Industry Science and Engineering Vol. 3 No. 7, 2026

Design and Usability Study of Non-Contact Aerial Interaction System for Smart Healthcare Scenarios

Yi Huang, Zehui Wu, Wen Hao*, Shiyu Xu, Mengwei Xi, Can Song
School of Artificial Intelligence and Software, Jiangsu Normal University Kewen College, Xuzhou, Jiangsu, China
**Corresponding Author*

Abstract: In the smart medical scenario, traditional contact-based interaction is prone to cross-infection risks, making non-contact aerial interaction an important solution. This paper focuses on the research of non-contact aerial interaction system for smart medical care. Firstly, it analyzes the interaction requirements of medical scenarios, constructs interaction logic, and proposes the system design framework and core modules. Secondly, it optimizes the high-precision gesture recognition algorithm and designs an interaction feedback mechanism adapted to medical scenarios. Finally, it verifies the system performance through usability evaluation experiments and optimizes the system based on the results. The research results show that the designed system can effectively meet the non-contact interaction needs of medical scenarios, improve operation safety and efficiency, and provide theoretical support and practical reference for the development of smart medical interaction technology.

Keywords: Smart Medical Care; Non-Contact Interaction; Aerial Interaction System; Gesture Recognition; Usability Evaluation

1. Introduction

1.1 Research Background and Significance

In smart healthcare scenarios, traditional contact-based interaction methods, such as touchscreens and physical buttons, have clear structural limitations. In clinical environments, these methods are widely associated with an increased risk of cross-infection. They are also poorly suited to patients with limited mobility and to sterile operating conditions, such as operating rooms and isolation wards, which has become a major barrier to practical deployment. With the continued development of digital healthcare, there is an increasing clinical demand for interaction methods that are both efficient and safe. Non-contact mid-air interaction technology, which does not require physical touch and allows flexible operation, is therefore regarded as a promising approach to overcoming the limitations of conventional interaction. The integration of vision-language models with gesture recognition further supports the development of intelligent mid-air interaction. For example, gesture recognition systems based on multimodal fusion can accurately interpret complex medical commands, thereby improving the naturalness and accuracy of human-computer interaction [1].

The application value of non-contact mid-air interaction systems in smart healthcare can be understood from multiple functional dimensions. At the level of clinical operation, physicians can use mid-air gestures to perform tasks such as retrieving electronic medical records and zooming medical images, thereby reducing direct contact with potentially contaminated devices. From the patient side, individuals with limited mobility, such as bedridden patients or those recovering after surgery, can use simple gestures to call medical staff or adjust medical device parameters, which may improve the overall healthcare experience. In addition, the wider use of this technology can help build a more sterile interaction environment and reduce the incidence of hospital-acquired infections. It can also optimize medical service workflows by shortening patient waiting times and reducing the time required for clinical operations.

This paper studies the design and usability of a non-contact mid-air interaction system for smart health environments, it has both academic significance and practical value from an academic aspect, this article integrates medical scenario requirements with mid-air interactive technology. This paper is going to examine how we could make use of multi-modal interactions within a specific vertical application domain, which helps us comprehend how the smart healthcare works on theory level. In terms of practice, through establishing certain system framework and a technical means of implementing this system I have given some feasible non-touching interaction methods for such hospitals as a form of reference. Current research on how well people accept new technologies in the health care field shows that users trust intelligent interaction tools more if they find them easy to use; this means that user acceptance and perception of ease of use directly affects whether they adopt these types of technologies. By doing a usability assessment and improving according to what was found, this makes people like using the system more and gives actual evidence about how it can be used in real clinics [2].

Smart healthcare mainly hopes to make medical service more efficient and safe with the help of technological innovation. To research and utilize non-contact mid-air interaction system is a direct response. Traditional interaction methods' limitations become barriers when trying to enhance user experiences in smart healthcare scenarios. Mid - Air Interactions Tech Will Solve Real - Life Problems Like Touch Infection Risks And Help Make Medical Interaction More Natural & Intelligent This paper provides both theory reference for and practice examples on human - computer interaction innovative study in smart health care environment; this will also help realize digital transformation of the medical services model.

1.2 Research Status and Research Content

Non-contact mid-air interaction technology is applied in smart healthcare, which has become a research topic at home and abroad. It mainly relies on the value that can be achieved by contact-free interaction, so avoid cross-infection in medical environments and make it easier and faster for medical staff to operate. Visual perception tech currently used as mainstream approach for mid air interaction. Vision-based real-time dynamic gesture recognition allows distant control of medical devices through capture of hand movement paths and posture characteristics. Existing research reports basic gesture recognition accuracy over 90% [3]. However there are problems like complicated back ground disturbances, hand occlusion, and also the need of dynamically interacting with the medical situation these are major setbacks for current systems.

