Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 11 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i11.62126

PAPER

Collaborative Optimization of Human–Computer Interaction Efficiency and Cognitive Load in Mobile Auditing

Hongling Zhang(✉)

Shangqiu Polytechnic, Shangqiu, China

sqzyshenji@163.com

ABSTRACT

Mobile auditing has been increasingly recognized as a critical direction in the digital transformation of auditing practices. However, field auditing scenarios are constrained by limited device resources, sensitive data privacy requirements, unstable network conditions, and elevated cognitive load among auditors, all of which substantially hinder human–computer interaction efficiency and audit quality. To address these challenges, a multi-module collaborative optimization framework for human–computer interaction was proposed. Four core technologies were integrated into the framework: non-intrusive cognitive load quantification, lightweight on-device cognitive inference, dynamic user interface information density reconfiguration, and adaptive computation offloading under weak network conditions. Through this integration, end-to-end coordination was achieved, encompassing cognitive state awareness, adaptive interface adjustment, and computational task optimization. To validate the effectiveness of the proposed framework, controlled dual-task experiments were conducted, simulating both static and dynamic interference conditions commonly encountered in real-world field auditing. Performance comparisons between the optimized system and a baseline system demonstrated that cognitive load was significantly reduced, while interaction efficiency and task accuracy were markedly improved. Furthermore, stable system performance was maintained under dynamic interference conditions, alongside lightweight deployment and enhanced privacy preservation capabilities. The proposed approach provides a practical technical pathway for optimizing human–computer interaction in mobile professional productivity tools, enriches interdisciplinary research at the intersection of mobile computing and cognitive engineering, and offers substantial academic and engineering value.

KEYWORDS

mobile computing, audit human–computer interaction, cognitive load awareness, on-device edge inference, dynamic user interface reconfiguration

Zhang, H. (2026). Collaborative Optimization of Human–Computer Interaction Efficiency and Cognitive Load in Mobile Auditing. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(11), pp. 89–102. https://doi.org/10.3991/ijim.v20i11.62126

Article submitted 2026-02-24. Revision uploaded 2026-04-20. Final acceptance 2026-04-27.

1 INTRODUCTION

With the continuous advancement of audit informatization [1, 2], mobile auditing has emerged as the dominant mode of field auditing operations [3]. While spatial constraints have been effectively alleviated, mobile auditing environments remain subject to a combination of device, task, and environmental constraints. The limited screen size of mobile terminals [4] and their constrained computational resources [5] are insufficient to support the high-density data inspection requirements inherent in auditing tasks. Meanwhile, the sensitive nature of audit data imposes stringent requirements on data security and privacy protection during processing [6, 7]. In addition, unstable network conditions in field environments further exacerbate the challenges associated with working paper loading and data synchronization [8]. As a consequence of these constraints, the direct transplantation of desktop-based auditing tools to mobile platforms has resulted in reduced human–computer interaction efficiency, excessive cognitive load, and heightened privacy risks, thereby impeding the large-scale deployment of mobile auditing systems. Existing studies on mobile human–computer interaction optimization have predominantly focused on general consumer scenarios, with limited attention given to domain-specific requirements in auditing. Moreover, end-to-end collaboration across cognitive load awareness, interaction adaptation, and computational optimization has not been achieved, revealing a clear research gap.

Significant limitations remain in current approaches when applied to mobile auditing scenarios. At the level of cognitive load awareness, existing methods largely rely on subjective rating scales or intrusive wearable devices [9, 10], both of which impose operational burdens and lack real-time responsiveness. Non-intrusive quantification approaches that can be executed in real time on-device, without requiring additional hardware, remain underexplored, thus failing to meet the lightweight requirements of field auditing. At the user interface adaptation level, conventional mobile auditing applications have predominantly adopted scaled-down versions of desktop interfaces [11, 12]. Such approaches fail to dynamically adjust information density based on device characteristics, user cognitive state, and handheld interaction posture. Consequently, issues such as information overload, high mis-touch rates, and difficulty in information localization are frequently observed. At the level of computation and synchronization, full data transmission and local processing strategies under weak network conditions consume substantial bandwidth and may block the main processing thread, thereby increasing cognitive load. Existing task offloading and synchronization mechanisms have not achieved coordinated integration of semantic-level querying, incremental synchronization, and data consistency maintenance [13, 14]. Finally, in terms of validation methodologies, most existing studies have relied on single static tasks and subjective evaluations [15, 16], without adequately simulating complex real-world conditions such as dynamic interference and network fluctuations in field auditing. Data collection has remained limited in scope, making it difficult to quantitatively evaluate the effectiveness of collaborative optimization strategies and consequently reducing the practical applicability of reported findings.

To address the constraints inherent in mobile auditing scenarios and the limitations of existing studies, a human–computer interaction collaborative optimization framework is developed, with joint consideration of privacy preservation and lightweight deployment. End-to-end coordination across cognitive load awareness, user interface adaptation, and computational offloading is achieved, enabling the reduction of auditors' cognitive load while enhancing interaction efficiency and task accuracy, thereby supporting the scalable deployment of mobile auditing systems.

