Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 15 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i15.62709

PAPER

An Interactive Mobile Computing and Analytical Framework for Complex Financial Data with Applications to Audit Tasks

Hongling Zhang(✉)

Shangqiu Polytechnic, Shangqiu, China

sqzyshenji@163.com

ABSTRACT

The growing need for real-time interactive analysis of financial data in mobile on-site auditing is hindered by mobile computational limits and poor network conditions, leading to high multidimensional query latency, low fraud detection accuracy, and weak data synchronization reliability. Existing end-edge-cloud frameworks are not tailored to financial auditing, failing to simultaneously meet real-time interactivity, low terminal power consumption, and high-precision risk detection. In this study, an end-edge-cloud three-layer collaborative interactive mobile computing and analytical framework was constructed for financial auditing, in which dynamic and coordinated scheduling of multi-tier resources was achieved through a task offloading agent. To address domain-specific challenges, a set of tailored optimization mechanisms was proposed: multidimensional financial data query efficiency on mobile devices was enhanced through user behavior prediction and incremental visualization strategies; a two-stage inference architecture combining coarse-grained screening on the end side and fine-grained judgment on the edge side was designed, where global cost optimization and proportional-integral-derivative-based adaptive threshold control were integrated to achieve a dynamic balance between detection accuracy and device energy consumption; a multi-constrained ternary task offloading model and a dependency-aware scheduling strategy were developed to accommodate the tidal workload patterns of audit tasks; a network-graded differential synchronization mechanism was established to guarantee the consistency of financial data synchronization under weak network conditions; and for core audit scenarios, a dedicated algorithmic system incorporating heterogeneous anomaly scoring, graph convolutional fund flow mining, and dual-encoder evidence matching was constructed. This study advances the theory of mobile edge collaborative computing in the financial auditing domain and provides an efficient and feasible technical solution for the large-scale deployment of intelligent mobile auditing.

KEYWORDS

mobile edge computing, end-edge-cloud collaboration, multidimensional financial online analytical processing, adaptive task offloading, interactive mobile data analysis, financial anomaly auditing

Zhang, H. (2026). An Interactive Mobile Computing and Analytical Framework for Complex Financial Data with Applications to Audit Tasks. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(15), pp. 94–108. https://doi.org/10.3991/ijim.v20i15.62709

Article submitted 2026-04-18. Revision uploaded 2026-06-08. Final acceptance 2026-06-15.

1 INTRODUCTION

Digital transformation has driven financial auditing to shift from traditional post-hoc verification toward real-time on-site mobile analysis [1–3], with mobile operations having become the core mode of enterprise field auditing. Massive amounts of multidimensional financial data accumulated by enterprises exhibit complex hierarchies and strong correlations, imposing stringent demands on the efficiency of interactive mobile analysis and the capability of intelligent risk identification. Constrained by the limited computational capacity and energy resources of mobile devices, as well as dynamic network fluctuations in field environments, existing mobile auditing practices generally suffer from high latency in multidimensional data queries and delays in risk warning [4], thereby failing to support real-time on-site decision-making. Although end-edge-cloud collaborative mobile computing technologies have advanced rapidly and been widely applied in general-purpose scenarios such as the Internet of Things and multimedia, the relevant architectures and optimization strategies have not been tailored to the characteristics of multidimensional financial data structures or the operational constraints of audit risk control [5, 6], resulting in a lack of dedicated collaborative computing systems for the financial auditing domain. Against this background, research on a customized mobile computing framework is conducted in this study. Theoretically, this work can refine the theory of mobile-end collaborative interactive computing for structured industry big data and address the theoretical shortcomings in intelligent financial auditing; practically, it can effectively resolve the real-time analysis challenges in field auditing and provide reliable technical support for intelligent on-site auditing in large-scale enterprises.

At the current stage, substantial research findings have been accumulated in the fields of mobile data analysis, end-edge collaborative inference, mobile task scheduling, and data synchronization. However, significant limitations remain when these approaches are adapted to the specific scenario of financial auditing. In the area of mobile multidimensional data analysis [7, 8], existing mobile online analytical processing optimization methods are mostly focused on static storage pruning and fixed caching strategies. The spatiotemporal characteristics of user interactions have not been leveraged for behavior prediction, and dynamic prefetching and incremental update mechanisms are absent. As a result, network round-trip delays caused by random queries cannot be eliminated, making it difficult to meet the operational requirements of high-frequency multidimensional drill-down analysis in auditing.

In the area of end-edge collaborative inference, mainstream hierarchical inference schemes split models and allocate tasks solely based on computational resources, using fixed thresholds to screen anomalous samples. The cost of missed detection risks in financial fraud detection has not been considered [9, 10], nor have terminal power consumption and edge load been incorporated into dynamic parameter tuning. Therefore, an optimal balance among detection accuracy, device energy consumption, and system load cannot be achieved. In the area of task scheduling and data synchronization [11, 12], traditional offloading strategies mostly adopt an end-cloud binary scheduling model, which fails to fully leverage the collaborative scheduling capabilities of edge nodes and cannot adequately accommodate the tidal workload characteristics of audit tasks. Moreover, existing data synchronization methods are predominantly based on full synchronization [13], which cannot adapt to weak network conditions and network fluctuations in field environments, making it difficult to simultaneously ensure synchronization efficiency and multi-terminal data consistency. In addition, most studies have focused only on the simulation validation of individual algorithms [14], without forming a full-link audit solution covering anomaly

detection, evidence collection, and working paper compilation. Consequently, a disconnect exists between theoretical research and engineering implementation.

