# M-Ring: A Lightweight Ring for Gesture Interaction and User Authentication With Multimodal Fusion

Shilin Chen, Yuqi Cheng, Yuling Tan, Chaonan Tang, Qian Zheng, Yongpan Zou, *Member, IEEE*, Victor C. M. Leung, *Life Fellow, IEEE*, and Kaishun Wu, *Fellow, IEEE*

***Abstract*—In extended reality (XR) virtual interaction environments, both flexible user interaction and reliable authentication are essential. However, current wearable systems relying on single-sensor data suffer from insufficient data dimensionality, often failing to simultaneously achieve fine-grained gesture recognition and continuous authentication. Multisensor systems, while more capable, tend to be bulky, comfort-compromising, and computationally demanding for edge deployment. Furthermore, inefficient parallel processing and the decoupled pipeline for gesture and authentication tasks degrade computational efficiency, ultimately hindering practical adoption. To overcome these limitations, we propose M-Ring, a lightweight ring-based system equipped with a compact, optimized-sensor layout that synchronously captures impedance and inertial measurement unit (IMU) signals. By fusing complementary physiological and behavioral cues, M-Ring enables accurate hand-state perception. We design an adaptive-scale convolutional network with a progressive layered extraction (PLE) architecture, which effectively captures intrinsic correlations within multimodal data and enhances the synergy between gesture recognition and user authentication, all while minimizing computational overhead. For fine-grained authentication, we propose a dual-factor framework that integrates Siamese networks with contrastive learning to enhance interuser discriminability. By jointly modeling physiological activations and kinematic patterns, our approach enables both seamless gesture-based interaction and nonintrusive continuous authentication. Experiments demonstrate that M-Ring achieves 97.8% gesture recognition accuracy and 97.5% identity authentication accuracy under complex conditions, while maintaining low enrollment cost, power efficiency, low latency, and strong edge deployment capability.**

Received 9 October 2025; accepted 13 November 2025. Date of publication 26 November 2025; date of current version 5 February 2026. This work was supported in part by NSFC under Grant 62172286 and Grant 62472366; in part by the Key Program of Department of Education of Guangdong Province (DEGP) under Grant 2024ZDZX1011, Grant 2024GCZX003, and Grant 2023KCXTD042; in part by Guangdong Provincial Key Laboratory of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things under Grant 2023B1212010007; in part by Shenzhen Science and Technology Program under Grant JCYJ20250604181429038; in part by Shenzhen Science and Technology Foundation under Grant ZDSYS20190902092853047; and in part by the 111 Center under Grant D25008. *(Corresponding author: Yongpan Zou.)*

Shilin Chen, Yuqi Cheng, Yuling Tan, Chaonan Tang, Qian Zheng, and Yongpan Zou are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China (e-mail: shilinchen2023@163.com; cyuqi2090@gmail.com; tanyulingszu@qq.com; chaonantang2002@163.com; qianzhengxx@szu.edu.cn; yongpan@szu.edu.cn).

Victor C. M. Leung is with the Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen, Guangdong 518172, China (e-mail: vleung@ieee.org).

Kaishun Wu is with the Information Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong 510530, China (e-mail: wuks@ust-gz.edu.cn).

Digital Object Identifier 10.1109/JIOT.2025.3637178

***Index Terms*—Gesture, identity authentication, multitask learning (MTL), virtual/augmented reality, wearable computers.**

## I. Introduction

Modern extended reality (XR) systems are increasingly applied in sensitive domains such as financial transactions, digital assets management, and healthcare, highlighting the urgent need for secure and efficient user authentication during interaction [1]. However, to meet demands for portability and compact design, these systems often compromise between precision and security [2]. Reliable authentication not only safeguards user data but also enables personalized services and seamless connections between devices. Pursuing continuous, context-aware, unobtrusive user authentication is a major hurdle, thus requiring a passive, enduring mechanism to stabilize interactions and enhance security.

Current commercial XR interfaces primarily rely on vision-based tracking and handheld controllers. Vision-based methods, which use external or embedded cameras, are vulnerable to environmental conditions such as lighting, occlusion, and limited field of view (FOV) [3]. Meanwhile, handheld controllers suffer from complex button mappings, cognitive overload, and grip fatigue [4]. Most importantly, these systems often depend on traditional authentication methods (e.g., passwords or biometric scans), which demand explicit user operation, undermining the immersive experience that lies at the heart of XR's goals.

Wearable devices, with their low power consumption, low latency, and robustness to environmental interference, provide a promising alternative. They enable more intuitive and natural interactions by supporting diverse sensing modalities with cost-effective deployment. Advances in microelectromechanical systems (MEMS) and flexible hybrid electronics further promote the multisensor integration and miniaturization of wearable devices. Among them, ring-shaped devices stand out due to their inconspicuous design and anatomical compatibility with the finger. They ensure continuous and stable wearability without sacrificing dexterity, and their close fit to the body enables higher fidelity data collection, making them suitable for prolonged use and multifunctional sensing.

Nevertheless, achieving accurate gesture recognition and user authentication on compact ring-shaped devices remains a significant challenge. Prior wearable systems often rely on single sensing modalities, which limits their ability to simultaneously ensure robust interaction and reliable authentication. Furthermore, many solutions focus on a single task (e.g., gesture recognition or heart rate monitoring), neglecting the integration of interaction and identity features. The lack of co-design between sensing and computation leads to inefficient resource use, such as redundant data collection, sequential task processing, and isolated sensor operation. More critically, existing methods fail to exploit the spatiotemporal coupling patterns within sensor data, which serve as key indicators for distinguishing legitimate users from impostors. These limitations reflect a fundamental tension between the constrained computational and physical resources of wearable XR systems and the growing demand for seamless, secure, and multifunctional interactions. Addressing this tension requires a unified hardware–software framework capable of integrating multimodal sensing, supporting concurrent tasks, and operating reliably under real-world constraints.

Such a system must overcome several technical challenges. The limited form factor restricts sensing capability, reducing signal richness. Single-modality sensors, when used in isolation, suffer from noise, environmental sensitivity, or low resolution. For instance, skin impedance sensors effectively capture fine-grained physiological traits but are susceptible to external interference [5], while inertial measurement units (IMUs) exhibit low signal-to-noise ratios in fine motion detection, requiring physiological signals to distinguish intentional from involuntary actions. Moreover, constrained resources on edge devices challenge the feasibility of real-time multitask learning (MTL). Gesture recognition benefits from shared motion patterns across users, while authentication relies on distinctive individual traits. Performing gesture recognition and user authentication requires a balance between interuser consistency and intrauser uniqueness.

To address these issues, we present M-Ring, a joint hardware–software system designed for multitask collaborative sensing as shown in Fig. 1. At the hardware level, M-Ring integrates complementary impedance sensors and IMUs in parallel. Impedance sensors capture detailed physiological signals to compensate for IMU's limited motion resolution, while IMUs enhance robustness to environmental variation. The preprocessed data are input to a multitask network [6], [7] that jointly performs gesture recognition and user authentication. The shared and task-specific expert modules extract general and task-relevant features first. These experts, composed of dynamically scalable convolutional layers, extract multiscale spatiotemporal patterns while suppressing noise. A gating mechanism then fuses high-frequency bioimpedance and low-frequency motion signals with adaptive modality weighting. Progressive routing further optimizes cross-task feature allocation, and task-specific subnetworks generate final predictions. For authentication, a Siamese network [8] is employed to generate gesture-based user embeddings, which are refined through supervised contrastive learning [9] to cluster intrauser features and separate interuser distributions, enhancing generalizability for passive authentication. This dual-network architecture supports seamless, multitask operation without explicit user input.

M-Ring introduces a unified, lightweight architecture for simultaneous gesture recognition and continuous user authentication by leveraging three core innovations. First, it fuses high-frequency bioimpedance and low-frequency motion data through a dynamic modality weighting mechanism, effectively capturing both physiological and behavioral traits for robust multimodal representation. Second, an efficient MTL framework with shared expert networks and adaptive convolutional layers extracts multiscale spatiotemporal features while minimizing redundant computation, enabling parallel inference with minimal increase in model complexity. Third, contrastively trained Siamese embeddings support implicit, continuous two-factor authentication during natural interaction.

We validate M-Ring through comprehensive experiments including cross-user gesture recognition, robustness to environmental variations, authentication scalability, ablation studies, edge deployment efficiency, and immersive XR user experience. Our system achieves 97.8% gesture recognition accuracy and 97.5% authentication accuracy, effectively identifying thumb–index interactions and distinguishing authorized users from impostors. Moreover, M-Ring demonstrates low latency and power consumption, supporting its viability for practical XR interaction.

The main contributions of this article are as follows.

1) We propose M-Ring, a novel multimodal ring-sensing system that enables robust fine-grained gesture recognition and user authentication using a single ring device.
2) We design and implement an M-Ring prototype, integrating bioelectrical and motion sensing within a multitask architecture. This design optimizes multimodal feature extraction while minimizing computational overhead, balancing task functionality with efficient edge deployability.
3) We demonstrate M-Ring's real-world effectiveness through 30 user studies, showcasing superior multitask performance, low latency, and power efficiency in gesture recognition and user authentication.

The rest of this article is structured as follows. Section II surveys relevant previous research. Section III details the M-Ring system. Section IV outlines data collection and preprocessing steps. Section V shows evaluation results across different experimental setups. Section VI discusses the system's extensibility, identifies current limitations, and outlines directions for future research. Finally, Section VII wraps up our work.

## II. Related Works

### A. Wearable Interaction Technologies

Gesture recognition gains renewed momentum with the emergence of virtual reality technologies. Early approaches based on visual tracking [10], [11] or handheld controllers, such as MediaPipeHands [12], which uses RGB cameras for hand tracking, suffer from lighting sensitivity, background complexity, occlusion, limited interaction range, high cognitive load, and computational overhead, reducing their robustness and immersive quality. In contrast, wearable devices offer advantages such as low power consumption, low latency, and robustness to environmental interference, providing more stable and continuous sensing solutions for mobile applications. Early systems like DataGlove [13] use bending sensors to track gestures but are bulky and restrictive. Driven by the pursuit of enhanced usability, numerous recent studies focus on novel sensing modalities and ring-shaped form factors. Electrical signal-based systems, such as the MYO wristband [14] using electromyography (EMG) to capture muscle activity, and ring-based systems like Z-Ring [15], Z-Pose [16], EF-Ring [17], eRing [18], PeriSense [19], and ElectroRing [20], utilize various sensing techniques such as electric field changes, swept-frequency RF, self-capacitance measurements, and bioimpedance for accurate, low-power posture and gesture recognition. However, these systems face challenges like electrode placement, individual variability, and sensitivity to muscle fatigue and skin perspiration [21], [22]. At the same time, in wearable dynamic behavior tracking, IMUs, which consist of accelerometers, gyroscopes, and magnetometers, offer portability, low cost, and high accuracy, providing rich kinematic data for hand and finger movements, and making them widely adopted. For example, Madgwick et al. [23] used a wrist-worn IMU with an RNN model for precise 3-D joint tracking. FinGTrAC [24] and SignRing [25] leverage IMU-based rings for robust sign language recognition. However, IMU-based systems often struggle to detect subtle or low-amplitude movements, leading to reduced recognition accuracy in fine-grained gestures.

It is worth noting that in recent years, in order to meet the sensing requirements of higher accuracy and a wider range of motion, the development of gesture recognition systems and technologies that integrate multiple sensing technologies and collect multiple features attracts widespread attention. Systems such as the Myo bracelet [21] integrate EMG and IMU sensors, while Zhao et al. [26] combined EMG with A-mode ultrasound, improving recognition accuracy by 20.6%.

The proposed M-Ring system integrates complementary skin impedance sensors and an IMU, overcoming the limitations of single-sensor approaches, while maintaining a lightweight and comfortable form factor.

### B. Wearable Authentication Technologies

As AR devices increasingly store sensitive personal information, robust data security mechanisms are essential. Biometric authentication using wearable sensors to capture unique physiological or behavioral traits provides enhanced security and intuitive interaction, drawing increasing attention from researchers [27], [28].