The primary academic contributions are summarized below. First, a customized non-intrusive cognitive load quantification method for mobile auditing scenarios is proposed. Multi-dimensional features are captured and quantified in real time using native sensors on mobile devices, eliminating the need for additional wearable equipment. As a result, lightweight cognitive load awareness and privacy-preserving data acquisition in field auditing contexts are effectively achieved. Second, a lightweight on-device cognitive state inference engine is designed. Through model optimization and quantization-based compression, real-time classification of cognitive states is enabled, overcoming constraints related to limited computational resources and latency on mobile terminals. Third, a state machine–driven dynamic user interface information density reconfiguration algorithm is introduced. By jointly incorporating cognitive state and handheld interaction posture, adaptive interface adjustment is achieved, addressing the limitations associated with direct desktop interface transplantation and filling the gap in coordinated optimization between cognitive awareness and user interface adaptation in domain-specific applications. Fourth, a computation offloading and incremental synchronization mechanism for weak network conditions is developed. By integrating on-device database optimization with semantic-level data transmission, the conflict between constrained computational resources and limited network bandwidth is mitigated, thereby improving interaction fluency. Fifth, a multimodal controlled experimental framework is established to simulate realistic field auditing scenarios. The effectiveness of the proposed optimization framework is quantitatively validated, providing an experimental paradigm and benchmarking approach for the optimization of mobile professional productivity tools.

2 METHODOLOGY

The overall architecture of the proposed collaborative optimization framework is illustrated in Figure 1. The system is hierarchically structured into four layers: the perception layer, the inference layer, the adaptation layer, and the coordination layer. Through the closed-loop integration of on-device multimodal feature acquisition, lightweight cognitive state inference, dynamic user interface information density reconfiguration, and computation offloading under weak network conditions, coordinated optimization of human–computer interaction efficiency and cognitive load is achieved. This is accomplished while ensuring the privacy and security of audit data throughout the processing pipeline.

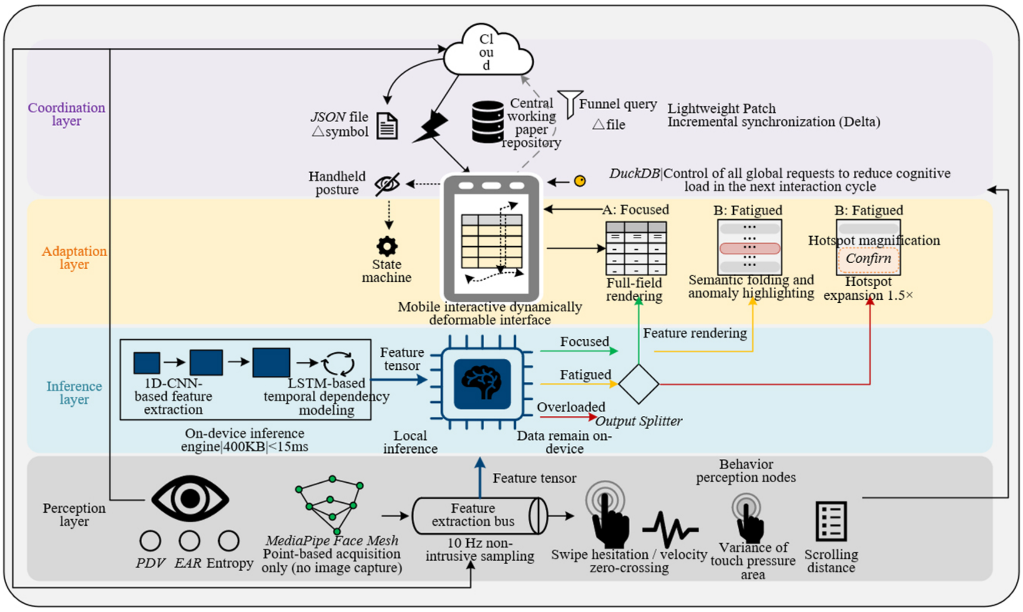

Fig. 1. Overall architecture of the collaborative optimization framework for human–computer interaction in mobile auditing

2.1 Non-intrusive cognitive load quantification method in mobile auditing scenarios

To address the requirements of high-density data inspection and real-time cognitive assessment in mobile auditing tasks, a non-intrusive feature acquisition framework is established based on native front-facing red, green, and blue (RGB) cameras and touch sensors of mobile devices. Cognitive load is quantitatively characterized from both behavioral and physiological dimensions. At the behavioral level, four quantitative indicators are defined: effective operation path length, information retrieval scrolling distance, swipe hesitation, and variance of touch contact area. Swipe hesitation is directly quantified by the frequency of velocity zero-crossing points within a single drag gesture, reflecting the degree of decision-making hesitation. The variance of touch contact area is employed to capture the dispersion of touch regions caused by micro-hand tremors. At the physiological level, key ocular features are extracted. The pupil diameter variability is computed as the ratio of the standard deviation to the mean of time-series pupil diameters, expressed as PDV = σ_D/μ_D. The eye aspect ratio is constructed based on the spatial coordinates of eye landmarks and is expressed as EAR = $(y_2 - y_6 + y_3 - y_5)/(2(x_4 - x_1))$. In addition, the spatial entropy of pupil movement trajectories is utilized to quantify the degree of disorder in gaze distribution over high-density data regions. All extracted features are normalized using min–max normalization to ensure dimensional consistency, with the normalization formula defined as $\hat{x} = (x - x_{min})/(x_{max} - x_{min})$. The processed multidimensional features are subsequently fused into a fixed-dimensional input tensor, which is utilized for downstream cognitive state inference.

