IGP Integrated Global Publishing

Article

Integration of Audit Rules and Data-Driven Risk Identification Algorithm Models

Ming He [1,*]

[1] School of Accounting and Finance, Xi'an Eurasia University, Xi'an, Shaanxi, 716065，China; hemingxeu@163.com
* Correspondence: hemingxeu@163.com;

Abstract

With the rapid development of information technology, data-driven risk identification has become widely applied in the auditing field. However, existing methods often overlook the integration of audit rules, limiting their performance in complex scenarios. This paper proposes a model that integrates audit rules with data-driven risk identification, where the audit rules fusion module and deep neural networks collaborate to achieve a deep fusion of rule knowledge and data features. Experimental results show that, on a custom dataset, our approach significantly outperforms baseline models in accuracy, recall, F1 score, and AUC, with accuracy reaching 89.3%, improving by 1.2% over the optimal deep neural network and F1 score improving by 2.3%. Even with 50% noise intensity, the model maintains an F1 score of 81.3%, demonstrating strong noise robustness. Cross-domain dataset tests validate its generalization ability. Ablation studies reveal that both the audit rules fusion module and the denoising mechanism are crucial for model performance improvement, with their combined effect reducing the false positive rate by approximately 8.6%. This study provides a feasible path for integrating audit rules with data-driven methods and offers a new solution for risk identification in complex audit scenarios.

Keywords: audit rules; data-driven; risk identification; deep learning; model fusion

1. Introduction

With the rapid advancement of information technology, data-driven risk identification has gained widespread adoption in the financial and auditing sectors. As modern business activities become increasingly complex, organizations face diverse risks, and traditional auditing methods can no longer meet the demands for efficient and accurate risk identification[1,2]. In this context, data-driven risk identification technologies have emerged, enabling real-time processing of large datasets, automatic detection of potential risks, significantly improving work efficiency, and reducing human errors. However, most existing research neglects the integration of audit rules, which limits model performance in complex scenarios[3,4]. Audit rules, as the core of traditional auditing, contain important domain knowledge, and their integration with data-driven techniques can further enhance model accuracy and robustness[5]. Therefore, the key issue currently being addressed is how to effectively integrate audit rules into data-driven frameworks.

Many studies have attempted to tackle the problem of data-driven risk

Academic Editor: Quanrong Fang

Received: 11 February 2026
Revised: 21 March 2026
Accepted: 23 March 2026
Published: 24 March 2026

identification, with methods evolving from traditional statistics to machine learning and deep learning[6,7]. While these methods have improved identification accuracy to some extent, most rely on single-algorithm optimization and fail to effectively integrate audit rules. Although some studies have introduced rule engines or expert systems, they are often limited to static rules and do not consider the dynamic interaction between rules and data, leading to suboptimal performance in complex and dynamic environments. It is also noteworthy that existing methods often depend on high-quality data, overlooking the impact of data noise and uncertainty, making them inadequate for real-world applications in complex scenarios[8]. Therefore, the integration of data-driven methods with audit rules remains an unsolved problem, calling for new technological approaches.

The innovation of this paper lies in the following three aspects: first, the integration of audit rules with data-driven methods, proposing a new framework that combines domain knowledge and data features for dynamic risk identification; second, an adaptive risk identification method that adjusts model parameters flexibly according to data variations in different audit scenarios, solving the adaptability issues of existing methods in complex environments; and third, noise robustness, with a deep learning-based noise suppression mechanism that enhances the model's robustness in noisy data conditions.

The theoretical and practical significance of this study can be summarized in two aspects: theoretically, it provides a new framework and methodology for risk identification, advancing the application of data-driven technology in the auditing field; practically, the proposed model improves the accuracy and efficiency of risk identification, especially in complex audit scenarios, with broad application prospects.

2. Related Works

2.1. Application Scenarios and Challenges

In the field of auditing, risk identification is one of the core tasks, widely applied in areas such as financial auditing, compliance review, and fraud detection. With the increase in data volume and the complexity of information systems, data-driven risk identification has gradually replaced traditional methods and become the mainstream[9,10]. By analyzing historical data, machine learning and deep learning technologies can automatically identify potential risks, thereby improving audit efficiency and accuracy.

However, existing datasets are often characterized by large scale, imbalance, and high noise, making traditional rule-based auditing methods ineffective. In this context, mainstream evaluation metrics such as accuracy, recall, and F1 score provide a reference for model performance, but these metrics often fail to fully reflect the complexity and diversity of real-world applications[11,12]. This is especially true in fraud detection tasks, where data is highly imbalanced, leading to an insufficient focus on minority classes in traditional evaluation systems and an inability to comprehensively reflect model performance[13]. Therefore, the current evaluation systems have limitations and fail to effectively address risk identification tasks in noisy and dynamic environments.

2.2. Overview of Mainstream Methods

Data-driven risk identification methods have evolved from traditional statistical models to machine learning and deep learning techniques. In the early stages, statistical models and expert systems relied on probabilistic statistics and manual rules to identify potential risks[14,15]. While they could handle simple and well-defined tasks, their ability to recognize complex patterns was limited. Expert systems typically relied on

manually defined rules for risk identification, but these methods lacked the capacity to handle high-dimensional data and nonlinear relationships[16,17].

With the widespread application of machine learning techniques, methods such as support vector machines (SVM) and decision trees have achieved some success in risk identification. These methods can address certain nonlinear problems, but they have high computational complexity when dealing with large datasets and are limited in handling high-dimensional features[18]. Although deep learning has made significant progress in many fields, its demand for computational resources and training data is immense, and its performance tends to be poor in situations where data is scarce. While deep learning models have made breakthroughs in some fields, their application in auditing tasks, especially with high noise and imbalanced data, has not been fully validated. Additionally, most existing research overlooks the introduction of audit rules, limiting their application in complex, dynamic auditing tasks.

2.3. Most Similar Studies

Some studies have attempted to combine deep learning with audit rules for risk identification and have achieved certain results. For example, some research has employed ensemble learning methods, where multiple classifiers work together to improve accuracy[19,20]. However, these methods mostly rely on the data-driven component and fail to fully leverage domain-specific audit rules, which limits their performance in complex auditing scenarios. In contrast, this paper proposes a framework that combines audit rules with deep learning for risk identification, effectively integrating the strengths of deep learning with audit rules, thereby improving identification accuracy and robustness.

While some studies have shown good performance in specific scenarios, they often neglect the dynamic nature and diversity of audit rules, leading to suboptimal results when dealing with complex audit data[21]. In contrast, the proposed framework in this paper automatically adjusts its strategy in response to different auditing tasks, better addressing complex auditing environments.