Physiological signal-based authentication utilizes signals such as ECG, blood volume pulse (BVP), and electroencephalogram (EEG) that exhibit individual differences, liveness characteristics, and resistance to forgery [29], [30]. Behavioral biometric authentication, based on traits like gait, voice, touch gestures, and breathing patterns, is noninvasive and easy to collect, making it suitable for wearable applications [27], [28], [31]. For example, prior studies achieve robust continuous authentication across diverse scenarios using touch and voice [27] or breathing patterns [28]. However, physiological traits can be affected by motion artifacts, emotional states, and health conditions, whereas behavioral traits are vulnerable to external interference, leading to limitations in biometric authentication when relying on single-sensor systems. Multisensor fusion addresses these issues by combining complementary physiological and behavioral data, improving specificity and accuracy [32]. For instance, Arteaga-Falconi et al. [33] combined fingerprints with ECG for enhanced verification, and VAuth matches body surface vibrations with voice commands to ensure commands originate from the device owner [31].

Continuous and passive user authentication emerges as an important research direction for wearable devices, offering enhanced security and user experience without requiring explicit user actions [27], [31]. Systems like GlassGuard and VAuth exemplify this by leveraging wearable sensors for continuous verification [27], [31].

The proposed M-Ring system captures physiological and behavioral features from natural touch gestures (e.g., single-finger taps and swipes) during daily interactions, enabling seamless, context-aware authentication for long-term wear without explicit user actions.

### C. Multitask Learning

MTL [34] enables the simultaneous learning of multiple related tasks by sharing parameters or learning processes within a model. By leveraging latent task correlations, MTL improves overall performance while reducing the computational cost of training separate models for each task [35]. The shared-bottom model, one of the earliest MTL architectures, shares underlying network parameters across tasks [36]. Although simple to implement, it often suffers from task conflicts, where competing optimization objectives degrade overall performance [37]. To address this, the mixture of multiplexed experts (MMoE) model [38] introduces task-specific experts guided by a gating network, offering flexibility for tasks with varying correlations. For instance, DeepSeek-V3 employs the mixture of experts (MoE) architecture, selectively activating parameters to reduce computational costs while maintaining performance, providing a novel solution for high training and inference costs in large language models (LLMs) [39]. Tencent's progressive layered extraction (PLE) model [7] further enhances MMoE by incorporating multilevel expert networks, improving task adaptability, and mitigating conflicts through layered information extraction.

The proposed M-Ring system leverages MTL to simultaneously address gesture recognition and user authentication. Its core innovation lies in an adaptive-scale convolutional network based on parameter separation and progressive routing mechanisms. This design dynamically allocates resources, minimizes task interference, and uncovers intrinsic relationships within multimodal data, enabling efficient collaborative processing tailored to task characteristics while ensuring high performance and reduced computational overhead.

## III. System Overview

This section presents the design details of M-Ring. In this section, the system architecture is first presented to

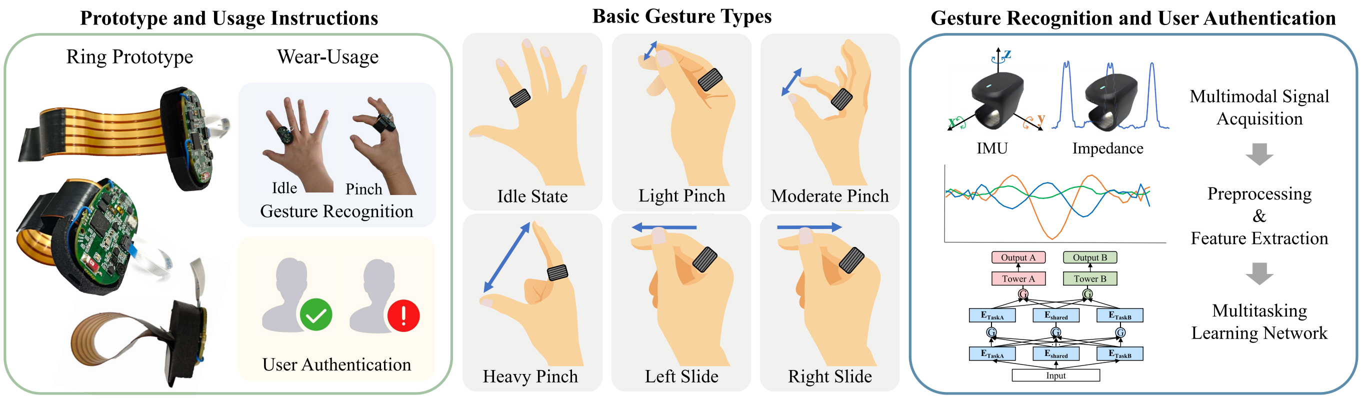



Fig. 1. Proposed M-Ring prototype, usage, basic gestures, and the gesture recognition and user authentication process.

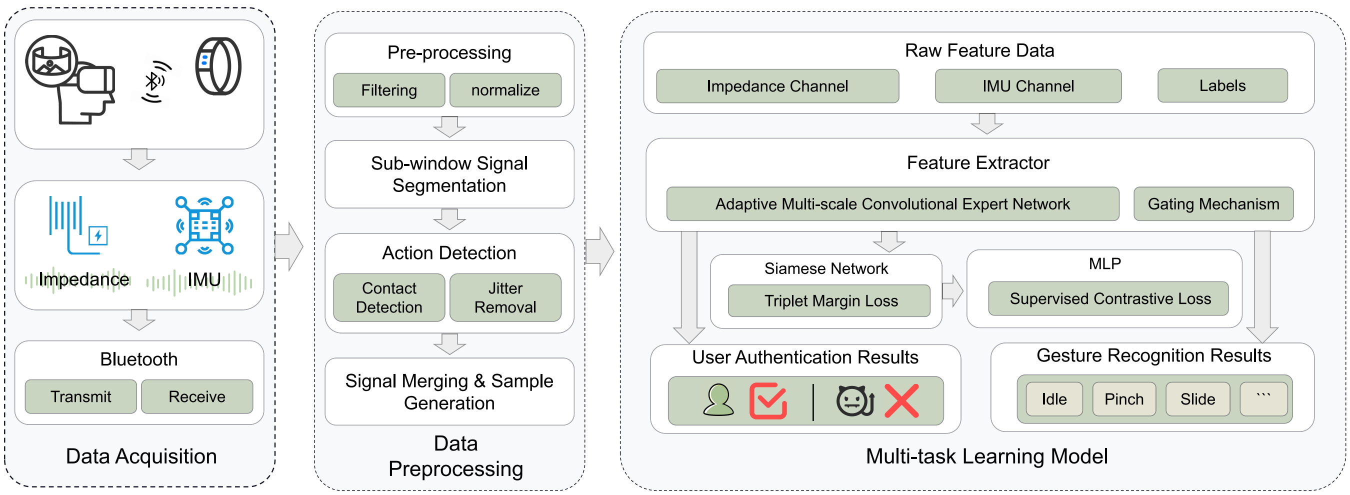



Fig. 2. Workflow architecture includes multimodal data acquisition, preprocessing, and MTL network.

contextualize the research objective. Subsequently, the prototype's wearing configuration and multimodal sensing mechanism are detailed, followed by a comprehensive description of the hardware design, signal acquisition, and transmission pipeline. Finally, we provide a detailed description of the model architecture, highlighting its distinctive design choices, and summarize the training and evaluation procedures.

The system architecture is illustrated in Fig. 2. The proposed miniature ring device continuously acquires impedance and IMU signals, which are transmitted via Bluetooth to the XR host for further processing. The signal preprocessing pipeline consists of filtering, normalization, and subwindow segmentation. To detect user actions, a coarse-grained detection module leverages the high temporal sensitivity of the impedance signal to locate candidate event windows. These windows are then merged and structured into feature-aligned samples. For downstream recognition tasks, we adopt an MTL framework comprising an adaptive multiscale convolutional expert network. This network employs a gating mechanism to extract scale-aware features tailored to specific task characteristics. The resulting representations are fed into task-specific branches: a Siamese network with triplet margin loss for user authentication and an MLP enhanced with supervised contrastive loss for gesture classification. Finally, the system outputs the identity verification result along with the predicted gesture category.

### A. Problem Definition

We aim to construct a unified framework for joint gesture recognition and user authentication using multimodal time-series data collected from wearable sensors.

Let $\mathbf{B} \in \mathbb{R}^{T\times d_b}$ be the bioimpedance signals and $\mathbf{I} \in \mathbb{R}^{T\times d_i}$ be the inertial signals, both sampled over a time window of length $T$. These are combined into a single input tensor $\mathbf{X}_i \in \mathbb{R}^{T\times M}$, where $M = d_b+d_i$. Each instance is associated with two labels: gesture category ($y^{\mathrm{gr}}$) and user identity ($y^{\mathrm{auth}}$).

To extract a joint representation from heterogeneous sensor inputs, we first apply a fusion function $\Phi(\cdot)$ to produce a shared latent representation $\mathbf{H}$

$$\mathbf{H} = \Phi\left(\mathbf{B} \oplus \mathbf{I}\right). \tag{1}$$

This representation is then used to optimize the multitask objective

$$\mathcal{L} = \alpha\mathcal{L}_{\mathrm{gr}} + \beta\mathcal{L}_{\mathrm{auth}} \tag{2}$$

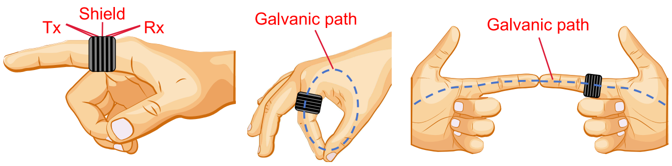



Fig. 3. Electrode layout and touch-detection galvanic path.

where $\mathcal{L}_{\text{gr}}$ and $\mathcal{L}_{\text{auth}}$ denote the losses for gesture recognition and user authentication, respectively, and $\alpha$ and $\beta$ are the weighting coefficients.

### B. Sensing Mechanism

The M-Ring prototype is worn at the base of the index finger, as illustrated in Fig. 3. This location inherently captures rich and representative hand motion, making it particularly suitable for gesture recognition [40]. Furthermore, the device's lightweight and compact design enhances user comfort, minimizes interference with adjacent fingers.

To comprehensively capture both physiological and behavioral aspects of gestures, M-Ring incorporates two complementary sensing modalities: bioimpedance and inertial measurement. The selection of these sensors is motivated by their respective strengths and their combined capability to support both gesture recognition and user authentication tasks effectively. As illustrated in Fig. 3, the impedance sensor injects a weak high-frequency alternating current into the skin via a pair of electrodes and measures the resulting voltage across a separate receiving electrode pair. A centrally placed shielding electrode outputs a dc voltage to suppress direct capacitive coupling between the source and receiving electrodes, forming a compact five-electrode configuration that minimizes the device form factor. The acquired signal reflects the complex bioimpedance of subdermal tissues, which is influenced by the composition, structure, and hydration of biological layers such as skin, muscle, and fat. Subtle physiological changes—such as finger pressing, skin deformation, and microcirculatory activity—modulate the impedance in both amplitude and phase domains. These variations are highly individual-specific and exhibit low intrauser variability across sessions, making bioimpedance a stable and distinctive biometric signal [41]. Furthermore, impedance sensing demonstrates strong robustness to environmental factors such as lighting changes or visual occlusion, and enables passive, continuous data collection during natural finger gestures. Its high sensitivity to low-force and static gestures—such as pinching or touching—also complements inertial sensing. These characteristics collectively highlight impedance sensing as a reliable and informative modality for wearable identity authentication systems.

