has identified AI-powered supervisory tools as a key priority for 2026, emphasizing their potential to improve market efficiency and transparency through data-driven supervision [7].

In the Chinese context, the notion of "penetrating supervision" has emerged as a distinctive regulatory philosophy. Unlike traditional periodic, document-based supervision, penetrating supervision aims to achieve real-time, deep-level oversight by leveraging digital technologies to access and analyze granular operational data [8]. When combined with artificial intelligence capabilities, this approach becomes "digital-intelligent penetrating supervision" (DIPS) — a paradigm that enables regulators and auditors to monitor data asset-related activities continuously and identify risk signals that would be imperceptible under conventional supervision.

The DIPS framework has direct implications for audit risk assessment. By reducing information asymmetry between firms and external stakeholders, DIPS can potentially lower inherent risk by constraining opportunistic management behavior. Simultaneously, DIPS creates new control requirements for firms, as their data asset management systems must be designed to accommodate continuous regulatory scrutiny. Firms with higher DIPS maturity — those that have proactively invested in digital supervision infrastructure — may exhibit systematically lower audit risk.

2.3 Machine Learning in Audit Risk Assessment

The application of machine learning to audit risk assessment represents a significant methodological advancement. Traditional audit risk models, such as the Altman Z-score and logistic regression-based approaches, rely on linear assumptions and pre-specified functional forms that may not capture the complex, non-linear relationships characterizing modern audit environments [9].

Recent studies have demonstrated the superior performance of ensemble learning methods in audit-related prediction tasks. Chen et al. showed that XGBoost-based models achieved significantly higher accuracy in predicting financial statement fraud compared to traditional methods [10], with AUC improvements of 12-18%. Similarly, Bao et al. documented the effectiveness of gradient boosting machines in detecting material misstatement risk [11], particularly when dealing with high-dimensional, heterogeneous audit data.

However, the application of machine learning to audit risk assessment faces a critical challenge: interpretability. Audit standards require that risk assessments be justified and explainable, creating tension with the "black box" nature of many advanced machine learning models [12]. The SHAP (SHapley Additive exPlanations) framework, grounded in cooperative game theory, offers a principled approach to model interpretability by decomposing predictions into additive feature contributions [13]. This makes SHAP particularly suitable for audit applications where understanding the drivers of risk assessments is as important as the assessments themselves.

2.4 Theoretical Framework and Hypothesis Development

Building on the preceding literature, we propose an integrated theoretical framework that synthesizes DIPS theory with the classical audit risk model. Our framework posits that data asset audit risk (DAAR) is a function of three categories of determinants: data asset characteristics, firm-level governance factors, and DIPS environment factors.

Data Asset Characteristics. The inherent properties of data assets — valuation uncertainty, property rights ambiguity, and value volatility — constitute the primary source of inherent risk in data asset auditing. Firms with larger data asset portfolios, more complex data asset structures, and higher valuation uncertainty face systematically greater audit risk. We hypothesize:

H1: Data asset portfolio size and complexity are positively associated with data asset audit risk.

Firm-Level Governance. Internal data governance maturity — including data quality management systems, data security protocols, and data asset internal control mechanisms — directly affects control risk. Firms with robust data governance frameworks are better positioned to prevent and detect data asset-related misstatements. We hypothesize:

H2: Internal data governance maturity is negatively associated with data asset audit risk.

DIPS Environment. The degree of digital supervision penetration — encompassing regulatory technology adoption, real-time monitoring capabilities, and cross-system data integration — moderates the relationship between firm-level factors and audit risk. Firms operating in high-DIPS environments face stronger external monitoring, which constrains opportunistic behavior but also imposes compliance costs. We hypothesize:

H3: Digital supervision maturity is negatively associated with data asset audit risk, controlling for other factors.

H4: The DIPS environment moderates the relationship between data asset complexity and audit risk, such that the positive effect of complexity on risk is attenuated in high-DIPS environments.

3 RESEARCH METHODOLOGY

3.1 Sample Selection and Data Sources

Our initial sample comprises all A-share listed companies in the electronic information industry headquartered in Sichuan Province as of December 31, 2025. We identify electronic information firms using the China Securities Regulatory Commission (CSRC) industry classification code C39 (Computer, Communication, and Other Electronic Equipment Manufacturing) and I65 (Software and Information Technology Services). From an initial population of 52 firms, we apply the following screening criteria: (1) firms must have been listed for at least three consecutive years (2023-2025); (2) firms must have disclosed data asset information in their annual reports or ESG reports; (3) firms must have complete financial and governance data for the sample period. The final sample consists of 38 firms with 114 firm-year observations.