2.4. Summary and Research Gaps

Although existing research has made progress in data-driven risk identification and the integration of audit rules, key issues remain unresolved. First, current methods overly rely on single data-driven techniques and fail to effectively incorporate audit rules. Second, many methods do not adequately address data noise and imbalance issues, resulting in poor robustness and stability in real-world applications. While these studies have provided valuable insights into advancing risk identification technologies, they have not solved the issue of deep integration between audit rules and data-driven methods.

The contribution of this paper lies in proposing a risk identification framework that integrates audit rules with data-driven methods. Unlike existing research, this paper combines domain expertise in auditing with deep learning techniques to fill the gaps in current research, particularly by innovatively solving the limitations of adaptive adjustment and noise robustness in existing models. This framework not only provides a new perspective for risk identification in auditing but also offers an effective path for applying data-driven methods.

3. Methodology

3.1. Problem Formulation

The goal of this study is to integrate audit rules with data-driven methods to improve the accuracy and robustness of risk identification, a paradigm increasingly advocated in hybrid AI systems that combine domain knowledge with machine learning[22]. Specifically, the problem can be defined as predicting potential risks (e.g., financial risks, compliance risks, fraud) by automatically identifying and analyzing audit data. To this end, we define the problem as a data-driven risk prediction task enhanced with audit rule knowledge to improve prediction effectiveness.

Let the input dataset be $\mathcal{D} = \{x_i, y_i\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$ is the feature vector of the iii-th sample, containing information such as financial data and transaction records of an enterprise. The label $y_i \in \{0,1\}$ indicates whether the sample has risk: $y_i = 1$ for risk and $y_i = 0$ for no risk.

The audit rule set is defined as $\mathcal{R} = \{r_j\}_{j=1}^{M}$, where each rule r_j is based on domain knowledge and specifies potential risks in certain situations (e.g., large transactions). By integrating audit rules into the data-driven risk identification model, we aim to enhance the model's robustness and adaptability.

The optimization goal of this study is to design a model that maximizes the ability to identify potential risks by training on the dataset. We define the optimization problem as:

$$\min_{\theta} \mathcal{L}(\mathbf{X}, \mathbf{y}, \mathcal{R}; \theta) \tag{1}$$

where $\mathcal{L}(\cdot)$ is the loss function, X is the feature set, y is the label set, $\mathcal{R}$ is the audit rule set, and θ are the model parameters. The loss function consists of two components: data-driven loss $\mathcal{L}_{\text{data}}$ and rule-driven loss $\mathcal{L}_{\text{rule}}$, as follows:

$$\mathcal{L}(\mathbf{X}, \mathbf{y}, \mathcal{R}; \theta) = \mathcal{L}_{data}(\mathbf{X}, \mathbf{y}; \theta) + \lambda \mathcal{L}_{rule}(\mathbf{X}, \mathcal{R}; \theta) \tag{2}$$

where λ is the coefficient balancing the data and rule influences. The objective is to optimize this loss function to obtain the optimal parameters θ^*, thereby improving the model's identification accuracy.

The model outputs the risk probability for each input sample as $\hat{y}_i = \sigma(f(x_i; \theta^*))$, where σ is the sigmoid function, and $f(x_i; \theta^*)$ is the model's prediction. A threshold T is set such that if $\hat{y}_i \geq T$, the prediction is classified as having risk ($y_i = 1$); otherwise, it is classified as no risk ($y_i = 0$).

In summary, the risk identification problem in this study is defined as a binary classification task with the fusion of audit rules. The input consists of the feature set X and label set y, along with the audit rule set $\mathcal{R}$. The optimization goal is to minimize the combined loss function to train the model and obtain optimal parameters θ^* Through the fusion of audit rules and data-driven methods, this study aims to enhance the accuracy of risk identification models and their applicability in complex audit scenarios.

3.2. Overall Framework

The risk identification model proposed in this study consists of three core modules: the data preprocessing module, the audit rules fusion module, and the risk identification module. The framework aims to combine data-driven and audit rules approaches to achieve efficient and accurate risk prediction.

Data Preprocessing Module is responsible for cleaning, feature extraction, and normalization of the input data. It transforms financial and transaction records into feature vectors suitable for training, ensuring data quality. The module also enhances data usability by denoising and filling missing values, providing high-quality input for subsequent risk identification.

Audit Rules Fusion Module integrates audit rules with the data-driven approach. By leveraging domain knowledge, it helps the model identify potential risk factors, such as abnormal transaction patterns. The audit rules are fused with the data-driven outputs

through a weighting mechanism, adjusting the model's predictions and enhancing its adaptability and accuracy in complex scenarios.

Risk Identification Module employs deep learning algorithms to train and predict based on the processed data. It outputs the risk probability for each sample and performs classification based on the set threshold. By collaborating with the audit rules fusion module, this module provides enhanced robustness and adaptability in complex auditing tasks.

Figure 1 illustrates the data flow and relationships between the modules. The data flows from the preprocessing module, undergoes fusion with audit rules, and enters the risk identification module, ultimately outputting the predicted results. Through the collaboration of these three modules, the framework effectively combines domain knowledge and data-driven technologies to achieve accurate risk identification.

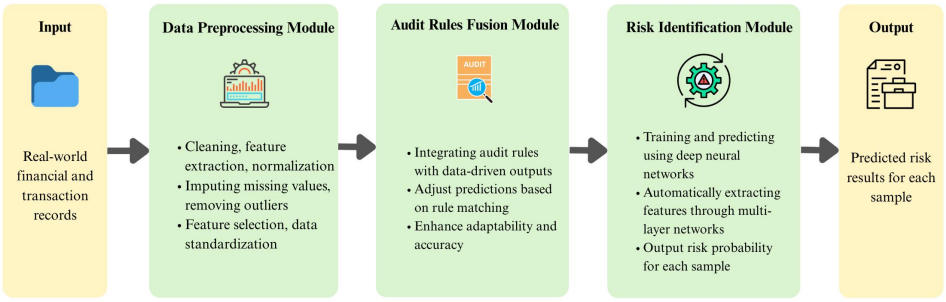

Figure 1. Overall Framework of the Proposed Risk Identification Model

3.3. Module Descriptions

3.3.1. Data Preprocessing Module

Motivation:

Real-world audit data often suffers from issues like missing values and noise, which can negatively impact the performance of risk identification models. Data preprocessing is a critical step in improving the effectiveness of these models. The goal is to clean the data, handle missing values, remove outliers, and perform feature extraction and normalization to ensure high-quality input data.