As further shown in Fig. 3, finger–thumb contact forms a closed-loop circuit, and the resulting voltage differential serves as an indirect indicator of impedance variation. As illustrated in Fig. 4, in the absence of contact, bioimpedance signals across electrodes remain similar and stable, exhibiting a consistent baseline under static conditions. When a finger movement initiates skin contact and forms a current path, the impedance signal deviates from the baseline, producing sharp spikes that correspond one-to-one with the motion events. That is, the sensor demonstrates high sensitivity to contact-induced impedance changes. Based on this property, we develop a coarse-grained motion detection method using thresholding. The baseline is defined as the average of the previous four samples, and the difference between the latest sample and this baseline is computed. A motion event is detected if the difference exceeds a positive threshold or falls below a negative one; otherwise, the system remains in a static state. The thresholds are optimized through a 2-D grid search to minimize false positives and false negatives, with the optimal values being 30 mV for contact onset and −20 mV for release detection.

The IMU module comprises a three-axis accelerometer and gyroscope, capturing both translational and rotational movements of the finger. It excels at identifying dynamic gestures such as swiping or flicking, providing continuous motion profiles essential for time-series gesture classification.

As shown in Fig. 4, for fine-grained gestures such as light pinching, impedance signals exhibit significant changes, while the accelerometer and gyroscope signals from the IMU remain extremely weak—highlighting the limitations of IMUs in capturing low-amplitude gestures. In contrast, IMU signals show greater specificity when distinguishing between different gesture types, whereas impedance signals vary less distinctly across gestures. This indicates that impedance sensing is well-suited for capturing subtle physiological interactions, particularly sensitive to finger-to-finger contact, while IMUs excel at tracking motion dynamics. The fusion of these complementary modalities provides a more comprehensive and discriminative representation of gestures. This multimodal integration enables the model to learn complex spatiotemporal patterns, thereby enhancing recognition accuracy and robustness.

### C. Prototype Design and Implementation

The M-Ring prototype is a compact wearable device integrating multiple sensors and electronic components for unobtrusive signal acquisition and processing, weighing only 6.1 g, as shown in Figs. 1 and 5. In this section, we elaborate on the design and manufacture of prototype equipment from four aspects: electrode design, signal generation and processing, data transmission, and overall system integration.

*1) Electrode Design and Fabrication:* The M-Ring employs a flexible printed circuit board (PCB) as the electrode substrate, ensuring adaptability to different finger sizes and enhancing user comfort. As illustrated in Fig. 3, the electrode layout comprises five internally arranged electrodes. The two proximal electrodes emit a 10.7-MHz ac signal into the user's body, while the central shielding electrode suppresses direct capacitive coupling between the source and receiving electrodes. The two distal electrodes measure the voltage gradient along the finger. This electrode-distribution design, similar to the concept proposed by Kienzle et al. [20] in ElectroRing, optimizes electrode layout and amplifier topology, enabling efficient impedance signal acquisition via a single-point device. As illustrated in Fig. 5, each electrode

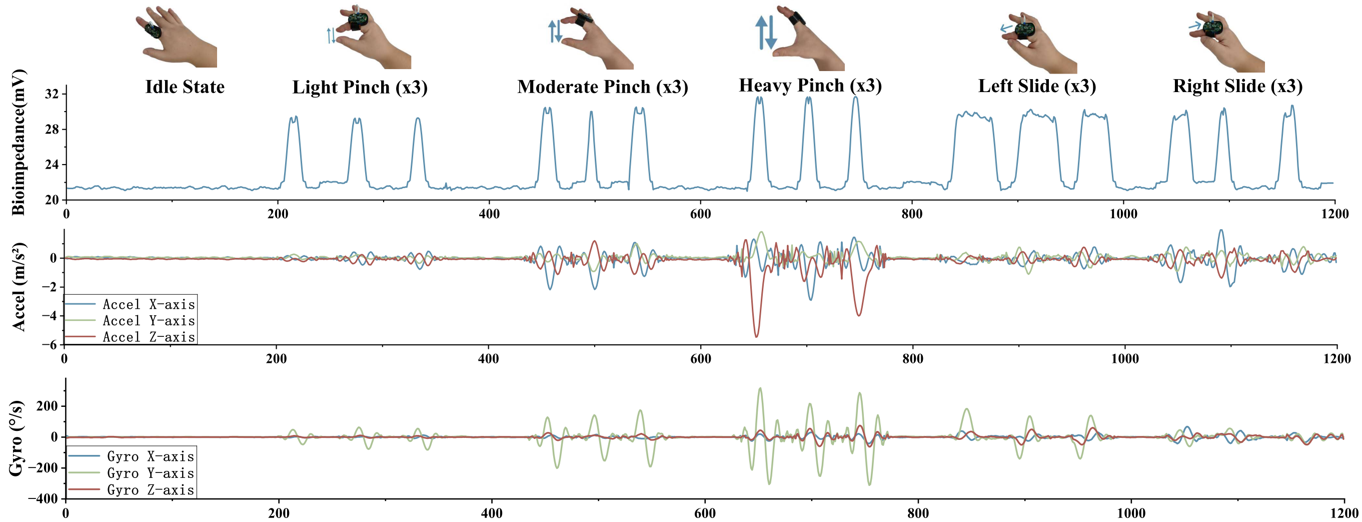



Fig. 4. Gesture waveforms from bioimpedance, accelerometer, and gyroscope sensors.

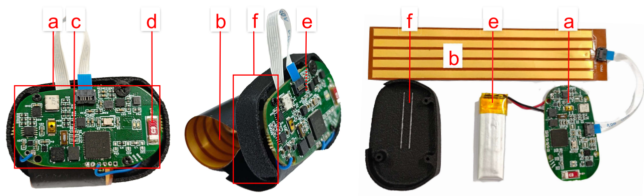


Fig. 5. Labeling of the front and side views and components of hardware equipment. a: PCB; b: impedance electrode; c: IMU sensor; d: communication module; e: power management module; and f: mechanical housing.

consists of an 85 × 3 mm exposed copper region on a 0.02-mm-thick, 20-mm-wide polyimide substrate. The electrodes are spaced 1 mm apart and plated with electroless nickel immersion gold (ENIG) to prevent oxidation. The flexible structure allows it to wrap around the finger, with a five-core flexible flat cable connecting it to the processing unit. Dual-sided adhesive and Velcro fasteners provide secure attachment, accommodating different users.

*2) Signal Generation and Processing:* The signal generation circuit employs an MEMS oscillator (DSC6003) to generate a 10.7-MHz square wave, which is converted into a differential signal using unity-gain buffers and inverting amplifiers (dual OPA2836) before being capacitively coupled to the transmitting electrodes. The received signal is amplified 30-fold, converted to a single-ended signal using an instrumentation amplifier (3× LTC6253), and filtered via a narrowband ceramic filter to suppress noise aliasing. The Cypress PSoC 6 system-on-chip (SoC) manages data acquisition, touch-state detection, and Bluetooth communication. The onboard ADC undersamples at 1.6 ks/s, while a Goertzel filter extracts the amplitude of the aliased 10.7-MHz signal, optimizing sampling frequency to enhance signal quality.

*3) IMU Sensing and Data Transmission:* Meanwhile, the prototype integrates an IMU sensor to capture motion dynamics. Processed impedance and IMU data are fused into 25-Hz frames and transmitted in real time via Bluetooth low energy (BLE) with a 20-m range. A custom Python application on the host device manages the BLE connection while receiving, analyzing, and visualizing the streaming sensor data. Through system evaluation and sampling rate tuning, optimal frequencies are set to 25 Hz for the IMU and 1600 Hz for the impedance sensor, effectively reducing computational load. At the transmission layer, a frame-based local buffering mechanism is employed, packaging data at 25 Hz. Each packet contains one IMU sample and 64 impedance samples (totaling 70 data points). Due to the inherent unreliability of Bluetooth transmission, occasional packet loss is observed during data acquisition, leading to fluctuations in the effective frame rate between 20 and 25 Hz. While this variability cannot be completely eliminated, it reflects the constraints commonly found in real world, resource-constrained wearable scenarios. Rather than relying on idealized data, our system embraces these variations and demonstrates strong adaptability. To further enhance robustness under transmission variability, we employ data augmentation techniques during training, including random cropping, noise injection, and temporal window shifting, as detailed in Section III-D.

*4) System Integration and Power Management:* As illustrated in Fig. 5, designed for compactness, the signal processing PCB adopts an elliptical form factor (65 × 20 mm), with one end slightly widened for orientation guidance. The system is powered by an 80-mAh lithium-polymer battery, delivering approximately 220 mW of power, supporting continuous operation for 1.5 h. The enclosure, constructed from lightweight 3-D-printed material, is secured with microscrews to enhance structural integrity. It features a compact form factor measuring 4.0×2.6×0.7 cm, designed to ensure comfort during long-term wear.

## *D. Model Design*

As shown in Fig. 6, the M-Ring system is designed to perform two key tasks: gesture recognition and user authentication. The multimodal data from impedance and IMU sensors exhibit varying patterns, scales, and complex correlations,

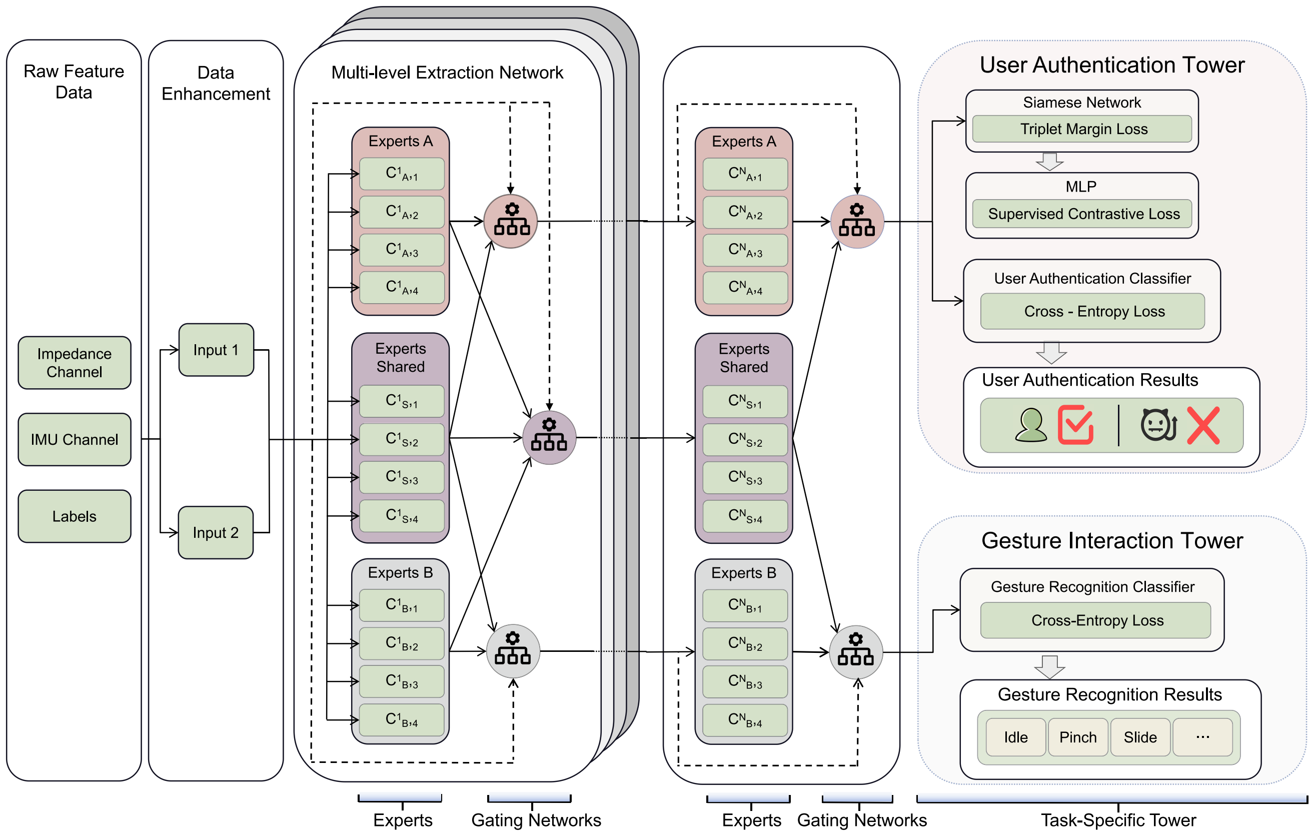



Fig. 6. Model architecture with parameter separation and progressive routing for multimodal fusion.

necessitating a model design that can effectively capture cross-modal dependencies and temporal dynamics. Second, while these tasks are distinct, they are interconnected. Gesture recognition focuses on extracting features related to hand motion and posture, while user authentication distinguishes unique biometric and behavioral traits. Traditional single-task learning methods face limitations in this context, as they often treat data and tasks in isolation, failing to leverage the synergies between them. This leads to suboptimal performance, especially when processing multimodal data, where information from different sources complements each other.