The proposed cognitive load quantification method adopts an on-device localized feature extraction paradigm. During the data acquisition stage, only raw sensor signals and human key point coordinate information are parsed, while no raw image pixel data from auditing scenarios are stored or transmitted. The data acquisition process is executed at a fixed sampling frequency of 10 Hz. All feature computation and data processing tasks are performed within background threads on the mobile device, thereby avoiding resource contention with the application's user interface main thread. The feature acquisition and computation pipeline operates without reliance on any external wearable devices. Data collection is achieved solely through the device's built-in image capture module and touch sensing module. The generation and handling of feature tensors are completed entirely on-device, without involving any cloud-based data interaction. All computational logic is integrated within the underlying perception module of the mobile application and is executed in a decoupled and parallel manner relative to the auditing workflow.

Figure 2 summarizes the core elements of the experimental design. A 2 × 2 mixed experimental design is adopted, in which the system type is treated as a within-subject variable, and the level of environmental interference is treated as a between-subject variable. Three categories of metrics are synchronously collected, including interaction efficiency, cognitive load, and system performance.

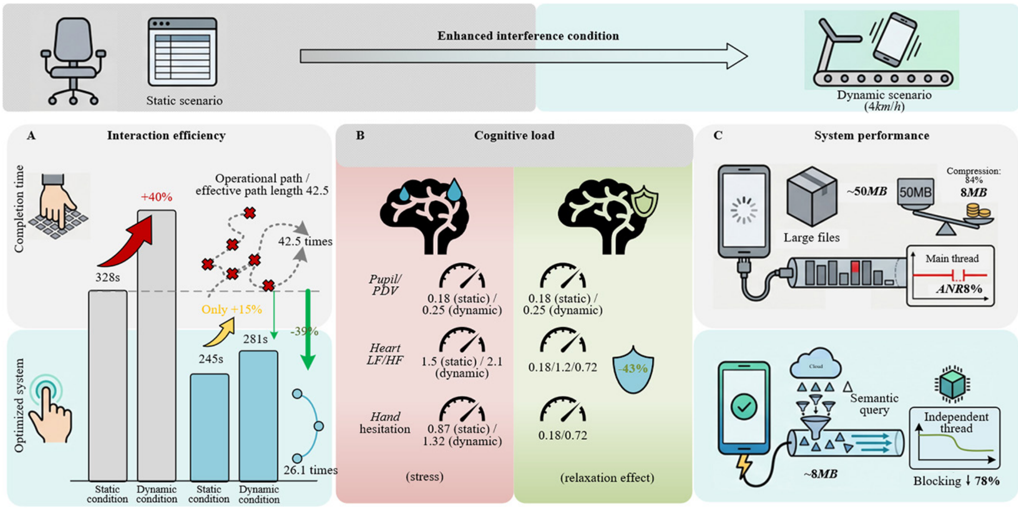

Fig. 2. Dual-task controlled experimental design and multimodal data acquisition framework

2.2 Design of the lightweight on-device cognitive state inference engine

To satisfy the privacy constraints requiring fully localized processing of audit data, a lightweight on-device cognitive state inference engine is constructed on mobile terminals. All feature parsing and model inference processes are executed locally, without any data interaction with public cloud interfaces. The inference engine adopts a compact hybrid architecture combining one-dimensional convolutional neural networks and long short-term memory networks. The one-dimensional convolutional layer performs local receptive field operations on the input temporal feature tensor, enabling the extraction of short-term variation patterns and correlations among features such as pupil diameter and touch interaction behaviors. The convolutional output can be expressed as:

$$x_{conv}[t] = f\left(\sum_{k=1}^{K} w_k \cdot x_{feat}[t-k+1]+b\right) \tag{1}$$

where, w_k denotes the convolution kernel weights, x_{feat} represents the normalized input features, K is the receptive field length, and f denotes a nonlinear activation function. The long short-term memory layer receives the feature sequence generated by the convolutional operation and captures the temporal dependencies of cognitive states as the task progresses through a gating mechanism. While preserving temporal representation capability, only the fundamental gating structure is retained to reduce the number of parameters. The overall network architecture is intentionally constrained to a single convolutional branch and a single long short-term memory unit, thereby avoiding excessive computational overhead associated with deep network designs.

The inference engine employs 8-bit integer quantization to achieve model compression and accelerated inference. Within the TensorFlow Lite framework, floating-point parameters are mapped into an integer numerical space, where the quantization mapping relationship is defined as follows:

$$q = round\left(\frac{x}{s} + z\right) \tag{2}$$

where, x denotes the floating-point feature, s represents the quantization scale, z is the zero-point offset, and q corresponds to the quantized integer value. Following quantization, the model size is constrained to within 400 KB. The inference pipeline is further optimized at the underlying level using the Android neural networks application programming interface, ensuring that the latency of a single forward inference pass does not exceed 15 ms. As a result, stable operation can be achieved on mobile devices equipped with mid- to low-power processors.