To address the aforementioned research gaps, systematic innovations are developed in this work at both the theoretical modeling and engineering implementation levels. At the theoretical level, an interaction-intent-driven incremental update model for multidimensional financial data is constructed, an end-edge collaborative inference optimization framework incorporating miss detection penalties is established, a multi-constrained end-edge-cloud ternary task offloading model is formulated, and a set of network-adaptive data synchronization principles is derived. At the engineering level, a resource-aware collaborative scheduling agent is developed, lightweight intelligent algorithms are deployed for core audit scenarios, and an integrated prototype system for mobile audit analysis is built, thereby enabling full-link lightweight intelligent audit operations.

2 AN END-EDGE-CLOUD INTERACTIVE MOBILE COMPUTING AND ANALYTICAL FRAMEWORK FOR MOBILE FINANCIAL AUDITING

2.1 Overall three-layer hierarchical architecture and design of the task offloading agent for resource-aware scheduling

To address the industry problems of delayed interactive response in mobile financial auditing, insufficient adaptability of general-purpose collaborative architectures, and tidal workload fluctuations in audit tasks, a three-layer hierarchical end-edge-cloud collaborative computing framework tailored to the operational characteristics of financial verification is constructed in this study, thereby overcoming the homogenized design limitations of traditional general-purpose collaborative architectures originally intended for the Internet of Things and multimedia scenarios. Customized functional divisions across the hierarchical layers are carried out based on the specific requirements of multidimensional financial analysis and intelligent risk auditing. Figure 1 illustrates the proposed end-edge-cloud three-layer interactive collaborative computing framework for mobile financial auditing. Mobile terminals are focused on lightweight interaction and front-end computational tasks, including user interaction behavior sensing, incremental rendering of multidimensional financial data, dynamic data prefetching, and preliminary screening of transaction anomalies, thereby ensuring real-time interactive experience in field auditing. Edge computing nodes undertake the core functions of high-precision risk inference, local financial data aggregation, and regional task scheduling, leveraging the advantage of proximate computational power to alleviate the resource constraints of terminals and the remote transmission latency of the cloud. The cloud platform is responsible for persistent storage of global financial data, global offline multidimensional analysis, intelligent voucher verification, and model retraining and updating, thereby supporting systematic audit analysis and continuous algorithm optimization.

To achieve dynamic and efficient collaboration among the three heterogeneous tiers, a dedicated task offloading and scheduling agent is designed, which continuously collects multidimensional resource status parameters, including terminal computational load, device power consumption, network round-trip latency, edge node queue congestion, and cloud interface response latency. Unlike fixed-period resource monitoring approaches, this agent adaptively adjusts its sampling and decision update frequency according to the tidal workload characteristics of audit tasks: scheduling responsiveness is enhanced during peak business hours,

whereas unnecessary resource overhead is reduced during off-peak hours. Ultimately, globally balanced scheduling of computational, network, and storage resources across the end, edge, and cloud tiers is achieved.

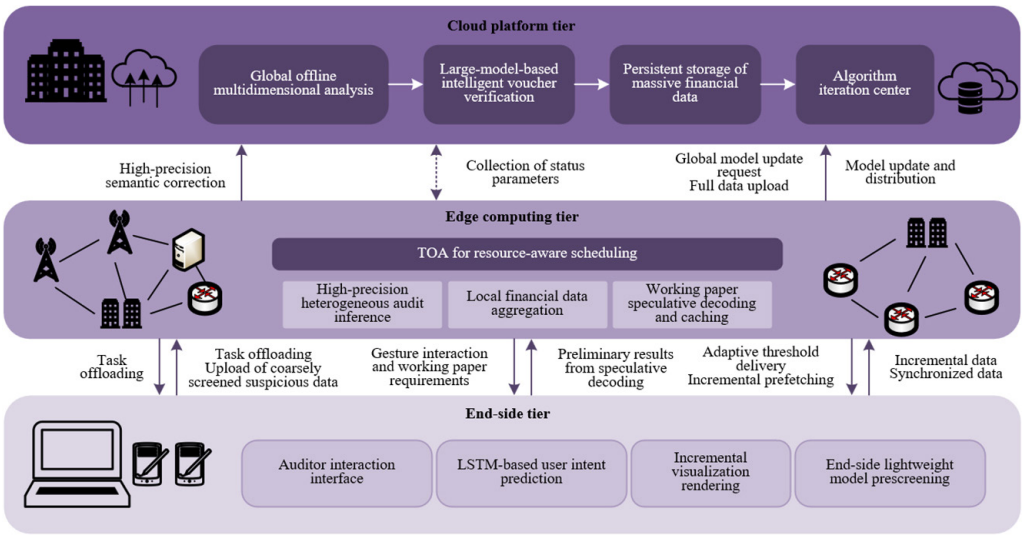

Fig. 1. End-edge-cloud three-layer interactive collaborative computing framework for mobile financial auditing