In contrast, an MTL architecture is well-suited for this scenario. It enables simultaneous handling of multiple tasks, allowing the model to learn both shared and task-specific features from multimodal data. By sharing feature representations across tasks, MTL reduces the risk of overfitting, particularly when training data for individual tasks is limited. It also enhances model generalization, as knowledge learned from one task transfers to another.

In the M-Ring system's model, features related to hand movement in gesture recognition provide additional context for user authentication, and vice versa. The MTL architecture enables the model to effectively capture and leverage these cross-task relationships, optimizing performance in both gesture recognition and user authentication.

*1) Key Model Structures:* The model design of M-Ring encompasses the following key structures.

*a) Experts:* The model employs a multilevel expert architecture, as depicted by the small dark rectangles in Fig. 6. The red and gray colors represent task-specific experts for Authentication Task A and Gesture Recognition Task B, respectively, while the purple color denotes shared experts. These experts are hierarchically organized, with lower layers capturing shared representations across tasks, and deeper layers progressively specializing toward task-specific features. This hierarchical design facilitates both effective feature separation and multimodal representation learning, enabling simultaneous optimization for gesture recognition and user authentication.

*b) Gating Networks:* Task-specific gating networks dynamically aggregate expert outputs based on input features and task context, as illustrated by the small dark circles in Fig. 6. The red and gray colors represent task-specific gating units for Authentication Task A and Gesture Recognition Task B, respectively, while the purple color denotes shared gating units. Each gating unit computes soft attention weights over all experts, enabling adaptive feature fusion tailored to the demands of each task. By incorporating signals from lower layer outputs, the gating mechanism supports selective knowledge sharing and improves generalization across tasks with varying relatedness.

*c) Task-Specific Tower:* Each task is assigned a dedicated output tower. The *gesture recognition tower*, shown as the red block in the lower right part of Fig. 6, employs a cross-entropy-based classifier to categorize gestures. The *user authentication tower*, shown as the gray block in the upper right part of Fig. 6, is composed of a Siamese network (for triplet margin loss), an MLP (for supervised contrastive loss), and a final classifier (for cross-entropy loss). These modules, including the Siamese network and contrastive MLP, jointly project input features into a discriminative embedding space before classification. This projection ensures intraclass similarity and interclass separation, facilitating more accurate classification. Although these modules do not directly predict user identities, their optimization of the embedding space significantly enhances the robustness and reliability of the final classification layer.

*2) Training Architecture Overview:* The overall architecture of the training phase is shown in Fig. 6. The raw multimodal input data first undergo random clipping and noise injection for data augmentation, generating positive and negative sample pairs. The model architecture begins with a feature extraction module composed of four identical layers. Each layer consists of three parallel expert networks: Expert A, Shared Expert, and Expert B, each containing four convolutional layers, such as $C^1_{A,1}$ to $C^1_{A,4}$ for Expert Network A. A gating mechanism dynamically assigns weights to each expert based on input characteristics, enabling adaptive feature extraction across different data modalities. The Shared Expert captures common characteristics between tasks, while Expert A and Expert B focus on task-specific representations. This hierarchical structure facilitates multigranular feature learning, ultimately mapping the augmented data into an embedding space that balances noise robustness and discriminative sample separation. Following feature extraction, the model branches into two main task-specific pathways.

For gesture recognition, the extracted features are input directly into the gesture recognition classifier. The goal is to train the network to embed unique features for each gesture. Cross-entropy loss is used, defined as

$$L_{\text{CEgesture}} = -\sum_{i=1}^{n} t_i \log(p_i) \tag{3}$$

where $n$ is the number of gesture classes, $t_i$ is a binary indicator (0 or 1) representing whether the sample belongs to the $i$th class (true gesture label), and $p_i$ is the predicted probability of the sample belonging to the $i$th class, obtained from the Softmax output.

For user authentication, the features are input into the user legitimacy classifier, which also uses cross-entropy loss as in the gesture recognition task. The final output is the user authentication result.

For the authentication task, the feature data first passes through the Siamese network. The Siamese network head aims to extract discriminative features that can distinguish genuine users from impostors. The triplet margin loss function is used, defined as

$$L_{\text{TM}}(a; p; n) = \max\left(d(a_i, p_i) - d(a_i, n_i) + \text{margin}, 0\right) \tag{4}$$

where $a_i$ represents the anchor sample (a reference user sample), $p_i$ represents the positive sample (another sample from the same user as the anchor), and $n_i$ represents the negative sample (a sample from a different user). The function $d(x_i, y_i) = \|x_i - y_i\|_p$ denotes the distance metric used to measure similarity, typically the $L_p$ norm (e.g., Euclidean distance when $p = 2$). The term margin is a predefined constant that controls the separation between positive and negative pairs.

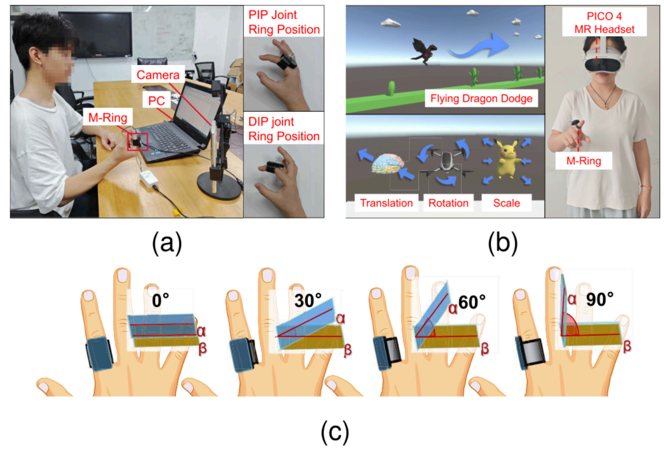



Fig. 7. Experimental scenarios. (a) Data acquisition scenario. (b) XR interactive evaluation. (c) Direction variation test ($\alpha$: back of hand and $\beta$: sensor plane).

The Siamese network head's embeddings then pass through an MLP. To enhance robustness against adversarial examples and noise, a supervised contrastive pretraining method is applied. The MLP head learns to bring the embeddings of samples and their augmentations from the same user closer while pushing apart those from different users. The final branch of the user authentication task connects to the user authentication classifier.

As mentioned earlier, cross-entropy loss is used in the user authentication classifier, defined as

$$L_{\text{CEuser}} = -\sum_{i=1}^{n} t_i \log(p_i) \tag{5}$$

where $n$ is the number of user classes (e.g., legitimate users and impostors), $t_i$ is a binary indicator representing whether the sample belongs to the $i$th class (true user label), and $p_i$ is the predicted probability of the sample belonging to the $i$th class, obtained from the Softmax output.

The total loss is computed as

$$L_{\text{total}} = L_{\text{CEgesture}} + L_{\text{CEuser}} + L_{\text{TM}} + L_{\text{SC}} \tag{6}$$

where $L_{\text{CEgesture}}$ is the cross-entropy loss for gesture recognition, $L_{\text{CEuser}}$ is the cross-entropy loss for user authentication, $L_{\text{TM}}$ is the triplet margin loss for learning user identity representations, and $L_{\text{SC}}$ is the supervised contrastive loss for improving robustness against noise and adversarial attacks.

## IV. Data Collection and Processing

We conduct a study with 30 participants (13 females and 17 males) aged 30.1 on average, divided into four age groups [three teenagers ($\leq$ 18), 21 youth (18–29), three middle-age (30–49), and three elderly ($\geq$ 50)], to evaluate the performance of gesture recognition and user authentication. The scene of data acquisition setup is illustrated in Fig. 7(a).

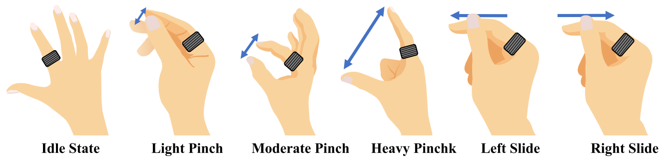



Fig. 8. Basic gesture defined in the system, completed through thumb and forefinger coordinated movements.

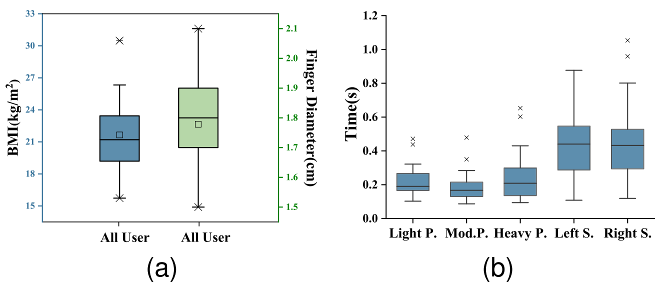



Fig. 9. Data diversity. (a) Distribution of user BMI and finger diameter. (b) Time spent on different gestures by users. Light $P$.: light pinch; Mod. $P$.: moderate pinch; Heavy $P$.: heavy pinch; Left $S$.: left swipe gesture; and Right $S$.: right swipe gesture.

### *A. Data Collection*

Each participant completes all tasks within a single session lasting approximately 1 h. Participants are seated in front of a computer screen that provides visual instructions for performing the required gestures. As shown in Figs. 7 and 8, they are instructed to perform five basic gestures using thumb and index finger interactions, and they use another hand without a ring to tap the designated keyboard to label the current gesture sample. To correctly guide participants in performing the experimental tasks, we provide a detailed description of the gesture actions: The idle state is defined as gestures other than noncontact and sliding behaviors, serving as the reference state. The light pinch means the web space between the thumb and index finger opens to an angle of less than 30°, then the thumb and index finger are in small-area, minimum-pressure contact, showing a significant spike in bioimpedance value and almost no change in IMU. The moderate pinch is when the web space opens between 45° and 90°, then the thumb stably touches the index finger's pulp with a bit of force. There is a significant spike in bioimpedance value and a slight but perceptible response in IMU. The heavy pinch is defined as the web space opening at an angle greater than 90°, then the thumb and index finger pinching firmly. It has the largest contact area, a significant pressure increase, the maximum impedance peak, and drastic fluctuations in IMU acceleration and angular velocity. The left slide refers to the fingertip of the thumb sliding from right to left along the side of the index finger wearing the ring, presenting a long impedance pulse accompanied by a horizontal negative displacement or angular velocity detected by IMU. The right slide is similar. The fingertip slides from left to right, showing an opposite long impedance pulse and a positive inertial change in the IMU direction.

These gestures are selected for several reasons: First, thumb–index finger interactions are ergonomic, commonly used, and involve minimal hand strain, making them suitable for long-term use. They have a small range of motion and are versatile, supporting compound gestures like double-click, triple-click, and long-press. Second, they are practical and directly applicable to various interactive interfaces, such as single-click for selection and sliding for scrolling or page switching. To ensure realism and ecological validity, participants are instructed to perform gestures in a natural and unconstrained manner, without strict enforcement of execution speed or precision. All recorded samples—including ambiguous, imprecise, or borderline gestures—are retained to reflect the variability inherent in real-world human interactions. The idle gesture, representing arbitrary or unrelated hand states, further enables evaluation of the system's ability to distinguish intentional input from casual or incidental hand movements. The experiment is conducted indoors with appropriate lighting, and all actions are recorded. To evaluate the system's robustness in complex environments, participants are instructed to perform gestures under varying conditions, including arm swinging, indoor walking, and using wet fingers. In addition, to assess performance across different wearing positions, they are asked to remove and rewear the ring, switch the ring to different finger joints, and adjust the ring to three distinct wearing angles. Each participant completes 2700 gestures, covering six gesture types and five challenging environments, with approximately 300 repetitions and 3-min duration per gesture type. The collected data are used for both gesture recognition and user authentication analysis.