During the inference stage, discretized three-level cognitive states are output, corresponding to adaptive interface control commands, rather than continuous cognitive load values. A confidence-based decision rule is incorporated, whereby a valid state output is generated only when the model confidence is no less than 0.85. Otherwise, the previously stable state is retained to reduce the probability of misclassification. All inference processes are executed within an independent computation thread on the mobile device, operating asynchronously with respect to the application's main thread. Consequently, interface rendering and business operations remain unaffected, ensuring both real-time cognitive state estimation and overall system stability.

2.3 State machine–driven dynamic user interface information density reconfiguration algorithm

A state machine–driven dynamic user interface information density reconfiguration algorithm is developed, in which cognitive state and handheld posture are jointly utilized as inputs. Through the use of a finite-state automaton, precise triggering and dynamic adjustment of reconfiguration strategies are achieved, addressing issues of information overload and mis-touch that arise from the direct transplantation of desktop auditing interfaces to mobile platforms. The core architecture of the algorithm is formulated as a finite-state automaton, whose state space is defined as the combination of cognitive states and handheld posture conditions. Input variables include the cognitive state labels generated by the on-device inference engine and the handheld posture data acquired from the device gyroscope. Posture confidence is computed based on the variance of angular velocity across three axes, expressed as:

$$C = exp\left(-\frac{\sigma_{\omega}^{2}}{\sigma_{0}^{2}}\right) \tag{3}$$

where, σ_{ω}^2 denotes the real-time variance of the tri-axial angular velocity, and σ_0^2 represents a predefined posture stability threshold. The confidence value C, within a range of [0, 1], is utilized to distinguish between different posture scenarios, such as single-hand operation and dual-hand landscape usage. State transitions within the automaton are governed by a transition function that enables seamless switching between reconfiguration strategies under varying conditions. The transition logic is determined through the joint evaluation of cognitive state and posture confidence, thereby ensuring both the specificity and real-time responsiveness of user interface adaptation.

The reconfiguration strategy adopts a state-specific precise adaptation mechanism, in which corresponding strategies are triggered based on the combined outputs of cognitive state and posture confidence. In the focused state, the user interface maintains the original information density, and full-field rendering of detailed audit ledger tables is preserved to support precise data inspection requirements. In the fatigued state, a semantic folding strategy is activated. A lightweight rule-based summarizer is employed to hierarchically compress audit account labels, while anomalous data rows are selectively highlighted. Anomaly detection follows the criterion $|x - \mu| > 2\sigma$, where x represents the value of the current data row, μ denotes the historical mean of the corresponding account, and σ is the standard deviation. For non-anomalous rows, font opacity is reduced to 40%, thereby decreasing visual information load. In the cognitive overload state, a hotspot self-expansion protection mechanism is triggered. Based on a heatmap of touch interaction points over the most recent 5 seconds, high-frequency interaction regions are identified, and their responsive areas are expanded by a factor of 1.5. Simultaneously, the scrolling friction coefficient of the list is adjusted, and content transition intensity is reduced by lowering scroll damping, thereby alleviating cognitive overload.

All decision-making logic of the algorithm is deployed within an independent asynchronous thread, ensuring resource isolation from the user interface main thread and preventing competition for rendering resources. The Android Choreographer mechanism is utilized to synchronize vertical refresh signals, aligning user interface reconfiguration operations with the screen refresh cycle. As a result, a stable frame rate exceeding 60 frames per second is maintained, effectively preventing frame drops. Reconfiguration commands are implemented using an incremental update strategy, in which only the user interface components requiring modification are selectively refreshed, rather than performing full interface redraws. This approach reduces system resource consumption and ensures smooth auditing operations during dynamic reconfiguration. Consequently, uninterrupted cognitive flow is preserved for auditors, enabling coordinated optimization of cognitive load and interaction efficiency.

2.4 Auditing computation offloading and incremental synchronization mechanisms under weak network conditions

An auditing computation offloading and incremental synchronization mechanism based on reactive synchronization is developed to address issues such as working paper loading latency and main thread blocking in mobile auditing under weak network conditions through task redistribution and fine-grained data transmission, achieving efficient coordination between on-device processing and backend systems. An on-device instance of DuckDB compiled via WebAssembly is deployed to replace the traditional SQLite database, thereby improving the performance of complex aggregation queries. The WebAssembly instance is adapted to mobile central processing unit architectures through bytecode compilation and can be executed directly in both browser environments and native applications without additional dependency libraries. Compared with SQLite, query execution efficiency is improved by more than threefold, enabling efficient handling of complex auditing queries, including multi-table joins and multi-year comparative analyses. A semantic-level computation offloading strategy is adopted, based on a lightweight "query instruction–result feedback" interaction paradigm. When an audit query is initiated on the mobile device, only semantic instructions containing query conditions and required fields are transmitted to the backend, rather than the complete audit working papers. After performing full data association and computation, the backend returns only the first-screen view metadata and incremental update packets to the device. Incremental updates are encoded using a delta encoding scheme, with the core formulation defined as:

$$\Delta X = X_{new} - X_{old} \tag{4}$$

where, X_{new} represents the updated dataset at the backend, and X_{old} denotes the cached historical data on the device. Through delta encoding, only the changed data are transmitted, and the compression ratio of data transmission satisfies the expression below, thereby substantially reducing bandwidth consumption under weak network conditions.