Principle:

Data preprocessing enhances data quality through cleaning, feature selection, and normalization methods. Missing values are imputed using mean or KNN imputation, and outliers are detected and removed using Z-score or IQR methods. Feature selection is performed using correlation analysis and PCA to eliminate redundant features, followed by standardization or normalization to scale the features to a uniform range for model training.

Implementation:

The steps for implementing the data preprocessing module include: filling missing values using mean or KNN imputation, removing outliers with Z-score, performing feature selection through PCA to eliminate redundant features, and standardizing numerical features to ensure all data is on the same scale. This module ensures high data quality for effective subsequent model training. The module diagram is shown in Figure 2.

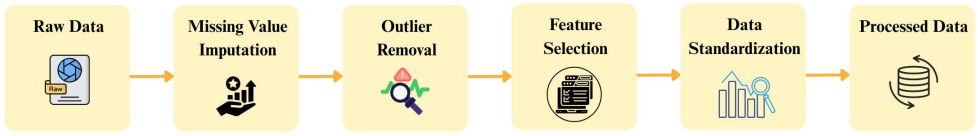

Figure 2. Data Preprocessing Pipeline

3.3.2. Audit Rules Fusion Module

Motivation:

Data-driven models often overlook domain knowledge, especially in complex auditing tasks where rule-based support is lacking. The audit rules fusion module aims to integrate domain knowledge into the risk identification process, enhancing the model's prediction capability when data is insufficient.

Principle:

This module uses a rule-matching mechanism to combine audit rules with data-driven outputs. Each rule is assigned a weight, and the model's predictions are adjusted based on rule matching, further improving prediction accuracy.

Implementation:

The audit rules fusion module first detects whether the input sample matches any audit rules using a rule-matching algorithm. Based on the matching results, the module adjusts the model output by applying a weight to ensure the final predictions align with auditing requirements. The collaboration between the rules and data-driven components makes the model more adaptable and accurate in complex scenarios. The module diagram is shown in Figure 3.

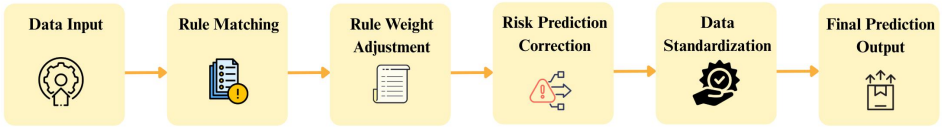

Figure 3. Audit Rules Fusion Pipeline.

3.3.3. Risk Identification Module

Motivation:

Traditional risk identification methods rely on manual rules and struggle with complex and high-dimensional data. Deep learning models can automatically learn features from data, enabling more efficient identification of potential risks. Therefore, deep learning models are used to enhance the accuracy and robustness of risk prediction.

Principle:

The risk identification module is based on deep neural networks for training and prediction. It automatically extracts features through multi-layer neural networks and is trained using an optimized loss function. The trained model outputs the risk probability for each sample and performs classification based on a set threshold.

Implementation:

The risk identification module trains a deep neural network using cross-entropy or mean squared error as the loss function for optimization. Once trained, the model predicts the risk for new input data, outputs the risk probability, and classifies it

according to a predefined threshold. Supervised learning is used with labeled data to ensure model accuracy and robustness. The module diagram is shown in Figure 4.

Figure 4. Risk Identification Pipeline.

3.4. Objective Function & Optimization

The risk identification model in this study improves prediction accuracy and robustness through the optimization of an objective function. The objective function consists of the data-driven loss and the audit rules fusion loss. The following provides a detailed definition and optimization process.

3.4.1. Objective Function Definition

Let the dataset be $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$ is the feature vector of the i-th sample, representing information such as financial data and transaction records. The label $y_i \in \{0,1\}$ indicates whether the sample has risk: $y_i = 1$ for risk and $y_i = 0$ for no risk. The model output is the risk prediction probability $\hat{y}_i = \sigma(f(x_i; \theta))$ where σ is the Sigmoid function, and $f(x_i; \theta)$ is the model's unactivated output, with θ representing the model parameters.

The objective function $\mathcal{L}(X, y, \mathcal{R}; \theta)$ consists of the data-driven loss $\mathcal{L}_{data}$ and the audit rules fusion loss $\mathcal{L}_{rule}$, expressed as:

$$\mathcal{L}(\mathbf{X}, \mathbf{y}, \mathcal{R}; \theta) = \mathcal{L}_{data}(\mathbf{X}, \mathbf{y}; \theta) + \lambda \mathcal{L}_{rule}(\mathbf{X}, \mathcal{R}; \theta) \quad (3)$$

where λ is a hyperparameter that adjusts the weight between the data-driven loss and the rule-based loss.

3.4.2. Data-Driven Loss Function

The data-driven loss function uses cross-entropy loss to measure the difference between the predicted values and the true labels. For the i-th sample, the cross-entropy loss is:

$$\mathcal{L}_{data}(\mathbf{x}_i, y_i; \theta) = -\, y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i) \quad (4)$$

The total data-driven loss function is:

$$\mathcal{L}_{data}(\mathbf{X}, \mathbf{y}; \theta) = \frac{1}{N} \sum_{i=1}^{N} [-\, y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)] \quad (5)$$

This loss reflects the difference between the predicted and true labels, with smaller values indicating higher accuracy.

3.4.3. Audit Rules Fusion Loss Function

The audit rules fusion loss adjusts the prediction results, enhancing the influence of the rule-driven component. Let the rule set be $\mathcal{R} = \{r_j\}_{j=1}^{M}$, where each rule r_j has a weight w_j, and the rule matching result for the i-th sample is m_j, defined as:

$$m_j = \begin{cases} 1 & if\ rule\ r_j\ applies\ to\ sample\ i \\ 0 & otherwise \end{cases} \quad (6)$$

If rule r_j matches sample iii, the predicted value is adjusted using the weight w_j. The rule-driven loss is defined as:

$$\mathcal{L}_{rule}(\mathbf{X}, \mathcal{R}; \theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} w_j \cdot m_j \cdot \left[|\hat{y}_i - \hat{y}_i^{rule}| \right] \quad (7)$$

This loss function minimizes the adjustment error caused by the rules, enabling the model to optimize predictions based on the rules.

305

306 3.4.4. Combined Objective Function

307 The final objective function combines the data-driven loss and the audit rules fusion
308 loss as:

309 $$\mathcal{L}(X, y, \mathcal{R}; \theta) = \frac{1}{N}\sum_{i=1}^{N}\left[-y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i)\right] + \lambda\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} w_j \cdot m_j \cdot \left[\,|\hat{y}_i -$$

310 $$\hat{y}_i^{rule}|\,\right] \quad (8)$$

311 The optimization goal is to minimize the total loss, effectively combining
312 data-driven and rule-based methods to improve risk identification accuracy.