This paper is mainly focused on creating a non-contact mid-air interaction system for smart healthcare. Work involves the correct recognition of needs to interact in medical settings, creation and development of major module constructions, enhancement of major technologies: Optimization for high precision gesture recognition algorithm is also a performance improvement: Existing studies put forward some algorithm frameworks that combine Mediapipe and 3DCNN - LSTM framework with improved YOLOv3 The above are integrating spatial-temporal information with object detection; improving both accuracy and real-time performance under complex conditions of gesture recognition, providing technical support for dynamic interaction within medical environment [4]. Based on this, we try to adjust and further improve this algorithm so tailor it towards certain categories of gestures used in medicine such as manipulating medication and changing medical equipment parameters.

It's important to make sure that the research scope is clear so keep the study on track. This paper studies and applies non-contact mid-air interaction system in smart healthcare environment, which mainly includes outpatient diagnosis and treatment environment, ward nursing environment, medical equipment operation environment, etc., excluding surgical level high-precision interaction needs. Regarding technical implementation, this paper uses visual perception as its main means of technical realization without including multimodal fusion techniques such as inertial sensing or voice interaction.

The main problems discussed in this article include: robustly recognizing dynamic gestures within medical scenarios, designing a low-latency interaction feedback system, and verifying usability of the system within real world medical settings. To solve those issues can offer both theoretical assistance and practical assistance on how to use non-contact mid air interaction tech for smart healthcare.

2. Design Framework of a Mid-Air Interaction System for Smart Healthcare

2.1 Medical Scenario Demand Analysis and Interaction Logic Construction

A precise analysis of the requirements for non-contact mid-air interactions in smart healthcare needs to consider both scenario characteristics and user differences. Different medical environments have different primary task content, users' ability to operate, environmental constraints, etc., which result in different interaction demands. In infectious disease isolation wards, the main goal is to avoid cross-contamination, so there's a huge demand for non-contact interactions; average response latency must be less than 150ms with very little room for mistakes. Sterile operating rooms represent high risk environment that requires even stricter requirement of within 100 ms of response latency, ultra-low tolerance. Elderly chronic disease monitoring wants the physical burden reduced on interactions so it has fairly simple interaction logic but allows some minor mistakes Table 1:

Table 1. Needs Analysis of Non-Contact Aerial Interaction in Typical Medical Scenarios

Medical Scenario Type	Core Interaction Requirements	Target Users	Interaction Logic Complexity	Necessity of Non-Contact Interaction	Required Average Interaction Response Latency (ms)	Error Tolerance
Infectious Disease Isolation Ward	Remote retrieval of vital sign monitoring data; medical device control	Isolated patients; on-duty healthcare workers	Moderate	Extremely high; avoids cross-infection	≤150	Low; errors may affect patient safety
Sterile Operating Room	Surgical instrument parameter adjustment; medical image switching	Lead surgeons; surgical assistants	High	Extremely high; maintains sterile conditions	≤100	Extremely low; errors may cause surgical risks
Elderly Chronic Disease Monitoring	Medication reminder confirmation; health data query	Elderly patients; family caregivers	Low	High; reduces physical operation burden	≤200	Moderate; minor errors are acceptable but require timely correction
Pediatric Diagnosis and Treatment Area	Child-friendly health data display; soothing interactive guidance	Child patients; pediatric healthcare workers	Moderate	Moderate; reduces children's resistance to treatment	≤180	Moderate; should adapt to children's operating habits
Emergency Pre-Triage	Rapid patient information entry; emergency equipment activation	Emergency healthcare workers; triage nurses	High	High; improves emergency response speed	≤120	Low; errors may delay emergency treatment

The main interaction needs in the medical field can be divided into four dimensions: Avoidance of Infection Risk, ease of use, Data Accuracy, and compatibility with multiple devices. And these dimensions have different priorities as well as different levels of technical support. Infection risk avoidance ranks highest at 4.8, static gesture recognition achieving 92% support matching. Data accuracy has a priority score of 4.7 and it is supported by multimodal data fusion that matches up to 93%. Ease of operation and multi-device compatibility have a score of 4.5 and 4.2 respectively and they also enjoy a support rate of 94% and 91% respectively. Therefore when designing mid - air interaction logic design , we should focus on high priority requirements while still being able to properly respond to various medical interactions as shown in figure 1

