Research article

Advancing biomedical engineering through a multi-modal sensor fusion system for enhanced physical training

Yi Deng[1], Zhiguo Wang[1], Xiaohui Li[2,3], Yu Lei [4,*] and Owen Omalley[5,*]

[1] College of Physical Education, Hunan International Economics University, Changsha 410012, China

[2] Department of Wushu and China, Songshan Shaolin Wushu College, Zhengzhou 452470, China

[3] Department of History and Pakistan, University of the Punjab, Lahore 54000, Pakistan

[4] College of Humanities and Arts, Hunan International Economics University, Changsha 410012, China

[5] NexTech Innovations, Mountain View, California 94043, USA

* **Correspondence:** Email: xiaoyu9767@gmail.com; owen.omalley7@gmail.com

Abstract: In this paper, we introduce a multi-modal sensor fusion system designed for biomedical engineering, specifically geared toward optimizing physical training by collecting detailed body movement data. This system employs inertial measurement units, flex sensors, electromyography sensors, and Microsoft's Kinect V2 to generate an in-depth analysis of an individual's physical performance. We incorporate a gated recurrent unit- recurrent neural network algorithm to achieve highly accurate body and hand motion estimation, thus surpassing the performance of traditional machine learning algorithms in terms of accuracy, precision, recall, and F1 score. The system's integration with the PICO 4 VR environment creates a rich, interactive experience for physical training. Unlike conventional motion capture systems, our sensor fusion system is not limited to a fixed workspace, allowing users to engage in exercise within a flexible, free-form environment.

Keywords: biomedical engineering; physical training; sensor fusion; virtual reality; multi-modal system; motion estimation; body movement data; gated recurrent unit; recurrent neural network

1. Introduction

Physical training, an integral aspect of modern life, encompasses various systematic and targeted exercise activities aimed at improving one's physical fitness, overall health and well-being [1]. The significance of physical training extends beyond individual benefits, as it not only helps to prevent

various health issues such as obesity, cardiovascular diseases, and diabetes, but it also enhances mental health, boosts self-esteem and fosters social connections [2–5]. From a societal perspective, promoting physical training contributes to reduced healthcare costs, increased productivity and improved quality of life, making it a vital factor in the well-being of communities and nations alike [6]. As the demand for efficient and personalized physical training methods grows, there is an emerging need to leverage advanced technologies to optimize training outcomes and maximize the benefits derived from exercise.