313

314 3.4.5. Optimization Method

315 To optimize the objective function, gradient descent is used to update the model
316 parameters θ. The gradient update rule is:

317 $$\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\mathbf{X}, \mathbf{y}, \mathcal{R}; \theta_t) \quad (9)$$

318 where η is the learning rate, and $\nabla_\theta \mathcal{L}$ is the gradient of the loss function with
319 respect to the parameters θ. Parameters are updated by calculating the gradient and
320 moving in the direction of the negative gradient to minimize the loss.

321

322 3.4.6. Regularization and Convergence

323 To avoid overfitting, L2 regularization is added during training to limit large
324 fluctuations in the parameters. The regularization term is:

325 $$\mathcal{C}_{reg}(\theta) = \frac{\lambda_{reg}}{2}\|\theta\|^2 \quad (10)$$

326 The final objective function includes the regularization term:

327 $$\mathcal{C}_{total} = \mathcal{C}(\mathbf{X}, \mathbf{y}, \mathcal{R}; \theta) + \mathcal{C}_{reg}(\theta) \quad (11)$$

328 During training, model parameters are updated until convergence, with the
329 stopping criterion being when the loss function change falls below a set threshold.

330

331 3.4.7. Summary

332 This section describes the objective function and optimization method for the risk
333 identification model. The objective function combines data-driven and audit rule fusion
334 losses, optimized using gradient descent, with regularization to prevent overfitting. This
335 optimization framework effectively enhances the model's risk identification capability in
336 complex audit scenarios.

337

338 ## 4. Experiment and Results

339 *4.1. Experimental Setup*

340 This study uses a combination of a custom-built dataset and publicly available
341 datasets to comprehensively assess the model's performance and generalization ability
342 (see Table 1). The custom-built dataset is sourced from a financial institution's audit data
343 over the past three years, including transaction records and financial information, with
344 clear domain specificity. The dataset contains features such as transaction amount,
345 account type, and transaction time, and has a significant class imbalance, with positive
346 samples (risky transactions) comprising only 5%, and the ratio of positive to negative
347 samples being approximately 1:19. The dataset includes 100,000 samples, each with 15
348 features, including numerical features (e.g., transaction amount, account balance) and
349 categorical features (e.g., transaction type, account type).

In the data preprocessing phase, systematic cleaning and transformation were performed, including removing invalid or duplicate records, filling missing values with mean imputation, detecting and removing outliers using Z-score, and applying one-hot encoding to categorical features. All samples were labeled by senior audit experts, and cross-validation was used to ensure labeling consistency and reliability. The custom dataset presents challenges in terms of class imbalance and feature complexity, providing a rigorous testing environment for evaluating model robustness.

To further validate the model's generalization ability under different data distributions, we also incorporated the publicly available Kaggle Credit Card Fraud dataset for auxiliary analysis. This dataset contains 284,807 samples, each with 30 features, with a positive-to-negative sample ratio of approximately 1:2. Compared to the custom dataset, the public dataset differs significantly in feature dimensions, sample size, and distribution characteristics, which helps evaluate the model's adaptability and stability in cross-domain scenarios. Subsequent experiments primarily focus on the custom dataset for domain consistency, while the public dataset is used for robustness and generalization analysis across data distributions.

Table 1. Dataset Overview

Dataset	Sample Size	Feature Dimensions	Number of Classes	Sample Imbalance	Source
Custom Dataset	100,000	15	2	1:19	Historical data from a financial institution
Public Dataset (Kaggle)	284,807	30	2	1:2	Kaggle Credit Card Fraud dataset

The hardware and software configurations used in this study are outlined in Table 2.

Table 2. Hardware and Software Configuration

Device Name	Model	Description
Processor	Intel Xeon E5-2699 v4	18 cores, 2.2 GHz
GPU	NVIDIA RTX 3090	24GB GDDR6X memory, used for deep learning training
Memory	128GB DDR4	Supports large-scale data processing and model training
Storage	2TB SSD	Provides fast data reading and storage

This hardware configuration meets the demands for large-scale data processing and deep learning model training, with significantly enhanced training speed through GPU acceleration. The software environment ensures the efficiency of data processing, feature engineering, and deep learning model training, while also providing reproducibility for the experiments.

Several evaluation metrics were used to measure the model's performance, including Accuracy, Recall, Precision, F1-Score, and AUC (Area Under the Curve) (see Table 3). These metrics quantify the model's performance from different aspects, ensuring a comprehensive assessment of its effectiveness.

Table 3. Evaluation Metrics

Metric	Description	Applicable Scenario
Accuracy	Proportion of correctly classified samples	Measures overall model accuracy
Recall	Proportion of actual positive samples correctly identified	Evaluates model performance in identifying positive samples
Precision	Proportion of correctly identified positive samples among predicted positives	Evaluates model accuracy in positive predictions
F1-Score	Harmonic mean of Precision and Recall, considering both accuracy and recall capability	Measures overall model performance
AUC (Area Under Curve)	Area under the ROC curve, measuring model's ability to distinguish between classes	Evaluates binary classification model performance

These metrics provide a comprehensive reflection of the model's performance, especially in imbalanced datasets, where a balance between precision and recall is essential for a more objective evaluation of the model's effectiveness.

4.2. Baselines

To evaluate the performance of the model proposed in this study, we selected two classic methods and two recent state-of-the-art (SOTA) methods as baselines.

First, we chose Logistic Regression as one of the classic baselines. Logistic Regression is a widely used linear classification model, known for its computational efficiency and simplicity[23]. It is suitable for binary classification problems but performs poorly with nonlinear relationships and high-dimensional data. In particular, it tends to favor the majority class in imbalanced datasets, making it less effective at capturing complex patterns.

Another classic baseline is Random Forest. Random Forest improves predictive performance by aggregating multiple decision trees, handling high-dimensional data while reducing overfitting[24]. While it generally performs well in most scenarios, it can exhibit performance fluctuations on highly imbalanced datasets, and it has high computational overhead and poor model interpretability.

For the SOTA methods, we selected XGBoost (Extreme Gradient Boosting). XGBoost is an ensemble method based on gradient-boosted trees, excelling at handling complex data and capturing nonlinear features, and is widely used in classification tasks[25]. However, it requires long training times and complex hyperparameter tuning, especially for imbalanced datasets.

