EAI Endorsed Transactions
on Scalable Information Systems
Research Article EAI.EU

AI-Driven Financial Fraud Auditing and Evidence Chain Construction in Smart Manufacturing

Zhaojun Hou[1*]

[1]Shandong Normal University Jinan 250358, Shandong, China

Abstract

INTRODUCTION: In smart manufacturing enterprises, procurement, production, warehousing, sales, and financial settlement data are distributed across multiple business systems. Financial fraud increasingly involves cross-entity collusion, nested business chains, and fragmented evidence, making it difficult for traditional sampling audits to simultaneously ensure accurate risk identification, efficient verification, and traceable evidence support.
OBJECTIVES: To address fragmented cross-system associations among manufacturing audit data, difficulties in tracing multi-hop fraud propagation across business entities, and insufficient evidence support for model-generated risk alerts, this study develops an AI-driven financial fraud auditing framework tailored to smart manufacturing processes. The framework links fraud risk identification, dynamic audit-path optimization, and verifiable evidence-chain construction within a unified closed-loop procedure.
METHODS: Multi-source heterogeneous audit samples are constructed through entity matching, temporal alignment, and data-quality filtering. Business process constraints are introduced to encode inconsistencies in transaction amounts, material quantities, event timing, and entity relationships. A heterogeneous graph neural network jointly identifies high-risk entities, anomalous transactions, and fraud categories while tracing cross-entity risk propagation paths. The audit process is further formulated as a sequential decision-making task, in which deep reinforcement learning dynamically selects verification actions according to risk information gain, execution cost, and evidence sufficiency. Original records, risk relationships, audit actions, and review statuses are subsequently organized into a structured and traceable evidence chain.
RESULTS:Simulation experiments show that the proposed method achieves an F1 score of 92.7% and an Area Under the Curve (AUC) of 98.4% for fraud detection, together with a 90.7% recall rate for risk propagation paths. Evidence completeness reaches 95.4%, while the average number of audit steps required to establish a complete evidence chain decreases to 6.9. Direct real-world testing was further conducted on three anonymized manufacturing enterprises containing 715,220 raw cross-system records and 3,660 independently audited business chains. With all trained parameters and decision thresholds fixed, the framework identified 271 of 307 confirmed anomalous chains and achieved a pooled Precision of 89.74%, Recall of 88.27%, F1 score of 89.00%, AUC of 0.956, and false alarm rate of 0.92%. Risk-propagation-path recall and evidence completeness reached 86.64% and 92.18%, respectively, while the average number of audit steps was 7.5.
CONCLUSION: Integrating business-process-constrained heterogeneous graph reasoning, risk-feedback-driven audit decision-making, and structured evidence verification enables complex financial fraud to be identified, localized, and substantiated within a unified audit workflow. The proposed framework extends conventional fraud detection from isolated risk classification toward traceable cross-entity reasoning and evidence-supported audit decision-making in smart manufacturing environments.

Keywords: Smart manufacturing, financial fraud auditing, heterogeneous graph neural networks, deep reinforcement learning, evidence chain construction

Received on 02 September 2026, accepted on 06 November 2026, published on 20 November 2026

Copyright © 2026 Zhaojun Hou *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.14854

*Corresponding author. Email: houzhaojun_sdnu@163.com

1. Introduction

Recent advances in machine learning, graph neural networks, and reinforcement learning have expanded the capabilities of intelligent risk analysis and automated decision-making. Sahoo and Lo reviewed the technological basis of intelligent manufacturing and showed that AI, IA, IE and AI were gradually integrated into production activities, from product design to manufacturing and quality control [1]. In intelligent manufacturing companies, the procurement, manufacturing, storage, sales, and financial clearing activities are spread across the Enterprise Resource Planning System (ERP), Manufacturing Execution System (MES), Warehouse Management System, Supply Chain Platform, and Financial System. As a result, financial fraud has moved from dealing with individual documents to coordinating transactions covering orders, invoices, logistical records, manufacturing actions, funds transfers and related parties. Traditional sampling audits and fixed-rule approaches have a hard time reconstructing these inter-system relationships, which often leads to delays in identifying risks, lengthy verification processes, and incomplete chain of evidence.

At present, there are four kinds of techniques: Machine Learning Classification, Time Sequence Behavior Modeling, Graph Relation Reasoning, and Process Intelligence Audit. Sood et al. reviewed the applications of artificial intelligence and natural language processing (NLP) in financial fraud detection and showed that NLP can extend the utilization of unstructured financial information. However, there is no uniform representation of text descriptions and structured records like orders, logistic events, invoices, and funds transfers [2]. Li et al. examined trustworthy artificial intelligence in human-centric smart manufacturing from the perspectives of safety, explanation, accountability, uncertainty, and human participation; however, their framework focuses on trusted human-machine collaboration throughout the production cycle, not on the traceability of conclusions and audit evidence of financial fraud [3]. Imran et al. demonstrated that by reconstructing business execution trajectories from event logs, process mining can bring automation, transparency, and efficiency to process audits; however, there is still a limitation on the representation of hidden related party relationships and cross-entity funds flows that do not follow explicit business sequences [4]. Yoo et al. showed that the relationship between health providers, recipients, and doctors is a useful tool for detecting fraud; however, their node semantics do not include production work orders, material consumption, inventory movement, and production cost transfer [5]. Zioviris et al proposed a new approach to detect fraud in asymmetric transaction data, but this model is not able to explain the fraud chains formed by coordinated interactions between multiple entities [6]. More specifically, several graph-based fraud detectors have addressed limitations that are directly relevant to the present task. GraphConsis mitigates context, feature, and relation inconsistency by introducing context representations, consistency-based neighbor sampling, and relation-aware aggregation. CARE-GNN focuses on fraudster camouflage and uses label-aware similarity, reinforcement-learning-based neighbor selection, and relation-aware aggregation to suppress misleading neighboring information . PC-GNN addresses severe class imbalance through label-balanced node sampling and neighborhood selection before graph aggregation . These methods substantially improve node-level fraud classification, but their optimization targets remain different from cross-system manufacturing auditing. They do not explicitly encode consistency constraints among purchase orders, production work orders, warehouse movements, invoices, payments, and accounting vouchers, nor do they convert graph risk outputs into a sequence of executable audit actions followed by evidence sufficiency verification. The present framework therefore does not treat the heterogeneous graph neural network as the sole contribution. Its distinction lies in coupling business-process residuals with relationship reasoning, jointly predicting entity risk, anomalous transactions, and fraud types, using the resulting risk subgraph to drive sequential audit decisions, and organizing newly collected records into a traceable evidence chain. Leng et al. summarized the implementation principles, enabling technologies, and development paths of IE in Industry 5.0, with emphasis on collaborative intelligence and self-learning decision making; however, no attention was paid to translating model risk information into executable audit procedures [7]. Nadim et al proposed a new approach to extract dynamic causal relationships from industrial event data, which can be used to reconstruct abnormal process trajectories. However, their approach does not optimise verification activities in terms of audit costs, risk mitigation and evidence adequacy [8]. As a result, the existing research has extended the fraud analysis from the static to the text, the temporal, the relation, and the process, but the identification of the risk, the choice of the audit, and the organization of the evidence remain separate. It is not possible to convert risk probabilities directly into verifiable and traceable audit evidence chains.

Accordingly, an AI-driven financial fraud auditing framework is developed for smart manufacturing rather than a standalone graph classifier. Manufacturing enterprises, suppliers, customers, accounts, orders, invoices, materials, production work orders, warehouse records, and accounting vouchers are represented in a heterogeneous audit graph. Business-process residuals describing amount, quantity, timing, and entity-relation inconsistencies are incorporated into edge representation and relationship-aware aggregation. Multi-task graph inference produces entity-risk probabilities, transaction anomalies, fraud-type probabilities, and candidate propagation paths. These outputs initialize a sequential audit decision process that selects verification actions according to information gain, execution cost, uncertainty, and evidence sufficiency. Verified source records, risk relationships, executed actions, and review results are subsequently connected through explicit support, validation, and refutation relations. The resulting contribution is an end-to-end transition from graph-based fraud screening to process-constrained relational reasoning, dynamic audit verification, and traceable evidence-chain construction.

2. Principles of Intelligent Inference in Multi-Source Audit Data Representation Graphs

Audit data in smart manufacturing enterprises is dispersed across enterprise resource planning (ERP) systems, manufacturing execution systems (MES), warehouse management systems (WMS), supply chain platforms, tax systems, and bank settlement systems. For the same business object, there are typically discrepancies in field naming, coding inconsistencies, and misalignments in time granularity across different databases [10]. Audit data representation requires converting purchase orders, production work orders, warehouse receipt records, sales contracts, invoices, payment transaction logs, and accounting vouchers into unified event units, and performing cross-system entity alignment based on enterprise codes, account identifiers, material numbers, document numbers, and timestamps. Given missing fields and interface noise, entity matching cannot rely on a single primary key; the degree of association between two records can be expressed as:

$$M_{ab} = \sum_{k=1}^{K} w_k \, s_k(a,b) + w_t \, exp\left(-\frac{|t_a - t_b|}{\tau}\right) \quad (1)$$

where M_{ab} is the matching score between record a and record b,$s_k(a,b)$ represents the similarity of the kth business field, w_k is the field weight, t_a and t_b are the event times, τ is the time decay scale, and w_t is the time consistency weight. Records with a matching score exceeding the threshold are merged into the same business entity, while low-confidence records retain their source labels and quality weights to prevent erroneous associations from being amplified during subsequent graph propagation. After entity unification is complete, the audit subject can be represented as a heterogeneous attribute graph G = (V, E, R). The node set V includes companies, suppliers, customers, accounts, orders, materials, invoices, work orders, and vouchers; the edge set E describes business relationships such as procurement, sales, payments, invoicing, warehousing, controls, and cash inflows; and R is the set of relationship types. In addition to the relationship category, each edge must also encode the transaction amount, occurrence time, quantity deviation, data quality, and process constraint residuals. The edge confidence weight is expressed as:

$$w_{ij} = \sigma\left(\theta^T \left[e_{ij} \| \Delta t_{ij} \| \Delta a_{ij} \| q_i \| q_j\right]\right) \quad (2)$$

where $x_{ij} = \left[e_{ij} \| \Delta t_{ij} \| \Delta a_{ij} \| q_i \| q_j\right]$ denotes the concatenated edge-quality vector and θ is a trainable projection vector. Rather than being manually specified or optimized using an independent edge label, θ is included in the complete parameter set Θ in Eq. (11) and is jointly optimized with the HGNN encoder. Because w_{ij} directly participates in the subsequent relationship-aware attention calculation, gradients generated by the entity-risk, transaction-anomaly, fraud-type, and graph-consistency objectives are propagated through w_{ij} to θ. Its update at training iteration ss can be expressed as

$$\theta^{(s+1)} = \theta^{(s)} - \eta \frac{\partial L}{\partial \theta},$$

$$\frac{\partial L}{\partial \theta} = \sum_{(i,j)\in E} \frac{\partial L}{\partial w_{ij}} w_{ij}\left(1 - w_{ij}\right) x_{ij} + 2\eta_5 \theta. \quad (3)$$

Here, η denotes the learning rate, and the second term originates from the L_2 regularization in Eq. (11). Therefore, the edge-confidence mapping is learned end-to-end from downstream fraud-supervision signals rather than determined by fixed empirical coefficients.