Physical training has evolved significantly. Figure 1 illustrates the technological advancements in fitness equipment over time. In the 18th and early 19th centuries, physical training primarily served military purposes, emphasizing the development of physical prowess for combat readiness [7]. The inception of the modern Olympic Games in the early 20th century marked a turning point, as various forms of exercise and fitness training emerged, including aerobic exercise, strength training, yoga and Pilates [8]. The advent of modern technology has further revolutionized physical training, employing tools such as a smart bracelet and intelligent body fat scales [9]. However, current physical training methodologies face notable challenges. Most devices on the market are limited to single-modality monitoring, providing an incomplete understanding of an individual's overall physical condition [10–12]. Addressing this gap, multi-modal sensor fusion systems that combine data from various sensors may serve as solutions to provide a comprehensive picture of an individual's fitness [13]. These systems have the potential to significantly enhance the effectiveness of physical training by offering more accurate, personalized feedback and recommendations based on a holistic understanding of the trainee's physical state.
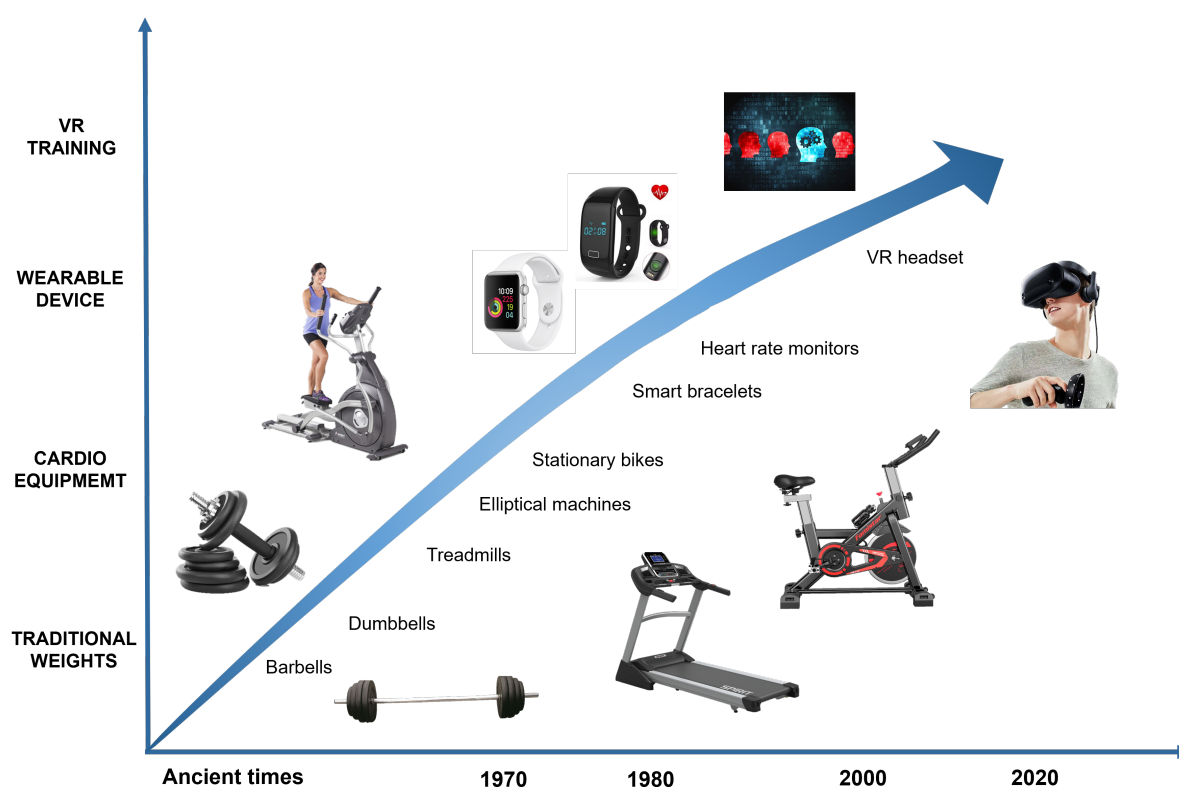

Figure 1. The digitization process for fitness equipment.

In light of these challenges, the primary objective of this research was to validate the effectiveness of the proposed multi-modal sensor fusion system in enhancing physical training performance. The developed system integrates a diverse range of sensors, including inertial measurement units (IMUs), heart rate sensors, electromyography (EMG) sensors, pressure sensors, GPS sensors, cameras, computer vision systems and environmental sensors. The research methodology here combines principles from geometry, biology and kinematics to develop a comprehensive understanding of an individual's physical state and performance during training. Figure 2 presents the design idea of the system and the core technologies involved.
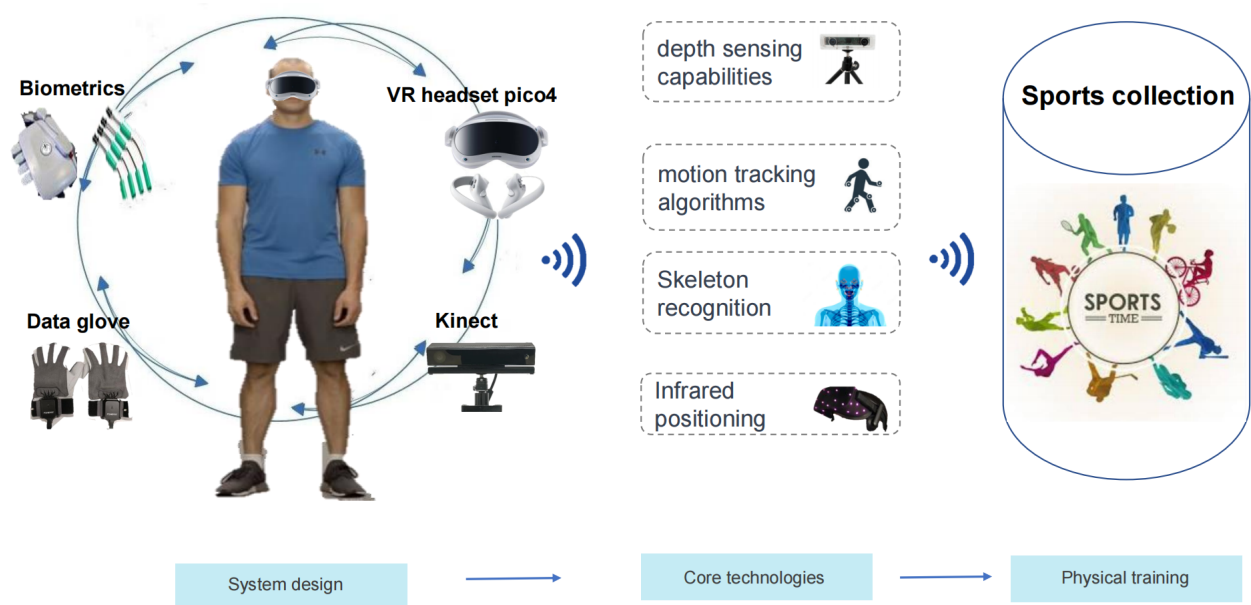