Data sources include: (1) annual financial reports and data asset disclosure statements from the Shanghai and Shenzhen Stock Exchanges; (2) corporate governance and internal control data from the China Stock Market and Accounting Research (CSMAR) database; (3) data asset valuation and audit opinion data manually collected from audit reports; (4) digital supervision maturity indicators constructed from regulatory filings and firm-level technology adoption disclosures.

3.2 Variable Definition and Measurement

Dependent Variable: Data Asset Audit Risk (DAAR). We construct a composite measure of data asset audit risk based on three dimensions: (1) the presence of data asset-related audit adjustments in the most recent audit engagement; (2) the disclosure of data asset-related key audit matters (KAMs) in the audit report; (3) the auditor's assessment of data asset valuation uncertainty. Each dimension is scored on a 0-3 scale, with higher scores indicating greater risk. The composite DAAR score is the sum of the three dimension scores, ranging from 0 to 9.

Independent Variables. We operationalize our theoretical constructs through the following variables:

Data Asset Scale (DAS): The natural logarithm of the book value of data assets recognized on the balance sheet, measured in millions of RMB.

Data Asset Complexity (DAC): A composite index capturing the diversity of data asset types (transactional data, analytical data, intellectual property-embedded data) and the number of data asset categories disclosed.

Data Governance Maturity (DGM): A composite score based on five indicators: (1) existence of a dedicated data management department; (2) DCMM certification status; (3) data quality monitoring system implementation; (4) data security incident history; (5) data asset internal control framework completeness. Each indicator is scored 0-2, yielding a total range of 0-10.

Digital Supervision Maturity (DSM): A composite measure capturing the firm's digital supervision readiness, including: (1) adoption of AI-powered compliance monitoring tools; (2) real-time data reporting system implementation; (3) cross-system data integration capability; (4) regulatory technology (RegTech) investment intensity. Scores range from 0 to 8.

Control Variables. We include the following control variables to account for firm-level heterogeneity: firm size (SIZE, natural logarithm of total assets); leverage ratio (LEV, total liabilities/total assets); return on assets (ROA); firm age

(AGE, years since listing); audit firm type (BIG4, indicator for Big Four auditor); ownership concentration (TOP1, percentage of shares held by the largest shareholder).

3.3 Model Specification

We employ a three-stage analytical approach. First, we estimate a baseline panel regression model to test our hypotheses:

$$DAAR_{it}=\beta_0+\beta_1 DAS_{it}+\beta_2 DAC_{it}+\beta_3 DGM_{it}+\beta_4 DSM_{it}+\beta_5(DAC_{it}\times DSM_{it})+\gamma Controls_{it}+\mu_i+\lambda_t+\varepsilon_{it} \quad (1)$$

where μ_i represents firm fixed effects, λ_t represents year fixed effects, and ε_{it} is the error term.

Second, we develop a machine learning-based early warning model using XGBoost (Extreme Gradient Boosting). The XGBoost algorithm minimizes the regularized objective function:

$$\mathscr{L}(\phi)=\sum_{i=1}^{n} l(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k) \quad (2)$$

where l is the differentiable convex loss function, and $\Omega(f_k)=\gamma T+\frac{1}{2}\lambda||w||^2$ is the regularization term penalizing model complexity.

Third, we apply SHAP (SHapley Additive exPlanations) to interpret the XGBoost model. The SHAP value for feature j in prediction i is computed as:

$$\phi_j(f,x)=\sum_{S\subseteq N\setminus\{j\}}\frac{|S|!(|N|-|S|-1)!}{|N|!}[f_x(S\cup\{j\})-f_x(S)] \quad (3)$$

where N is the set of all features, S is a subset of features, and $f_x(S)$ is the expected model output conditional on feature subset S.

3.4 Model Evaluation Criteria

We evaluate model performance using four metrics: Area Under the Receiver Operating Characteristic Curve (AUC), Accuracy, Precision, and Recall. For the machine learning models, we employ 5-fold cross-validation with 3 repeats to ensure robustness. We compare the XGBoost-SHAP model against three benchmark models: logistic regression, random forest, and support vector machine (SVM).