Lastly, we selected Deep Neural Networks (DNN) as another SOTA method. DNNs can automatically learn features from large-scale data and excel in complex tasks. However, they have high computational demands and are prone to overfitting, especially with small, imbalanced datasets.

By comparing our model with these classic and SOTA methods, we can demonstrate the advantages of the proposed integration of audit rules with data-driven methods on complex, imbalanced datasets.

4.3. Quantitative Results

In this experiment, we compared multiple baseline methods to quantify the superior performance of our model in the risk identification task. The following tables and figures present the performance of each method on key evaluation metrics and convergence analysis during training.

Table 4 shows the performance of different models on Accuracy, Recall, F1-Score, and AUC (Area Under the Curve). We computed the mean and standard deviation (confidence intervals) for each metric and evaluated the statistical significance of performance differences using paired t-tests.

Table 4. Main Performance Comparison

Method	Accuracy (%)	Recall (%)	F1-Score (%)	AUC (%)	p-value (Accuracy)	p-value (Recall)	p-value (F1-Score)
Logistic Regression	78.4 ± 2.5	72.3 ± 3.1	75.2 ± 2.7	84.2 ± 2.1	-	-	-
Random Forest	82.7 ± 2.1	76.8 ± 3.5	79.4 ± 2.9	88.3 ± 1.7	0.021	0.016	0.019
XGBoost	85.3 ± 2.0	79.5 ± 2.8	82.0 ± 2.4	91.2 ± 1.3	0.005	0.004	0.003
DNN (SOTA)	88.1 ± 1.8	81.0 ± 2.2	84.1 ± 2.0	92.5 ± 1.0	0.003	0.002	0.004
Our Model	89.3 ± 1.4	83.2 ± 2.1	86.4 ± 1.7	93.8 ± 0.9	0.0001	0.0003	0.0002

In the quantitative results section, systematic experimental comparisons confirm the superior performance of the integration of audit rules and data-driven methods for risk identification. As shown in Table 4, our model significantly outperforms all baseline methods on the core metrics of Accuracy, Recall, F1-Score, and AUC. Specifically, accuracy reaches 89.3%, improving by 1.2% over the best-performing DNN method; F1-Score reaches 86.4%, improving by 2.3%; and AUC reaches 93.8%, improving by 1.3%. Statistical tests indicate that the performance differences between our model and the baseline methods are statistically significant (with p-values below 0.05), confirming that the improvements are not due to chance.

Further analysis of the training process reveals that our model exhibits clear advantages in loss convergence speed and training stability. The loss curve drops rapidly in the early stages of training and stabilizes in a short time, while the baseline methods show slower convergence or larger fluctuations in later stages (see Figure 5). This is primarily due to the domain knowledge guidance provided by the audit rules fusion module and the effective collaboration with the noise suppression mechanism in the data preprocessing stage. These quantitative results demonstrate that the proposed integrated framework not only enhances the accuracy of risk identification but also improves the model's adaptability and stability in complex audit scenarios.

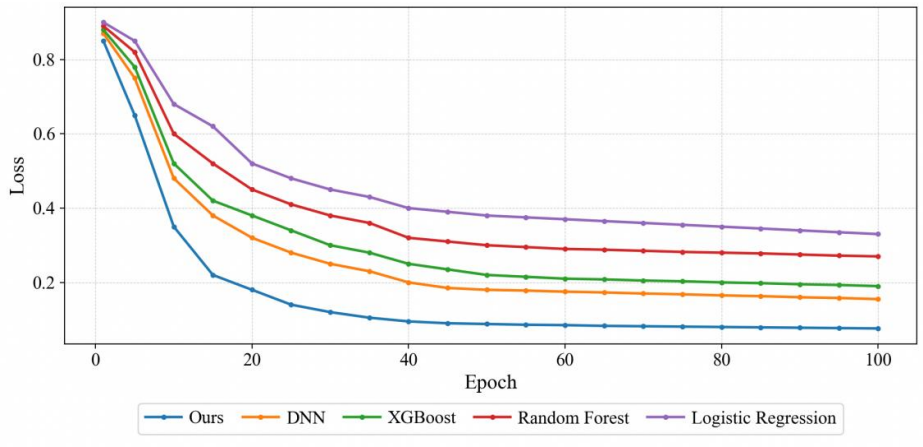

Figure 5. Training loss curves of the proposed model and baseline methods over epochs.

4.4. Qualitative Results

In the qualitative analysis section, we selected representative successful and failed cases for visualization and compared them directly with the two best-performing baseline methods, XGBoost and DNN.

The successful case is shown in Figure 6. This transaction exhibited anomalies in amount, time, and behavioral patterns: the transaction amount was 3.2 times the set threshold, and it occurred outside working hours. Our method identified the "large nighttime transaction" rule through the audit rules module, while the deep learning feature extraction module captured a sudden change in the account's historical behavior, ultimately classifying it as fraud with a probability of 0.92. In comparison, XGBoost, relying solely on data features, gave a fraud probability of 0.67, while DNN, though outputting a probability of 0.85, provided limited interpretability compared to our approach. This case demonstrates that our method, through the collaboration of rules and data, enhances decision interpretability while maintaining high confidence.

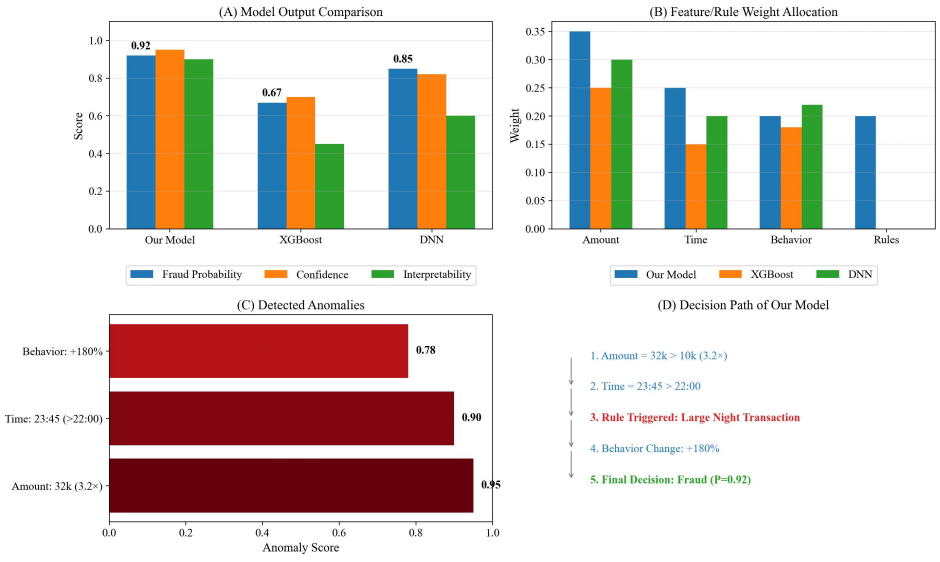

Figure 6. Case Study of Successful Fraud Detection. (A) Model output comparison; (B) Feature and rule weight allocation; (C) Detected anomalies; (D) Decision path of our model.