$$R = \left(1 - \frac{|\Delta X|}{|X_{old}|}\right) \times 100\% \geq 80\% \tag{5}$$

User operation synchronization under weak network conditions is implemented using lightweight JSON Patch instructions. Each instruction contains only the operation type, data row identifier, and timestamp, with the size of a single instruction constrained to within 1 KB. Best-effort transmission is achieved through adaptive switching between WebSocket and Message Queuing Telemetry Transport protocols. The protocol switching logic is determined based on the real-time packet loss rate. The packet loss rate ρ is calculated from the reception status of the most recent 10 data packets as $\rho = 1010 - N_{recv}$, where N_{recv} denotes the number of successfully received packets. When $\rho > 0.3$, the communication protocol is automatically switched to Message Queuing Telemetry Transport, leveraging its publish–subscribe model to enhance transmission stability under weak network conditions. Upon network recovery, a Conflict-Free Replicated Data Type-based algorithm is employed to achieve eventual consistency between on-device data and the central working paper repository. Conflict resolution follows a timestamp-priority rule: for conflicting operations on the same data item, the operation with the larger timestamp t overrides the one with the smaller timestamp. This process can be expressed as:

$$O_{final} = \begin{cases} O_1 & t_1 > t_2 \\ O_2 & t_1 \leq t_2 \end{cases} \tag{6}$$

where, O_1 and O_2 denote two operations on the same data item, and t_1 and t_2 represent their respective timestamps. Through this mechanism, the integrity of audit operation records and the consistency of data are effectively ensured.

Main thread protection is achieved through asynchronous task scheduling and a multi-level caching mechanism. All computation offloading and data synchronization tasks are executed in independent background threads, ensuring resource isolation from the user interface main thread. A task priority queue management strategy is adopted, in which audit business operations are assigned higher priority than computation and synchronization tasks, thereby preventing time-consuming processes from occupying main thread resources. A multi-level caching architecture is deployed on the device side, where frequently accessed audit details and first-screen metadata are cached locally. A Least Recently Used policy is employed for cache eviction. The cache hit rate H is calculated below, ensuring that it is maintained above 75% to reduce redundant network requests and computational overhead:

$$H = \frac{N_{hit}}{N_{hit} + N_{miss}} \times 100\% \tag{7}$$

In addition, cached data are updated using a scheduled incremental update strategy, with the update interval dynamically adjusted according to network conditions. This approach effectively mitigates loading latency under weak network environments, maintains smooth audit interaction, and indirectly reduces the cognitive load experienced by auditors due to system delays.

3 EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

To evaluate the effectiveness of the proposed collaborative optimization framework for human–computer interaction in mobile auditing, controlled experiments were conducted under conditions designed to closely replicate real-world field auditing scenarios. Through multimodal data acquisition and statistical analysis, the performance improvements of the proposed approach were quantitatively assessed across three dimensions: interaction efficiency, cognitive load, and system performance. These evaluations provide robust empirical support for both the practical deployment and academic validation of the proposed framework.

3.1 Experimental design

A dual-task controlled laboratory experiment with a 2 × 2 mixed design was conducted to simulate complex real-world field auditing scenarios. The within-subject variable was the system type, which included a baseline version and an optimized version. The baseline system refers to a mobile auditing application without any optimization modules, whereas the optimized system integrates four core components: non-intrusive cognitive load quantification, lightweight on-device cognitive state inference, dynamic user interface information density reconfiguration, and computation offloading with incremental synchronization under weak network conditions. The between-subjects variable was the level of environmental interference, which consisted of two conditions: a static, seated scenario and a dynamic interference scenario. The dynamic condition was simulated using a treadmill operating at a constant speed of 4 km/h to replicate body motion disturbances encountered during field inventory tasks. A total of 20 participants with accounting backgrounds and practical auditing experience were recruited. All participants were screened to exclude visual or motor impairments, ensuring the validity of the experimental data.

Each participant was required to complete a standardized audit evidence collection task under both environmental conditions. Specifically, three groups of fictitious duplicate payment clues were identified from a dataset containing 200 procurement records with embedded noise. The task complexity was designed to closely match real-world field auditing data inspection scenarios, thereby ensuring the transferability of the results.

A multimodal data acquisition approach was employed to comprehensively evaluate three categories of key performance metrics: interaction efficiency, cognitive load, and system performance. Interaction efficiency metrics included task completion time, number of erroneous touches, effective operation path length, and information retrieval scrolling distance. Cognitive load was assessed using pupil diameter variability, heart rate variability expressed as the low-frequency to high-frequency ratio, and swipe hesitation. The low-frequency to high-frequency ratio is used to quantify sympathetic nervous system activity and indirectly reflect cognitive load levels. System performance metrics included model inference latency, data transmission volume, main thread blocking duration, and the incidence rate of application-not-responding events. All data were collected using standardized mobile devices configured with the Android 13 operating system and a Qualcomm Snapdragon 870 processor to ensure computational consistency. Heart rate variability was measured using a Polar H10 chest strap, ocular features were captured via the device's front-facing camera, and behavioral as well as system performance data were recorded in real time through embedded application logging mechanisms. All data streams were sampled at a uniform frequency of 10 Hz to ensure temporal synchronization and measurement accuracy.