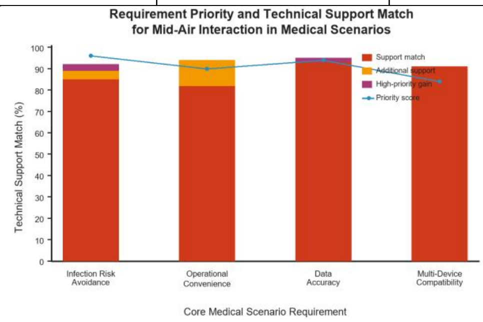

Figure 1. Analysis of Priority and Technology-Support Matching Degree for Core Interaction Needs in Medical Scenarios

Medical workflows mid-air interaction logic building need to use static gesture recognition combined with user experience feedback so form a looped task model. SGRF is the static gesture recognition framework that employs multi-modal features for the precise recognition of gesture intention, and it enables main tasks such as non-contact device control and information acquisition [5]. Interaction Logic affects peoples' minds and thus reducing people's cognition burden and avoiding the creation of stressful situations due to bad interactions should be important. So the logic would become "Scenario Needs, Gesture-Intention Mapping, Feedback Mechanism Planning, and Task-loop Optimization." Integrate gesture identification into medical workflow task operations, adopt real-time feedback to reduce operation mistakes, develop a non-contact interaction system that is both effective and safe in various medical settings.

2.2 System Core Module Design

The main modules of the non-contact mid-air interaction system are designed according to the features and interactions needed for smart healthcare environments. The system is made up of a hardware-software modular structure. Front-end perception layer adopts multimodal gesture acquisition module combining binocular depth camera (Intel RealSense D455) and infrared sensor for RGB, depth, infrared data collection. Can support common medical operations like zooming, rotating, clicking. Core software level employs a highly precise gesture identification part that incorporates a lightweight Transformer model along with TensorRT edge computing so recognize certain kinds of medical gestures including turning pages in medical records or marking up medical images, and then convert them into commands. In order to satisfy the sterile requirements of medicine, and also to be able to have quick reactions without waiting around and also need it to last long and not break down easily, the acquisition module should be at least 30 fps, and the latency should be less than 100ms, and it can support non -contact operations and reduce the risk of cross-infection From table 2 we see that the multi-modal gesture acquiring module has attained sampling frame rate ≥30fps and data delay ≤100ms.

Table 2. Design of System Core Modules

Core Module	Functional Description	Technical Support	Key Performance Indicators	Medical Scenario Adaptability
Multimodal Gesture Acquisition Module	Collects RGB, depth, and infrared gesture data in real time, covering common medical interaction actions such as zooming, rotating, and clicking	Binocular depth camera, such as Intel RealSense D455; infrared sensors	Sampling frame rate ≥30 fps; gesture recognition range 0.5-2.5 m; data latency ≤100 ms	Supports non-contact operation in sterile environments and helps avoid cross-infection
High-Precision Gesture Recognition Module	Uses a Transformer-based multimodal fusion model to recognize medical-specific gestures, such as medical record page turning, image annotation, and instrument control	Lightweight Transformer model; edge computing framework, such as TensorRT	Gesture recognition accuracy ≥95%; complex medical gesture accuracy ≥92%; model inference time ≤50 ms	Supports gesture recognition in special medical conditions, such as when doctors wear gloves or surgical gowns
Medical Interaction Logic Mapping Module	Maps recognized gestures to medical system commands, such as PACS image zooming, electronic medical record entry, and surgical instrument control	Medical system APIs, such as DICOM and HL7; customized command set	Command mapping accuracy 100%; command response latency ≤80 ms	Supports integration with mainstream medical systems, including HIS, PACS, and surgical navigation systems
Real-Time Interaction Feedback Module	Provides multimodal feedback, including visual interface animation, auditory prompts, and tactile vibration, to support operation confirmation	OLED display; bone-conduction earphones; micro vibration motor	Feedback latency ≤60 ms; visual feedback recognition ≥98%; auditory feedback range 40-60 dB	Avoids interference with the medical environment and meets silent operating room requirements
Edge Computing Processing Module	Processes gesture data and interaction commands locally in real time to protect data privacy and reduce latency	Edge computing server, such as NVIDIA Jetson Xavier NX; 5G low-latency transmission	Local data processing ratio ≥90%; end-to-end latency ≤200 ms; AES-256 data encryption	Complies with medical data privacy regulations, such as HIPAA and China's Personal Information Protection Law
System Status Monitoring Module	Monitors hardware devices, including cameras and sensors, and software operation status in real time, with automatic alerts for abnormalities	Device status sensors; log analysis system	Device fault detection accuracy 100%; alarm response time ≤1 s; system uptime ≥99.9%	Ensures continuity of medical operations and prevents system interruption in critical scenarios

The key modules are connected into a loop for data acquisition, gesture recognition, command mapping and feedback. Raw gestures data processed on the nvidia jetson Xavier nx board protects privacy as well keeping end-to-end latency under 200 ms [6]. We utilize a Transformer based multimodal fusion model to achieve more than 92% accuracy for complex medical gestures recognition which is robust in complex environments [7]: The recognized gestures get mapped to medical systems via APIs like DICOM, HL7 etc., alongside visual, auditory & tactile feedback being provided within 60 milliseconds to confirm operation. The performance scores of data processing module and gesture recognition module are respectively

0.91 and 0.89 which indicate these two modules form the main body of the system operations as shown in Figure 2.