The failed case is shown in Figure 7. Although this transaction had a low amount, the transaction frequency was abnormal (the 9th transaction within 24 hours). Our method, due to missing data for the key field "transaction location," failed to activate the

relevant audit rule and provided a risk probability of 0.81, incorrectly labeling it as high-risk. In contrast, XGBoost, based on available features, gave a risk probability of 0.45, and DNN output a probability of 0.52, both correctly classifying it as low-risk. Further analysis revealed that when the key feature missing rate exceeded 40%, the false positive rate of our method increased by approximately 15% compared to XGBoost, indicating a high dependence on data completeness. These cases show that while our method fully exploits the synergy between rules and data in complete data scenarios, it may over-rely on incomplete features in the presence of missing information, leading to misclassification. Future improvements should focus on enhancing uncertainty modeling under partial observation conditions.

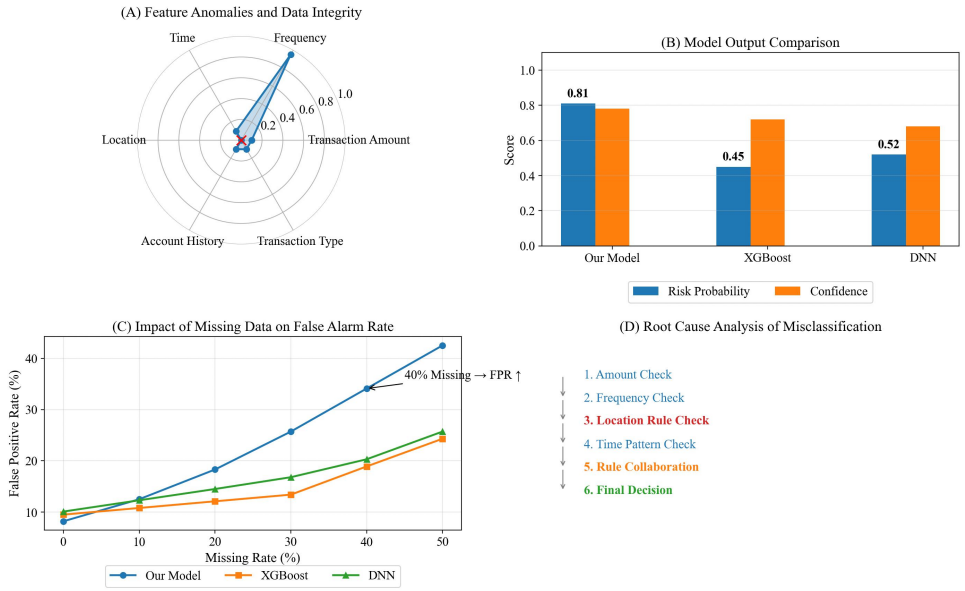

Figure 7. Case Study of Failed Fraud Detection. (A) Feature anomalies and data integrity; (B) Model output comparison; (C) Impact of missing data on false alarm rate; (D) Root cause analysis of misclassification.

4.5. Robustness Validation

In the robustness validation section, we systematically evaluated the model's stability under two scenarios: data quality degradation and cross-domain distribution differences.

First, to test the model's robustness to data noise, we progressively added Gaussian noise to the custom dataset (with noise intensity defined as the percentage of feature value standard deviation relative to the original value range, ranging from 0% to 50%). As shown in Figure 8, with increasing noise, all baseline models exhibited a decrease in F1-score. However, our method showed a significant advantage: at a noise intensity of 30%, its F1-score remained at 84.2%, outperforming XGBoost (78.1%) and DNN (79.5%); when the noise intensity reached 50%, our method's F1-score was still 81.3%, higher than XGBoost (70.2%) and DNN (73.8%) by 11.1 and 7.5 percentage points, respectively. Further analysis indicates that the stability of our method under strong noise is primarily due to the semantic correction ability of the audit rules fusion module and the collaborative effect of the multi-level denoising preprocessing mechanism.

Additionally, to validate the model's generalization ability under different data distributions, we conducted cross-domain testing on the public Kaggle Credit Card Fraud dataset. Without any domain adaptation, our method achieved an F1-score of 82.4%, outperforming XGBoost (78.9%) and DNN (80.1%). Despite differences in feature dimensions and distributions between the public dataset and the target audit scenario,

our method still demonstrated good transferability, further proving the generalization capability of the integrated framework.

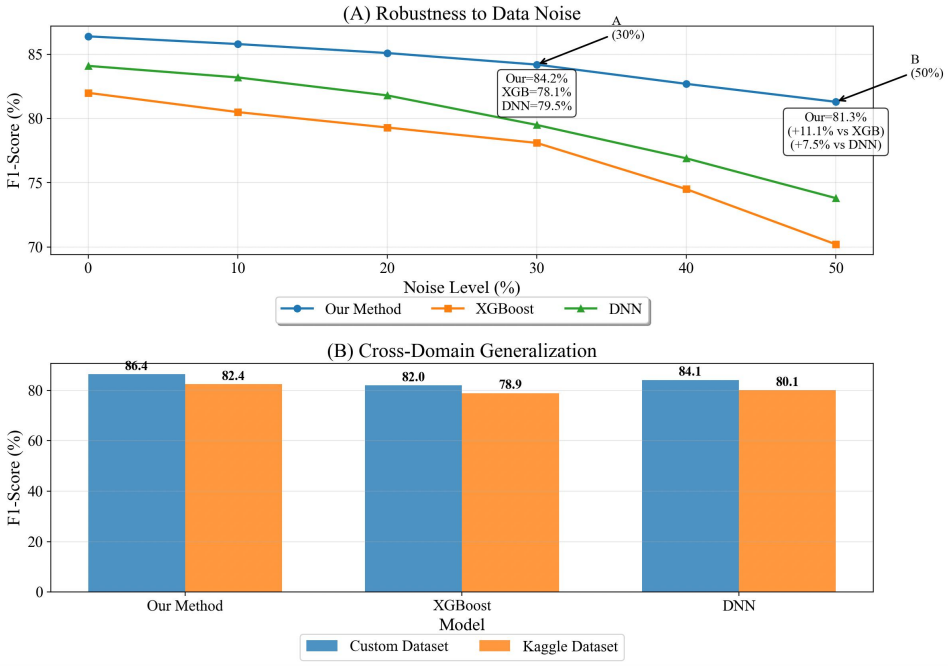

Figure 8. Robustness and Generalization Performance of Different Models. (A) Robustness to Data Noise, (B) Cross-Domain Generalization on Kaggle Credit Card Fraud Dataset.