This representation distinguishes between normal business connections and abnormally strong associations, allowing closed-loop funding, fraudulent procurement, duplicate invoicing, and abnormal cost transfers to be preserved as local subgraphs. Graph-based intelligent reasoning aggregates neighborhood information through relationship-aware attention, extending the analysis of individual transaction anomalies to cross-entity risk assessments. When node i receives information from neighbor j, it simultaneously considers node attributes, relationship types, edge confidence, and process constraint deviations; its attention coefficient is expressed as:

$$\alpha_{ij} = \frac{exp\left(Leaky\ Re\ LU\left(a^T\left[W_r h_i \| W_r h_j \| Ue_{ij}\right] + \beta w_{ij} + \gamma c_{ij}\right)\right)}{\sum_{k\in N_r(i)} exp\left(Leaky\ Re\ LU\left(a^T\left[W_r h_i \| W_r h_k \| Ue_{ik}\right] + \beta w_{ik} + \gamma c_{ik}\right)\right)} \quad (4)$$

where h_i and h_j are node features; W_r is the mapping matrix corresponding to relationship type r; U is the edge attribute transformation matrix; c_{ij} is the process constraint residual; β and γ are regularization parameters; and $N_r(i)$ is the set of neighbors of node i under relationship r. High-value deviations, abnormal time intervals, and fund return relationships are assigned higher propagation weights, and risk information is aggregated layer by layer along the chains of orders, invoices, accounts, and vouchers. The processes of multi-source data cleansing, entity mapping, event segmentation, and quality weighting are shown in Figure 1. Based on these, the subsequent model generates node risk representations, transaction anomaly probabilities, fraud category distributions, and interpretable risk propagation paths.

3. AI-Driven Financial Fraud Audit Methodology

3.1. Construction of multi-source, heterogeneous audit samples for smart manufacturing

The results of cross-system entity alignment still exist in the form of discrete records and cannot directly reflect complete

business activities; therefore, audit samples must be organized according to the "business entity—business document—execution event—financial outcome" framework [11].Data sources include purchase orders and accounting vouchers from the enterprise resource planning (ERP) system; production work orders and material consumption records from the manufacturing execution system (MES); inventory receipt and shipment data from the warehouse management system; contracts and logistics information from the supply chain platform; invoice records from the tax system; and fund transaction records from the banking system [12].Order numbers, contract numbers, material codes, account identifiers, and voucher numbers are used as deterministic join keys. In the absence of a uniform code, approximate matching is carried out with the entity name, the quantity range, the transaction date, and the summary description, and all the events are mapped to the same time line. Every sample covers the entire business cycle — whether it is paying for purchase, producing, selling, or paying for expenses — while preserving the sequence of events, source systems, and related entities. Data quality affects the reliability of anomaly features; therefore, a record quality score is calculated based on field completeness, cross-table consistency, source reliability, and temporal continuity:

$$q_n = \lambda_1 c_n + \lambda_2 s_n + \lambda_3 r_n + \lambda_4 t_n, \quad \sum_{k=1}^{4} \lambda_k = 1 \qquad (5)$$

Where q_n is the quality score for the nth record; c_n, s_n, r_n , and t_n represent completeness, consistency, source reliability, and temporal continuity, respectively; and λ_k is the corresponding weight. Missing fields are filled in using preceding and subsequent records within the same business chain; fields that cannot be reliably reconstructed are not subject to forced interpolation but are instead marked with a missing value indicator. Figure 1 illustrates the processes of raw data ingestion, field standardization, entity matching, event sorting, quality scoring, and sample partitioning. Fraud-label definitions and simulation templates were established with reference to confirmed fraud cases, internal audit conclusions, and historical review records. These materials were used to define fraud logic, abnormal business-process relationships, and evidence dependencies rather than being directly incorporated as raw samples into the experimental dataset. The generated business chains were subsequently divided into training, validation, and test sets in chronological order to prevent the same business chain or associated entity records from appearing in different subsets.

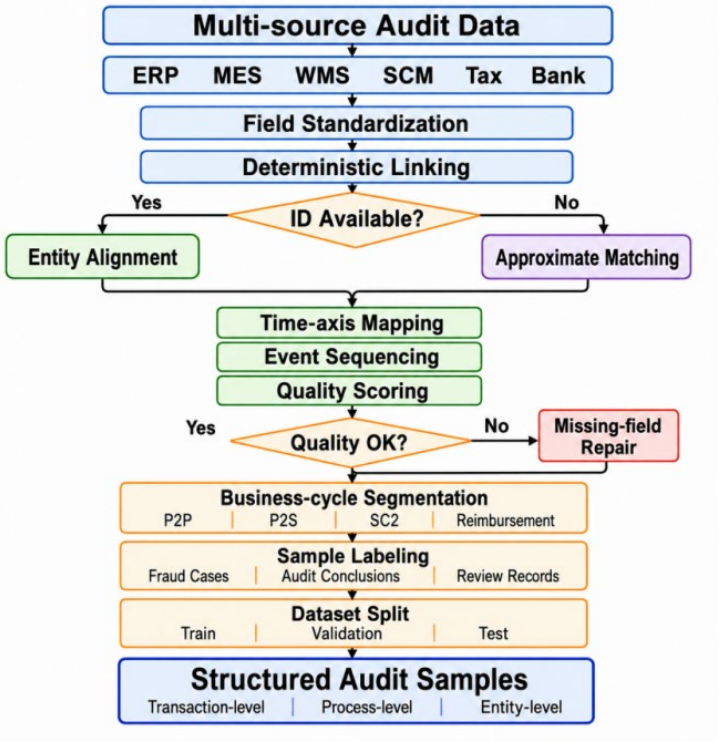


Figure 1. Flowchart for constructing multi-source, heterogeneous audit samples in smart manufacturing

3.2. Business process constraints driving the coding of financial anomaly features

After segmenting the business cycle, anomaly detection should not rely solely on individual transaction amounts or financial ratios; instead, it should examine the degree of business closure among orders, contracts, invoices, logistics, goods receipt, payments, and accounting entries [13]. In the procurement payment process, deviations are identified between order amounts, goods received quantities, invoice amounts, and actual payment amounts; the production-to-warehouse process correlates the bill of materials, actual material withdrawals, equipment output, and completed quantities [14]; the sales-to-collection process verifies the sequence of contract dates, shipment times, revenue recognition, and fund receipt; and the cost-allocation process analyzes variances between standard costs, actual consumption, and period-based allocations. To eliminate the impact of different metric units, the residual of the mth process constraint is expressed as:

$$c_m = \frac{|y_m - \hat{y}_m|}{MAD(y_m) + \varepsilon} \qquad (6)$$

where y_m is the business record value, $\hat{y}_m$ is the expected value calculated based on associated documents and process rules, $MAD(\cdot)$ is the median absolute deviation, and ε is used to prevent the denominator from becoming zero. This treatment reduces the interference of extreme amounts on the

standardized results and converts time inversions, quantity mismatches, and entity conflicts into continuous anomaly intensities. Structured numerical features, summary text semantics, event temporal features, and process residuals are processed through multi-layer perceptrons, pre-trained text encoders, and dilated temporal convolutions, respectively, and then fused using quality gating:

$$z_t = \sum_{d=1}^{D} g_t^{(d)} h_t^{(d)}, \quad g_t^{(d)} = \frac{exp\left(w_d^T h_t^{(d)} + \rho q_t^{(d)}\right)}{\sum_{l=1}^{D} exp\left(w_l^T h_t^{(l)} + \rho q_t^{(l)}\right)} \quad (7)$$

where $h_t^{(d)}$ is the encoding result for data of class d at time t, $g_t^{(d)}$ is the gating weight, $q_t^{(d)}$ is the data quality score, and ρ is the quality adjustment coefficient.

3.3. Heterogeneous graph neural networks for financial fraud risk identification

3.3.1. Construction of multi-entity interconnected audit graphs

Business process features can describe the consistency within a single business chain, but coordinated fraud involving affiliated companies, shared accounts, and circular capital flows requires representation through a graph structure [15]. The audit graph designates manufacturing enterprises, suppliers, customers, bank accounts, and actual controllers as entity nodes, and orders, contracts, invoices, production work orders, warehouse records, and accounting vouchers as business nodes. Edge types include signing, procurement, sales, invoicing, payments and receipts, warehousing, material issuance, bookkeeping, guarantees, control, and account sharing; different edges carry attributes such as amount, quantity, time, location, source system, and process residuals. Considering that transactional relationships change over time, dynamic graph snapshots are constructed according to accounting periods. The connection strength between node i and node j in the rth type of relationship is defined as:

$$w_{ij}^r = q_i q_j \, exp\left(-\frac{|t_i - t_j|}{\tau_r}\right)\left(1 + \mu_r c_{ij}\right) \quad (8)$$

In the equation, w_{ij}^r represents the relationship strength; q_i and q_j represent the data quality weights of the two end nodes; $|t_i - t_j|$ represents the time interval between associated events; τ_r represents the decay parameter corresponding to the relationship type; c_{ij} represents the process constraint residual; and μ_r represents the anomaly adjustment coefficient. Regular purchase relations typically feature stable order, logistic and payment intervals; short term centralized billing, return of funds to a single account, and reconciliation of amounts between entities lead to increased connectivity power. Graph images are constantly updated with a sliding window; long-term entity properties are maintained in global nodes, and new business events are added to the local sub-graphs, creating a multi-entity association audit graph suitable for batch training and on-line reasoning.