Our participant cohort exhibits substantial diversity in both physiological characteristics and interaction styles. Fig. 9(a) presents the distribution of body mass index (BMI) and finger diameter (in cm), which, respectively, reflect participants' overall body composition and hand morphology. These attributes significantly influence both impedance measurements and motion dynamics. Further diversity arises from gesture execution patterns. The time distribution for different gesture types performed by users is illustrated in Fig. 9(b). The distinct positions and spread of the box plots across gesture categories indicate substantial variation in execution time. Moreover, the range within each gesture type reveals significant intraclass variability among participants, reflecting differences in motor habits and gesture fluency.

### *B. Data Processing*

We employ a two-stage approach to enhance system reliability. In the first stage, we develop a touch-detection algorithm based on the observation that impedance signals exhibit significant rapid rises and falls during touch and release events, independent of fingertip pressure and finger speed; detailed analysis is provided in Section III-B. Therefore, we roughly determine whether a user performs a touch operation by analyzing raw impedance sensor data. Specifically, the algorithm first applies a 40-ms window moving median filter to minimize noise and motion artifacts in real-time impedance data. Then, it calculates the signal change between the current sample and the previous three samples. A change exceeding 30 mV is recorded as a touch event, while a change below −20 mV is recorded as a release event. In addition, to reduce jitter, any touch events occurring within 60 ms (i.e., 2–3 samples) are ignored. Experimental results demonstrate that this method achieves a recognition accuracy of 99% [20].

In the second stage, we implement a fine-grained gesture category recognition and user authentication algorithm, detailed analysis is described in Section III-D. This algorithm defines gesture event windows based on touch events detected in the first stage and groups impedance and motion data within the windows into 1.2-s windows (approximately 30 samples at a 25-Hz sampling rate). Subsequently, the algorithm vertically stacks all samples within each window to construct input samples for the multitask network. In addition, before feeding raw samples into the MTL model for feature extraction and model training, we perform data augmentation by sliding a 30-ms window between −600 and + 600 ms to simulate gestures at different time points. Finally, the augmented data are combined with the original data for model training.

## V. Experimental Evaluation

### A. Experimental Setup

This section provides a unified setup and implementation protocol applicable to all evaluation tasks. The experiments are categorized into five aspects: gesture recognition, user authentication, ablation studies, edge deployment capability, and usability. Performance is measured using precision, recall, accuracy, specificity, and $F1$ score. To mitigate temporal biases such as fatigue, each 3-min recording is segmented into discrete samples, which were randomly shuffled to construct nonoverlapping training and testing datasets.

*1) Gesture Recognition Experiments:* To evaluate gesture recognition performance under various conditions, we adopt the following testing protocols.

*User-Dependent:* The entire dataset comprising all users is partitioned into nonoverlapping subsets, which are used for training and testing the model in a disjoint manner.

*User-Independent:* To evaluate the performance of cross-user generalization, we use leave-one-user-out (LOUO) protocol for testing.

*Rapid Personalization:* To test the performance of rapidly adapting to new users, we randomly select three users, each using varying numbers of data samples, to fine-tune the model pretrained in user-independent experiments.

*Robustness Testing:* To simulate real-world use, we evaluate the system under six representative environments.

1) *Ideal:* Static seated posture without external interference.
2) *Hand Wave:* Users swing arms while performing gestures (upper limb motion noise).
3) *Walking:* Gestures performed while walking indoors (whole-body locomotion noise).
4) *Position Switch:* The ring is repositioned from the proximal interphalangeal (PIP) joint to the distal interphalangeal (DIP) joint of the index finger, as shown in Fig. 7(a).
5) *Wet Hand:* Participants perform gestures immediately after handwashing to simulate skin moisture caused by sweating.
6) *Direction Shift:* Users rotates the finger ring to multiple angles to perform gestures (sensor directional noise), with four angular settings (0°–90° between sensor plane $\beta$ and hand-back plane $\alpha$) shown in Fig. 7(c).

Each condition is evaluated under both user-dependent and user-independent protocols.

*2) User Authentication Experiments:* To assess the authentication capabilities of M-Ring, we use an unseen impostor evaluation protocol.

*Authentication Performance:* Each round designates one user as legitimate, 28 as impostor candidates, and one as an unseen impostor. A personalized binary classifier is trained using the legitimate user's data and negative-class samples. The test set includes unseen data from both the legitimate and impostor users to evaluate generalization.

*Enrollment Effort Analysis:* We vary the number of registered samples per user (50–300) to investigate the impact of enrollment size on authentication accuracy.

*3) Ablation Experiments:* To analyze the contributions of core components and modalities in M-Ring, we conduct controlled ablations.

*w/o IC:* It removes the impedance signal channel.

*w/o IM:* It removes inertial motion signal channel.

*w/o MSCE:* It excludes the multiscale convolutional expert module.

*w/o SCL:* It excludes the supervised contrastive learning MLP structure.

*w/o SN:* It excludes the Siamese network structure.

Each variant is evaluated using the same recognition protocol to isolate the performance contribution of each component.

*4) Latency and Power Analysis:* We evaluate the system's runtime performance in terms of end-to-end latency and energy efficiency, which are critical for continuous user interaction in immersive XR environments. The M-Ring system is deployed on the ByteDance PICO4 mixed reality headset (Snapdragon XR2 processor) with a dual-thread architecture: one thread for display rendering, and the other for ring-based gesture processing and authentication, enabling parallel core operations.

*a) System Latency:* Raw sensor data are collected via a 25-sample sliding buffer (1-s window, updated every 0.04 s, and 25-Hz sampling rate). Latency is evaluated across 100 independent trials, primarily focusing on signal preprocessing and deep learning inference, as the time cost of signal acquisition and Bluetooth transmission is minimal.

*b) Power Consumption:* The wearable ring is powered by a 3.7-V, 80-mAh battery. Its average power consumption is measured using a dc power analyzer connected in series with the battery circuit. For the PICO4 host, the built-in background power monitoring function is utilized to measure the baseline power consumption during idle operation and the power consumption when the inference pipeline is active, respectively. The difference between the two is calculated as the measured power overhead of the M-Ring inference pipeline, to verify the M-Ring system's capability of fast response and autonomous operation capabilities.

*5) Usability Evaluation in XR Scenarios:* We evaluate usability via a gesture-controlled XR app on the PICO4

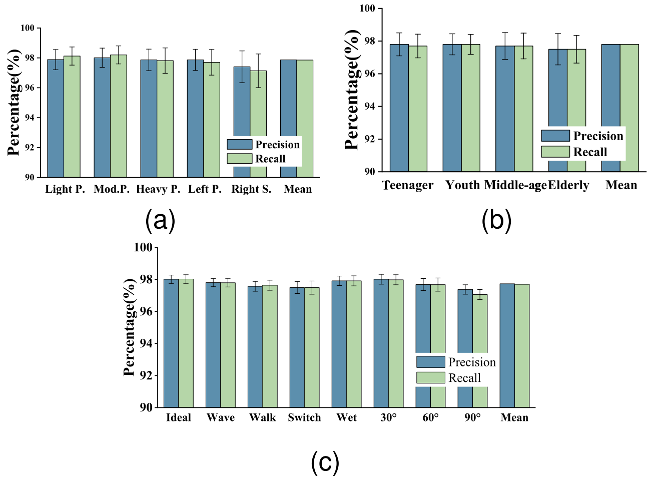



Fig. 10. Results of gesture recognition with user-dependence. (a) Per gesture. (b) Per age group. (c) Per experimental condition. Mean: average of all categories; Ideal: ideal; Wave: hand wave; Walk: walking; Switch: position switch; Wet: wet hand; and 30°, 60°, and 90°: ring angle offsets.

headset. As shown in Fig. 7(b), 20 users complete two interaction tasks using the ring as the sole input device.

1) *Dynamic Obstacle Avoidance:* Participants control a virtual flying character using continuous pinch gestures. The altitude of the character is modulated in real time by adjusting the duration of the pinch, enabling obstacle avoidance under varying spatial conditions.
2) *3-D Object Manipulation:* Participants interact with virtual objects using translation, rotation, and scaling operations. Object selection is based on spatial proximity, determined via Euclidean distance, while manipulation is achieved through hand motion and gesture input. Once a target object is selected, pinch gestures trigger specific transformations mapped to real-time hand displacements.

After completing both tasks, participants fill out the system usability scale (SUS) and provide open-ended feedback regarding comfort, responsiveness, and ease of use.

### B. Experimental Results and Analysis

*1) Gesture Recognition Experiments:* The experimental setup is described in Section V-A1.

*a) User-Dependent Test:* The recognition result for each gesture is presented in Fig. 10(a) and the upper part of Table I. The model demonstrates high and consistent precision across all gesture types, achieving an average precision of 97.8%. The moderate pinch gesture achieves the highest precision of 98.0%, likely due to its distinctive force profile and brief execution duration. In contrast, the left swipe gesture shows a slightly lower precision of 97.4%, which may result from temporal and angular similarities to the right swipe gesture, potentially leading to subtle misclassification. Despite these minor variations, both precision and recall remain consistently high, indicating the model's balanced ability to distinguish between gesture types while maintaining sensitivity to gesture occurrences.

TABLE I

AVERAGE GESTURE RECOGNITION RESULTS PER GESTURE

| Gesture | Precision(%) | Recall(%) | Accuracy(%) | Specificity(%) | F1 Score |
|---|---|---|---|---|---|
| **User-Dependent** | | | | | |
| Light P. | 97.9 | 98.1 | 98.1 | 99.1 | 0.980 |
| Mod.P. | 98.0 | 98.2 | 98.2 | 99.1 | 0.981 |
| Heavy P. | 97.8 | 97.8 | 97.8 | 99.0 | 0.978 |
| Left S. | 97.8 | 97.7 | 97.7 | 99.0 | 0.977 |
| Right S. | 97.4 | 97.1 | 97.1 | 99.0 | 0.972 |
| Mean | 97.8 | 97.8 | 97.8 | 99.0 | 0.978 |
| **User-Independent** | | | | | |
| Light P. | 87.8 | 88.4 | 88.5 | 97.2 | 0.880 |
| Mod.P. | 76.1 | 75.1 | 75.3 | 93.8 | 0.756 |
| Heavy P. | 76.5 | 74.1 | 73.7 | 93.5 | 0.753 |
| Left S. | 73.6 | 71.1 | 72.3 | 93.2 | 0.724 |
| Right S. | 75.1 | 74.7 | 74.2 | 93.5 | 0.749 |
| Mean | 77.8 | 76.7 | 76.8 | 94.2 | 0.772 |

TABLE II

AVERAGE GESTURE RECOGNITION RESULTS PER AGE GROUP

| Group | Precision(%) | Recall(%) | Accuracy(%) | Specificity(%) | F1 Score |
|---|---|---|---|---|---|
| **User-Dependent** | | | | | |
| Teenagers | 97.8 | 97.7 | 97.7 | 99.1 | 0.977 |
| Youth | 97.8 | 97.8 | 97.8 | 99.1 | 0.978 |
| Middle-age | 97.7 | 97.7 | 97.7 | 99.0 | 0.977 |
| Elderly | 97.5 | 97.5 | 97.5 | 99.0 | 0.975 |
| Mean | 97.8 | 97.8 | 97.8 | 99.1 | 0.978 |
| **User-Independent** | | | | | |
| Teenagers | 77.54 | 74.37 | 74.43 | 93.60 | 0.759 |
| Youth | 78.40 | 77.44 | 77.27 | 94.31 | 0.779 |
| Middle-age | 76.95 | 74.61 | 73.57 | 93.42 | 0.758 |
| Elderly | 75.14 | 73.78 | 72.79 | 93.16 | 0.745 |
| Mean | 77.01 | 75.05 | 74.51 | 93.62 | 0.760 |

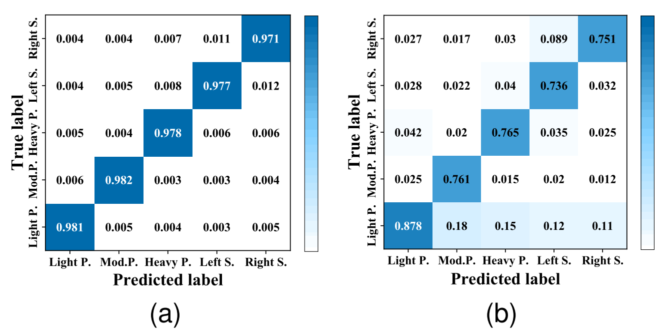



Fig. 11. Confusion matrix-based gesture recognition results. (a) User-dependent test. (b) User-independent test.