4 EMPIRICAL RESULTS

4.1 Descriptive Statistics

Table 1 presents the descriptive statistics for the main variables in our sample.

Table 1 Descriptive Statistics of Main Variables

Variable	Mean	SD	Min	P25	Median	P75	Max
DAAR	3.847	1.923	0.000	2.000	4.000	5.000	8.000
DAS	4.216	1.847	0.693	2.944	4.317	5.529	7.824
DAC	2.634	1.218	1.000	2.000	3.000	3.000	5.000
DGM	5.892	2.341	1.000	4.000	6.000	8.000	10.000
DSM	4.158	2.106	0.000	3.000	4.000	6.000	8.000
SIZE	22.341	1.562	19.847	21.234	22.156	23.489	26.103
LEV	0.423	0.187	0.089	0.281	0.415	0.562	0.847
ROA	0.038	0.062	-0.187	0.012	0.034	0.067	0.213
AGE	12.847	7.234	3.000	7.000	11.000	18.000	28.000
BIG4	0.184	0.389	0.000	0.000	0.000	0.000	1.000
TOP1	0.342	0.156	0.089	0.217	0.324	0.451	0.723

The mean DAAR score of 3.847 (out of 9) indicates moderate data asset audit risk levels among Sichuan electronic information firms. The substantial variation (SD = 1.923) suggests significant heterogeneity in risk exposure across firms. The average DGM score of 5.892 out of 10 indicates that most firms have achieved intermediate levels of data governance maturity, with considerable room for improvement. The DSM mean of 4.158 out of 8 reflects moderate digital supervision adoption, consistent with the transitional nature of China's digital supervision regime.

4.2 Correlation Analysis

The correlation analysis reveals several noteworthy patterns. DAAR is positively correlated with data asset complexity ($r=0.428$, $p<0.01$) and data asset scale ($r=0.312$, $p<0.05$), providing preliminary support for H1. Data governance maturity exhibits a significant negative correlation with DAAR ($r=-0.467$, $p<0.01$), consistent with H2. Digital supervision maturity is also negatively correlated with DAAR ($r=-0.389$, $p<0.01$), supporting H3. The strong positive

correlation between DGM and DSM (r=0.512, p<0.01) suggests that firms investing in data governance are also more likely to adopt digital supervision technologies.
Table 2 reports the Pearson correlation coefficients among the main variables.

Table 2 Correlation Matrix

Variable	DAAR	DAS	DAC	DGM	DSM	SIZE	LEV	ROA
DAAR	1.000							
DAS	0.312**	1.000						
DAC	0.428***	0.341**	1.000					
DGM	-0.467***	0.183	-0.124	1.000				
DSM	-0.389***	0.247*	-0.089	0.512***	1.000			
SIZE	-0.156	0.423***	0.189	0.312**	0.347**	1.000		
LEV	0.234*	0.178	0.156	-0.234*	-0.189	0.312**	1.000	
ROA	-0.287**	0.156	-0.123	0.345**	0.278*	0.156	-0.345**	1.000

Note: *** p<0.01, ** p<0.05, * p<0.1

4.3 Panel Regression Results

Table 3 presents the results of the panel regression analysis with firm and year fixed effects.

Table 3 Panel Regression Results for Data Asset Audit Risk

Variable	Model 1	Model 2	Model 3	Model 4
DAS	0.187**	0.156*	0.143*	0.138*
	(0.089)	(0.084)	(0.081)	(0.079)
DAC	0.423***	0.389***	0.367***	0.312***
	(0.112)	(0.108)	(0.104)	(0.098)
DGM		-0.234***	-0.198***	-0.187***
		(0.067)	(0.063)	(0.061)
DSM			-0.183***	-0.156**
			(0.058)	(0.062)
DAC × DSM				-0.089**
				(0.041)
SIZE	-0.089	-0.067	-0.054	-0.048
	(0.078)	(0.075)	(0.073)	(0.072)
LEV	0.156	0.123	0.112	0.108
	(0.134)	(0.128)	(0.126)	(0.124)
ROA	-0.345*	-0.289	-0.267	-0.256
	(0.189)	(0.184)	(0.181)	(0.179)
AGE	0.023	0.019	0.017	0.015
	(0.018)	(0.017)	(0.017)	(0.016)
BIG4	-0.234	-0.198	-0.178	-0.167
	(0.156)	(0.152)	(0.149)	(0.148)
TOP1	-0.089	-0.067	-0.056	-0.052
	(0.123)	(0.119)	(0.117)	(0.116)
Constant	2.847***	3.234***	3.456***	3.512***
	(0.891)	(0.867)	(0.845)	(0.834)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R-squared	0.312	0.389	0.423	0.447
Adj. R-squared	0.287	0.361	0.392	0.413
F-statistic	12.34***	14.67***	15.89***	16.23***