2.2 Interaction-intent-driven end-side multidimensional analysis engine

To address the network round-trip latency problem caused by drill-down, slicing, and pivoting operations on multidimensional financial data on mobile devices, an interactive analysis prediction system tailored to the hierarchical data structure of financial data is constructed in this study. First, a hierarchical mathematical definition of the multidimensional financial data cube is provided, through which a standard set of analysis dimensions and the hierarchical membership structure of each dimension are established. Then, by incorporating the spatiotemporal continuity characteristics of auditors' touchscreen operations, a hidden Markov model is introduced to characterize the dynamic evolution of user behavior, and a state transition probability model integrating operation type, dimension hierarchy, and gesture features is formulated. The transition probability of a user's next operation can be defined as:

$$P(o_{t+1} \mid o_t, C) \tag{1}$$

where, o_t and o_{t+1} denote the data analysis operation at the current moment and the moment to be predicted, respectively, and C represents the set of contextual features consisting of dimension hierarchy, gesture sliding speed, and movement direction. To accommodate the low-latency and low-power constraints of mobile devices, a lightweight long short-term memory network is employed to mine latent dependencies in continuous operation sequences. Taking historically encoded operation vectors and touch motion features as inputs, the probability distribution over various subsequent interaction operations is produced as output. The model is embedded and deployed on the end side using TensorFlow Lite Micro, with single inference latency below 5 ms and computational power consumption controlled within 10 mJ, thereby enabling high-precision real-time prediction of user interaction intent without occupying excessive computational resources on the terminal.

Based on the prediction results of user interaction intent, high-confidence pending operations are selected by the system, and the corresponding dimension combinations to be prefetched are determined, after which target aggregated data are actively loaded from edge nodes. In this way, the interaction mode of mobile-side querying is upgraded from passive querying to active prefetching. To overcome the shortcomings of high redundancy and slow response in traditional full-refresh visualization mechanisms, a hierarchical incremental visualization update mechanism is designed, through which only the newly added data subset derived from the current dimension node is loaded and locally rendered, thereby completely avoiding the redundant overhead of full data retransmission and global redrawing. The data volume of incremental updates can be quantified using the hierarchical characteristics of financial dimensions:

$$\left|\Delta R\right| = \sum_{v \in \text{descednants}(u)} \text{count}(v) \approx \text{fanout}(u) \cdot \text{avg\_count\_per\_child} \tag{2}$$

where, u denotes the current visualization dimension node, descednants(u) denotes the set of lower-level dimension members corresponding to the current node, fanout(u) represents the fan-out coefficient of the dimension hierarchy, and avg_count_per_child is the average data volume of child dimension nodes. Based on this quantitative model, the transmission optimization gain can be further derived. Compared with traditional full-query schemes, the data transmission compression ratio achieved by this mechanism is $1 - \left|\Delta R\right| / \left|R_{new}\right|$. In conventional financial analysis scenarios, dimension fan-out coefficients are generally maintained within a relatively small range. This incremental update strategy can reduce ineffective data transmission by more than 80%, effectively eliminate interaction stuttering caused by network round-trip latency, and significantly improve the interactive real-time performance of multidimensional financial data analysis on mobile devices.

2.3 Two-stage audit model collaborative inference and proportional-integral-derivative-based adaptive threshold optimization

To simultaneously satisfy the real-time inference requirements of mobile devices and the discrimination reliability of financial anomaly auditing, a two-tier collaborative inference system is constructed in this study, through which hierarchical screening and precise verification of financial transaction samples are achieved. Given the inherent characteristics of financial transaction data—namely, convergent numerical distributions and sparse anomalous patterns—a customized 8-bit integer quantization strategy is employed to compress the lightweight audit model. Unlike global uniform quantization schemes used in general-purpose models, this approach preserves higher-precision weight distributions for financially sensitive feature dimensions. Consequently, the model volume is reduced to one-quarter of the original floating-point model, and inference speed is improved by a factor of three to five, while information loss in key fraud features is effectively suppressed. The end-side lightweight model M_{light} is responsible for prescreening all transaction samples, producing a normalized anomaly score for each sample within a bounded interval. Samples whose scores exceed a preset threshold are uploaded as suspicious candidate sets to edge nodes. At the edge side, a high-precision heterogeneous audit model integrating multiple discrimination mechanisms is deployed. Only the limited candidate samples screened by the end side undergo in-depth feature analysis and precise risk determination, thereby substantially reducing the global computational overhead of the high-precision model and enabling efficient reuse of computational resources. Figure 2 presents a

flowchart of the "end-side coarse screening and edge-side fine judgment" two-stage collaborative inference and dynamic adaptive regulation process.

To quantitatively balance the multiple constraints of inference latency, terminal energy consumption, and audit miss detection risk, a system-level global cost function integrating multidimensional loss indicators is constructed in this study, through which hardware operational overhead and business risk overhead are uniformly incorporated. The specific expression is given as:

$$C_{\text{total}} = \sum_{x \in \text{all}} \left[\alpha \cdot 1_{\text{edge}}(x) \cdot \left(d_{\text{edge}}^{\text{inf}} + d_{\text{net}} \right) + \beta \cdot E_{\text{end}}(x) \right] + \gamma \cdot Miss(C) \tag{3}$$

where, α and β denote the weighting coefficients for system latency and terminal energy consumption, respectively; $1_{\text{edge}}(x)$ is the sample offloading decision function; $d_{\text{edge}}^{\text{inf}}$, d_{net}, and $E_{\text{end}}(x)$ correspond to edge inference latency, network transmission latency, and end-side inference energy consumption, respectively; and $\gamma \cdot Miss(C)$ represents the miss detection penalty term introduced specifically for audit task characteristics, which quantifies the loss caused by fraudulent samples being missed due to threshold deviation. To overcome the limitation of fixed thresholds being unable to adapt to dynamic workloads, a proportional-integral-derivative-based closed-loop regulation mechanism is constructed based on actual anomaly detection feedback, through which the end-side screening threshold is dynamically and iteratively updated. The iterative update rule is expressed as:

$$\tau_{\text{light}}(k+1) = \tau_{\text{light}}(k) + K_p \cdot (\rho_{\text{target}} - \rho(k)) + K_i \sum_{j=0}^{k} (\rho_{\text{target}} - \rho(j)) \tag{4}$$

where, ρ_{target} denotes the target precision rate for candidate anomalous samples, $\rho(k)$ is the actual precision rate at the current moment, and K_p and K_i are the proportional control coefficient and the integral control coefficient, respectively. Through this mechanism, the screening criteria can be dynamically calibrated according to the tidal workload characteristics of audit tasks; the quality of the candidate anomaly set is continuously optimized; task congestion at edge nodes and missed detections of fraudulent samples are effectively avoided; and a dynamic balance among system energy consumption, inference latency, and audit accuracy is achieved.

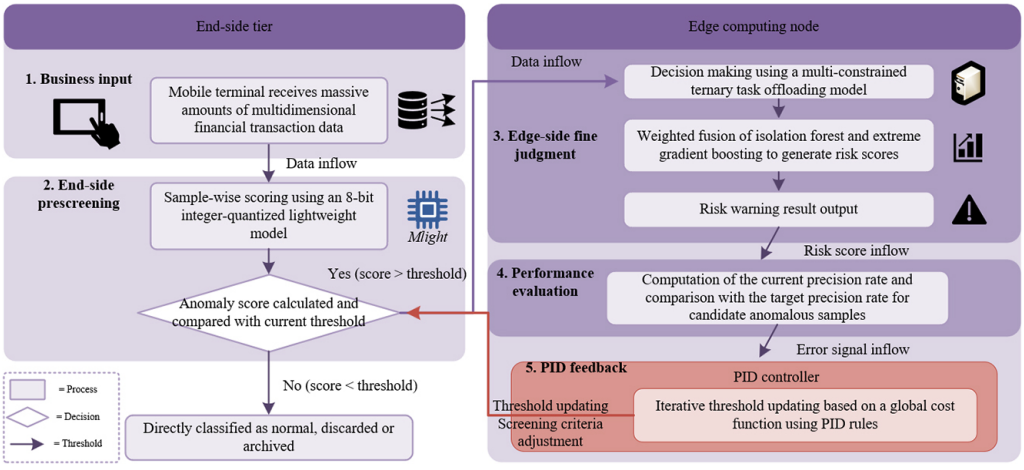

Fig. 2. Flowchart of the "end-side coarse screening and edge-side fine judgment" two-stage collaborative inference and dynamic adaptive regulation process

2.4 Multi-constrained adaptive ternary task offloading and graded differential data synchronization under weak network conditions

Given the business characteristics of financial audit tasks—namely, significant tidal workloads, high coupling of data dependencies, and stringent real-time constraints—traditional end-cloud binary offloading strategies cannot fully exploit the collaborative scheduling value of edge computational resources and fail to meet the global optimization requirements of multi-tier heterogeneous computing resources. In this study, an end-edge-cloud ternary task scheduling mechanism is established, in which a decision variable ($x_i \in \{0,1,2\}$) is defined to distinguish the task execution tier, corresponding to local execution on the mobile terminal, offloaded execution to an edge node, and offloaded execution to the cloud, respectively. Based on the differences in computational capacity, transmission bandwidth, and communication latency across tiers, a refined mathematical model of task execution latency and mobile terminal energy consumption is formulated, through which the resource consumption characteristics of different execution schemes are precisely characterized. Subsequently, with the minimization of the weighted sum of task latency and device energy consumption as the optimization objective, and with the introduction of task deadline constraints, a multi-constrained integer programming scheduling model is constructed as follows:

$$\min_{\{x_i\}} \sum_i \left(w_L \cdot L_i^{x_i} + w_E \cdot E_i^{x_i} \right) \quad \text{s.t. } L_i^{x_i} \le \mathrm{T}_i^{\mathrm{dealline}} \tag{5}$$

where, w_L and w_E denote the weighting coefficients for system latency and terminal energy consumption, respectively; $L_i^{x_i}$ and $E_i^{x_i}$ represent the completion latency and terminal energy consumption of the task under the corresponding execution tier; and $T_i^{dealline}$ is the hard completion deadline of the audit task. To address the additional overhead caused by cross-tier scheduling of dependent tasks, a cluster-based collaborative scheduling strategy is developed based on task dependency topologies. Audit tasks exhibiting strong data coupling are aggregated into unified scheduling units and assigned collectively to the same computing tier. Meanwhile, through a pipelined prefetching mechanism, intermediate computation results are streamed in a push manner, thereby effectively eliminating cross-tier task blocking latency and improving the overall efficiency of multi-task concurrent scheduling.

To address the problems of inefficient data synchronization and data inconsistency caused by dynamic network fluctuations in field auditing scenarios, a quantifiable network quality assessment model and an adaptive graded differential synchronization mechanism are constructed in this study, through which dynamic matching between synchronization strategies and network conditions is achieved. The quantitative criterion for comprehensive network quality scoring is defined as:

$$Q_{\mathrm{net}} = \min\left(1, \frac{B_{\mathrm{up}}}{B_{\mathrm{ideal}}}\right) \cdot \left(1 - \frac{\mathrm{packet\_loss}}{0.1}\right) \tag{6}$$

where, B_{up} and B_{ideal} denote the real-time uplink bandwidth and the ideal reference bandwidth, respectively, and packet_loss is the packet loss rate of the current link. Differentiated synchronization modes are selected based on network quality score thresholds: under high network quality conditions, full real-time synchronization is performed; under medium network conditions, incremental differential synchronization is adopted to reduce transmission volume; under extreme weak

network conditions, only local operation records are retained, and batch data merging is performed after network recovery. The synchronization process integrates version vector timestamping, Bloom filter-based conflict detection, and conflict-free replicated data type merging mechanisms, through which update differences across multiple terminals are precisely identified and concurrent conflicts are resolved. Consequently, while significantly reducing transmission overhead under weak network conditions, eventual consistency of global financial audit data is guaranteed.