3.3.2. Relationship-aware graph attention feature aggregation

The semantic differences among edges in multi-entity audit graphs are significant: procurement relationships reflect business dependencies, payment relationships represent fund transfers, and control relationships depict potential associations. Uniform neighborhood aggregation weakens the ability to identify critical relationships [16]. The model configures an independent mapping matrix for each relationship type, feeding node attributes, edge attributes, and process residuals into the attention layer. The relationship attention weight of node i for neighbor j is calculated as:

$$\alpha_{ij}^r = \frac{exp\left(a_r^T [W_r h_i \| W_r h_j \| U_r e_{ij}] + \beta_r w_{ij}^r\right)}{\sum_{k \in N_r(i)} exp\left(a_r^T [W_r h_i \| W_r h_k \| U_r e_{ik}] + \beta_r w_{ik}^r\right)} \quad (9)$$

where h_i represents node features, e_{ij} represents edge attributes, W_r and U_r are relationship mapping matrices, a_r is the attention parameter, and $N_r(i)$ is the set of neighbors under relationship r. Unlike GraphConsis and CARE-GNN, the proposed relationship weighting does not remove neighbors solely according to feature or label similarity. Process-constraint residuals and source-quality information are jointly encoded into edge attention so that low-similarity but audit-critical links, such as abnormal fund-return, quantity-mismatch, or cross-system timing links, can still be retained when they provide strong process evidence. Multi-head attention separately learns patterns such as cash flow reversal, abnormal invoicing, related-party control, and logistics discrepancies; the node state is updated as:

$$h_i^{(l+1)} = \sigma\left(W_0 h_i^{(l)} + \sum_{r \in R} \sum_{j \in N_r(i)} \alpha_{ij}^r h_j^{(l)}\right) \quad (10)$$

where l is the network layer, W_0 preserves the node's own information, and $\sigma(\cdot)$ is a nonlinear activation function.

3.3.3. Multi-task joint fraud risk prediction

After completing relationship aggregation, single-entity classification cannot simultaneously meet the requirements for risk screening, anomaly detection, and fraud type determination [17].The shared graph encoder outputs entity node, business node, and local subgraph representations. The entity prediction branch calculates risk probabilities for enterprises, suppliers, and accounts; the transaction branch identifies anomalous orders, invoices, payments, and vouchers; and the type branch identifies fraudulent procurement, premature revenue recognition, inflated inventory, cost manipulation, related-party transactions, and misappropriation of funds [18].The three tasks share an underlying relationship representation, while the high-level prediction heads maintain independent parameters to avoid direct conflicts arising from different label granularities. PC-GNN alleviates fraud-label imbalance through label-balanced node sampling and neighborhood sampling. In the present setting, directly resampling isolated nodes may alter the original proportions and dependencies of orders, invoices, production events, payments, and vouchers within a complete business chain. The proposed model therefore preserves chain-

level process structure and introduces class-balance, hard-sample, and graph-consistency terms at the loss-function level. Given the low proportion of fraud samples, conventional cross-entropy is prone to favoring normal categories. The joint training objective incorporates constraints for class balance, hard-sample adjustment, and graph structural consistency:

$$L = \eta_1 L_{node} + \eta_2 L_{edge} + \eta_3 L_{type} + \eta_4 \sum_{(i,j)\in E_f} \left\| \hat{p}_i - \hat{p}_j \right\|_2^2 + \eta_5 \|\Theta\|_2^2 \quad (11)$$

Where L_{node}, L_{edge}, and L_{type} represent the entity risk, transaction anomaly, and fraud type losses, respectively; E_f denotes the associated edges in confirmed fraud chains; $\hat{p}_i$ represents the node risk probability; Θ denotes the model parameters; and η_k denotes the task weights. During the training phase, task weights are dynamically adjusted based on the uncertainty in the validation set; the entity task, which has relatively complete risk labels, bears the primary gradient, while the type tasks with sparse labels obtain supplementary information through shared representations. The model outputs a ranking of entity risks, a list of anomalous transactions, the probability of fraud types, and a local risk subgraph.

For distributed fraud types, the fraud-type branch operates on a local-subgraph representation rather than relying on a single transaction node. Given a candidate local subgraph $G_s = (V_s, E_s)$, node representations from the final graph layer are aggregated through attention pooling:

$$h_s = \sum_{v\in V_s} w_v \, h_v^{(L)}, \quad \sum_{v\in V_s} w_v = 1 \quad (12)$$

and the subgraph-level fraud probability is calculated as:

$$p_s = soft\,max(W_s h_s + b_s) \quad (13)$$

For cost-transfer fraud, the local subgraph is centered on work-order and accounting-period nodes and jointly aggregates material-consumption records, labor-hour allocation, production-cost postings, period-end transfers, and accounting vouchers within the graph receptive field. Consequently, multiple individually weak deviations can jointly produce a high subgraph-level risk even when no single transaction has a sufficiently high anomaly score. The fraud-type loss L_{type} in Eq. (11) is calculated from this subgraph-level prediction for distributed fraud categories.

3.3.4. Extraction of Interpretable Risk Propagation Paths

Risk prediction results must be traced back to specific business events; therefore, starting from high-risk subject nodes, we search the local audit graph for propagation paths that support the prediction conclusions [19]. Candidate paths consist of order, invoice, account, payment, and voucher nodes, and path selection simultaneously considers relationship attention, node risk increment, edge credibility, and process residual. The interpretive contribution of a path PPP is defined as:

$$S(P) = \sum_{(i,j,r)\in P} \alpha_{ij}^r \, w_{ij}^r \left| \hat{p}_j - \hat{p}_i \right| + \kappa \sum_{(i,j)\in P} \varepsilon_{ij} - R(P)$$

$$R(P) = \sum_{v\in P}^{\sum_{[0,n_v(P)-1]} \sum_{(i,j)\in P} 1(c_{ij}<\tau_c)} max$$

$$c_{ij} = \alpha_{ij}^r w_{ij}^r \left| \hat{p}_j - \hat{p}_i \right| + \kappa \varepsilon_{ij} \quad (14)$$

where $S(P)$ denotes the contribution of candidate path P, c_{ij} is the marginal evidence contribution of an edge, and $R(P)$ is a redundancy term rather than a direct path-length penalty. Repeated nodes and low-contribution hops are penalized, whereas additional hops supported by sufficiently strong relationship attention, edge credibility, risk increments, and process residuals are retained. This design avoids systematically suppressing legitimate multi-hop fraud chains. The maximum search depth is used only to control computational complexity and is not included directly in the path score. Figure 2 shows the path of spreading the risk, which is composed of high risk businesses, affiliated suppliers, shared accounts, unusual invoices, and returned funds. Not size is the risk probability, the margin width is the explanatory contribution, and the path tags record the variation of the quantity, the time interval, and the relation type. The output generates explanatory entries organized by entity, transaction, relation, and process break points, while retaining the original record index, which allows tracking of each risk assessment to a particular contract, order, invoice, transaction history, and accounting document.

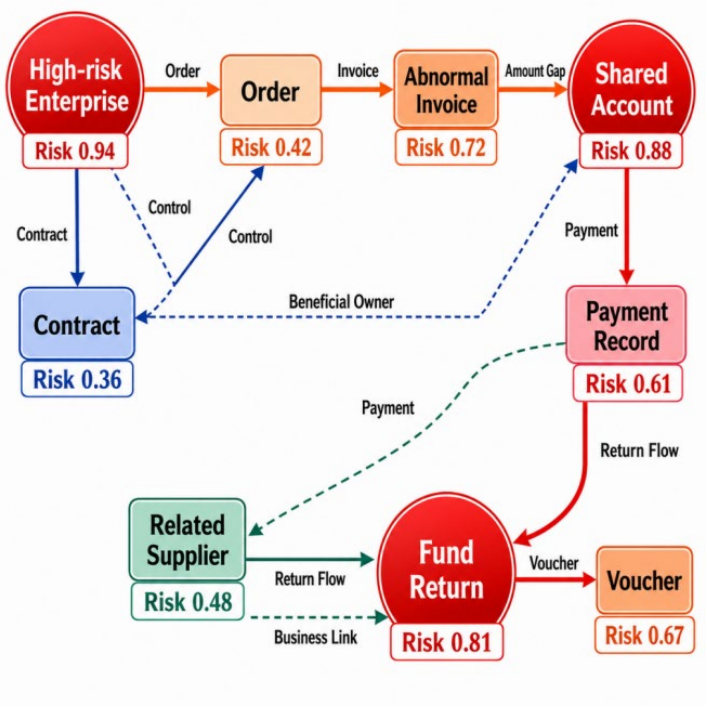


Figure 2. Schematic diagram of financial fraud risk propagation pathways

3.4. Deep reinforcement learning-driven audit path optimization

3.4.1. Modeling sequential decision states in auditing

Graph models may generate risk entities and candidate propagation paths, but actual audits still require the establishment of a validation sequence with limited resources [20]. In each phase, the system reads out the outcome of the present risk, the evidence collected, the remaining budget, and the uncertainty of the model, and then selects the next audit activity [21]. To avoid the infinite number of states, the data is compressed to a fixed dimension by attention pooling. The audit state at time t is expressed as:

$$s_t = [z_t^r \| z_t^e \| b_t \| u_t \| h_t^a] \quad (15)$$

where z_t^r represents high-risk entities and propagation paths, z_t^e represents acquired evidence, b_t represents the remaining audit budget, u_t represents model prediction uncertainty, and h_t^a represents the historical audit action code. Following the execution of an action, the new evidence updates the entity risk probability, path contribution, and evidence coverage, forming the next state. Actions that don't exist, don't have data access rights, or have already been performed are excluded through the feasibility mask. The decision-making process ends when the risk falls below the termination threshold, the evidence sufficiency meets the requirements, or the audit budget is exhausted, outputting the actual execution path and the corresponding set of evidence.

3.4.2. Design of a multi-objective audit reward function

Audit path optimization cannot be limited to the sole objective of detecting fraud; otherwise, the strategy may repeatedly select high-cost verification actions [22]. The reward function simultaneously measures risk information gain, coverage of key entities, evidence sufficiency, execution cost, and operational repetition rate [23]. Information gain is calculated as the difference in entropy between the risk distributions before and after an action is executed, while the adequacy of evidence is updated based on whether a closed-loop system involving contracts, logistics, funds, and supporting documents has been established for key business processes. The reward for a single step is defined as:

$$r_t = \lambda_g \Delta I_t + \lambda_c \Delta C_t + \lambda_e \Delta E_t - \lambda_k K(a_t) - \lambda_d D(a_t) - \lambda_l \quad (16)$$

where ΔI_t is the risk information gain, ΔC_t is the new coverage rate of high-risk entities and transactions, ΔE_t is the increment in evidence sufficiency, $K(a_t)$ is the action cost, $D(a_t)$ is the degree of repetition with historical actions, λ_l is the path length penalty, and λ_g, λ_c, λ_e, λ_k, and λ_d are weights. In high-budget scenarios, the weight for evidence sufficiency is increased; in resource-constrained scenarios, the penalty for action cost is increased; when key evidence shifts the risk conclusion from uncertain to stable, an additional termination reward is granted. The reinforcement-learning audit environment was constructed from the simulated business chains described in Section 4.1. Confirmed-case summaries, internal audit conclusions, and historical review records were used only to calibrate fraud-pattern rules and the relative probability of evidence discovery for different audit actions; they were not directly mixed into the reported training, validation, or test datasets. Unconfirmed states generated controllable feedback through business-process and evidence-consistency constraints.