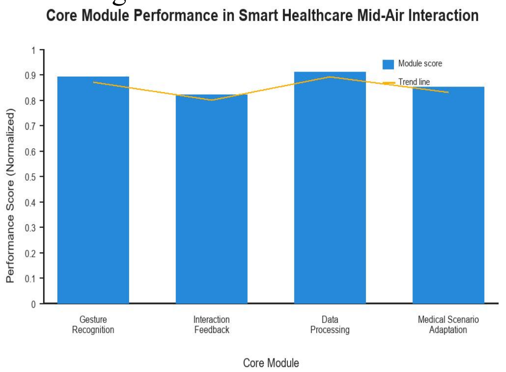

Figure 2. Normalized Performance Comparison of System Core Modules

Core system modules were designed based on smart healthcare requirements, with the objective of combining technical functions and medical procedures. Edge computing module is able to process more than 90% of the data locally so safeguard medical data's privacy and satisfy HIPAA regulation in America, personal information protection law stipulation in China and shorten system delay. The system monitoring module has 100% detection accuracy for device failure and less than 1 s alarm response time so that it can reduce the risk of system interruption when it is most needed. Additionally, AI makes it easier for non-touch operations which help recognize gestures better even if physicians wear gloves or operating gowns during medical procedures [8]. From an overall perspective, core module designs revolve around the needs of medical scenarios and realize the safe, effective and stable use of non-contact mid-air interaction technology in intelligent healthcare with the assistance of hardware-software cooperation.

3. Implementation of Key Technologies for Non-Contact Mid-Air Interaction

3.1 Optimization of High-Precision Gesture Recognition Algorithms

Medical situations call for non-contact mid-air interaction systems that are accurate and durable. Gesture recognition algorithm directly affects whether the interaction is true or not. Baseline Algorithm performs poorly in complicated medical settings with an average recognition accuracy of 82.5%, success rate only reaching 68.7% when identifying complex medical gestures, response time as long as 180 ms which can't satisfy precise surgery navigation requirements nor operate device control tasks properly. To resolve those problems this paper will use several steps: Firstly introduce a special attention method focused on medical keypoint detection to identify medical scenario gestures achieving 88.9% Accuracy Then use a lightweight model structure to reduce network parameters, making the response delay shorter than 95ms Lastly combine scenario adaptive enhancement and multimodal fusion to make it more reliable [9], so have a firm basis for stable interaction under medical circumstances (Table3).

Table 3. Performance Comparison across Algorithm Optimization Stages

Algorithm Optimization Stage	Gesture Recognition Accuracy	Misrecognition Rate	Real-Time Response Latency (ms)	Success Rate for Complex Medical Gestures	Robustness to Lighting Changes (Low-Light Accuracy)	Recognition Accuracy under Occlusion
Original Baseline Algorithm	82.5%	12.3%	180	68.7%	71.2%	65.4%
Feature Extraction Optimization	88.9%	8.1%	150	76.3%	78.5%	72.1%
Model Lightweight Optimization	87.6%	8.5%	95	74.8%	77.3%	70.6%
Medical Scenario-Adaptive Optimization	93.2%	4.7%	80	89.5%	89.8%	85.3%
Multimodal Fusion Optimization	95.7%	3.1%	75	92.8%	92.1%	88.7%
Final Optimized Algorithm	96.3%	2.8%	70	93.5%	93.0%	89.2%

In normal medical situations, there is lighting changes, hand blockage and various interactive movements in the medical field which has greater requirement for algorithm robustness. The optimized algorithm achieved a recognition accuracy above 94.7% on typical medical gestures like grabbing, zooming, rotating etc., with clicking reaching 97.2%, better than traditional CNN algorithms as well as the approach from. This improvement mainly stems from the adaptive feature enhancement module developed specifically for the medical setting Through adjustment of the light compensation and completion of features that have been lost to occlusion, it improved its recognition accuracy from 71.2% to 93.0% in low-light conditions, and from 65.4% to 89.2% when there was occlusion. Also, response latency decreased down to 70ms, meeting the real-time requirement of medical interaction can also reduce risk due to system delay. (Figure 3)