2.5 Customized deployment algorithms for mobile auditing scenarios

To address the audit challenges of limited discriminative accuracy in single-transaction anomaly detection and the high concealment of related-party fund fraud, a heterogeneous ensemble anomaly detection system tailored to the characteristics of multidimensional financial data is constructed in this study, through which the technical advantages of supervised and unsupervised learning are integrated. The model combines the unsupervised outlier mining capability of isolation forest with the supervised classification advantage of extreme gradient boosting. Through a dynamic weighting mechanism, the anomaly scoring results of the two types of models are fused, and a standardized risk score is uniformly output as follows:

$$S(x) = \lambda \cdot f_{\mathrm{ISO}}(x) + (1 - \lambda) \cdot f_{\mathrm{XGB}}(x) \tag{7}$$

The weight coefficient is no longer set to a fixed value but is adaptively and dynamically updated based on the prediction variances of the two models on the validation set. The update rule satisfies:

$$\lambda = \frac{\sigma_{\mathrm{XGB}}^2}{\sigma_{\mathrm{ISO}}^2 + \sigma_{\mathrm{XGB}}^2} \tag{8}$$

Thus, the contribution ratio of each model can be adaptively adjusted according to the sufficiency of sample labeling, effectively accommodating the inherent characteristic of imbalanced sample distribution in financial auditing. To identify concealed related-party transactions and closed-loop fund fraud behaviors, business entities and their fund flow relationships are constructed as a topological transaction network, upon which a graph convolutional network is employed to perform deep aggregation and iterative learning of node features. The feature update formula is given as:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{9}$$

Through this structure, implicit fund flow associations among entities can be fully exploited, thereby overcoming the technical limitation of traditional single-transaction detection models that cannot identify collusive fraud, and enabling fine-grained risk discrimination across the global transaction chain.

To meet the requirements of interactive mobile auditing operations, a gesture-driven intelligent evidence retrieval mechanism and an edge-cloud collaborative intelligent working paper generation system are constructed in this study. In response to auditors' interactive behavior of circling anomalous data, the system captures the spatiotemporal coordinate features of gestures and generates query vectors through a lightweight temporal network encoding. By means of a dual-encoder structure,

unified embedding representations of interaction features and audit evidence texts are respectively generated, and the relevance between them is quantified via cosine similarity, thereby achieving precise evidence matching:

$$sim(q,d) = \cos(q,d) = \frac{q \cdot d}{\|q\| \cdot \|d\|} \tag{10}$$

The model can output the Top-k most relevant vouchers and transaction records ranked by similarity, enabling rapid linkage and aggregation between anomalous points and audit evidence. To address the problems of high inference latency of cloud-based large models and low editing efficiency of working papers on mobile devices, a collaborative generation strategy combining edge caching and speculative decoding is constructed in this study. Standardized working paper templates for high-frequency audit scenarios are pre-stored at edge nodes, allowing routine text completion requests to be directly satisfied. For complex audit description scenarios, speculative decoding is preferentially performed at edge nodes to produce preliminary results, while the cloud simultaneously performs high-precision semantic correction and content optimization. Through this edge-cloud dual-link collaborative mode, both text generation quality and interactive real-time performance are balanced, effectively meeting the requirements of working paper compilation in mobile on-site auditing operations.

3 EXPERIMENTAL DESIGN AND RESULT ANALYSIS

3.1 Experimental environment and datasets

A physical experimental platform was built based on an end-edge-cloud hardware cluster. An Android 14 smart terminal was used as the mobile end device, an NVIDIA Jetson Orin was deployed as the edge node, and a storage and analysis environment for massive financial data was established on the cloud using the Spark distributed computing framework and the ClickHouse columnar database. The experimental datasets consisted of two parts: first, a desensitized financial audit dataset from a provincial-level power group spanning three consecutive years, containing a total of 5 million transaction records and covering 30 business dimensions and 12 types of measure indicators; second, the publicly available CEFS listed company fraud dataset, which was used to supplement fraudulent anomaly samples and to improve the validation of anomaly detection performance.

3.2 Experimental results of interactive multidimensional online analytical processing query performance on mobile devices

This set of experiments was conducted to validate the optimization effects of the interaction-intent prefetching and incremental visualization modules. The controlled variables included the dimension fan-out coefficient and four types of standard audit interaction operations. The observed metrics comprised average touch response latency, prefetch hit rate, and data volume transmitted per query. The experimental results are shown in Table 1.

An increase in the dimension fan-out coefficient resulted in more dimension branches and a larger volume of queried data. Consequently, latency and transmission overhead increased slightly for all compared schemes. The cloud-only scheme,

constrained by round-trip communication overhead throughout the entire process, exhibited the highest interaction latency across all experimental groups. The end-only and single-layer edge schemes relied solely on static caching and lacked user behavior prediction capability, achieving prefetch hit rates below 65%. By leveraging intent prediction-based prefetching combined with the incremental refresh mechanism, the proposed scheme consistently achieved prefetch hit rates above 85%. Compared with full queries, the data transmission volume was reduced by more than 80%, query latency was stably controlled within 0.4 seconds, and the bandwidth waste caused by full data retransmission was effectively avoided by the incremental rendering mechanism. The optimization effect remained stable even as dimension complexity increased.