3.4.3. Risk feedback-driven audit action selection

Since the action space consists of discrete audit procedures, and the costs and execution conditions of different actions vary significantly, a dual-competition deep Q-network is employed to estimate state-action values [24]. Value branches evaluate the overall audit value of the current state, while dominance branches distinguish the incremental benefits of each candidate action; the target network is periodically synchronized to reduce value overestimation. The model introduces a graph-based uncertainty reward during action selection, ensuring that business chains with lower prediction confidence but higher potential risk are given additional verification opportunities. The action selection rule is expressed as:

$$a_t = arg\max_{a \in A_t} [Q_\theta(s_t, a) + \zeta U(s_t, a) - \chi K(a)] \quad (17)$$

where A_t is the set of candidate actions after applying the feasibility mask, $Q_\theta(s_t, a)$ is the action value, $U(s_t, a)$ is the expected reduction in model uncertainty after executing the action, K(a) is the action cost, and ζ and χ are adjustment coefficients. During the initial training phase, a decay-based exploration strategy is used to expand action coverage; in the stable phase, the system primarily executes verification procedures with higher value. The risk-feedback-driven decision-making process is illustrated in Figure 3. After the graph model outputs candidate risk paths, the policy network selects an audit action. The audit environment returns new evidence, execution costs, and risk adjustment results; after updating the state, the decision-making process continues. If bank transaction tracking has already confirmed the return of funds, subsequent strategies will reduce the probability of repeating fund verification and shift to verifying contract authenticity, logistics records, or control relationships, ultimately recording the sequence of actions, state changes, and risk convergence trajectories.

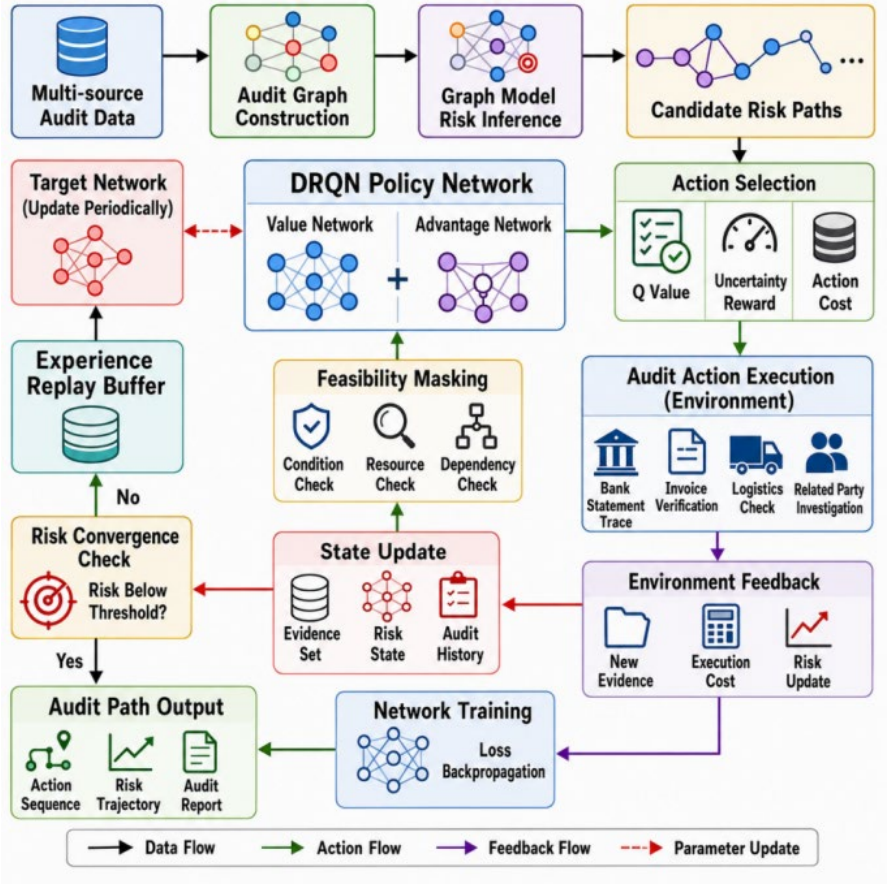


Figure 3. Flowchart of the deep reinforcement learning audit path decision-making process

3.4.4. Closed-loop generation of a trusted chain of evidence

Contracts, invoices, transaction histories, logs, and inventory records generated by audit actions must be organized into structured evidence rather than remaining as isolated attachments [25]. The evidence chain is represented as a directed attribute graph. The original records, anomaly characteristics, risk entities, audit actions, and audit findings are identified as separate nodes. Edge types include Source, Mapping, Validation, Support, and Refutation. Each Proof Node records the source, acquisition time, operator, file hash, model contribution, and review status. The reliability of the evidence node e is calculated on the basis of authenticity, cross-source consistency, conclusion relevance, independent support level, and manual review status:

$$T(e) = \delta_1 A(e) + \delta_2 C(e) + \delta_3 R(e) + \delta_4 S(e) + \delta_5 H(e), \quad \sum_{K=1}^{5} \delta_k = 1 \tag{18}$$

where A(e) is source authenticity, C(e) is cross-system consistency, R(e) is risk conclusion relevance, S(e) is the degree of independent support from other evidence for this node, H(e) is the manual review status, andδ_k are weights. Unreliable evidence does not go straight into the end chain, but triggers additional validation actions; conflicting records are kept as rebuttal nodes, and the validation process is re-selected by the policy network. A sufficient sub-graph is extracted from the candidate evidence graph, which covers key fraud phases and contains fewer nodes. It outputs fraud perpetrators, anomalous transactions, fund flows, business breakpoints, execution procedures, original record indices, and review status, thus forming a closed-loop audit record that spans from multi-source data inputs to financial fraud.

3.5. End-to-End AI-Driven Financial Fraud Auditing Algorithm

To provide an implementation-oriented summary of the complete framework, the procedures described in Sections 3.1–3.4 are integrated into an end-to-end auditing algorithm. The algorithm starts from multi-source business records, performs entity alignment and business-chain reconstruction, encodes process deviations, and constructs the heterogeneous audit graph. Fraud risks and propagation paths are then inferred by the graph model. These outputs initialize the audit decision state, after which the reinforcement-learning policy repeatedly selects feasible verification actions until the risk converges, the evidence sufficiency threshold is satisfied, or the audit budget is exhausted. The collected records and

verification results are finally organized into a structured evidence chain.

Algorithm 1. End-to-End AI-Driven Financial Fraud Auditing and Evidence Chain Construction

Input: Multi-source audit records D; business process rules R_b; audit budget B; risk threshold τ_r; evidence sufficiency threshold τ_e; trained heterogeneous graph neural network (HGNN) Θ; trained audit policy π.

Output: High-risk entities V_r; anomalous transactions T_a; fraud-type probabilities P_f; risk propagation paths P_r; trusted evidence chain G_e.

1) Multi-source audit records are collected from Enterprise Resource Planning (ERP), Manufacturing Execution System (MES), Warehouse Management System (WMS), supply-chain platforms, tax systems, and bank settlement systems to form the initial audit data pool.
2) The collected records are standardized by unifying field formats, removing duplicate entries, correcting temporal inconsistencies, and calculating record-quality scores according to completeness, consistency, source reliability, and temporal continuity.
3) Cross-system entities are matched using deterministic identifiers such as order numbers, contract numbers, material codes, and account identifiers. Approximate matching is applied when unified identifiers are unavailable, and all associated events are aligned on a common timeline.
4) The aligned records are segmented into complete business chains according to business entities, business documents, execution events, and financial outcomes, while the corresponding source-system information and entity-link relationships are retained.
5) Deviations in transaction amounts, material quantities, event timing, and entity relationships are calculated according to the business process rules R_b. Structured attributes, textual information, temporal characteristics, and process residuals are subsequently encoded into unified audit features.
6) A heterogeneous audit graph G=(V,E,R)is constructed using the aligned entities, business documents, transactions, and relationship types. Relationship-aware node and edge representations are then generated to preserve transaction attributes, process deviations, and data-quality information.
7) The trained heterogeneous graph neural network Θ is applied to the audit graph to jointly estimate high-risk entities V_r, anomalous transactions T_a, and fraud-type probabilities P_f.
8) Local high-risk subgraphs are searched from the identified risk entities. Candidate paths are scored according to relationship attention, edge credibility, adjacent-node risk increments, and process residuals. Redundant node revisits and low-information hops are penalized using $R(P)$, whereas path length itself is not directly penalized. Candidate paths with high evidence contribution are retained for subsequent masking-based verification.
9) The initial audit decision state is constructed using the identified risks, candidate propagation paths, collected evidence, remaining audit budget, model uncertainty, and historical audit actions.
10) While the remaining audit budget is greater than zero, the estimated risk has not fallen below τ_r, and the evidence sufficiency remains below τ_e, the audit decision process is iteratively executed. Feasible verification actions are generated while unavailable, unauthorized, or previously completed actions are masked. The policy πthen selects the next verification action according to the estimated action value, expected uncertainty reduction, and execution cost. After the selected action is executed, new audit evidence is collected, and the entity risk, path contribution, evidence coverage, model uncertainty, and remaining budget are updated. The authenticity, cross-source consistency, conclusion relevance, independent support level, and review status of the newly obtained evidence are subsequently evaluated. If conflicting evidence is detected, the corresponding record is retained as a refutation node and an additional verification action is triggered.
11) Original records, anomaly features, high-risk entities, executed audit actions, and review findings are organized into the directed evidence graph G_e, with explicit links describing source, mapping, validation, support, and refutation relationships.
12) Redundant, duplicate, and low-credibility evidence nodes are removed from the candidate evidence graph. The minimum sufficient evidence subgraph is then retained only if it covers the key stages of the suspected fraud process and preserves both supporting and conflicting evidence required for audit verification.
13) The algorithm outputs the identified high-risk entities, anomalous transactions, fraud-type probabilities, risk propagation paths, executed audit actions, and the trusted evidence chain. Each final audit conclusion is linked to its original records and review status to ensure end-to-end traceability.