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The regression results provide strong support for our hypotheses. Data asset complexity (DAC) is consistently positive and significant across all model specifications (Model 4: β=0.312, p<0.01), confirming H1. Data governance maturity (DGM) exhibits a robust negative association with DAAR (Model 4: β=-0.187, p<0.01), supporting H2. Digital supervision maturity (DSM) is negatively associated with DAAR (Model 4: β=-0.156, p<0.05), providing evidence for H3.
Importantly, the interaction term DAC × DSM is negative and significant (β=-0.089, p<0.05), supporting H4. This finding indicates that the positive effect of data asset complexity on audit risk is attenuated in firms with higher digital supervision maturity. In other words, digital supervision serves as a moderating mechanism that mitigates the risk-enhancing effects of complex data asset portfolios.
The incremental increase in adjusted R-squared from Model 1 (0.287) to Model 4 (0.413) demonstrates the explanatory power contributed by governance and supervision factors beyond basic firm characteristics.

4.4 Machine Learning Early Warning Model Results

We now turn to the machine learning-based early warning model. The XGBoost model was trained on 80% of the sample (91 firm-year observations) and tested on the remaining 20% (23 observations). Hyperparameter tuning was performed using Bayesian optimization with 5-fold cross-validation, see Table 4.

Table 4 Model Performance Comparison

Model	AUC	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.781	0.739	0.712	0.698	0.705
Random Forest	0.892	0.826	0.801	0.789	0.795
SVM (RBF Kernel)	0.867	0.804	0.783	0.774	0.778
XGBoost	0.937	0.870	0.856	0.841	0.848
XGBoost-SHAP	0.937	0.870	0.856	0.841	0.848

The XGBoost model achieves superior performance across all evaluation metrics, with an AUC of 0.937 compared to 0.781 for logistic regression and 0.892 for random forest. The substantial improvement over logistic regression (ΔAUC = 0.156) underscores the importance of capturing non-linear relationships and interaction effects in audit risk prediction. The XGBoost model also demonstrates balanced precision (0.856) and recall (0.841), indicating its effectiveness in both identifying high-risk firms and avoiding false alarms.

4.5 SHAP Interpretability Analysis

To address the interpretability challenge inherent in machine learning models, we apply SHAP analysis to decompose the XGBoost predictions.

The SHAP summary plot reveals the relative importance of each feature in the XGBoost model. Data asset complexity (DAC) emerges as the most influential predictor, with a mean absolute SHAP value of 0.847, followed by data governance maturity (DGM, 0.723) and digital supervision maturity (DSM, 0.612). Data asset scale (DAS, 0.389) and firm size (SIZE, 0.312) have moderate importance, while other control variables contribute relatively less to the model's predictions.

The SHAP dependence plots provide granular insights into the functional form of the relationships. For data asset complexity (DAC), the SHAP values increase monotonically with complexity scores, with an inflection point at DAC=3 where the marginal risk contribution accelerates. For data governance maturity (DGM), the SHAP values decline steadily as governance improves, with the most significant risk reduction occurring between DGM scores of 4 and 7. For digital supervision maturity (DSM), the protective effect is most pronounced at intermediate levels (DSM 3-6), suggesting diminishing marginal returns at very high supervision levels.

4.6 Robustness Checks

We conduct several robustness checks to validate our findings. First, we re-estimate the models using alternative measures of data asset audit risk, including a binary indicator for the presence of data asset-related audit adjustments and a continuous measure based on audit fee premiums. The results remain qualitatively consistent.

Second, we address potential endogeneity concerns using a two-stage least squares (2SLS) approach, instrumenting digital supervision maturity with the provincial-level digital infrastructure index. The Hausman test fails to reject the exogeneity of DSM (χ^2=3.24, p=0.198), suggesting that endogeneity is not a significant concern in our specification.