Figure 2. The system design and core technologies.

This study had the following goals. First, we introduce a multi-modal sensor fusion system that revolutionizes the approach to physical training. Our system has been designed, implemented and thoroughly evaluated to provide a distinctive and innovative approach, surpassing the capabilities of conventional single-modality systems. Integrating multiple sensors enables the acquisition of richer and more comprehensive data and insights, facilitating a deeper understanding of an individual's physical condition during training.

Second, the system incorporates advanced sensor technology and employs sophisticated data processing algorithms. This integration ensures the system's ability to monitor various aspects of an individual's physical condition with high accuracy and reliability. Through the utilization of cutting-edge sensors and advanced algorithms, we enhance the precision of data collection and analysis, enabling more effective monitoring of the trainee's physical state.

Last, the proposed multi-modal sensor fusion system showcases its potential to significantly enhance physical training outcomes. By leveraging the comprehensive data and insights obtained from multiple sensors, the system provides tailored feedback and recommendations. This

personalized approach, rooted in a comprehensive understanding of the trainee's fitness level and performance, has the potential to optimize training routines and unlock improved outcomes in physical training.

In essence, this paper delves into the utilization of multi-modal sensor fusion systems for physical training. The subsequent sections are organized as follows. Section 2 provides an extensive literature review, examining prior work in the field. Building upon this foundation, section 3 outlines the proposed methodology, elucidating the system design. In Section 4, the implementation of our multi-modal sensor fusion system is presented. Section 5 verifies the effectiveness of our approach through experimental demonstrations. Finally, Section 6 concludes the paper by summarizing the principal contributions and suggesting future research directions.

2. Related work

2.1. Single-modality monitoring for physical training

Within the realm of physical training, various single-modality monitoring approaches have been utilized, such as force sensors, depth cameras, IMUs and heart rate monitors, to track different facets of human performance. However, these individual methods are not without their limitations, including decreased precision and unreliable data. To address these shortcomings, a wearable sensing system was developed to enhance rehabilitation for neurological patients by employing an electronic glove equipped with compact force sensors and a modified interface to measure fingertip force [14]. Additionally, a novel teaching methodology was introduced, employing a gamified learning process that promotes physical activity through the use of interactive balls in educational activities [15]. Notably, this system utilized a Kinect depth camera for image recognition, enabling the detection of ball hits on activity projections. In [16], to prevent cognitive and physical decline in older individuals with mild cognitive impairment, an exergame was designed. This exergame employed IMUs worn on the wrists and feet of the participants to track their movements as they navigated through the game environment. Within the field of rehabilitation training, Batalik et al. [17] explored the potential of telerehabilitation with wrist heart rate monitors as an alternative to traditional outpatient cardiac rehabilitation. The study demonstrated significant improvements in physical fitness for both regular outpatient training and interventional home-based telerehabilitation groups. Most recently, Bo and Sun [18] developed a heart-rate monitoring platform for college sports training, utilizing wireless networks. This platform featured multi-level link equalization, a transmission numerical model for heart-rate monitoring channels, data redundancy verification and a topology stage structure.