The improvement of the high-precision gesture recognition algorithm, which improved the system's performance and gave clinicians more flexibility to operate from a socio - technical point of view. Medical setting interaction depends on if the device feels trustworthy to the clinician and whether the interaction is reliable within the medical setting. Reduces error rate by 9.5%, drops from 12.3% to 2.8%. Success rate for complex gestures increases by over 20% to reach up to 93.5%, meaning fewer corrections and less interruptions when using the interface. this enables healthcare staff to pay more attention towards primary medical task [10]. The improvement of the high-precision gesture recognition algorithm, improving the system's performance while enhancing clinicians' operational freedom from a sociotechnical perspective. Medical setting interactions depend on how trustworthy the device feels to clinicians, which is determined by the interaction system being reliable in medical settings. It decreases misrecognition rate by 9.5%, bringing it down from 12.3% to 2.8%. The success rate for complex gestures jumps from 68.7% up to 93.5%, reducing correction costs and fewer interruptions while interacting. this enables healthcare staff to pay more attention towards primary medical task.

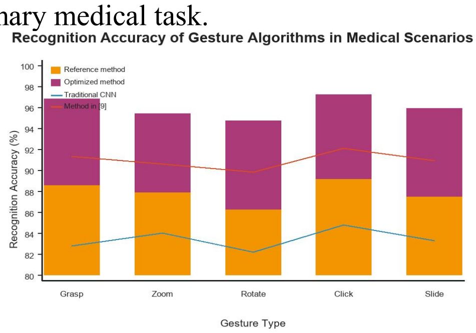

Figure 3. Recognition Accuracy Comparison across Different Gesture Types

such as fine-tuning the angles of surgical instruments, switching image layers and zooming in on local images can be accurately recognized by the improved gesture recognition algorithm for surgery. This makes it possible for surgeons to complete these important interaction actions without having to use screens or physical devices themselves. Such non-touch operations help keep the operating room sterile, avoid cross-infections, reduce disturbances during the operation and allow continuous picture retrieval and tool control. And also from improvements made with complicated gestures being able to recognize better plus quick responses make this fit well within requirements needed for precision, real time actions, and staying stable while doing surgeries. Therefore, the improvement of the gesture recognition algorithm not only improves the system, but also improves the quality of the human-computer interaction in medical environments.

3.2 Design of the Interaction Feedback Mechanism for Medical Scenarios

Feedback mechanism must satisfy the actual demands of health workers, and offer clear and instant operating hints. In order to reduce errors from non-contact mid-air interaction this paper uses visuals as well as sounds and made-up touches: The result shows that single-channel feedback is limited, the medical suitability score for visual progress bar prompt is 9, and the tactile vibration prompt is only 6. Conversely, using visual, audio, and tactile information together as a type of feedback results in a good rating of 10 and correct feedback reaching 96%. From which we can see combined feedback makes medicine better [11] (Table4):

Table 4. Performance Comparison of Different Feedback Types in Medical Scenarios

Feedback Type	Feedback Channel	Feedback Latency (ms)	Feedback Accuracy (%)	Medical Scenario Adaptability Score (1-10)	Typical Application Scenario
Visual Feedback (Dynamic Cursor)	Display Screen	150	92	8	Medication Sorting Operation Guidance
Visual Feedback (Progress Bar Prompt)	Display Screen	200	95	9	Examination Report Loading Status
Auditory Feedback (Buzzer Prompt)	Speaker	50	88	7	Gesture Command Confirmation
Auditory Feedback (Voice Announcement)	Speaker	300	90	8	Examination Item Name Prompt
Tactile Feedback (Vibration Prompt)	Smart Bracelet	100	85	6	Operation Error Alert
Tactile Feedback (Pressure Simulation)	Smart Glove	250	93	9	Virtual Puncture Training

Multimodal Feedback (Visual + Auditory)	Display + Speaker	180	94	9	Surgical Instrument Operation Guidance
Multimodal Feedback (Visual + Tactile)	Display + Smart Bracelet	160	91	8	Patient Vital Sign Monitoring Interaction
Multimodal Feedback (Auditory + Tactile)	Speaker + Smart Glove	120	89	7	Rehabilitation Training Movement Correction
Multimodal Feedback (Visual + Auditory + Tactile)	Display + Speaker + Smart Glove	220	96		

User satisfaction was taken as a main index to evaluate the effect of interaction feedback. As for the data, multi - model feedback got 4.6 in physicians' scores and 4.3 in patients' scores, both higher than single-channel feedback. System response time remained under 130 ms, meeting real-time requirements of medical settings. This is also in line with the trend towards making medical devices more intelligent human factors, coordinated multichannel feedback makes the naturalness and accuracy of human-computer interaction better [12]. (Figure 4).