3.2 Results analysis

Repeated-measures analysis of variance was employed to statistically evaluate the experimental data. The main effects of system type and environmental interference level, as well as their interaction effects, were examined. The significance level was set at $\alpha = 0.05$, and differences are considered statistically significant when $P < 0.05$. Through this approach, the effectiveness of the proposed optimization framework was quantitatively validated, ensuring the scientific rigor and reliability of the experimental findings.

Following statistical analysis, the results for the three categories of metrics—interaction efficiency, cognitive load, and system performance—are summarized in Tables 1, 2, and 3, respectively. Across all metrics, the optimized system demonstrates significant advantages over the baseline system, with more pronounced improvements observed under dynamic interference conditions.

Table 1. Statistical results of interaction efficiency metrics (mean ± standard deviation)

Metric	Baseline (Static)	Optimized (Static)	Baseline (Dynamic)	Optimized (Dynamic)	P-Value
Task completion time (s)	328.6 ± 27.5	245.3 ± 21.8	460.1 ± 35.2	281.5 ± 25.7	<0.05
Error clicks	8.7 ± 2.3	3.3 ± 1.5	15.2 ± 3.1	5.7 ± 1.8	<0.05
Effective operation path length	42.5 ± 6.8	26.1 ± 5.3	61.3 ± 8.2	37.8 ± 6.5	<0.05
Information retrieval scrolling distance (px)	8960 ± 1250	5520 ± 980	12840 ± 1560	7230 ± 1120	<0.05

The results in Table 1 indicate that, under static conditions, the optimized system significantly reduces task completion time and the number of error clicks. The effective operation path length and information retrieval scrolling distance are reduced by 38.6% and 38.4%, respectively, compared with the baseline system. Under dynamic interference conditions, the task completion time of the baseline system increases by 40.0% relative to the static condition, whereas the optimized system exhibits an increase of only 14.8%. In addition, error clicks, effective operation path length, and information retrieval scrolling distance are reduced by 62.5%, 38.3%, and 43.7%, respectively. These results demonstrate that the optimized system effectively mitigates the impact of dynamic interference and significantly enhances interaction efficiency in mobile auditing scenarios.

The results in Table 2 indicate that cognitive load is significantly reduced by the optimized system under both static and dynamic interference conditions. Under static conditions, pupil diameter variability, the low-frequency to high-frequency ratio, and swipe hesitation are reduced by 27.8%, 33.3%, and 44.8%, respectively, compared with the baseline system. Under dynamic interference conditions, all cognitive load metrics of the baseline system increase substantially. In contrast, the optimized system achieves reductions of 28.0% in pupil diameter variability, 42.9% in the low-frequency to high-frequency ratio, and 45.5% in swipe hesitation relative to the baseline system. Moreover, the low-frequency to high-frequency ratio is maintained below 1.2. These findings demonstrate that the optimized system effectively alleviates cognitive pressure induced by information overload and dynamic interference, thereby maintaining a high-efficiency cognitive state for auditors during mobile auditing tasks.

Table 2. Statistical results of cognitive load metrics (mean ± standard deviation)

Metric	Baseline (Static)	Optimized (Static)	Baseline (Dynamic)	Optimized (Dynamic)	P-Value
Pupil diameter variability	0.18 ± 0.04	0.13 ± 0.03	0.25 ± 0.05	0.18 ± 0.04	<0.05
Heart rate variability (low-frequency to high-frequency ratio)	1.5 ± 0.4	1.0 ± 0.3	2.1 ± 0.5	1.2 ± 0.3	<0.05
Swipe hesitation (count/s)	0.87 ± 0.15	0.48 ± 0.11	1.32 ± 0.21	0.72 ± 0.14	<0.05

Table 3. Statistical results of system performance metrics (mean ± standard deviation)

Metric	Baseline (Static)	Optimized (Static)	Baseline (Dynamic)	Optimized (Dynamic)	P-Value
Model inference latency (ms)	–	12.3 ± 2.1	–	13.5 ± 2.4	<0.05
Data transmission volume (MB)	48.6 ± 5.2	7.9 ± 1.3	52.3 ± 6.1	8.5 ± 1.5	<0.05
Main thread blocking duration (ms)	286 ± 45	62 ± 18	423 ± 58	95 ± 22	<0.05
Application not responding occurrence rate (%)	8.0 ± 3.5	0.0 ± 0.0	15.0 ± 4.2	0.0 ± 0.0	<0.05

The results in Table 3 indicate that superior performance is consistently achieved by the optimized system across all evaluated metrics. Model inference latency is maintained within the range of 10–15 ms, satisfying the requirements for real-time cognitive state awareness and user interface adaptation. Data transmission volume is reduced by 83.7% compared with the baseline system, effectively alleviating bandwidth pressure under weak network conditions. In addition, the main thread blocking duration is reduced by 78.3%, and no application-not-responding events are observed in the optimized system. These findings demonstrate that, through computation offloading and asynchronous task scheduling, system stability and operational fluency are effectively ensured. Consequently, a robust system-level foundation is provided for enhancing interaction efficiency and reducing cognitive load in mobile auditing scenarios.