Table 1. Performance comparison of multidimensional online analytical processing interactive queries

Fan-Out Coefficient	Scheme	Average Response Latency (s)	Prefetch Hit Rate (%)	Data Volume Per Query (KB)
Fan-out = 3	Cloud-only	1.81	–	1286
	End-only	0.92	52.3	452
	Single-layer edge	0.75	64.8	387
	Proposed framework	0.38	86.2	72
Fan-out = 5	Cloud-only	1.95	–	1542
	End-only	1.05	49.6	516
	Single-layer edge	0.83	61.5	421
	Proposed framework	0.40	85.1	89

3.3 Comparative experiments on anomaly detection accuracy and mobile terminal energy consumption in two-stage hierarchical inference

Two transaction sample batches of 5,000 and 10,000 records were selected for the experiments. The fixed-threshold strategy and the proposed proportional-integral-derivative-based dynamic threshold regulation scheme were compared. The evaluated metrics included anomaly detection F1-score, fraud miss rate, and mobile terminal energy consumption per batch. The experimental results are shown in Table 2.

The fixed-threshold scheme could not dynamically adjust the screening criteria in response to changes in the total sample size, leading to a continuously increasing miss rate as the sample size expanded. By leveraging the feedback of the anomaly proportion from the edge side, the proposed framework performed closed-loop parameter correction using a proportional-integral-derivative controller, thereby dynamically optimizing the end-side screening threshold. Under both sample size conditions, the F1-score was stably maintained at 0.91, and the miss rate was fixed at 4.2%. In terms of energy consumption, compared with the cloud-only scheme, which required continuous high-frequency data backhaul, the average reduction in mobile terminal power consumption was 18%. Compared with the end-only architecture, which performed full local computation, the reduction in energy consumption reached 32%. Through the hierarchical inference architecture, a large number of normal samples were filtered at the terminal, and only suspicious samples were uploaded to the high-precision edge model, thereby achieving mobile terminal power consumption control while maintaining audit detection accuracy.

Table 2. Comparison of anomaly detection accuracy and mobile terminal energy consumption

Sample Size	Scheme	F1-Score	Miss Rate (%)	Energy Consumption Per Batch (mJ)
5,000	Cloud-only	0.84	7.1	186
	End-only	0.86	6.3	92
	Fixed-threshold end-edge	0.88	5.5	113
	Proposed framework	0.91	4.2	102
10,000	Cloud-only	0.83	7.8	245
	End-only	0.85	6.9	165
	Fixed-threshold end-edge	0.89	5.1	184
	Proposed framework	0.91	4.2	158

3.4 Experimental results of adaptive ternary task offloading optimization under audit tidal workloads

The experiments were conducted under two levels of task dependency complexity (low and high), with two sets of concurrent task volumes (30 and 100) used to simulate audit off-peak and busy-peak workload conditions. The evaluated metrics included the task deadline violation rate and the average utilization rates of the three-tier hardware resources. The results are shown in Figure 3.

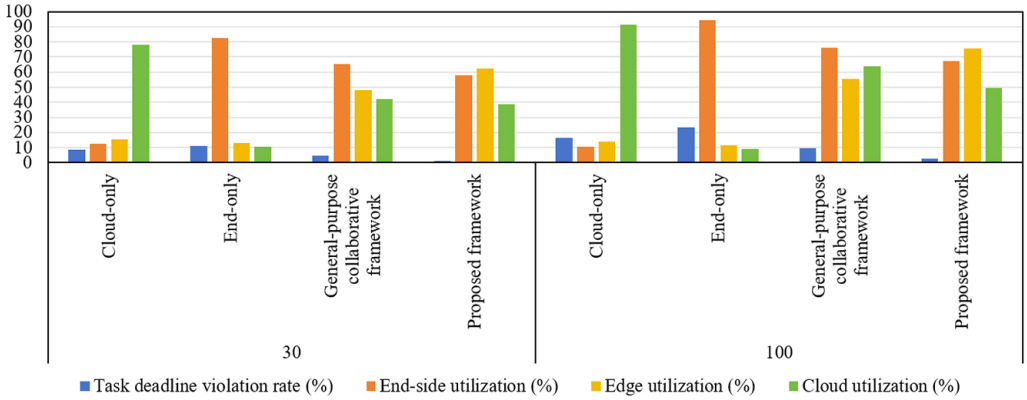

Fig. 3. Comparison of task scheduling performance under tidal workload scenarios

Traditional binary scheduling strategies tend to result in polarization of computational resources. In the cloud-only scheme, the workload is concentrated on cloud servers, while mobile terminal computational resources remain largely idle. In the end-only scheme, under high-concurrency tasks, the terminal becomes fully loaded, leading to widespread task deadline violations. General-purpose commercial scheduling does not account for task dependency topologies; cross-tier scheduling of highly coupled tasks introduces additional transmission latency and results in relatively high deadline violation rates. By adopting ternary offloading combined with dependency-aware clustering scheduling, the proposed framework packs dependent tasks and assigns them to the same computational tier, thereby reducing cross-node data interaction overhead. Under peak workload conditions, the task deadline violation rate

was only 2.8%, and the utilization ratios of end, edge, and cloud resources were well balanced, effectively mitigating the problems of resource idleness and localized overload caused by tidal workloads.