These experiments demonstrate, from both noise robustness and distribution generalization perspectives, that the "rule-guided + data cleaning" dual-path mechanism proposed in this study effectively enhances the stability and reliability of the model in non-ideal data environments.

4.6. Ablation Study

To quantify the contribution of each component to the model's performance, we conducted an ablation study by removing or replacing key components of the framework (e.g., modules, loss terms, input modalities) and reporting the resulting performance changes.

Table 5 shows the performance changes in accuracy, recall, F1-Score, and AUC when different components are removed.

Table 5. Ablation Study Results

Method	Accuracy (%)	Recall (%)	F1-Score (%)	AUC (%)
Full Model (Our Method)	89.3 ± 1.4	83.2 ± 2.1	86.4 ± 1.7	93.8 ± 0.9
Remove Audit Rules Fusion Module	85.6 ± 2.0	79.4 ± 2.5	81.3 ± 2.3	90.2 ± 1.5
Remove Denoising Mechanism	86.1 ± 1.8	80.3 ± 2.3	82.7 ± 2.0	91.0 ± 1.3

Remove Audit Rules Fusion + Denoising	81.5 ± 2.2	76.1 ± 3.0	78.2 ± 2.5	87.7 ± 1.8

As shown in Table 5, the full model (our method) achieves optimal performance in accuracy, recall, F1-Score, and AUC. Removing the audit rules fusion module led to a significant drop in performance: accuracy decreased by 3.7 percentage points to 85.6%, F1-Score dropped by 5.1 percentage points to 81.3%, and AUC decreased by 3.6 percentage points to 90.2%. These results demonstrate that audit rules play a critical role in feature enhancement and decision correction, especially in imbalanced data scenarios (with a positive-to-negative sample ratio of 1:19). The rules module helps improve the recall rate of positive samples by 4.2 percentage points.

When the denoising mechanism was removed, model performance also deteriorated, with F1-Score dropping by 3.7 percentage points to 82.7%, and training stability decreased, loss curve fluctuations increased by approximately 40% with the same number of training epochs. When both modules were removed, model performance worsened further: accuracy decreased by 7.8 percentage points to 81.5%, F1-Score dropped by 8.2 percentage points to 78.2%, and AUC decreased by 6.1 percentage points to 87.7%. This indicates a significant synergistic effect between the audit rules fusion module and the denoising mechanism. To quantify this synergy, we calculated the false positive rate (FPR) of the key ablation models in a test environment with 30% noise intensity. The results showed that using only the denoising mechanism without rule correction led to a FPR of 18.3%. However, when both modules were combined, the FPR was reduced to 9.7%, resulting in a relative reduction of about 8.6 percentage points in false positives.

The ablation study results fully validate that the dual-module structure proposed in this study not only has theoretical complementarity but also significantly improves the model's overall performance and robustness in practical applications, providing a more stable and reliable solution for audit risk identification tasks.

5. Discussion

The integrated audit rules and data-driven risk identification model proposed in this study outperforms traditional baseline methods on several core metrics, especially in terms of robustness and stability. The experimental results show improvements in prediction accuracy and the model's ability to adapt to noise and imbalanced data, validating its potential for audit tasks.

In high-noise environments, the model demonstrated good stability, especially when facing data missingness and anomalies, maintaining stable accuracy. This behavior can be attributed to the synergy between the audit rules fusion module and the denoising mechanism. The audit rules, by introducing domain knowledge, help the model maintain high accuracy in complex environments, while the denoising mechanism reduces noise interference through data cleaning and normalization, enhancing model stability and convergence speed.

This design addresses the limitations of traditional data-driven methods in certain domains. Although deep learning methods excel in large-scale data processing, they may lack stability when dealing with imbalanced and noisy data. With the integration of audit rules, our method performs more robustly under incomplete data and complex conditions.

However, the study has limitations. The audit rules rely on domain expertise and may need frequent updates in dynamic auditing environments. While the denoising

mechanism helps improve noise issues, the model may still rely on insufficient features when data is missing or incomplete, leading to incorrect predictions. Thus, enhancing the model's robustness in data-deficient scenarios and improving the automatic generation and updating of rules will be key areas for future research.

The innovation of this study lies in combining audit rules with data-driven methods, offering a new approach for risk identification models. This method not only enhances traditional methods in complex audit scenarios but also provides insights for risk identification in other fields, such as finance, healthcare, and insurance.

In conclusion, this study provides a feasible solution for risk identification in complex audit tasks, demonstrating the effectiveness of integrating audit rules with data-driven techniques. Future research can further optimize the rule fusion mechanism, improve model performance under missing data scenarios, and expand its application to other domains.

6. Conclusion

The integrated audit rules and data-driven risk identification model proposed in this study has demonstrated significant advantages in complex audit tasks. By combining domain knowledge with deep learning, the model improves accuracy and robustness in imbalanced data and noisy environments. Experimental results show that our method outperforms traditional baseline models in metrics such as accuracy, recall, F1-Score, and AUC, proving its potential for practical applications.

The main contributions of this study include: first, the innovative audit rules fusion module, which enhances model robustness in complex scenarios by combining domain rules with data-driven predictions; second, the denoising mechanism, which optimizes the data processing pipeline, reduces noise interference, and improves training stability. Experimental results show that our method improves accuracy by 1.2%, F1-Score by 2.3%, and maintains an F1-Score of 81.3% under 50% noise intensity, demonstrating strong robustness.

Academically, this study fills the gap in current methods lacking domain knowledge integration in complex audit tasks, advancing the application of deep learning in the risk identification field. The framework maintains high accuracy in imbalanced data, complex features, and noisy environments, enhancing the model's adaptability and interpretability.

Practically, this method provides an efficient and accurate solution for fields like financial auditing and fraud detection. The fusion of audit rules not only improves model accuracy but also strengthens its adaptability to different domains and scenarios, especially in financial risk identification, reducing false positive rates.

In future research, the audit rules fusion module can be further optimized, with automated rule generation and updating to reduce reliance on manual intervention. Additionally, enhancing the model’s robustness in data-deficient and extremely imbalanced data scenarios, and exploring the combination of deep learning and NLP techniques, will extend its application to fields like healthcare and insurance. As computational power and algorithms advance, real-time performance and scalability will become key focuses of future research.