Algorithm 1 clarifies the information flow among the major modules. The output of graph-based fraud reasoning is not treated as the final audit conclusion; instead, it is converted into the initial state of the sequential audit decision process. Newly acquired evidence continuously updates risk estimates and path contributions, while unreliable or conflicting evidence triggers additional verification. Therefore, the final output contains not only fraud probabilities but also the corresponding propagation paths, executed audit procedures, original record indices, and review statuses, enabling each audit conclusion to be traced back to its supporting evidence.

4. Experimental Analysis

4.1. Construction of the smart manufacturing audit dataset

Based on the business logic of procurement payment, production and warehousing, sales and collection, cost allocation, and fund transfer, a simulated audit dataset representing 12 discrete manufacturing enterprises from January 2022 to December 2024 was constructed. The raw records were generated for enterprise resource planning (ERP), manufacturing execution, warehouse management, electronic invoicing, and bank settlement modules, resulting in 1,223,560 business records. Cross-system records were linked using order numbers, contract numbers, account identifiers, material codes, and timestamps. After deduplication, field standardization, abnormal timestamp correction, and quality scoring, 82,560 complete business chains were obtained. For model development, 6,840 fraud chains and 75,720 normal chains were intentionally retained, corresponding to an enriched positive ratio of 8.28%. This value is not interpreted as the operational prevalence of financial fraud. The enriched development distribution was used to provide sufficient minority samples for supervised optimization and fraud-type learning. To evaluate sensitivity to substantially lower operational prevalence, three additional business-chain-disjoint synthetic evaluation sets were generated using independent random seeds and identical normal-process and fraud-template rules. The corresponding fraud ratios were fixed at 1.00%, 0.50%, and 0.20%. None of these prevalence-controlled sets was used for parameter training, validation-based model selection, or threshold optimization, such as Table 1.

Table 1. Composition of Prevalence-Controlled Synthetic Audit Datasets

Dataset	Total business chains	Fraud chains	Normal chains	Fraud ratio	Purpose
Development set	82,560	6,840	75,720	8.28%	Training, validation and primary controlled testing
Operational set A	82,560	826	81,734	1.00%	Low-prevalence robustness test
Operational set B	82,560	413	82,147	0.50%	Low-prevalence robustness test
Operational set C	82,560	165	82,395	0.20%	Extreme low-prevalence robustness test

The three operational sets preserve the same six fraud mechanisms and business-process dependencies as the development data while changing only the positive-class prevalence and the underlying random realization. This design separates the need for sufficient positive samples during model development from the substantially lower fraud prevalence expected in practical auditing. Precision, recall, F1 score, false-positive rate, and Precision–Recall Area Under the Curve (PR-AUC) should therefore be reported separately on the prevalence-controlled sets rather than extrapolated directly from the enriched development distribution.

Fraud samples were generated using chain-level pattern injection rather than random field perturbation. Six fraud templates were defined according to the fraud categories evaluated in this study. Fraudulent procurement was simulated by introducing inconsistencies among suppliers, purchase orders, invoices, goods-receipt records, and payment records. Revenue manipulation was generated by altering the temporal or quantitative consistency among contracts, shipments, invoicing, revenue recognition, and fund receipts. Cost-transfer fraud was constructed by reallocating material consumption or production costs across work orders or accounting periods, thereby producing distributed low-intensity anomalies across multiple records. Inventory inflation was simulated by introducing inconsistencies among production output, warehouse receipts, inventory movements, and accounting records. Related-party transactions were generated by establishing abnormal control relationships, shared accounts, repeated transactions, or circular fund flows among associated entities. Misappropriation of funds was simulated through abnormal payment destinations, split transfers, fund-return relationships, and inconsistencies between payment records and accounting vouchers. Only the records and relations required by the corresponding fraud template were modified, while unrelated fields in the same business chain retained their normal process structure. The composition of the dataset is shown in Table 2.

The fraud templates were formulated with reference to confirmed-case patterns, internal audit conclusions, and historical review records used in the sample-labeling procedure. To independently evaluate the realism of the simulated audit chains, a blinded expert review was introduced before the final model evaluation. A separate validation sample of 300 business chains was generated using a random seed not used for model training or testing. The sample contained 120 normal chains and 180 fraud chains, with 30 chains drawn from each of the six fraud templates. To independently evaluate the realism of the simulated audit chains, a blinded expert review was introduced before the final model evaluation. A separate validation sample of 300 business chains was generated using a random seed not used for model training or testing. The sample contained 120 normal chains and 180 fraud chains, with 30 chains drawn from each of the six fraud templates. Three manufacturing-audit

professionals who were not involved in model development independently reviewed the anonymized cross-system records. Before participation, the reviewers were informed of the purpose of the study, the review procedure, the types of information to be evaluated, the intended academic use of their ratings, and their right to discontinue participation without consequence. Informed consent for participation and use of the resulting anonymized ratings was obtained from all reviewers. No reviewer names, employer identifiers, contact information, or other directly identifying information were recorded in the analytical dataset. Each reviewer determined whether a chain represented a recognizable audit anomaly, assigned a fraud category when applicable, and rated business-process plausibility, cross-system consistency, and evidence traceability on a five-point scale. Fleiss' kappa was used to evaluate categorical agreement among the three reviewers. A template was retained only when the mean plausibility score was at least 4.0 and the inter-reviewer agreement exceeded 0.75. Expert judgments were used solely to assess the realism of the simulation templates and were not incorporated into model-training labels.

development_8p28 total = 82560 fraud = 6840 fraud rate = 8.285 %
operational_1p00 total = 82560 fraud = 826 fraud rate = 1.000 %
operational_0p50 total = 82560 fraud = 413 fraud rate = 0.500 %
operational_0p20 total = 82560 fraud = 165 fraud rate = 0.200 %

Table 2. Composition of the smart manufacturing audit dataset

Data Category	Number of Records	Number of Linked Entities	Number of Anomalous Records	Number of Fraud-Related Records
Purchase Orders	182,640	3,148	6,472	1,426
Invoice Records	169,380	3,526	7,188	1,735
Bank Statements	236,520	5,904	9,846	2,608
Accounting Vouchers	198,460	2,816	5,924	1,284
Production Orders	142,870	1962	4,116	936
Warehouse Records	174,930	2,408	5,368	1,102
Contract Logistics	118,760	2,774	3,984	824
Total	1,223,560	22,538	42,898	9,915

Transaction chains were divided chronologically into training, validation, and test sets in the proportions of 70%, 15%, and 15%, respectively; the same entity and its associated documents were assigned to only one data subset. The audit graph contained 96,784 nodes and 2,184,630 typed edges, and the fraud test set contained 1,026 transaction chains. The experimental platform is configured with an Intel Core i9-14900K processor, 64 GB of RAM, and an NVIDIA RTX 4090 graphics card. Model training was performed using Python 3.11, PyTorch 2.4, and CUDA 12.1, with a learning rate of 0.001, a hidden dimension of 128, three graph-attention layers, and a batch size of 64. All trainable parameters, including the edge-confidence parameter θ in Eq. (2), relationship-specific mappings, attention parameters, and multi-task prediction heads, were jointly updated using the loss in Eq. (11). The parameter θ was therefore not independently fitted or manually tuned.The complete random seeds, prevalence settings, normal-chain generation procedure, six fraud-injection functions, and dataset-export code are provided in Supplementary Script S1. The script generates the enriched 8.28% development set and the 1.00%, 0.50%, and 0.20% prevalence-controlled evaluation sets without manual modification of individual test samples.

In addition to the controlled datasets, an independent real-world test set was constructed from anonymized historical audit records of three manufacturing enterprises for the external validation reported in Section 4.7. The real-world dataset contained 715,220 raw cross-system records and 3,660 independently audited business chains, including 307 confirmed anomalous chains. These records were completely excluded from graph-model training, validation-based model selection, reinforcement-learning policy optimization, and decision-threshold calibration. Before inference, only deterministic schema mapping, identifier anonymization, timestamp normalization, and measurement-unit conversion were performed. Therefore, the results reported in Section 4.7 represent direct out-of-sample application of the frozen framework rather than enterprise-specific retraining.

4.2. Comparison of financial fraud detection performance

To evaluate the joint detection capabilities of business process constraints and heterogeneous relationship modeling, logistic regression, support vector machines (SVMs), random forests, XGBoost, LSTM, GCN, and heterogeneous graph neural networks were selected as comparison methods. All models used the same training and test sets, and evaluation metrics included accuracy, precision, recall, F1 score, and AUC. The test results for the different models are presented in Figure 4.

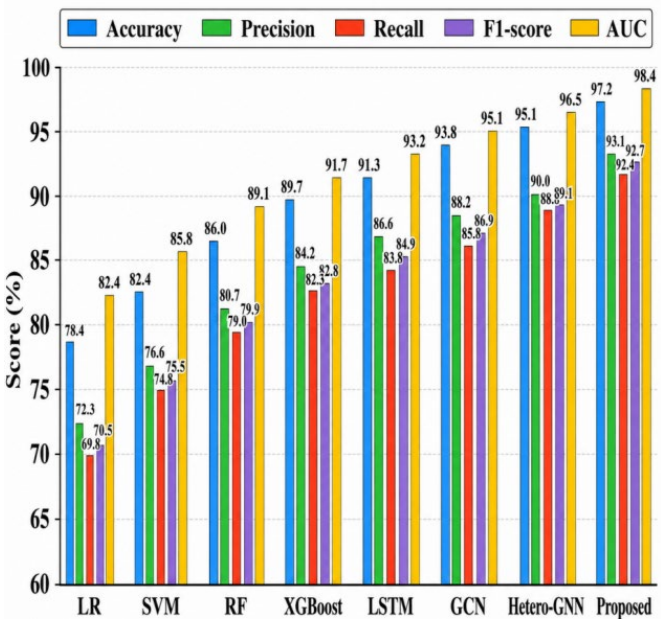


Figure 4. Comparison of financial fraud detection performance across different models

The F1 scores of 70.5% and 75.5% were obtained by the logistic regression and the Support Vector Machine, which showed that the static financial index was unable to detect cross-system cooperative anomalies. XGBoost raised the F1 score to 82.8%, but it was still missing a large number of cases related to fund reflows and control relationships between related entities. LSTM achieved an F1 score of 84.9% by means of transaction order modeling; its strengths lie in detecting end-period concentrated billing and successive split payments, but not enough to detect related fraud without immediate temporal proximity. The F1 scores of GCN and the heterogeneous graph neural network (HGNN) were 86.9% and 89.1%, respectively, which showed that the relation aggregation could complement the independent transaction characteristics. The proposed method achieved accuracy, precision, recall, F1 score, and AUC of 97.2%, 93.1%, 92.4%, 92.7%, and 98.4%, respectively, with a 3.6 percentage point increase in recall compared with the heterogeneous graph neural network (HGNN). The gains primarily stem from the explicit constraints imposed by process residuals on anomalous edges, as well as the joint correction of entity risk, anomalous transactions, and fraud types through multi-task prediction; the false positive rate is controlled at 3.8%.