3.5 Experimental results of data synchronization performance under multi-level weak network conditions

Three network environments were configured for the experiments: a good network, a moderately weak network, and an extremely weak network. Link bandwidth and packet loss parameters were adjusted to reproduce network fluctuations in field auditing scenarios. The observed metrics included average synchronization time, transmitted data volume, and data consistency error rate. The results are shown in Table 3.

All baseline schemes adopted a fixed full-synchronization strategy, with the transmitted data volume not being adjusted according to network conditions. Under weak network scenarios, transmission latency and the probability of data conflicts increased rapidly. By dynamically switching synchronization strategies based on the quantified network quality metric, the proposed framework performed full synchronization under high-quality links, enabled incremental differential transmission under medium network conditions, and cached local change records under extreme weak network conditions, with batch merging performed via the conflict-free replicated data type after link recovery. Compared with the general-purpose collaborative framework, the data transmission volume was reduced by 69.1% under moderate weak network conditions and by 88.2% under extreme weak network conditions. Synchronization latency was significantly reduced, and the multi-terminal data consistency error rate was consistently controlled within 0.25%.

Table 3. Comparison of data synchronization metrics under differentiated network environments

Network Condition	Scheme	Average Synchronization Time (s)	Data Transmitted (MB)	Consistency Error Rate (%)
Good network	Cloud-only	2.15	12.6	0.12
	End-only	1.82	8.3	0.45
	General-purpose collaborative framework	1.46	6.8	0.18
	Proposed framework	1.31	6.2	0.10
Moderate weak network	Cloud-only	4.86	12.6	0.86
	End-only	3.52	8.3	1.24
	General-purpose collaborative framework	2.73	6.8	0.52
	Proposed framework	1.65	2.1	0.15
Extreme weak network	Cloud-only	9.72	12.6	2.35
	End-only	7.15	8.3	3.12
	General-purpose collaborative framework	5.38	6.8	1.46
	Proposed framework	2.42	0.8	0.21

3.6 Comprehensive end-to-end performance evaluation of mobile auditing

The complete on-site auditing business chain was fully reproduced in the experiments, including anomaly localization on mobile devices, risk warning at the edge, intelligent evidence retrieval, and assisted working paper generation. The time consumption of each segment and the total end-to-end latency were measured. The experimental results are shown in Figure 4.

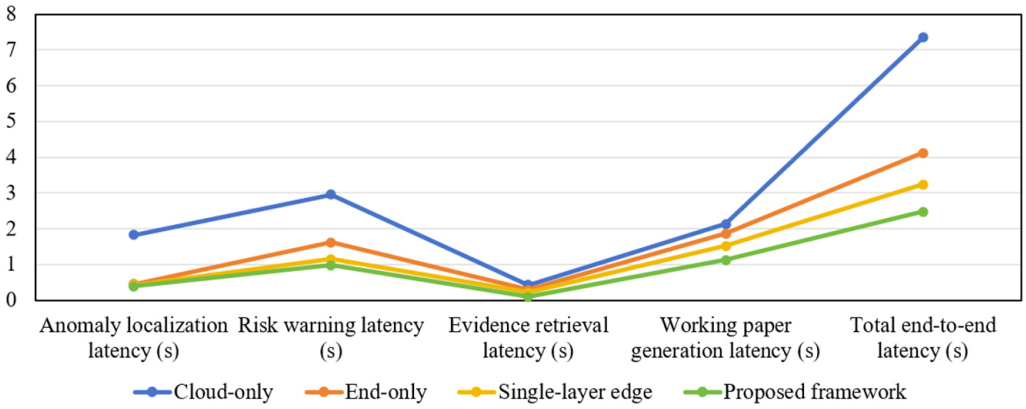

Fig. 4. Comparison of end-to-end latency across the full audit business chain

The end-to-end test results reflect the comprehensive benefits of synergistic optimization across all modules. The overall end-to-end audit latency achieved by the proposed scheme was 2.48 seconds, representing a 67% reduction compared with the cloud-only architecture. Anomaly localization time was shortened by end-side lightweight prescreening, while the deployment of retrieval and high-precision discrimination models at the edge avoided the overhead of long-distance cloud access. Working paper generation efficiency was further optimized through edge-cached templates combined with a speculative decoding mechanism. All baseline schemes, constrained by computational resource distribution and transmission architecture limitations, exhibited higher latency in individual segments or across the full chain compared with the proposed scheme, thereby validating the engineering adaptability of the complete framework for mobile auditing scenarios.

4 CONCLUSION AND FUTURE RESEARCH DIRECTIONS

An end-edge-cloud three-layer collaborative interactive mobile computing framework tailored for mobile financial field auditing was constructed in this study. Systematic mathematical modeling and technical implementation were carried out in the areas of mobile-end multidimensional query optimization, hierarchical intelligent inference, multi-constrained task scheduling, adaptive data synchronization, and scenario-specific audit algorithms. Through four key optimization schemes—intent-driven prefetching and incremental rendering, proportional-integral-derivative-based closed-loop threshold regulation, ternary dependency-aware offloading, and graded differential synchronization—the persistent problems of traditional mobile auditing, including interaction latency, high terminal power consumption, ineffective synchronization under weak network conditions, and insufficient anomaly detection accuracy, were systematically addressed. Multiple comparative experiments

and deployment data from the electric power industry demonstrated that the proposed framework significantly reduced the response latency of multidimensional analysis on mobile devices and lowered device power consumption. Performance improvements in anomaly detection and end-to-end audit tasks were achieved across all metrics compared with mainstream baseline schemes. Theoretically, the proposed framework complements the body of mobile edge collaborative computing for structured industry-specific big data; practically, it provides a mature and feasible technical paradigm for digital field auditing in corporate enterprises.