Figure 3 presents a comprehensive visualization of the distribution characteristics and optimization effects of three cognitive load metrics—pupil diameter variability, heart rate variability (low-frequency to high-frequency ratio), and swipe hesitation—across four experimental conditions: baseline versus optimized systems and static versus dynamic scenarios. A normalized heatmap representation is employed, in which the raw mean values of each metric under the four conditions are mapped onto the range of 0–1. A gradient color scale ranging from dark green (0–0.2) to dark red (0.8–1.0) is used to indicate relative levels of cognitive load. The heatmap reveals that, under the baseline dynamic condition, the normalized values of all three metrics reach or approach 1.00 (pupil diameter variability = 1.00, the low-frequency to high-frequency ratio ≈ 1.00, and swipe hesitation ≈ 1.00), with corresponding color blocks appearing in dark red. This indicates that dynamic interference significantly increases visual strain, sympathetic nervous system activity, and manual operation hesitation among auditors. In contrast, under the optimized dynamic condition, the normalized values decrease to 0.42, 0.18, and 0.29 for pupil diameter variability, the low-frequency to high-frequency ratio, and swipe hesitation, respectively. The corresponding color transitions from dark red to light green or light orange. Notably, the cognitive load levels under the optimized dynamic condition approach, or even surpass, those observed under the baseline static condition (normalized values of 0.42, 0.45, and 0.46, respectively). These findings provide strong evidence that the proposed collaborative optimization framework effectively suppresses multidimensional cognitive load under dynamic interference conditions. Furthermore, the observed variation in reduction magnitudes suggests that swipe hesitation exhibits the highest sensitivity to changes in user interface information density.

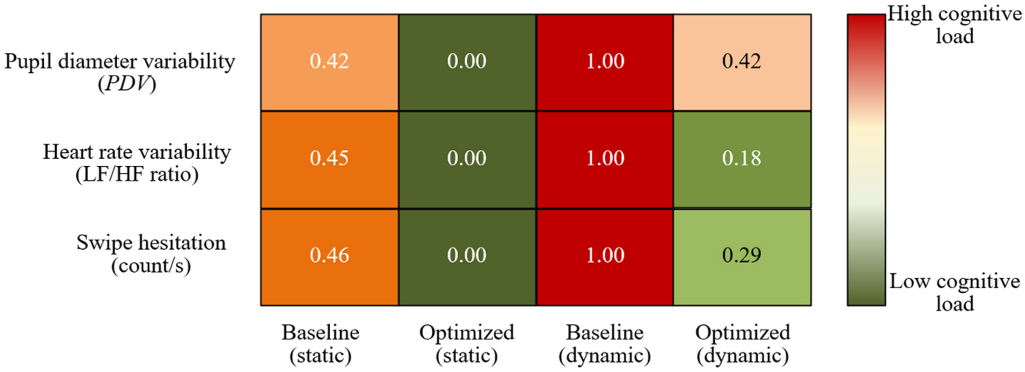

Fig. 3. Normalized heatmap of cognitive load metrics

4 CONCLUSION

To address the key challenges encountered in mobile auditing field scenarios, including device constraints, data privacy sensitivity, network instability, and elevated cognitive load, a collaborative optimization framework for human–computer interaction in mobile auditing environments was established. Through the coordinated integration of four core technologies, precise joint optimization of interaction efficiency and cognitive load was achieved, and the predefined research objectives were fully realized. The proposed non-intrusive cognitive load quantification method for mobile auditing scenarios enabled fine-grained acquisition and standardized fusion of physiological and behavioral features using native device sensors. Real-time cognitive load quantification was achieved without the need for additional hardware, effectively addressing the dual requirements of lightweight deployment and privacy preservation in field auditing contexts. The lightweight on-device cognitive state inference engine, enabled by architectural optimization and 8-bit integer quantization, overcame the limitations of computational resources and latency on mobile devices. Accurate classification and real-time output of cognitive states were achieved, providing reliable decision support for adaptive user interface reconfiguration. The state machine–driven dynamic user interface information density reconfiguration algorithm enabled adaptive interface adjustment by jointly incorporating cognitive state and handheld posture. Issues of information overload and mis-touch associated with the direct transplantation of desktop interfaces were effectively mitigated, thereby enhancing interaction precision and fluency. The audit computation offloading and incremental synchronization mechanism under weak network conditions significantly reduced bandwidth consumption and main thread blocking through semantic-level computation transfer, delta encoding–based data transmission, and Conflict-Free Replicated Data Type-based consistency merging. As a result, system stability was ensured under weak network and dynamic interference scenarios.