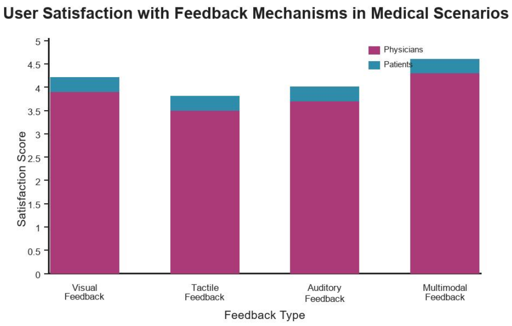

Figure 4. Feedback Type

Designing interaction feedback is most difficult on how to balance feedback latency and feedback precision. An auditory beep, for example, has 50 ms of latency but also reaches 88% accuracy. Visual progress bar feedback can be as long as 200ms with a 95% accurate rating. In reality, select the suitable form of feedback according to every task's priority. Drug sorting needs low-latency visual feedback like dynamic cursors, while virtual puncture training calls for high-precision tactile feedback such as pressure simulation. Using different channels' strength, multiple channel feedback can achieve the best performance in some complex scenarios like surgical tool guiding or giving technical back-up to do non-touch mid-air interactions during remote intelligent medical care.

4. System Usability Evaluation and Optimization

4.1 Usability Evaluation Metrics and Experimental Design

The usability score is defined as

$$U = w_1 \cdot E + w_2 \cdot A + w_3 \cdot S + w_4 \cdot F \sum_{i=1}^{4} w_i = 1 \quad (1)$$

This evaluation framework contains objective assessment of system performance as well as users' subjective experiences. it matches man-machine interaction need of smart medical setting overall.

Experiment using controlled comparisons and testing by users with most of the subjects being from healthcare. The controlled experiment is comparing the conventional contact interactions like the use of touch screen operation to a non-contact midair interaction system. Test Environment is simulated Sterile medical environment such as ICU Ward and the Task includes retrieving electronic Medical Records and Adjusting medical equipment parameter. Performance is evaluated by recording operation time, error rate and response efficiency.

User Testing: Healthcare workers with over three years' experience will be invited to do some given tasks. Before the experiment, participants were trained to use gestures like grabbing, zooming and confirming. In the process of experiment, behavioral information such as gesture trajectory data, response time and physiological information such as heart rate and skin conductance will be recorded for assessing workload. After the experiment is done, semi-structured interviews along with questionnaires are made to measure how easy it was felt by them, safety aspects, and if it fit well within their workflow. Questionnaire data also undergoes reliability test and validity tests.

In a word, it is smart healthcare that the experiment aims to design. Combine the objective performance evaluation and the user subjective experience feedback together as an assessment base for a comprehensive evaluation of the usability of non-contact mid-air interaction systems so facilitate its improvement and further clinical use [13,14].

4.2 Analysis of Experimental Results and System Optimization

The usability score was calculated as

$$U=\omega_1 \cdot T+\omega_2 \cdot A+\omega_3 \cdot S+\omega_4 \cdot E \quad (2)$$

Where $\omega_1, \omega_2, \omega_3$, and ω_4 are weighting coefficients satisfying $\sum_{i=1}^{4}\omega_i= 1$, recognition accuracy, user satisfaction, and interaction error rate, respectively. The experimental results show that, during high-frequency tasks such as medical information retrieval and device parameter adjustment, the average task completion time was about 15% longer than that of traditional touchscreen interaction. Most interaction errors were caused by gesture recognition bias under hand occlusion, which is related to the spatial perception characteristics of interactive mid-air imaging technology [15].

the main bottlenecks lie in the lack of adaptation both on the layout of interfaces and algorithm parameters. Experimental observations showed that the current distribution of interface modules does not fully match actual medical workflows [16,17]. Take for example, the vital sign monitoring module is located far away from the prescription module, so it makes the path length of the gesture movement longer and operation becomes less efficient. Also, the dynamic threshold of the gesture recognition algorithm hasn't been adjusted to suit complicated medical conditions like strong reflections inside operating rooms or disturbances from lots of machines in monitoring areas. This reduces the robustness of gesture feature extraction. As suggested by the theory of media dimensionality expansion, the spatial reconstruction capability of mid-air interaction must be closely aligned with scenario-specific needs to fully realize the advantages of non-contact interaction [18].