4.3 Analysis of complex interconnected fraud risk identification

A single classification metric cannot evaluate whether the model can identify fraudulent entities and reconstruct propagation chains. Therefore, the test samples were categorized into fraudulent procurement, revenue manipulation, cost shifting, inventory inflation, related-party transactions, and misappropriation of funds, and the hit rate for risk entities and the recall rate for propagation paths were calculated. The results for the identification and localization of various types of fraud are listed in Table 3.

Table 3. Identification results for different fraud types

Fraud Type	Number of Test Samples	Precision (%)	Recall Rate/%	F1 Score/%	Risk Entity Hit Rate (%)	Transmission Path Recall Rate (%)
Fake Procurement	188	94.2	93.5	93.8	95.1	91.7
Revenue Manipulation	154	92.1	90.8	91.4	93.6	88.9
Cost Transfer	132	90.4	88.7	89.5	91.8	86.5
Inflated inventory	147	91.8	89.9	90.8	92.7	87.6
Related-party transactions	182	94.8	93.9	94.3	96.2	92.8
Capital tied up	223	95.6	94.7	95.1	97.0	94.1

Cost-transfer fraud remained the most difficult category at the individual-transaction and path levels, with a transaction-level recall of 88.7% and a propagation-path recall of 86.5%. Unlike misappropriation of funds, which often produces concentrated abnormal payment destinations or explicit fund-return edges, cost-transfer anomalies are distributed across material consumption, work-hour allocation, production-cost postings, accounting periods, and period-end transfer nodes. Each individual edge may therefore remain below the transaction-level decision threshold. To examine whether these distributed low-intensity signals can be recovered collectively, an additional subgraph-level evaluation was performed using the work-order/accounting-period local subgraphs defined in Section 3.3.3. Subgraph Recall and Subgraph F1 were calculated against the confirmed cost-transfer chains. The comparison between transaction-level and subgraph-level results distinguishes failures caused by weak individual records from failures caused by incomplete relational aggregation.To determine whether path scoring introduces a systematic bias against multi-hop fraud structures,

confirmed fraud paths were further stratified by hop count into short (2–3 hops), medium (4–5 hops), and long (≥6 hops) groups. Path recall was calculated separately for each group using both the original linear length penalty and the revised redundancy-aware score. This experiment directly evaluates whether additional legitimate business links are incorrectly suppressed during path extraction.

4.4. Validation of audit path optimization results

After identifying high-risk entities, we set the maximum number of audit steps to 4, 6, 8, 10, and 12 to compare fixed rules, expert experience, greedy search, and deep reinforcement learning methods. Each scenario was repeated 30 times, and the effectiveness of path selection was evaluated based on high-risk transaction coverage, key evidence discovery rate, and the average number of steps required to establish a complete evidence chain. The results are shown in Figure 5.

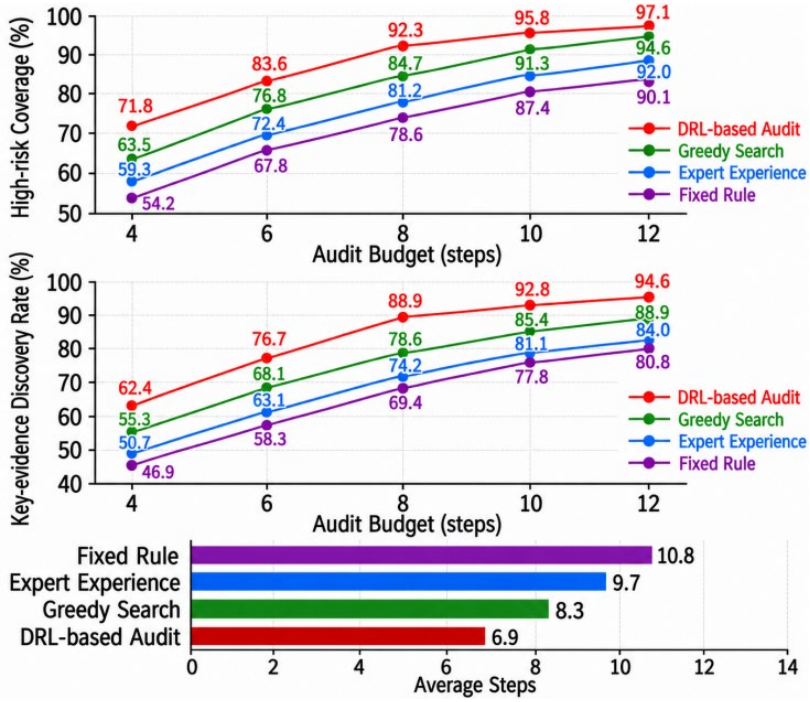


Figure 5. Performance curves of path optimization under different audit resource budgets

When the audit budget was set to 4 steps, the risk coverage rates for the fixed-rule and expert-based methods were 54.2% and 59.6%, respectively, while the proposed method achieved 71.8%, with a key evidence discovery rate of 62.4%.The differences observed under low budgets indicate that fixed-order methods tend to prioritize the examination of documents with low information gain, whereas the risk-feedback strategy can select bank statements, related entities, and logistics records based on relationship contribution and model uncertainty. When the budget was increased to 8 steps, the proposed method achieved risk coverage and key evidence discovery rates of 92.3% and 88.9%, respectively—representing improvements of 7.6 and 10.3 percentage points over greedy search. When the budget reached 12 steps, both metrics increased to 97.1% and 94.6%, respectively; the rate of increase slowed significantly after 10 steps, indicating that the primary fraud chains had been covered. The average number of audit steps required to construct a complete evidence chain decreased from 10.8 steps under the fixed-rule approach to 6.9 steps, and the proportion of duplicate verifications fell from 14.6% to 3.1%. This path compression did not come at the expense of high-risk transaction coverage.

4.5. Evaluation of evidence chain construction quality

The results of path optimization require further verification to determine whether the evidence can support stable audit conclusions. In the experiment, the evidence sets generated by fixed rules, association rules, graph search, greedy path, and the proposed method were submitted to three auditors for independent review. Evaluation criteria included evidence completeness, consistency, key evidence discovery rate, minimum number of evidence nodes, conclusion consistency rate, and manual review time. The results are shown in Table 4.

Table 4. Comparison of evidence chain construction quality across different methods

Method	Evidence Completeness (%)	Evidence Consistency (%)	Key Evidence Detection Rate (%)	Minimum Number of Evidence Nodes	Conclusion Consistency Rate (%)	Manual Review Time/min
Fixed-Rule Method	76.2	82.5	70.8	14.7	78.9	38.6
Association Rule Method	80.4	85.1	75.6	13.2	82.3	34.1
Graph Search Methods	87.6	89.3	84.2	11.4	88.7	27.5
Greedy Path Algorithm	90.1	91.6	87.9	10.2	91.3	23.8
Proposed Method	95.4	96.1	94.6	8.1	96.5	16.7

The fixed-rule method can preserve the underlying documents but struggles to handle conflicting records across systems, resulting in an evidence completeness rate of only 76.2% .Graph search improves association coverage but does not constrain evidence authenticity or duplicate support; the minimum number of evidence nodes remains at 11.4. The proposed method eliminates the low-contribution duplicate nodes by means of plausibility checks and triggers an additional audit of the contradictory records, which achieves the completeness and compliance rate of 95.4% and 96.1% respectively, respectively, which shows that the chain compression is mainly due to the removal of redundant elements and not the omission of key elements. Audit conclusion conformity rate increased to 96.5%, and the manual review time decreased to 16.7 minutes, showing that the risk path, audit activity, and original records were consistent.

4.6. Analysis of model ablation and deployment performance

To quantify the contribution and statistical robustness of each module, each ablation configuration and the complete model were independently trained using the same five random seeds. For every metric, paired differences between the ablated and complete models were calculated seed by seed. Results are reported as the mean change $\bar{\Delta}$, standard deviation (SD), and 95% confidence interval (CI):

$$\bar{\Delta} = \frac{1}{K}\sum_{k=1}^{K}\Delta_k,$$

$$SD = \sqrt{\frac{\sum_{k=1}^{K}(\Delta_k - \bar{\Delta})^2}{K-1}},$$

$$CI_{95\%} = \bar{\Delta} \pm t_{90.975,K-1}\frac{SD}{\sqrt{K}} \quad (19)$$

With K=5, $t_{0.975,4} = 2.776$. A two-sided paired t-test was retained because the complete and ablated models used matched random seeds. To control family-wise error caused by testing seven ablation configurations, raw p-values were adjusted using the Holm–Bonferroni procedure separately for each evaluation metric. Statistical significance was determined using the adjusted value $p_{Holm} < 0.05$ .The results are summarized in Figure 6.

	F1 Change	Recall Change	False Alarm Change	Path Recall Change	Evidence Integrity Change	Avg Steps Change	Latency Change
Without Process Constraints	-4.8	-5.1	+3.2	-4.3	-3.8	+1.7	+0.8
Without Temporal Encoding	-3.9	-4.2	+2.4	-4.8	-3.1	+1.3	+0.6
Without Graph Topology	-6.7	-6.1	+2.8	-11.4	-5.6	+1.9	+1.1
Without Relation Attention	-3.5	-3.2	+1.9	-5.9	-3.7	+1.1	+0.5
Without Multi-task Learning	-2.8	-2.5	+1.4	-3.6	-4.2	+0.9	+0.4
Without Risk Feedback	-0.4	-0.6	+0.5	-1.1	-2.8	+2.6	+0.3
Without Evidence Credibility	-1.1	-0.9	+1.3	-2.2	-9.7	+1.5	+0.7

Figure 6. Mean ablation-induced performance changes across five random seeds

Across repeated runs with different random seeds, the direction and relative magnitude of the ablation effects remained stable, indicating that the observed degradation was not dominated by a specific model initialization. The largest deterioration occurred after removing graph topology: the F1 score decreased by 6.7 percentage points, recall decreased by 6.1 percentage points, and propagation-path recall decreased by 11.4 percentage points. The repeated-run significance test confirmed that these reductions were statistically significant ($p < 0.05$). From an auditing perspective, this degradation is practically substantial because shared accounts, inter-entity transactions, and closed-loop fund flows cannot be adequately reconstructed using isolated transaction features alone.