The proposed framework advances the interdisciplinary field of mobile human–computer interaction and domain-specific auditing applications. The full-chain collaborative optimization paradigm integrating cognitive load awareness, user interface adaptation, and computation offloading enriches existing research on high cognitive load tasks in mobile environments and provides a novel technical pathway and experimental benchmark for the design of mobile professional productivity tools. Furthermore, the proposed approach demonstrates strong engineering applicability and can be directly integrated into existing mobile auditing applications without requiring large-scale hardware upgrades. This capability is expected to facilitate the digital and mobile transformation of auditing practices and to provide valuable insights for the broader application of mobile technologies in professional domains.

5 REFERENCES

[1] E. Liu, "Development and evaluation of a blockchain-based financial audit tracking system for mobile platforms," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 3, pp. 73–86, 2025. https://doi.org/10.3991/ijim.v19i03.53947

[2] O. Senturk, "AI-driven and data-intensive auditing: Enhancing sustainability and intelligent assurance," *Journal of Accounting, Finance and Auditing Studies*, vol. 11, no. 1, pp. 61–71, 2025. https://doi.org/10.56578/jafas110105

[3] T. H. Wu, S. Y. Huang, and C. W. Lin, "Relationship quality, computer auditing and information security for mobile industry," *International Journal of Mobile Communications*, vol. 18, no. 4, pp. 405–419, 2020. https://doi.org/10.1504/IJMC.2020.108293

[4] H. Li, Q. Liu, M. Hu, and X. Zhu, "Research of adaptive transformation method based on webpage semantic features for small-screen terminals," *ETRI Journal*, vol. 35, no. 5, pp. 900–910, 2013. https://doi.org/10.4218/etrij.13.0112.0834

[5] A. Samiee, Y. Sun, R. F. Demara, Y. Choi, and Y. Bai, "Energy efficient mobile service computing with differential spintronic-c-elements: A logic-in-memory asynchronous computing paradigm," *IEEE Access*, vol. 7, pp. 55851–55860, 2019. https://doi.org/10.1109/ACCESS.2019.2911098

[6] W. X. Zhao, W. Huang, and Y. J. Liao, "Improving privacy budget auditing of differentially private artificial intelligence models through variance of model parameters," *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 10391–10404, 2025. https://doi.org/10.1109/TIFS.2025.3613955

[7] D. Bernau, G. Eibl, and F. Kerschbaum, "Quantifying identifiability to choose and audit ε in differentially private deep learning," *Proceedings of the VLDB Endowment*, vol. 14, no. 13, pp. 3335–3347, 2021. https://doi.org/10.14778/3484224.3484231

[8] N. Milosevic, M. Stefanovic, and C. Stefanovic, "Performance analysis of interference-limited mobile-to-mobile κ–μ fading channel," *Wireless Personal Communications*, vol. 101, no. 3, pp. 1685–1701, 2018. https://doi.org/10.1007/s11277-018-5784-4

[9] G. Chen, "Design and application of scenario-based perception of smart wearable device interaction method," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 13, pp. 69–81, 2024. https://doi.org/10.3991/ijim.v18i13.49071

[10] D. K. J. B. Saini, C. Shieh, L. J. Sankpal, M. Mehrotra, K. S. Bhosale, and Y. Raut, "Deep learning-based optimized model for emotional psychological disorder activities identification in smart healthcare system," *Journal of Research, Innovation and Technologies*, vol. 4, no. 2, pp. 143–157, 2025. https://doi.org/10.57017/jorit.v4.2(8).02

[11] Y. Si, "Construction and application of enterprise internal audit data analysis model based on decision tree algorithm," *Discrete Dynamics in Nature and Society*, vol. 2022, no. 1, p. 4892046, 2022. https://doi.org/10.1155/2022/4892046

[12] L. J. Abbott, S. Parker, and G. F. Peters, "Internal audit assistance and external audit timeliness," *Auditing: A Journal of Practice & Theory*, vol. 31, no. 4, pp. 3–20, 2012. https://doi.org/10.2308/ajpt-10296

[13] C. M. Chen and Y. J. Lin, "Effects of different text display types on reading comprehension, sustained attention and cognitive load in mobile reading contexts," *Interactive Learning Environments*, vol. 24, no. 3, pp. 553–571, 2016. https://doi.org/10.1080/10494820.2014.891526

[14] B. Cheng, E. Lin, A. Wunderlich, K. Gramann, and S. I. Fabrikant, "Using spontaneous eye blink-related brain activity to investigate cognitive load during mobile map-assisted navigation," *Frontiers in Neuroscience*, vol. 17, p. 1024583, 2023. https://doi.org/10.3389/fnins.2023.1024583

[15] L. Cao, W. He, X. Guo, and T. Feng, "A scheme for verification on data integrity in mobile multicloud computing environment," *Mathematical Problems in Engineering*, vol. 2016, no. 1, p. 9267608, 2016. https://doi.org/10.1155/2016/9267608

[16] C. Lin, Z. D. Shen, and F. T. Sheldon, "A data integrity verification scheme in mobile cloud computing," *Journal of Network and Computer Applications*, vol. 77, pp. 146–151, 2017. https://doi.org/10.1016/j.jnca.2016.08.017

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).