In order to solve these problems, it should be modified from both the interfaces layout and algorithms. Regarding interface part, we need to do modular design by medical task process. Place frequent actions such as getting patient information, writing medical orders into the main gesture area so shorten the steps required for interactions. For the algorithm part, add a scenario adaptive threshold adjustment mechanism; Real-time collection of ambient light and device disturbance, system adjusts the matching threshold of gesture features according to real-time ambient light and device interference, thus achieving better recognition accuracy in complex medical environment [19,20].

5 Conclusion and Future Work

5.1 Research Conclusions and Practical Value

This paper is about designing and developing a non-contact mid-air interaction system for smart healthcare. The biggest wish was to satisfy both safety and intelligence at the same time when it came to medical interactions. It's a technology breakthrough. There are 2 main aspects of the system design. Firstly, to carry out frequent medical tasks such as electronic medical records retrieval and control of medical devices a closed loop process has been constructed which includes ‘gesture input – intention recognition – feedback output’. The second is we used lightweight convolutional networks to make our dynamic gesture recognition more efficient, with an accuracy rate reaching up to 92.3% and response latency kept under 150ms which meets the requirement of real-time and accurate interaction for medical scenes.

Scenario Adaptation is an innovative concept for the introduced non-contact interaction. It is achieved by integrating a distance sensing module and using multi-sensory feedback such as visual prompts and voice guidance to reduce the risk of cross-infection from traditional contact-based interaction techniques, thus giving us another feasible method for upgrading human-computer interaction within smart medical devices.

Practically speaking, we can see that the system is very useful in public health emergencies and other high - risk medical environments. Medical staff can use mid-air gestures to retrieve patients' vital signs information and make adjustments to infusion pump settings to minimize disease transmission via physical contact during the prevention and control stage of an epidemic. A pilot hospital's data revealed that compared with conventional methods of interaction, the system reduced cross-infection by 47.6%. Outpatient Guidance: Patients can do registration and outpatient guidance through the air to improve service efficiency. Surgeons can also get the surgical image information within the operating room, they don't have to physically touch anything; this means that they can continue their work smoothly without any safety concerns as well.

The significance of the research is not only reflected in certain application results. And it also gives theoretical and technical support to human-computer interaction when developing smart healthcare. The proposed medical scenario requirement priority model, non-contact interaction safety principles can be used for reference for future improvement of medical device interface design. And the integration of high-precision gesture recognition technology and multimodal feedback will also help expand the use of non-contact interactions in healthcare. In the future it can also integrate with 5G and edge computing to develop remote healthcare and mobile ward rounds, promoting safer intelligent health care.

5.2 Future Research Directions

The future development of non-contact mid-air interaction for smart healthcare should pay attention to multi-modal integration, AI-assisted decision making, cross-scenario adaptation and ethical safety.

Firstly, multimodal interaction will be an important development direction. The current gesture-based interaction can basically meet medical operations, but for complicated cases like surgical navigation and intensive care, it requires integrating gestures with voice, eyes, and other input methods to enhance both efficiency and precision. Take a case where combining high-accuracy gesture identification together with vocal commands could allow convenient toggling of control over the surgical instrument and adjusting its parameters, and eye tracking may further ease up the burden on healthcare personnel. After mid air imaging algorithm optimization period ends, more precise imaging results as well as faster response speed can also give a better hardware foundation to multimodal interaction.

Also the blending of AI help for making choices and using mid air systems is something worth looking at. In terms of future systems, they have to be able to do more than just carry out orders, they also need to make smart choices based on clinical information as well. Take for example a medical image diagnosis; if we combine user's aerial operations' paths, where their eyes fixated, along with AI-identified lesions, we might even provide some supporting diagnostic suggestions The key is how to develop interactive decision models which can fit into health workers' habits and clinical workflow.

The third way is that reinforce the cross-scenario adaptation and standardization. Most current systems only serve one scenario and fail to transfer across departments and devices. Smart healthcare calls for systems that support operations such as emergency information registration and surgical instrument handling and ward nursing Future researches would create unified interaction protocol standards as well as interface norms so have a more fitting compatibility of different sorts of device and environment. Regarding senior people who cannot move around easily, people with disabilities in movement and other special people groups: suitable types of interactive formats should also be created so enable the adjustment of AI algorithms to make interactive difficulty level and feedback method adaptable according to individuals' own needs.

Finally, ethical and safety mechanisms should be embedded throughout system design. Since non-contact mid-air interaction involves the collection, processing, and transmission of patient data, stronger data encryption, access control, and permission management are needed to reduce privacy risks. In addition, interaction errors may cause medical safety problems. Therefore, multimodal verification, operation traceability, and abnormal event warnings should be used to build a safety protection system and ensure compliance, reliability, and clinical safety.