Building upon the existing research outcomes, future work will continue to deepen and expand the framework by incorporating cutting-edge mobile computing technologies. Through the adoption of federated learning architectures on mobile devices, iterative updates of audit models will be enabled under the constraint that raw financial data does not leave its local domain, thereby addressing the data privacy limitations of the current framework. Task offloading constraints will be further refined by leveraging the low-latency and high-capacity transmission characteristics of 5G-Advanced air interfaces, further enhancing the real-time responsiveness of dynamic scheduling. A multi-terminal distributed collaborative analysis mechanism will be developed for multi-user collaborative audit scenarios, improving the collaborative resource scheduling logic among clustered mobile devices and extending the framework from individual field auditing to large-scale team-based collaborative audit scenarios.

5 REFERENCES

[1] C. H. Mao, J. H. Liu, and Z. F. Zhou, "Assessment mechanisms as drivers of university teachers' digital teaching competence in the context of digital transformation," *Education Science and Management*, vol. 3, no. 2, pp. 106–116, 2025. https://doi.org/10.56578/esm030203

[2] C. B. Xuan, L. H. Viet, and P. V. Thang, "Digital transformation in Vietnamese SMEs: Challenges, opportunities, and strategic implications," *International Journal of Interactive Mobile Technologies*, vol. 20, no. 7, pp. 84–96, 2026. https://doi.org/10.3991/ijim.v20i07.61093

[3] Q. Mpofu, "Digital transformation in the accounting profession in Sub-Saharan Africa: Challenges, opportunities, and strategic pathways," *Journal of Accounting, Finance and Auditing Studies*, vol. 11, no. 4, pp. 222–230, 2025. https://doi.org/10.56578/jafas110403

[4] X. Liu, Y. Luo, X. Yang, L. Wang, and X. Zhang, "Lattice-based proxy-oriented public auditing scheme for electronic health record in cloud-assisted WBANs," *IEEE Systems Journal*, vol. 16, no. 2, pp. 2968–2978, 2022. https://doi.org/10.1109/JSYST.2021.3138861

[5] X. Meng, P. Zhu, and Y. Zhang, "Can fintech improve audit efficiency? Empirical evidence from audit fees and corporate governance," *Applied Economics Letters*, vol. 32, no. 13, pp. 1869–1875, 2025. https://doi.org/10.1080/13504851.2024.2331657

[6] D. J. Rapp and J. Pampel, "Digitization of the audit of financial statements through process mining: Principles, areas of application, and (alleged) goal conflict in the audit of the internal control system," *BFuP-Betriebswirtschaftliche Forschung und Praxis*, vol. 73, no. 5, pp. 506–518, 2021. [Online]. Available: https://hal.science/hal-03363251/

[7] L. Yang, "Predictive modeling of tax compliance risks: A comparative study of machine learning approaches," *PLoS ONE*, vol. 20, no. 9, p. e0331715, 2025. https://doi.org/10.1371/journal.pone.0331715

[8] A. Dwekat, E. Seguí-Mas, G. Tormo-Carbó, and P. Carmona, "Corporate governance configurations and corporate social responsibility disclosure: Qualitative comparative analysis of audit committee and board characteristics," *Corporate Social Responsibility and Environmental Management*, vol. 27, no. 6, pp. 2879–2892, 2020. https://doi.org/10.1002/csr.2009

[9] V. Munteanu, M. R. Zuca, A. Horaicu, L. A. Florea, C. E. Poenaru, and G. Anghel, "Auditing the risk of financial fraud using the red flags technique," *Applied Sciences*, vol. 14, no. 2, p. 757, 2024. https://doi.org/10.3390/app14020757

[10] Y. S. Chen and Z. J. Wu, "Financial fraud detection of listed companies in China: A machine learning approach," *Sustainability*, vol. 15, no. 1, p. 105, 2023. https://doi.org/10.3390/su15010105

[11] L. Y. Liu, "Financial statement audits and data breaches," *Management Science*, vol. 71, no. 8, pp. 6340–6366, 2025. https://doi.org/10.1287/mnsc.2023.01357

[12] B. Z. Péter and I. Kocsis, "Privacy-preserving noninteractive compliance audits of blockchain ledgers with zero-knowledge proofs," *Acta Polytechnica Hungarica*, vol. 21, no. 11, pp. 7–27, 2024. https://doi.org/10.12700/aph.21.11.2024.11.1

[13] A. Shahim, "Security of the digital transformation," *Computers & Security*, vol. 108, p. 102345, 2021. https://doi.org/10.1016/j.cose.2021.102345

[14] C. Daah, A. Qureshi, I. Awan, and S. Konur, "Simulation-based evaluation of advanced threat detection and response in financial industry networks using zero trust and blockchain technology," *Simulation Modelling Practice and Theory*, vol. 138, p. 103027, 2025. https://doi.org/10.1016/j.simpat.2024.103027

6 AUTHOR

Hongling Zhang holds a Bachelor of Economics degree from Henan University of Economics and Law and currently works at Shangqiu Polytechnic. Her primary research interest is Economics (E-mail: sqzyshenji@163.com).