Removing business process constraints reduced F1 and recall by 4.8 and 5.1 percentage points, respectively, while increasing the false alarm rate by 3.2 percentage points. Relation-aware attention produced a similar effect on fraud-path localization, with propagation-path recall decreasing by 5.9 percentage points. These changes remained consistent across random seeds and were statistically significant ($p < 0.05$), demonstrating that process consistency and relationship-specific aggregation provide stable contributions rather than gains caused by random training fluctuations. Multi-task learning also reduced F1 by 2.8 percentage points and evidence integrity by 4.2 percentage points when removed.

The effects of risk feedback and evidence credibility verification were concentrated on audit efficiency and evidence quality rather than classification accuracy. Removing risk feedback reduced F1 by only 0.4 percentage points, but increased the average number of audit steps by 2.6, indicating that its primary contribution is to verification-path efficiency. Removing evidence credibility verification reduced evidence integrity by 9.7 percentage points while increasing the average audit steps by 1.5. These differences are operationally meaningful because additional verification steps directly increase audit workload, whereas lower evidence integrity weakens the reliability of the final evidence chain even when fraud classification accuracy changes only slightly.

The complete model contains 9.4 M parameters, requires 5.6 GB of GPU memory, and achieves an inference latency of 21.6 ms per business chain with an average throughput of 2,840 records/s. Therefore, the ablation results indicate that the major modules provide reproducible improvements in fraud identification, propagation-path localization, and evidence-chain quality, while the additional computational overhead remains compatible with batch-oriented audit processing. Overall, all seven core modules were independently ablated under identical experimental settings, and repeated-run significance testing was applied to distinguish systematic performance contributions from random training fluctuations.

To provide deployment-oriented evaluation criteria, three quantitative indicators are used to characterize operational reliability under the controlled experimental environment. The false alarm rate is calculated as $FAR=FP/(FP+TN)$, where FP denotes normal business chains incorrectly identified as fraudulent and TN denotes correctly identified normal chains. The missed alarm rate is defined as $MAR=FN/(TP+FN)=1-Recall$, where FN represents confirmed fraud chains not detected by the system. Resource overhead is evaluated using peak GPU memory consumption, average inference latency per business chain, and processing throughput under the fixed hardware configuration. These indicators provide measurable references for subsequent real-world evaluation and can be continuously monitored during future online or batch deployment.

4.7. Real-World Validation Using Anonymized Manufacturing Audit Data

To address the gap between controlled simulation and operational enterprise auditing, the trained framework was further evaluated directly on historical audit data obtained from three manufacturing enterprises. The data were provided for research use under enterprise-level authorization and were used exclusively for the external validation described in this section. The participating enterprises authorized the use of the historical records in anonymized and aggregated form for academic analysis and publication. No raw identifiable enterprise records were released publicly or transferred to parties outside the authorized research process. The real-world records were used exclusively for independent testing and were not incorporated into model training, validation-based model selection, loss-weight optimization, reinforcement-learning policy optimization, or decision-threshold calibration. All graph-model parameters, audit-policy parameters, fraud thresholds, and evidence-sufficiency thresholds were fixed using the controlled development data described in Section 4.1. No gradient update or enterprise-specific fine-tuning was conducted.

Before the records were made available for model inference, a multistage anonymization procedure was applied. Direct personal identifiers, including names, telephone numbers, personal identification information, employee identifiers, and other fields capable of directly identifying natural persons, were removed. Enterprise names, supplier and customer names, bank-account identifiers, internal account identifiers, contract numbers, order numbers, invoice identifiers, and other linkable business identifiers were replaced with randomly generated surrogate codes. The correspondence between original identifiers and surrogate codes was retained only within the source enterprise and was not supplied to the research dataset. Free-text fields containing commercially sensitive descriptions were removed or converted into predefined categorical attributes where required for process analysis. Rare categorical combinations and unnecessary metadata that could disclose enterprise or individual identity were suppressed before data transfer. Only variables required for business-chain reconstruction, process-consistency calculation, fraud identification, and evidence-path evaluation were retained.

A re-identification-risk assessment was conducted before external analysis by checking the remaining records for direct identifiers, unique or near-unique combinations of quasi-identifying attributes, rare entity categories, and sensitive free-text fragments. Records or fields presenting unnecessary identification risk were suppressed, generalized, or excluded. Cross-system linkage was subsequently performed only through the surrogate identifiers, thereby preserving relational structure without exposing the original identities. The research team did not receive the identifier-mapping tables and therefore could not reverse the surrogate codes to the original persons, accounts, suppliers, customers, or enterprise entities. These procedures were designed to satisfy the principle of data minimization while retaining the temporal and relational information required by the heterogeneous graph model.

The three enterprises covered automotive-component manufacturing, industrial-equipment manufacturing, and electronic-component assembly. Historical records were exported from Enterprise Resource Planning (ERP), Manufacturing Execution System (MES), Warehouse

Management System (WMS), electronic invoicing, accounting, and settlement systems. After anonymization, data-minimization filtering, schema harmonization, timestamp normalization, and measurement-unit conversion, 715,220 cross-system records were retained. Independent historical audit files were used to reconstruct 3,660 complete business chains, of which 307 had previously been confirmed as fraudulent or materially anomalous. Two professional auditors independently reviewed the candidate chains, and disagreements were adjudicated by a third senior auditor. The overall inter-reviewer Cohen's κ reached 0.84. The data characteristics are summarized in Table 5.

Table 5. Characteristics of the anonymized real-world manufacturing audit datasets

Dataset	Observation period	Raw records	Audited business chains	Confirmed anomalous chains	Anomaly ratio	Cell-level missing rate	Duplicate-record rate	Timestamp inconsistency	Unmatched cross-system records
Automotive-component manufacturer	Jan.2023–Dec.2024	284,615	1,420	126	8.87%	1.12%	0.34%	0.71%	2.86%
Industrial-equipment manufacturer	Apr.2023–Mar.2025	231,874	1,180	97	8.22%	0.76%	0.21%	0.48%	2.31%
Electronic-component manufacturer	Jan.2024–Jun.2025	198,731	1,060	84	7.92%	1.49%	0.42%	0.93%	3.18%
Total/weighted mean	—	715,220	3,660	307	8.39%	1.11%	0.32%	0.70%	2.78%

The real enterprise records showed substantially greater irregularity than the controlled dataset. Missing values mainly occurred in logistics identifiers, material batch numbers, auxiliary accounting descriptions, and payment-reference fields. Duplicate records primarily originated from repeated ERP exports and interface retransmission. Timestamp inconsistencies were concentrated around production completion, warehouse entry, invoice issuance, and accounting-posting events. Cross-system unmatched records were retained with missing-link indicators rather than discarded, allowing the model to be evaluated under incomplete entity alignment. The frozen model was applied directly to all 3,660 audited business chains. For the automotive-component manufacturer, 111 of 126 confirmed anomalous chains were correctly identified, with 12 normal chains incorrectly classified as anomalous. The corresponding confusion matrix contained 111 true positives, 12 false positives, 15 false negatives, and 1,282 true negatives. Precision, Recall, F1 score, false alarm rate, and missed alarm rate were therefore 90.24%, 88.10%, 89.16%, 0.93%, and 11.90%, respectively.

For the industrial-equipment manufacturer, 88 of 97 confirmed anomalous chains were detected. Eight normal chains generated false alarms, resulting in 88 true positives, 8 false positives, 9 false negatives, and 1,075 true negatives. Precision reached 91.67%, Recall reached 90.72%, and the F1 score was 91.19%. The false alarm rate and missed alarm rate were 0.74% and 9.28%, respectively.

The electronic-component dataset contained more missing cross-system identifiers and timestamp conflicts. Among the 84 confirmed anomalous chains, 72 were successfully detected. Eleven normal chains were incorrectly reported as anomalous. The resulting 72 true positives, 11 false positives, 12 false negatives, and 965 true negatives produced a Precision of 86.75%, Recall of 85.71%, F1 score of 86.23%, false alarm rate of 1.13%, and missed alarm rate of 14.29%. The direct real-world results are summarized in Table 6.

Table 6. Direct test results of the proposed framework on anonymized real enterprise data

Dataset	TP	FP	FN	TN	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC	FAR (%)	MAR (%)	Path Recall (%)	Evidence Completeness (%)	Average Audit Steps
Automotive-component manufacturer	111	12	15	1,282	98.10	90.24	88.10	89.16	0.958	0.93	11.90	86.51	92.18	7.4
Industrial-equipment manufacturer	88	8	9	1,075	98.56	91.67	90.72	91.19	0.969	0.74	9.28	89.69	93.49	7.1

Electronic-component manufacturer	72	11	12	965	97.83	86.75	85.71	86.23	0.943	1.13	14.29	83.33	90.67	7.8
Pooled results	271	31	36	3,322	98.17	89.74	88.27	89.00	0.956	0.92	11.73	86.64	92.18	7.5

The pooled results were calculated directly from the confusion matrix rather than by averaging the three enterprise-level scores. Specifically,

$$Precision = \frac{271}{271+31} = 89.74\%$$
$$Recall = \frac{271}{271+36} = 88.27\%$$
$$F1 = \frac{2 \times 89.74 \times 88.27}{89.74+88.27} = 89.00\%$$
$$FAR = \frac{31}{31+3322} = 0.92\%$$
$$MAR = \frac{36}{271+36} = 11.73\% \quad (20)$$

and the overall accuracy was:

$$Accuracy = \frac{271+3322}{3660} = 98.17\% \quad (21)$$

Risk-path reconstruction was evaluated against the audit-confirmed relational chains. Among the 307 confirmed anomalous business chains, the proposed framework correctly reconstructed the principal evidence-supported propagation path for 266 cases, corresponding to a path recall of 86.64%. The values for the three enterprises were 86.51%, 89.69%, and 83.33%, respectively. The lower value for the electronic-component manufacturer was primarily associated with incomplete material-batch identifiers and delayed synchronization between warehouse and accounting systems. These conditions interrupted several order–warehouse–invoice–payment links even when the final fraud classification remained correct.

Evidence-chain quality was also evaluated using the source documents retained in the historical audit files. Across the 307 confirmed anomalous chains, 1,803 key evidence items were identified by the auditors as necessary for supporting the corresponding audit conclusions. The framework retrieved and linked 1,662 of these items, producing an overall evidence completeness of 92.18%. Enterprise-level completeness rates were 92.18%, 93.49%, and 90.67%. The average number of verification actions required to form a sufficient evidence chain was 7.5, compared with 6.9 under the controlled experiment. Auditor conclusion consistency after model-assisted review reached 92.83%, and the mean manual review time was 19.4 min per anomalous chain, compared with 16.7 min under the controlled experiment.