The results averaged over five gestures for each age group are shown in Fig. 10(b) and the upper part of Table II. The results indicate that all age groups achieve a precision and recall rate of over 97%. The elderly group exhibits slightly lower performance, which may be attributed to factors such as gesture instability, irregular motion duration, or changes in sensor-skin contact. Importantly, no significant systematic performance degradation is observed across different user groups.

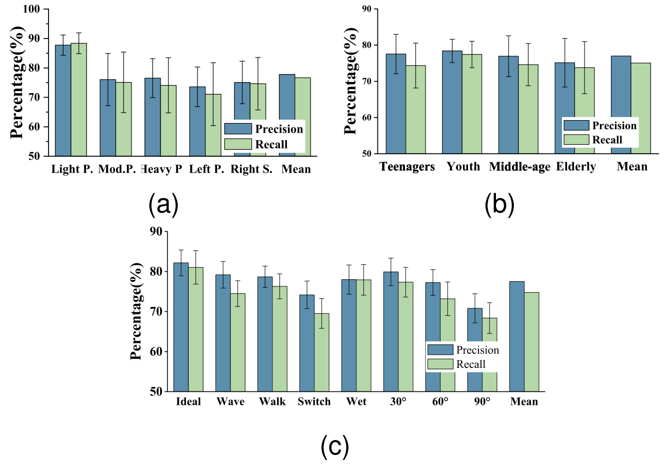



Fig. 12. Results of gesture recognition with user-independence. (a) Per gesture. (b) Per age group. (c) Per experimental condition.

The confusion matrix shown in Fig. 11(a) further supports the model's effectiveness. Diagonal values are consistently higher than 0.97, demonstrating strong classification performance. Nondiagonal entries remain extremely low, all below 0.7%, indicating minimal confusion among gesture categories. This result confirms that the model is capable of accurately distinguishing between similar gestures and maintaining robustness in the presence of intraclass variability.

*b) User-Independent Test:* To evaluate cross-individual generalization, a user-independent experiment is conducted. Recognition performance across different gesture categories is presented in Fig. 12(a), with corresponding metrics summarized in the lower part of Table I. The model achieves an average precision of 77.8%, indicating a moderate level of generalization when confronted with previously unseen users. Among all gestures, the light pinch gesture exhibits the highest recognition precision at 87.8%, likely due to its well-defined temporal profile and distinct motion features. In contrast, the left slide gestures show relatively lower precision at 73.6%, suggesting that gestures involving slide operation present more subtle signal characteristics, which may overlap with other gesture types and increase classification ambiguity. This observation underscores the inherent challenge of generalizing force-sensitive gesture recognition across users with diverse physical behaviors.

The average gesture recognition performance for each age group is visualized in Fig. 12(b), with corresponding numerical results summarized in the lower part of Table II. The variation in group results reflects the diversity of gesture execution styles across participants. Each age group achieves precision values above 75%. The lower precision can be attributed to differences in individual gesture habits, as their gesture patterns may differ significantly from those in the training set or be influenced by environmental factors.

A detailed view of the classification confusion is illustrated in Fig. 11(b). The left swipe gesture is the most frequently misclassified, often confused with the right swipe gesture. This is likely due to the inherent symmetry and overlapping motion trajectories of lateral swipe gestures, which can be further complicated by user-specific execution speed and angle.

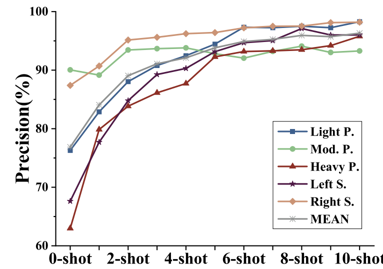



Fig. 13. Fine-tuning impact on cross-user recognition.

*c) Rapid Personalization:* Despite the relatively moderate performance in a fully user-independent setting, the model demonstrates strong potential for rapid personalization through lightweight fine-tuning. Fig. 13 reveals a clear upward trend in recognition precision as the number of adaptation samples increases. The average precision improves from below 80% in the zero-shot setting to over 90% with six-shot adaptation. Notably, even a single-shot update provides significant performance gains, underscoring the efficiency of few-shot transfer. For certain gesture types, precision approaches near-perfect levels after only a few samples, demonstrating the model's high plasticity. Fine-tuning time measurements on an NVIDIA GeForce RTX 2080 Ti indicate that the adaptation process completes in under 20 s, confirming the feasibility of real time, user-specific optimization in practical applications.

*d) Robustness Testing:* The recognition results under different conditions are shown in Figs. 10(c) and 12(c) and Table III.

1) *Hand Wave:* Precision reaches 97.8% (user-dependent) and 79.2% (user-independent), with only 0.2% and 3.0% drops, respectively. These results highlight the effectiveness of impedance signals in mitigating motion artifacts, supported by noise-suppression mechanisms that isolate gesture-specific signal components.
2) *Walking:* The system achieves 97.6% (user-dependent) and 78.7% (user-independent), showing reductions of 0.4% and 3.5%. These findings confirm the model's resilience to locomotion-induced signal variation and its suitability for mobile interaction contexts.
3) *Position Switch:* Recognition scores are 97.5% and 74.2%, showing reductions of 0.5% and 8%, indicating negligible impact from finger joint variation. This robustness to positional variance demonstrates the system's flexibility in accommodating diverse user pReferences and wearing habits to a certain extent.
4) *Wet Hand:* Even with surface moisture, precision remains at 97.9% and 78.0%, dropping by only 0.1% and 4.2%. These results affirm the system's electrical and algorithmic stability under varying skin conditions.
5) *Direction Shift:* Precision averages 97.7% (user-dependent) and 76.0% (user-independent), decreasing by 0.6% and 11.4% versus ideal conditions. This degradation can be attributed to the lack of rotational symmetry in the device, as the IMU sensor exhibits strong directional sensitivity. Angular shifts alter the spatial orientation of the IMU, leading to inconsistent signal patterns and reduced model generalizability.

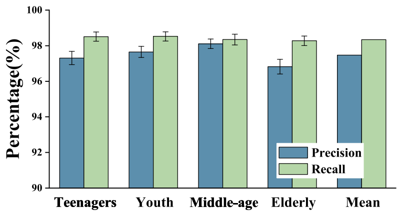
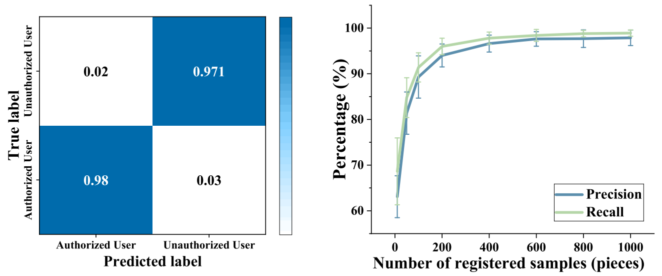



Fig. 14. Results per age group in user authentication.




(a) (b)

Fig. 15. User authentication results. (a) Confusion matrix. (b) Influence of sample quantity in registration on authentication results.

The consistently high performance across most scenarios confirms the system's robustness to common environmental perturbations. The model demonstrates resilience to motion artifacts, body movement, skin impedance variability, and sensor placement differences. These results highlight the feasibility of deploying M-Ring in diverse, unconstrained usage conditions. However, we also observe notable performance degradation under direction shift, highlighting the system's sensitivity to angular displacement. Future work may address this issue by incorporating angle-invariant signal processing techniques or redesigning the sensing layout to enhance geometric symmetry and reduce orientation dependence.

*2) User Authentication Experiments:* The experimental setup is described in Section V-A2.

*a) User Authentication Performance:* Fig. 14 and Table IV presents authentication performance across user. The model achieves an average precision of 97.5% and a recall of 98.3%, indicating high reliability in both accepting genuine users and rejecting impostors. Among all users, 18 achieve above-average precision, and recall remains consistently high across the board. The confusion matrix is shown in Fig. 15(a). The accuracy is 0.980 for authorized user prediction and 0.971 for unauthorized user rejection, confirming strong interuser discriminability. Lower precision in some cases may be attributed to limited intrauser variability or inconsistent gesture execution, which can be mitigated through diversified enrollment samples or feature enhancement strategies.

*b) Enrollment Effort Analysis:* The impact of registration sample size is plotted in Fig. 15(b). With only 50 samples, precision exceeds 85%. At 200 samples, both precision and recall exceed 95%, beyond which improvement saturates. This suggests that reliable authentication can be achieved with approximately 2 min of registration data.

*3) Ablation Experiments:* The experimental setup is described in Section V-A3. Ablation results are summarized in Fig. 16(a).

1) *M-Ring:* The complete model achieves the highest performance, with gesture recognition and user authentication both reaching around 98%, serving as the performance upper bound for comparison.
2) *w/o IC:* Gesture recognition is reduced to 87%, and user authentication is reduced to 81%, indicating that impedance contributes auxiliary discriminative cues—particularly for authentication robustness.
3) *w/o IM:* Gesture recognition drops to 37%, and user authentication dropped to 87%, representing a significant performance degradation. This highlights the dominant role of motion dynamics in both tasks, especially gesture recognition.
4) *w/o MSCE:* Gesture recognition is reduced to 72%, while user authentication remains at 82%. This demonstrates that multiscale features are critical for capturing the spatial–temporal complexity of gestures.
5) *w/o SCL:* Authentication accuracy drops to 90%. This result confirms its effectiveness in enforcing discriminative embeddings and improving fine-grained user separability. This demonstrates that contrastive learning enhances feature representation, improving classification by strengthening the distinction between different users.
6) *w/o SN:* Authentication performance is slightly reduced to 92%. This demonstrates that the Siamese structure enhances the pairwise comparison capability, reinforcing decision boundaries in user verification.

*4) Latency and Power Analysis:* The experimental setup is described in Section V-A4.

*a) System Latency:* The average time for signal preprocessing is measured at 4.9 $\pm$ 13.2 ms. The subsequent deep learning inference, including gesture recognition and identity verification, requires an average of 25.0 $\pm$ 1.3 ms. The total end-to-end latency is 29.9 ms, which is significantly lower than the commonly accepted 500-ms threshold for real-time interaction in XR systems. This confirms the suitability of M-Ring for responsive, low-latency user interaction.

*b) Power Consumption:* The measured average power consumption of the ring is 208 mW, resulting in a continuous battery-powered operational time of approximately 85 min. This indicates sufficient autonomy for short-session XR experiences and intermittent authentication scenarios. On the host device, executing the M-Ring inference pipeline yields an average power consumption overhead of 320 $\pm$ 7.6 mW. This level of energy demand is within the typical thermal and battery budget of a commercial XR Headset, indicating that the authentication algorithm can be sustained alongside other computational tasks without compromising user experience.