2.2. Multi-modal sensor fusion system

The utilization of multi-modal sensor fusion systems has garnered considerable attention in diverse domains, including affective computing, robotics and automated driving [19, 20]. These systems combine data from multiple sensors to provide a more comprehensive understanding of various phenomena. A crucial aspect of our research is the application of such systems in biomedical engineering, especially in enhancing physical training. In the domain of wearables, a wearable multi-modal biosensing system has been developed to collect, synchronize, record and transmit data from diverse biosensors, such as systems of PPG, EEG, eye-gaze headset, body motion capture and

GSR [21]. The performance of these sensors was evaluated by comparing them to standard research-grade biosensors. Comparatively, a study [22] yielded a scalable design model for artificial fingers in anthropomorphic robotic and prosthetic hands. This design integrated mechanical components, embedded electronics and a multi-modal sensor system. The fully parametric design enabled automated scaling, and interchangeable electronic modules facilitated mechanical adjustments. In the context of urban autonomous driving, Marco et al. [23] proposed a multi-modal sensor fusion scheme for accurately estimating 3D vehicle velocity, pitch and roll angles. The method simultaneously estimated gyroscope and accelerometer biases to enhance accuracy, proving itself to be effective during regular urban drives and collision avoidance maneuvers.

A multitude of multi-modal sensor fusion systems have emerged, aiming to augment physical training outcomes and yield an extensive comprehension of human performance. These systems amalgamate diverse sensor types, such as GPS, heart- rate monitor, IMUs and motion-capture devices, to capture a broad range of physical activity aspects. Noteworthy contributions include Ma and Hu's development [24] of a home-oriented cyber-physical system that enhances motion coordination capabilities via physical training. Their system employs a thermal camera to record leg and foot motions, as well as insole pressure sensors for plantar pressure measurement. Innovative algorithms for leg skeleton extraction and motion signal auto-segmentation, recognition and analysis were devised.

As early as 2016, Torreño et al. [25] conducted an investigation by utilizing GPS and heart- rate technology to explore the match running profiles, heart rates and performance efficiency indices of professional soccer players during official club-level games. Similarly, the utilization of a Polar GPS device alongside heart-rate sensors allowed for the determination of motion intensity in physical education settings [26]. To address the rehabilitation requirements, a novel virtual personal trainer system, tailored for rehabilitation purposes, was proposed. This system employed nine IMU sensors and a 3D camera for comprehensive full-body motion tracking [27]. The incorporation of a 3D camera mitigated IMU sensor output instability, counteracted gyroscopic drift and reduced electromagnetic interference. The resulting wireless full-body sensor array system provided a cost-effective alternative to on-site trainers. Stroke rehabilitation-focused physical training involved the development of a portable rehabilitation platform, encompassing both mental and physical training modalities [28]. This platform encompassed an EEG-based BCI system for mental training, a force sensor-embedded orthosis for elbow extension/flexion and an FES unit for hand extension. Later, a group [29] proposed a data fusion algorithm that combined skeletal data from three Kinect devices to overcome the limitations of single-device usage for human skeletons and motion tracking within sports medicine and rehabilitation. This approach offered comprehensive 3D spatial coverage of subjects and demonstrated an enhanced tracking accuracy improvement of 15.7%. As multi-modal sensor fusion technology advances, substantial strides are anticipated in the areas of data fusion algorithms and machine learning techniques, elevating the capabilities of these systems even further [30].