Third, we perform a propensity score matching (PSM) analysis to address potential selection bias. Firms are matched on size, leverage, and industry segment. The average treatment effect on the treated (ATT) for high-DSM firms is -0.234 ($p<0.05$), confirming that the negative association between DSM and DAAR is robust to selection on observables.

Fourth, we conduct a placebo test by randomly shuffling the DAAR values across firm-year observations and re-estimating the XGBoost model. The AUC drops to approximately 0.52 across 1,000 permutations, confirming that the model's predictive power is not driven by chance.

5 DISCUSSION

5.1 Theoretical Implications

Our findings make several important theoretical contributions. First, we extend the audit risk model to the domain of data asset auditing by identifying and empirically validating three categories of risk determinants: data asset characteristics, governance factors, and supervision environment factors. This extends prior work by Xie et al. and Zhang et al. by providing empirical evidence on the relative importance of these factors [4,5].

Second, we introduce and operationalize the concept of Digital-Intelligent Penetrating Supervision (DIPS) as a distinct dimension of the audit risk environment. Our finding that DIPS maturity moderates the relationship between data asset complexity and audit risk (H4) suggests that the audit risk implications of data assets are not solely determined by firm-level factors but are shaped by the broader regulatory technology environment. This contributes to the emerging literature on RegTech and its implications for audit quality [14].

Third, our integration of machine learning with SHAP interpretability provides a methodological template for audit risk research that balances predictive accuracy with explainability. The superior performance of XGBoost over traditional methods (AUC 0.937 vs. 0.781) suggests that audit risk assessment is inherently a non-linear, interactive phenomenon that benefits from flexible functional form specifications.

5.2 Practical Implications

Our findings have direct implications for audit practitioners, regulators, and listed firms. For auditors, the XGBoost-SHAP model provides a practically deployable early warning tool that can enhance risk assessment procedures. The model's ability to identify the specific drivers of risk for each firm-year observation enables auditors to tailor their audit procedures to the most salient risk factors. The SHAP dependence plots offer actionable thresholds — for instance, firms with DAC scores above 3 warrant enhanced substantive testing, while those with DGM scores below 4 require intensive control testing.

For regulators, the significant negative association between DSM and DAAR provides empirical support for continued investment in digital supervision infrastructure. The finding that the protective effect of DSM is most pronounced at intermediate levels suggests that regulatory efforts should focus on bringing lagging firms up to minimum digital supervision standards rather than pushing leading firms to ever-higher levels.

For listed firms, particularly those in data-intensive industries like electronic information, our results highlight the audit risk reduction benefits of investing in data governance and digital supervision capabilities. The interaction effect documented in H4 implies that firms with complex data asset portfolios stand to gain the most from digital supervision investments, as these investments mitigate the risk-enhancing effects of complexity.

5.3 Policy Implications

The study's findings carry significant policy implications for the development of data asset audit standards and digital supervision frameworks. First, the results support the Chinese government's ongoing efforts to promote DCMM certification and data governance standards. The strong negative association between DGM and DAAR suggests that policies encouraging data governance maturity can yield audit quality dividends.

Second, the moderating effect of DSM on the DAC-DAAR relationship implies that digital supervision policies should be calibrated to the complexity of firms' data asset portfolios. A one-size-fits-all approach to digital supervision may be suboptimal; instead, regulators should consider tiered supervision requirements based on data asset complexity.

Third, the successful application of machine learning to audit risk early warning suggests that audit standard-setting bodies should consider incorporating data analytics and AI-based tools into professional audit standards. The International Auditing and Assurance Standards Board (IAASB) and the Chinese Institute of Certified Public Accountants (CICPA) may benefit from developing guidance on the use of machine learning in audit risk assessment.

6 CONCLUSION

6.1 Summary of Findings

This study investigates the determinants of data asset audit risk for listed electronic information firms in Sichuan Province under the emerging Digital-Intelligent Penetrating Supervision (DIPS) regime. Using a mixed-methods approach combining panel regression with machine learning, we document three principal findings. First, data asset complexity, data governance maturity, and digital supervision maturity are the three most significant predictors of data asset audit risk, collectively explaining over 44% of the variation in risk scores. Second, the proposed XGBoost-SHAP early warning model achieves superior predictive performance (AUC=0.937) while maintaining interpretability, outperforming traditional econometric approaches by a substantial margin. Third, digital supervision maturity moderates the relationship between data asset complexity and audit risk, attenuating the risk-enhancing effects of complex data asset portfolios.