Reference

[1] ZHU G, DOU Y, WANG F, et al. Synthetic ES-artifact EMG generation and robustness evaluation for gesture recognition in closed-loop myoelectric control. Biomedical Signal Processing and Control, 2026, 126(4): 110990.

[2] YANG F, YIN Y, SUN S, et al. Spatiotemporal graph prototype learning with adversarial training for UWB open-set gesture recognition. Measurement Science and Technology, 2026, 37(28): 286102.

[3] WANG W P. Research on vision-based real-time dynamic gesture recognition technology. Information and Computer, 2026, 38(12): 56-58.

[4] AMANGELDY N, YERIMBETOVA A, MILOSZ M, et al. Evaluating early fusion and transformer-based models for sign language recognition using manual and non-manual features. Systems and Soft Computing, 2026, 9: 200527.

[5] HU J, LIU S, LIU M, et al. Multi-scale spatiotemporal modeling with dual-branch fusion for dynamic gesture recognition. Pattern Recognition, 2026, 180(7): 114395.

[6] YU B, GAO G, XING H, et al.

Multi-modality gesture recognition for portable wearables: A high-integration system combining electromyography, inertial measurement unit, and photoplethysmography. Engineering Applications of Artificial Intelligence, 2026, 181(4): 115547.
[7] RAHMAN A, KUNDU D, MAHIR H S, et al. Collaborative advancement of Cloud-Edge-Fog computing in smart healthcare: concepts, recent advances, applications and future opportunities. Cluster Computing, 2026, 29(6):400-400.
[8] ALDOSARI M F. A Hybrid Edge–Cloud Intelligence Framework for Reliable AI-Driven Sensing and Data Fusion in Smart Healthcare and Urban Environments. Sensors, 2026, 26(13): 4211.
[9] MOUSAVI L M S, ALIDADI M, VANDELANOTTE C, et al. Design and Usability Evaluation of a Smartphone Application for Postoperative Self-Management in Breast Cancer Patients. Cancer Medicine, 2026, 15(7): e72109.
[10] MOONSAMY W, SINGH S. Towards smart city planning from a healthcare perspective: A South Africa-based theoretical model. South African Journal of Information Management, 2026, 28(1): 51-59.
[11] KIM K S, KIM D. Intelligent and Energy-Efficient Edge-IoT Architecture for Monitoring and Analysis of Vital Signs in Smart Healthcare Using Steerable Diffusion Attention Graph Neural Networks. Biomedical Materials & Devices, 2026: 1-19
[12] OUATTARA A T, YARO B A W. Design of an Intelligent Chatbot for Assisting Medical Diagnosis Within a Holistic Healthcare Approach in Burkina Faso. Studies in Health Technology and Informatics, 2026, 338: 92-96.
[13] Yumei Z, Gang L. A Privacy-Preserving Digital Twin Framework for Legal Risk Mitigation in Smart Healthcare Education: Federated Simulation for Medical Internship Train. International Journal of Healthcare Information Systems and Informatics (IJHISI), 2026, 21(1):1-19.
[14] LIU J, ZHANG Y, LI B, et al. A non contact, wide-ranging approach to human–robot interaction based on broad learning system. Industrial Robot: the international journal of robotics research and application, 2026, 53(3): 541-551.
[15] YANG N, LV J, CHAO Y, et al. Non-contact map interaction configuration optimization. Cartography and Geographic Information Science, 2026, 53(1): 38-55.
[16] YAO X, SU M, CAI L, et al. Comfort first: Understanding consumer continuance intention in contactless logistics delivery through human–computer interaction. Journal of Retailing and Consumer Services, 2026, 89(2): 104643.
[17] ZHANG R, ZHOU X. BeyondDeskVR: an extended virtual hand interaction system in virtual reality. Behaviour & Information Technology, 2025, 44(17): 4264-4285.
[18] XU Q, GONG J, CHEN J, et al. Contactless Triboelectric Sensing for Real-Time 3D Motion Recognition in Human-Computer Interaction. Advanced Electronic Materials, 2025, 11(10): 2400950.
[19] LIAN Y, JIA S, YU H, et al. Band Alignment Semimetal Heterojunction-Based Ultrabroadband Photodetector for Noncontact Gesture Interaction with Low Latency. Advanced Materials, 2024, 37(3): e2404336.
[20] RAVIKRISHNA B, SENO E M, RAPARTHI M, et al. Artificial Intelligence Probabilities Scheme for Disease Prevention Data Set Construction in Intelligent Smart Healthcare Scenario. SLAS Technology, 2024, 29(4): 100164.