*5) Usability Evaluation in XR Scenarios:* The experimental setup is described in Section V-A5. As shown in the result of Fig. 16(b), the system achieves a score of 85.1, classified as "excellent" according to Bangor et al.'s scale [49]. Suggested improvements focus on refining the edge finishing of the flexible circuit board to enhance wearing comfort, simplifying the procedures for wearing and adjusting the device, and advancing the miniaturization and industrial design for improved esthetics and user acceptance. These findings confirm the system's usability and also inform further hardware refinement and ergonomic optimization.

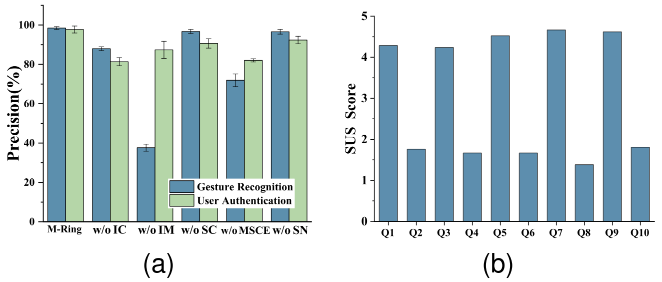



Fig. 16. (a) Ablation results. (b) SUS scores. M-Ring: complete model of M-Ring system; w/o: without; IC: impedance channel data; IM: IMU motion data; SC: supervised contrastive learning module based on MLP; MSCE: multiscale convolution expert structure; SN: Siamese network structure; and $Q$: question index of the SUS scale.

TABLE III
AVERAGE GESTURE RECOGNITION RESULTS PER CONDITION

| Condition | Precision(%) | Recall(%) | Accuracy(%) | Specificity(%) | F1 Score |
|---|---|---|---|---|---|
| **User-Dependent** | | | | | |
| Ideal | 98.0 | 98.0 | 98.0 | 99.1 | 0.980 |
| Wave | 97.8 | 97.8 | 97.8 | 99.1 | 0.978 |
| Walk | 97.6 | 97.6 | 97.6 | 99.0 | 0.976 |
| Switch | 97.5 | 97.5 | 97.5 | 99.0 | 0.975 |
| Wet | 97.9 | 97.9 | 97.9 | 99.1 | 0.979 |
| 30° | 98.0 | 98.0 | 98.0 | 99.5 | 0.980 |
| 60° | 97.7 | 97.7 | 97.7 | 99.0 | 0.977 |
| 90° | 97.4 | 97.1 | 97.1 | 99.0 | 0.972 |
| Mean | 97.7 | 97.7 | 97.7 | 99.1 | 0.977 |
| **User-Independent** | | | | | |
| Ideal | 82.2 | 81.0 | 81.1 | 95.3 | 0.815 |
| Wave | 79.2 | 74.5 | 74.5 | 93.6 | 0.768 |
| Walk | 78.7 | 76.3 | 75.8 | 94.1 | 0.774 |
| Switch | 74.2 | 69.5 | 69.6 | 92.4 | 0.718 |
| Wet | 78.0 | 77.9 | 77.9 | 94.5 | 0.780 |
| 30° | 79.9 | 77.3 | 76.8 | 94.2 | 0.786 |
| 60° | 77.3 | 73.2 | 71.9 | 93.0 | 0.752 |
| 90° | 70.8 | 68.4 | 67.6 | 91.9 | 0.696 |
| Mean | 77.5 | 74.8 | 74.4 | 93.6 | 0.762 |

TABLE IV
RESULTS FOR EACH USER IN USER AUTHENTICATION

| Group | Precision(%) | Recall(%) | Accuracy(%) | Specificity(%) | F1 Score |
|---|---|---|---|---|---|
| Teenagers | 97.3 | 98.5 | 98.1 | 98.5 | 0.979 |
| Youth | 97.6 | 98.5 | 98.2 | 98.5 | 0.980 |
| Middle-age | 98.1 | 98.0 | 98.0 | 98.0 | 0.980 |
| Elderly | 96.8 | 98.3 | 97.9 | 98.3 | 0.975 |
| Mean | 97.5 | 98.3 | 98.0 | 98.3 | 0.979 |

## VI. DISCUSSION AND FUTURE WORK

### A. Comparison of Similar Interaction Schemes

To assess the relative merits of M-Ring, we compare it against recent ring-based or vision interaction systems developed between 2021 and 2025, of which Ebrahimian et al. [48] is a nonring scheme, as summarized in Table V. Compared with most existing studies, M-Ring achieves superior gesture recognition across a wider range of gesture types while simultaneously supporting user authentication, whereas prior systems typically rely on unimodal sensing or support fewer gesture classes.

Noteworthy is that although studies like ATH-Rings [42] have achieved ultrahigh accuracy, they rely on bulky multimodal sensor arrays in a glove form factor and do not incorporate user authentication. Similarly, ElectroRing [20] focuses solely on simple pinching gestures, with a task complexity far lower than the six dynamic gestures supported by M-Ring. Unlike bulky or wired designs, M-Ring adopts a compact form factor ($4.0\times2.6\ \times0.7$ cm, 6.1 g) with Bluetooth connectivity, supporting real-time operation. It is also the only system that enables collaborative authentication during natural interaction, whereas other systems either omit authentication or implement it in isolation.

In ideal laboratory environments, vision systems (see [4], [47], [48]) achieve high accuracies of up to 98.5% for gesture recognition and 99.4% for authentication. However, vision-based solutions suffer from core limitations, including reliance on specialized imaging equipment and fixed environments, sensitivity to lighting variations and posture occlusions, privacy risks from excessive visual data collection, high computational demands that hinder mobile deployment, and weak resistance to static forgery. In contrast, M-Ring achieves superior robustness in complex daily scenarios by leveraging lightweight wearable hardware design for scene adaptability, employing multimodal sensing to suppress environmental interference, utilizing lightweight design to support mobile deployment, and enhancing antiforgery capabilities through physiological feature fusion.

### B. Generalizability and Potential Applications

While the M-Ring system is designed for XR interaction, its core capabilities—including multimodal sensing, collaborative authentication, and low-latency gesture recognition—offer strong potential for broader application domains, reducing reliance on touch-based interfaces. In automotive contexts, M-Ring may support in-cabin user verification and hands-free operation, improving both security and driving safety. In assistive communication and special education, real-time gesture sensing can facilitate gesture-based language input, enhancing accessibility for users with speech or motor impairments. In addition, in rehabilitation settings, fine-grained gesture tracking can be leveraged for interactive motor training and progress monitoring.

### C. From Simple to Sophisticated Gestures

The current system supports a set of fundamental gestures optimized for real-time recognition and authentication. Building on this foundation, it can be scaled to support more complex and dynamic gesture vocabularies, including long presses, continuous swipes, multitap sequences, and compound gestures. These extended gestures can be systematically mapped to application-level functions, such as adjusting volume, navigating interfaces, or controlling object parameters in 3-D space. Future research will explore bimanual interaction, such as defining functional zones on the left hand while using the right hand wearing the ring device for touch detection.

TABLE V
COMPARISON OF RING-BASED INTERACTION SYSTEMS

| Method | Sensor Type | Multi-task | Size | Power Consumption | Accuracy | Wireless | Real-time | Auth. |
|---|---|---|---|---|---|---|---|---|
| ATH-Rings [42] | Abundant | Independent | bulky | Self-powered | 14 classes 99.8% | Wireless | Yes | No |
| ElectroRing [20] | Impedance | No | small | 220mW | 2 classes 99.4% | Bluetooth | Yes | No |
| Zhao [43] | IMU | No | small | N/A | 26 classes 97.1% | Wireless | Yes | No |
| DualRing [44] | IMU$\times$2+Impedance | No | small | N/A | 4 classes 97% | Wired | Yes | No |
| SigningRing [45] | IMU | No | $3\times1.8\times1.5$ cm | N/A | Auth. 97.4% | Wireless | Yes | Dedicated |
| Iwakiri [46] | Active acoustics | No | small | N/A | Auth. 97.0% | Wired | Yes | Dedicated |
| Z-Ring [15] | Impedance | Independent | bulky | N/A | 5 classes 93.1% | Wired | No | Dedicated |
| EFRing [17] | E-field induction | No | small | N/A | 9 classes 89.5% | Bluetooth | Potential | No |
| Z-Pose [16] | Radio frequency | No | bulky | N/A | Pose 7.2mm | Wired | Potential | No |
| Wu [47] | DualCam-Wrist | No | $4.5\times2.5\times1.5$ cm | 410 mW | 21 classes 86.03% | Wired | Yes | No |
| Nguyen [4] | Wrist-wear camera | No | bulky | N/A | 12 classes 98.48% | Wired | No | No |
| Zahra [48] | Camera | No | N/A | N/A | Auth. 99.4% | N/A | Potential | Dedicated |
| M-Ring **(Ours)** | **Impedance & IMU** | **Sync** | **$10.0\times4.5\times4.5$ cm** | **320mW** | **6 classes 97.8%** | **Bluetooth** | **Yes** | **Collaborative** |

*Auth.*: Authentication, *Indep.*: Independently implemented, *Pose*: Position Tracking. *Real-time: "Yes" denotes interactive latency (<10ms) or system-claimed real-time capability; "Potential" indicates feasibility with optimization; classification relies on system descriptions when latency data is unavailable. Power Consumption:* Based on reported values or system features; "N/A" indicates data not provided.

### *D. Cross-Environment Robustness and System Optimization*

The system exhibits performance variation across users, reflecting limited generalization under user-independent conditions. As discussed in Section V-B1, while fine-tuning with few samples offers improvement, future work expands the dataset to include diverse hand morphologies, gesture styles, and environmental settings. Transfer learning and meta-learning approaches will be introduced to enhance adaptability with minimal calibration. Although inference is optimized for edge deployment, further work is needed on model compression, power-efficient algorithms, and BLE stability. Reliable impedance sensing also requires consistent skin contact, which may be affected by user behavior. To address this, hardware upgrades focus on refining circuit layout, adopting flexible biocompatible materials, and improving power efficiency. Based on user feedback in Section V-B5, enhancements include improved fixation, reduced skin friction, and enhanced comfort and esthetics for daily wear.

## VII. CONCLUSION

We present M-Ring, a compact finger-worn system that unifies gesture interaction and passive authentication via integrated impedance and IMU sensing. A MTL framework with adaptive-scale convolutional modules efficiently captures complementary spatiotemporal patterns across modalities, enhancing task synergy and reducing computational overhead. A Siamese contrastive branch further improves the discriminability of user-specific representations. Through hardware–software co-design, M-Ring achieves high-precision dual-purpose sensing with low latency and power consumption, where natural gestures simultaneously serve as interaction commands and identity cues, offering a scalable solution for secure and intuitive XR interaction.

## REFERENCES


[1] C. Warin and D. Reinhardt, "Vision: Usable privacy for XR in the era of the metaverse," in *Proc. Eur. Symp. Usable Secur.*, Karlsruhe, Germany, Sep. 2022, pp. 111–116, doi: 10.1145/3549015.3554212.

[2] T. Tan and G. Cao, "Thermal-aware scheduling for deep learning on mobile devices with NPU," *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 10706–10719, Dec. 2024, doi: 10.1109/TMC.2024.3379501.

[3] P. Streli, R. Armani, Y. F. Cheng, and C. Holz, "HOOV: Hand out-of-view tracking for proprioceptive interaction using inertial sensing," in *Proc. CHI Conf. Human Factors Comput. Syst.* Germany: ACM, Apr. 2023, pp. 1–16, doi: 10.1145/3544548.3581468.

[4] R. Nguyen, C. Gouin-Vallerand, and M. Amiri, "Hand interaction designs in mixed and augmented reality head mounted display: A scoping review and classification," *Frontiers Virtual Reality*, vol. 4, pp. 1–17, 2023.

[5] Y. Akamatsu et al., "Phase angle from bioelectrical impedance analysis is a useful indicator of muscle quality," *J. Cachexia, Sarcopenia Muscle*, vol. 13, no. 1, pp. 180–189, Feb. 2022, doi: 10.1002/jcsm.12860.