Deployment measurements were obtained using the same RTX 4090 platform reported in Section 4.1. Mean end-to-end inference latency increased from 21.6 ms per business chain in the controlled test to 24.9 ms for the real enterprise data because the real records contained more missing-link indicators and variable-size local subgraphs. Peak GPU memory consumption increased from 5.6 GB to 5.9 GB, while processing throughput decreased from 2,840 to 2,420 records/s. These changes did not require modification of the model architecture or audit policy.

Compared with the controlled test F1 score of 92.7%, the pooled real-world F1 score decreased by 3.70 percentage points, while propagation-path recall decreased from 90.7% to 86.64% and evidence completeness decreased from 95.4% to 92.18%. The degradation was concentrated in cases involving incomplete cross-system identifiers, inconsistent timestamps, and distributed cost-transfer anomalies. Nevertheless, the model retained an F1 score of 89.00%, an AUC of 0.956, and a false alarm rate below 1% without retraining or enterprise-specific threshold adjustment. These direct results provide preliminary evidence that the proposed heterogeneous graph reasoning, risk-feedback-driven audit policy, and evidence-chain construction procedure can operate on heterogeneous historical manufacturing data rather than only on controlled synthetic records.

4.8. Limitations and Generalizability

Several limitations should be considered when interpreting the results. Real manufacturing ERP and cross-system audit records contain missing fields, duplicate records, invalid values, class imbalance, incomplete identifiers, and cross-organizational distribution shifts. Although the controlled datasets enable reproducible comparison, their fraud patterns are generated from predefined business-process templates and cannot fully reproduce every form of operational manipulation or newly emerging fraud strategy. The three real-world enterprise datasets therefore provide an independent validation of transferability rather than evidence of universal deployment performance.

The real-world evaluation also remains subject to confidentiality and privacy constraints. Enterprise records were analyzed only after authorization, data minimization, identifier replacement, sensitive-text removal, and re-identification-risk screening. The research dataset did not contain the mapping information required to reverse surrogate identifiers to their original identities. Nevertheless, anonymization cannot be interpreted as eliminating all residual disclosure risk, particularly for rare transaction structures or unusual combinations of business attributes. The real-world records are therefore not publicly released at the record level, and the reported results are presented only in anonymized and aggregated form. Any subsequent cross-

enterprise replication should perform a new authorization and privacy-risk assessment rather than assuming that the current data-use arrangement applies to additional organizations. The expert evaluation of simulated business chains was conducted separately from model training. The participating audit professionals were informed of the purpose and procedure of the assessment, their evaluations were used only for simulation-realism validation, and their ratings were stored without direct personal identifiers. Their judgments were not used as model-training labels. This separation limits the possibility that reviewer participation directly influences the reported model-fitting results.

Direct testing on three anonymized manufacturing enterprises reduces the gap between controlled simulation and operational audit environments, but the real-world evaluation remains preliminary. The three datasets originate from discrete manufacturing scenarios and contain 3,660 independently audited business chains; therefore, the results should not be interpreted as evidence of universal cross-industry generalization. Performance degradation was mainly observed when cross-system identifiers were incomplete, timestamps were inconsistent, or fraud signals were distributed across multiple accounting periods. Future evaluations should expand the number of authorized enterprises, extend the observation period, examine prospective deployment under newly emerging fraud patterns and feature-distribution drift, and continue to assess privacy protection and re-identification risk when additional data sources are integrated.

5. Conclusion

Financial fraud in smart manufacturing environments is characterized by cross-system business chains, hidden entity relationships, and fragmented evidence. The proposed framework integrates business-process constraints, heterogeneous graph reasoning, risk-feedback-driven audit-path optimization, and evidence credibility verification within a unified workflow. Under the controlled experiment, the complete framework achieved an F1 score of 92.7%, an AUC of 98.4%, a propagation-path recall of 90.7%, and evidence completeness of 95.4%, while reducing the average number of audit steps to 6.9. Direct testing on three anonymized manufacturing enterprises further involved 715,220 raw records and 3,660 independently audited business chains. Without retraining or enterprise-specific threshold adjustment, the framework achieved a pooled Precision of 89.74%, Recall of 88.27%, F1 score of 89.00%, AUC of 0.956, false alarm rate of 0.92%, propagation-path recall of 86.64%, and evidence completeness of 92.18%. The decrease relative to the controlled experiment was primarily associated with missing cross-system identifiers, timestamp inconsistency, and distributed low-intensity anomalies. These results provide preliminary evidence that the complete auditing framework can be transferred from controlled data to heterogeneous historical manufacturing records while retaining fraud identification, path localization, and evidence-chain construction capabilities.

Ethics, Consent and Data Confidentiality Statement

The controlled experimental datasets used in this study were synthetically generated and did not contain identifiable information relating to natural persons. The historical enterprise records used for external validation were provided under authorization from the participating manufacturing enterprises and were processed under data-minimization and confidentiality requirements before research use. Direct personal identifiers and commercially sensitive free-text information were removed, linkable business identifiers were replaced with surrogate codes, and the original identifier-mapping information remained under the control of the source enterprises. Re-identification-risk screening was performed before the anonymized records were made available for analysis, and only anonymized or aggregated results are reported in this manuscript.
Three external manufacturing-audit professionals participated in the blinded assessment of simulation realism. They were informed of the purpose, procedure, intended use of the evaluation results, and voluntary nature of participation, and informed consent was obtained before the assessment. No directly identifying information relating to the reviewers was included in the analytical dataset or reported in the manuscript. Their evaluations were used only to assess simulation-template realism and were not incorporated into model-training labels.

References

[1] Sahoo S, Lo C Y. Smart manufacturing powered by recent technological advancements: A review[J]. Journal of Manufacturing Systems, 2022, 64: 236-250.

[2] Sood P, Sharma C, Nijjer S, et al. Review the role of artificial intelligence in detecting and preventing financial fraud using natural language processing[J]. International Journal of System Assurance Engineering and Management, 2023, 14(6): 2120-2135.

[3] Li D, Liu S, Wang B, et al. Trustworthy AI for human-centric smart manufacturing: A survey[J]. Journal of Manufacturing Systems, 2025, 78: 308-327.

[4] Imran M, Hamid S, Ismail M A. Advancing Process Audits with Process Mining: A systematic review of trends, challenges, and opportunities[J]. IEEE Access, 2023, 11: 68340-68357.

[5] Yoo Y, Shin J, Kyeong S. Medicare fraud detection using graph analysis: A comparative study of machine learning and graph neural networks[J]. IEEE Access, 2023, 11: 88278-88294.

[6] Zioviris G, Kolomvatsos K, Stamoulis G. An intelligent sequential fraud detection model based on deep learning[J]. The Journal of Supercomputing, 2024, 80(10): 14824-14847.

[7] Leng J, Zhu X, Huang Z, et al. Unlocking the power of industrial artificial intelligence towards Industry 5.0: Insights, pathways, and challenges[J]. Journal of Manufacturing Systems, 2024, 73: 349-363.

[8] Nadim K, Ragab A, Ouali M S. Data-driven dynamic causality analysis of industrial systems using interpretable machine learning and process mining[J]. Journal of Intelligent Manufacturing, 2023, 34(1): 57-83.

[9] Gandhar A, Gupta K, Pandey A K, et al. Fraud detection using machine learning and deep learning[J]. SN Computer Science, 2024, 5(5): 453.

[10] Talukder M A, Khalid M, Uddin M A. An integrated multistage ensemble machine learning model for fraudulent transaction detection[J]. Journal of Big Data, 2024, 11(1): 168.

[11] Meier S, Klarmann S, Thielen N, et al. A process model for systematically setting up the data basis for data-driven projects in manufacturing[J]. Journal of Manufacturing Systems, 2023, 71: 1-19.
[12] Innan N, Sawaika A, Dhor A, et al. Financial fraud detection using quantum graph neural networks[J]. Quantum Machine Intelligence, 2024, 6(1): 7.
[13] Devi R R, Raja J E, Chin Y B. Reinforcement learning with graph neural network (RL-GNN) fusion for real-time financial fraud detection: a context-aware community mining approach[J]. Scientific Reports, 2025, 15(1): 42953.
[14] Abdul Salam M, Fouad K M, Elbably D L, et al. A federated learning model for credit card fraud detection with data balancing techniques[J]. Neural Computing and Applications, 2024, 36(11): 6231-6256.
[15] Harish S, Lakhanpal C, Jafari A H. Leveraging graph-based learning for credit card fraud detection: a comparative study of classical, deep learning, and graph-based approaches[J]. Neural Computing and Applications, 2024, 36(34): 21873-21883.
[16] Weber P, Carl K V, Hinz O. Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature[J]. Management Review Quarterly, 2024, 74(2): 867-907.
[17] Zhang M, Tao F, Zuo Y, et al. Top ten intelligent algorithms towards smart manufacturing[J]. Journal of Manufacturing Systems, 2023, 71: 158-171.
[18] Puthanveettil Madathil A, Luo X, Liu Q, et al. A review of explainable artificial intelligence in smart manufacturing[J]. International Journal of Production Research, 2025, 63(23): 8654-8697.
[19] Meyers B, Vangheluwe H, Lietaert P, et al. Towards a knowledge graph framework for ad hoc analysis in manufacturing[J]. Journal of Intelligent Manufacturing, 2024, 35(8): 3731-3752.
[20] Li C, Zheng P, Yin Y, et al. Deep reinforcement learning in smart manufacturing: A review and prospects[J]. CIRP Journal of Manufacturing Science and Technology, 2023, 40: 75-101.
[21] Foalem P L, Silva L D, Khomh F, et al. Logging requirement for continuous auditing of responsible machine learning-based applications[J]. Empirical Software Engineering, 2025, 30(4): 97.
[22] Samsonov V, Hicham K B, Meisen T. Reinforcement Learning in Manufacturing Control: Baselines, challenges and ways forward[J]. Engineering Applications of Artificial Intelligence, 2022, 112: 104868.
[23] Compagnino A A, Maruccia Y, Cavuoti S, et al. An Introduction to Machine Learning Methods for Fraud Detection[J]. Applied Sciences, 2025, 15(21): 11787.
[24] Hacini A D, Benabdelouahad M, Abassi I, et al. LLM-Assisted Financial Fraud Detection with Reinforcement Learning[J]. Algorithms, 2025, 18(12): 792.
[25] Isaja M, Nguyen P, Goknil A, et al. A blockchain-based framework for trusted quality data sharing towards zero-defect manufacturing[J]. Computers in Industry, 2023, 146: 103853.