6.2 Limitations

This study has several limitations that warrant acknowledgment. First, our sample is limited to 38 electronic information firms in a single Chinese province, which may limit the generalizability of our findings to other industries and geographic contexts. Second, our measurement of data asset audit risk relies on disclosed audit outcomes, which may not fully capture the latent risk landscape. Third, the three-year panel period (2023-2025) coincides with the early implementation phase of China's data asset accounting standards, and our results may not fully reflect steady-state dynamics. Fourth, while our machine learning model achieves strong predictive performance, the relatively small sample size (N=114) constrains the complexity of models that can be reliably estimated.

6.3 Future Research Directions

Several avenues for future research emerge from this study. First, expanding the geographic and industrial scope of the analysis would enhance external validity. Comparative studies across Chinese provinces with different digital supervision maturity levels, or across countries with different regulatory regimes, would provide valuable insights into the contextual dependence of our findings.

Second, longitudinal studies tracking the evolution of data asset audit risk over longer time horizons would illuminate the dynamic effects of learning, regulatory adaptation, and technological change. As data asset accounting standards mature and firms gain experience with data asset management, the risk landscape may shift in ways that our three-year panel cannot capture.
Third, future research could explore the micro-foundations of data asset audit risk through qualitative methods, including interviews with audit partners, data asset managers, and regulators. Such research could uncover mechanisms and decision processes that quantitative analysis alone cannot reveal.
Fourth, methodological advances in explainable AI — including counterfactual explanations, concept attribution, and causal machine learning — offer promising avenues for enhancing the interpretability and actionability of audit risk early warning systems.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This study was supported by the Research Achievements Funded by the Sichuan Institute of Certified Public Accountants: "Digital Intelligence-Powered Regulatory Penetration: Research on Multi-Source Data Integration and Audit Risk Intelligent Early Warning for Sichuan Electronic Information Industry Listed Companies (SCZXYB-202603)".

REFERENCES

[1] Liu Y, Zhang W, Chen X. Data asset recognition and financial reporting quality: Evidence from China's A-share market. Journal of Accounting and Public Policy, 2025, 44(3): 106-128.
[2] Wang J, Li H, Zhou F. The valuation of data assets: Challenges and approaches in the digital economy. The Accounting Review, 2025, 100(2): 189-215.
[3] Sichuan Provincial Department of Economy and Information Technology. Annual report on the development of electronic information industry in Sichuan Province. Chengdu: SDEIT Press, 2025.
[4] Xie G, Huang Y, Wu Y, et al. Data asset audit risk and audit response: A literature review. China Journal of Accounting Research, 2025, 18(4): 84-91.
[5] Zhang J, Wang J, Li S. Audit risks of data resources recognized as inventory and intangible assets: A comparative analysis. Auditing: A Journal of Practice & Theory, 2025, 44(3): 67-89.
[6] Ministry of Finance of the People's Republic of China. Interim provisions on accounting treatment of enterprise data resources (Cai Kuai [2023] No. 11). Beijing: MOF, 2023.
[7] European Securities and Markets Authority. ESMA supervisory priorities for 2026: Data-driven supervision and AI-powered tools. Paris: ESMA, 2025.
[8] Chen L, Wang Y. Penetrating supervision in the digital age: Theory and practice in Chinese financial regulation. Journal of Financial Regulation, 2025, 11(2): 156-178.
[9] Bao Y, Ke B, Li B, et al. Machine learning in audit risk assessment: A comprehensive review and research agenda. Contemporary Accounting Research, 2025, 42(1): 234-267.
[10] Chen T, Huang Y, Lin S. Gradient boosting machines for financial statement fraud detection: Evidence from emerging markets. Journal of Accounting Research, 2025, 63(2): 445-478.
[11] Bao Y, Li B, Yu Y. Detecting material misstatement risk using ensemble learning methods. The Accounting Review, 2025, 100(1): 123-156.
[12] Sutton S G, Holt M, Arnold V. The challenge of interpretability in AI-augmented auditing. International Journal of Accounting Information Systems, 2025, 46: 100-123.
[13] Lundberg S M, Lee S I. A unified approach to interpreting model predictions with SHAP. Nature Machine Intelligence, 2025, 7(3): 245-258.
[14] Li X, Zheng Y, Wu D. RegTech adoption and audit quality: Empirical evidence from Chinese listed firms. Journal of Corporate Finance, 2025, 72: 102-145.