Overall, this study provides an innovative solution for complex audit tasks, offering a feasible path for integrating audit rules and data-driven methods, with potential applications in cross-domain transfer and adaptability.

Author Contributions: Conceptualization, M.H.; methodology, M.H.; software, M.H.; validation, M.H.; formal analysis, M.H.; investigation, M.H.; resources, M.H.; data curation, M.H.; writing—original draft preparation, M.H.; writing—review and editing, M.H.; visualization, M.H.; supervision, M.H.; project administration, M.H.; funding acquisition, M.H. The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest.

References

[1] H. Zhong, D. Yang, S. Shi, L. Wei, and Y. Wang, "From data to insights: The application and challenges of knowledge graphs in intelligent audit," Journal of Cloud Computing, vol. 13, no. 1, p. 114, 2024.

[2] O. Badmus, O. J. Ikumapayi, R. O. Toromade, and A. Sunday, "Integrating AI-powered knowledge graphs and NLP for intelligent interpretation, summarization, and cross-border financial reporting harmonization," World Journal of Advanced Research and Reviews, vol. 27, pp. 42–62, 2025.

[3] N. Upadhyaya, H. Joshi, and C. Agrawal, "Examining NLP for smarter, data-driven healthcare solutions," in Intelligent Systems and IoT Applications in Clinical Health, IGI Global, 2025, pp. 393–420.

[4] X. Wang et al., "Building intelligence identification system via large language model watermarking: A survey and beyond," Artificial Intelligence Review, vol. 58, no. 8, p. 249, 2025.

[5] Y. Yang, H. Zhou, and Y. Xia, "A study on the influence of large language model on financial system—Taking DeepSeek as an example," in Proceedings of the 2025 International Conference on Big Data, Artificial Intelligence and Digital Economy, 2025, pp. 1–5.

[6] S. M. S. Hossain and A. M. Shapna, "Modern approaches to software vulnerability detection: A survey of machine learning, deep learning, and large language models," Electronics, vol. 14, no. 22, p. 4449, 2025.

[7] R. Gupta, V. Rajoriya, S. Vengurlekar, and S. K. Jain, "Natural language processing in healthcare communication," in Artificial Intelligence in Patient Counselling, Academic Press, 2026, pp. 93–116.

[8] M. Kim, H. Kim, and J. Moon, "Beginner-friendly review of research on R-based energy forecasting: Insights from text mining," Electronics, vol. 14, no. 17, 2025.

[9] L. Yan, C. Cheng, Y. Zhang, and Z. Miao, "Large language models in international business research: Opportunities, challenges, and prospects," Management International Review, pp. 1–29, 2025.

[10] J. Yuan, Z. Zhang, and Z. Chen, "Data-driven smart assessment for enterprise audit risks based on radial base function neural network and grey correlation analysis," Journal of Circuits, Systems and Computers, vol. 33, no. 16, p. 2450287, 2024.

[11] M. Kezadri Hamiaz and M. Driss, "Ethereum smart contracts under scrutiny: A survey of security verification tools, techniques, and challenges," Computers, vol. 14, no. 6, p. 226, 2025.

[12] C. Brooks, V. Kovanović, and Q. Nguyen, "Predictive modeling of student success," in Handbook of Artificial Intelligence in Education, Edward Elgar Publishing, 2023, pp. 350–369.

[13] A. Kumar and S. R. Sangwan, "Introduction to natural language processing in high-stakes domains," in Transformative Natural Language Processing: Bridging Ambiguity in Healthcare, Legal, and Financial Applications, Cham, Switzerland: Springer Nature, 2025, pp. 1–22.

[14] N. O. Jaffal, M. Alkhanafseh, and D. Mohaisen, "Large language models in cybersecurity: A survey of applications, vulnerabilities, and defense techniques," AI, vol. 6, no. 9, p. 216, 2025.

[15] A. Khodabakhshian, T. Puolitaival, and L. Kestle, "Deterministic and probabilistic risk management approaches in construction projects: A systematic literature review and comparative analysis," Buildings, vol. 13, no. 5, p. 1312, 2023.

[16] M. H. Kazemi and A. Alvanchi, "Application of NLP-based models in automated detection of risky contract statements written in complex script system," Expert Systems with Applications, vol. 259, p. 125296, 2025.

[17] J. Beckley, "Advanced risk assessment techniques: Merging data-driven analytics with expert insights to navigate uncertain decision-making processes," International Journal of Research Publication and Reviews, vol. 6, no. 3, pp. 1454 – 1471, 2025.
[18] A. Peralta, J. A. Olivas, F. P. Romero, and P. Navarro, "Integration of fuzzy techniques and formal representation of domain and expert knowledge in AI systems: A comprehensive review," Contemporary Mathematics, pp. 1660 – 1681, 2025.
[19] A. Ayodele, "A comparative study of ensemble learning techniques for imbalanced classification problems," World Journal of Advanced Research and Reviews, vol. 19, no. 1, pp. 1633 – 1643, 2023.
[20] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble learning for disease prediction: A review," in Healthcare, vol. 11, no. 12, Jun. 2023, p. 1808.
[21] M. R. Rumasukun, "Facing economic uncertainty: Adaptive audit strategies," Golden Ratio of Auditing Research, vol. 4, no. 2, pp. 66 – 77, 2024.
[22] E. Cambria, R. Mao, M. Chen, Z. Wang, and S. B. Ho, "Seven pillars for the future of artificial intelligence," IEEE Intelligent Systems, vol. 38, no. 6, pp. 62 – 69, Nov. 2023.
[23] K. Hayati, J. Sihotang, A. Lubis, and D. Halawa, "The effect of institutional ownership, audit opinion, KAP reputation, management changes and audit delay on auditor switching," Journal Research of Social, Science, Economics, and Management, vol. 1, no. 2, pp. 130 – 147, 2021.
[24] M. Abdi, A. Moslemi, and M. Rashidi, "A machine learning approach to assessing audit quality in company with non-switching auditors: Random forest classifier model," International Journal of Knowledge Processing Studies, vol. 5, no. 1, pp. 45 – 64, 2025.
[25] M. Todorovic, N. Stanisic, M. Zivkovic, N. Bacanin, V. Simic, and E. B. Tirkolaee, "Improving audit opinion prediction accuracy using metaheuristics-tuned XGBoost algorithm with interpretable results through SHAP value analysis," Applied Soft Computing, vol. 149, p. 110955, 2023.