[6] J. Wang, Y. Sui, C. Liu, X. Shen, Z. Li, and D. Yu, "Multi-task learning with multi-gate mixture of transformer-based experts for predictive process monitoring in manufacturing," *Sci. Prog.*, vol. 107, no. 4, pp. 1–38, Oct. 2024, doi: 10.1177/00368504241292196.

[7] H. Tang, J. Liu, M. Zhao, and X. Gong, "Progressive layered extraction (PLE): A novel multi-task learning (MTL) model for personalized recommendations," in *Proc. 14th ACM Conf. Recommender Syst.*, Brazil, Sep. 2020, pp. 269–278, doi: 10.1145/3383313.3412236.

[8] G. Koch, R. S. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*. Lille, France, Jul. 2015, pp. 1–8.

[9] M. N. T. Chen, S. Kornblith, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn. (ICML)*, H. D. III and A. Singh, Eds., vol. 119, Jul. 2020, pp. 1597–1607.

[10] M. Kolsch, R. Bane, T. Hollerer, and M. Turk, "Multimodal interaction with a wearable augmented reality system," *IEEE Comput. Graph. Appl.*, vol. 26, no. 3, pp. 62–71, May 2006, doi: 10.1109/MCG.2006.63.

[11] V. Rantanen et al., "A wearable, wireless gaze tracker with integrated selection command source for human-computer interaction," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 5, pp. 795–801, Sep. 2011, doi: 10.1109/TITB.2011.2158321.

[12] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura, "Voice in human–agent interaction: A survey," *ACM Comput. Surveys*, vol. 54, no. 4, pp. 1–43, May 2021, doi: 10.1145/3386867.

[13] T. G. Zimmerman, J. Lanier, C. Blanchard, S. Bryson, and Y. Harvill, "A hand gesture interface device," *ACM SIGCHI Bull.*, vol. 17, no. SI, pp. 189–192, May 1986, doi: 10.1145/1165387.275628.

[14] Z. Zhu, X. Wang, A. Kapoor, Z. Zhang, T. Pan, and Z. Yu, "EIS: A wearable device for epidermal American sign language recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 1–22, Dec. 2018, doi: 10.1145/3287069.

[15] A. Waghmare, Y. B. Taleb, I. Chatterjee, A. Narendra, and S. Patel, "Z-ring: Single-point bio-impedance sensing for gesture, touch, object and user recognition," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Hamburg, Germany, Apr. 2023, pp. 1–18, doi: 10.1145/3544548.3581422.

[16] A. Waghmare, I. Chatterjee, and S. Patel, "Z-pose: Continuous 3D hand pose tracking using single-point bio-impedance sensing on a ring," in *Proc. 2nd Workshop Smart Wearable Syst. Appl.*, Atlanta, GA, USA, Oct. 2023, pp. 1–6, doi: 10.1145/3615592.3616851.
[17] X. Y. T. Chen, T. Li, and K. Zhu, "EFRing: Enabling thumb-to-index-finger microgesture interaction through electric field sensing using single smart ring," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 6, no. 4, pp. 1–31, Jan. 2023, doi: 10.1145/3569478.
[18] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst., Man, Cybern., A, Syst. Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011, doi: 10.1109/TSMCA.2011.2116004.
[19] M. Wilhelm, D. Krakowczyk, and S. Albayrak, "PeriSense: Ring-based multi-finger gesture interaction utilizing capacitive proximity sensing," *Sensors*, vol. 20, no. 14, p. 3990, Jul. 2020, doi: 10.3390/s20143990.
[20] W. Kienzle, E. Whitmire, C. Rittaler, and H. Benko, "ElectroRing: Subtle pinch and touch detection with a ring," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Yokohama, Japan, May 2021, pp. 1–12, doi: 10.1145/3411764.3445094.
[21] K. Rhee and H.-C. Shin, "Electromyogram-based hand gesture recognition robust to various arm postures," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 7, pp. 15–27, Jul. 2018, doi: 10.1177/1550147718790751.
[22] K. Suri and R. Gupta, "Continuous sign language recognition from wearable IMUs using deep capsule networks and game theory," *Comput. Electr. Eng.*, vol. 78, pp. 493–503, Sep. 2019, doi: 10.1016/j.compeleceng.2019.08.006.
[23] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabil. Robot.* Zurich, Switzerland: IEEE, Jun. 2011, pp. 1–7, doi: 10.1109/ICORR.2011.5975346.
[24] Y. Liu, F. Jiang, and M. Gowda, "Finger gesture tracking for interactive applications: A pilot study with sign languages," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–21, Sep. 2020, doi: 10.1145/3414117.
[25] J. Li et al., "SignRing: Continuous American sign language recognition using IMU rings and virtual IMU data," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 3, pp. 1–29, Sep. 2023, doi: 10.1145/3610881.
[26] T. Zhao, J. Liu, Y. Wang, H. Liu, and Y. Chen, "PPG-based finger-level gesture recognition leveraging wearables," in *Proc. IEEE Conf. Comput. Commun.*, Honolulu, HI, USA, Apr. 2018, pp. 1457–1465, doi: 10.1109/INFOCOM.2018.8486006.
[27] G. Peng, G. Zhou, D. T. Nguyen, X. Qi, Q. Yang, and S. Wang, "Continuous authentication with touch behavioral biometrics and voice on wearable glasses," *IEEE Trans. Hum.-Mach. Syst.*, vol. 47, no. 3, pp. 404–416, Jun. 2017, doi: 10.1109/THMS.2016.2623562.
[28] J. Chauhan, Y. Hu, S. Seneviratne, A. Misra, A. Seneviratne, and Y. Lee, "BreathPrint: Breathing acoustics-based user authentication," in *Proc. 15th Annu. Int. Conf. Mobile Syst., Appl., Services*. New York, NY, USA: ACM, Jun. 2017, pp. 278–291, doi: 10.1145/3081333.3081355.
[29] M. Martinho, A. Fred, and H. Silva, "Towards continuous user recognition by exploring physiological multimodality: An electrocardiogram (ECG) and blood volume pulse (BVP) approach," in *Proc. Int. Symp. Sens. Instrum. IoT Era (ISSI)*. Shanghai, China: IEEE, Sep. 2018, pp. 1–6, doi: 10.1109/ISSI.2018.8538075.
[30] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.* Tokyo, Japan: ACM, Oct. 2018, pp. 400–408, doi: 10.1145/3242969.3242985.
[31] H. Feng, K. Fawaz, and K. G. Shin, "Continuous authentication for voice assistants," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.* New York, NY, USA: ACM, Oct. 2017, pp. 343–355, doi: 10.1145/3117811.3117823.
[32] R. Ryu, S. Yeom, S.-H. Kim, and D. Herbert, "Continuous multimodal biometric authentication schemes: A systematic review," *IEEE Access*, vol. 9, pp. 34541–34557, 2021.
[33] J. S. Arteaga-Falconi, H. Al Osman, and A. El Saddik, "ECG and fingerprint bimodal authentication," *Sustain. Cities Soc.*, vol. 40, pp. 274–283, Jul. 2018, doi: 10.1016/j.scs.2017.12.023.
[34] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997, doi: 10.1023/A:1007379606734.
[35] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022, doi: 10.1109/TKDE.2021.3070203.
[36] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*.
[37] E. S. Olivas, J. D. M. Guerrero, M. Martínez-Sober, J. R. Magdalena-Benedito, and A. J. S. López, *Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques*. Hershey, PA, USA: IGI Global, 2009, doi: 10.4018/978-1-60566-766-9.
[38] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. London, U.K.: ACM, Jul. 2018, pp. 1930–1939.
[39] A. Liu et al., "DeepSeek-V3 technical report," 2024, *arXiv:2412.19437*.
[40] M. Göktürk and J. L. Sibert, "An analysis of the index finger as a pointing device," in *Proc. Extended Abstr. Human Factors Comput. Syst. (CHI)*. Pittsburgh, PA, USA: ACM, May 1999, p. 286, doi: 10.1145/632716.632892.
[41] H. W. Noh, C.-G. Ahn, H.-J. Kong, and J. Y. Sim, "Ratiometric impedance sensing of fingers for robust identity authentication," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Sep. 2019, doi: 10.1038/s41598-019-49792-9.
[42] Z. Sun, M. Zhu, X. Shan, and C. Lee, "Augmented tactile-perception and haptic-feedback rings as human-machine interfaces aiming for immersive interactions," *Nature Commun.*, vol. 13, no. 1, pp. 1–13, Sep. 2022, doi: 10.1038/s41467-022-32745-8.
[43] Y. Zhao et al., "A smart wireless IoT ring for real-time keystroke recognition using edge computing," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022, doi: 10.1109/TIM.2022.3214631.
[44] C. Liang, C. Yu, Y. Qin, Y. Wang, and Y. Shi, "DualRing: Enabling subtle and expressive hand interaction with dual IMU rings," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 3, pp. 1–27, Sep. 2021, doi: 10.1145/3478114.
[45] Y. Cao, A. Dhekne, and M. Ammar, "SigningRing: Signature-based authentication using inertial sensors on a ring form-factor," in *Proc. Workshop Body-Centric Comput. Syst.* South Korea: ACM, Jun. 2024, pp. 11–16, doi: 10.1145/3662009.3662019.
[46] S. Iwakiri and K. Murao, "User authentication method for wearable ring devices using active acoustic sensing," in *Proc. Int. Symp. Wearable Comput.*, Oct. 2023, pp. 17–21, doi: 10.1145/3594738.3611357.
[47] K. Wu, W. Zhang, Q. Zhang, and X. Hu, "Stretchable and wearable sensors for contact touch and gesture recognition based on poling-free piezoelectric polyester elastomer," *Polymers*, vol. 17, no. 8, p. 1105, Apr. 2025, doi: 10.3390/polym17081105.
[48] Z. Ebrahimian, S. A. Mirsharji, R. Toosi, and M. A. Akhaee, "Automated person identification from hand images using hierarchical vision transformer network," in *Proc. 12th Int. Conf. Comput. Knowl. Eng. (ICCKE)*, Nov. 2022, pp. 398–403.
[49] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *Int. J. Human-Computer Interact.*, vol. 24, no. 6, pp. 574–594, Jul. 2008, doi: 10.1080/10447310802205776.

**Shilin Chen** received the bachelor's degree from the School of Computer, University of South China, Hengyang, China, in 2023. He is currently pursuing the master's degree with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

His research focuses on mobile computing and human–machine interaction.

**Yuqi Cheng** received the bachelor's degree from the School of Mathematics and Physics, Wuhan Institute of Technology, Wuhan, China, in 2024. She is currently pursuing the master's degree with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

Her research interests include affective computing and mobile computing.

**Yuling Tan** is currently pursuing the bachelor's degree with the College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China.

Her research interests cover intelligent hardware and ubiquitous computing.

**Chaonan Tang** received the bachelor's degree from the College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China, in 2025.

His research focuses on hardware development and electrical signal acquisition.

**Qian Zheng** received the Ph.D. degree from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2022.

He is currently an Associate Researcher with the Intelligent Service Computing Research Center, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include neural oscillation separation, closed-loop neuromodulation, and embodied emotional interaction.

**Yongpan Zou** (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering (CSE), The Hong Kong University of Science and Technology, Hong Kong, in 2017.

He is currently an Associate Professor with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include ubiquitous sensing, mobile computing, and human–computer interaction.

**Victor C. M. Leung** (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering from The University of British Columbia, Vancouver, Canada, in 1981.

He is currently the Dean of the Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen, China. His research focuses on wireless networks and mobile systems with more than 60 000 citations.

Dr. Leung has received numerous accolades, such as the APEBC Gold Medal, NSERC Postgraduate Scholarships, and the IEEE Vancouver Section Centennial Award.

**Kaishun Wu** (Fellow, IEEE) received the Ph.D. degree from The Hong Kong University of Science and Technology, Hong Kong, in 2011.

He is currently a Professor with the Information Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. His research interests include wireless communications and mobile computing.

Dr. Wu won several best paper awards at international conferences, such as IEEE GLOBECOM 2012 and IEEE MASS 2014.