Multi-modal sensor fusion systems play a pivotal role in advancing research and technology [31]. Researchers have focused on various aspects of these systems, such as PRF-PIR, unified camera tracking approaches and novel methods for human activity recognition. In the realm of PRF-PIR, Yuan et al. [32] proposed a passive, multi-modal sensor fusion system named PRF-PIR. This system consists of a software-defined radio device and a novel passive infrared (PIR) sensor system. Using a

recurrent neural network (RNN) as the human identification and activity recognition model, the PRF-PIR system accurately and non-intrusively monitors human activity in indoor environments. The system's effectiveness was validated through data collection from 11 activities performed by 12 human subjects, employing explainable artificial intelligence methodologies. In [33], Oskiper et al. came up with a unified approach for a camera tracking system, employing an error-state Kalman filter algorithm. This approach utilizes relative and global measurements obtained from image-based motion estimation, landmark matching and radio frequency-ranging radios. The proposed approach demonstrated long-term stability and overall accuracy under vision-aided and vision-impaired conditions, as evidenced by the rendering of views from a 3D graphical model and actual video images. Regarding human activity recognition, an efficient method that fuses data from inertial sensors (IMUs), surface EMG, and visual depth sensors have been developed. This approach exhibits superior performance compared to single- or dual-sensor approaches, with high robustness against occlusions, data losses and other non-structured environmental events [34].

The integration of artificial intelligence into sensor fusion technology, encompassing data fusion algorithms and machine learning methods, holds immense potential for advancements [35–37]. Xu and his colleagues [38] posited that a self-adaptive wavelet transform-based data fusion algorithm, suitable for both static and dynamic systems, can achieve an optimal estimation of a measurand with minimum mean square error in multisensor systems. Similarly, the GAPSOBP algorithm, which intelligently combines a BP neural network, genetic algorithm and particle swarm optimization algorithm, exhibits efficiency in reducing the volume of data transmitted to a base station or sink node within wireless sensor networks, thus conserving network energy [39]. In the field of physical training, the utilization of data fusion algorithms and machine learning methods has substantial value. They enable the development of multi-modal sensor fusion systems that yield richer data and insights than those of conventional single-modality systems [40]. Furthermore, the synergy between multi-modal sensor fusion systems and virtual reality (VR) presents an opportunity to enhance various applications. Through the integration of data from diverse sensors, multi-modal sensor fusion systems offer a comprehensive and precise understanding of the physical environment and human behavior. In contrast, VR creates immersive and interactive virtual environments that simulate real-world scenarios, allowing users to engage with virtual objects and surroundings. The combination of these technologies results in a robust and authentic training or simulation platform, providing users with an immersive, interactive experience that delivers accurate and comprehensive feedback [41–43].

Table 1 provides a comprehensive overview of the diverse sensing modalities employed in mobile and wearable sensor-based multi-modal applications.

Table 1. Sensor modalities, applications and methods.

Applications	Refs.	Year	Multi-sensor	Methods
Neurologic Rehabilitation Trainings	[14]	2018	FSR，Wearable+force+embedded sensors	Rigid discs Electronic circuit
Physical Activity	[15]	2021	Kinect sensor	MOCAP TEL
Heart-Rate Monitor	[17]	2020	ECG Wearable sensors	Polar Flow web
	[18]	2023	Temperature+heart rate+acceleration+smoke sensors	Wireless Network
Bio signals	[21]	2018	PPG EEG Body motion capture GSR Inertial sensors	SSVEP
Vehicle Motion State Estimation	[23]	2020	ESC GNSS IMU+Series-grade chassis sensors	Intrinsic Euler angles Immersion transformation
Human Identification and Activity	[32]	2022	PIR PRF-PIR IMU MI-PIR	RGBD CNN NBC FMCW RNN
Infrastructure-free Localization in Vision-impaired Environments	[33]	2010	EO IMU RF	HMD Kalman filter

3. System design

This section presents the proposed methodology, expounding on the intricate configuration and execution of our multi-modal sensor fusion system. In Figure 3, an overview of the system's use in physical training is depicted. In the following subsections, we will describe each module in detail.

1. Skeleton tracker (Section 3.1), which integrates depth detection, bone tracking, face recognition and voice recognition using diverse cameras.

2. Body angle measurement (Section 3.2), which utilizes specialized equipment to collect EMG signals from various arm segments, providing real-time display and raw data export.

3. Finger motion monitor (Section 3.3), which captures real-time movement data from individual hand joints, utilizing non-sensing wear technology and facilitating data transfer.

4. VR headset (Section 3.4), allowing users to immerse themselves in a VR environment with the freedom to navigate and observe in any direction.
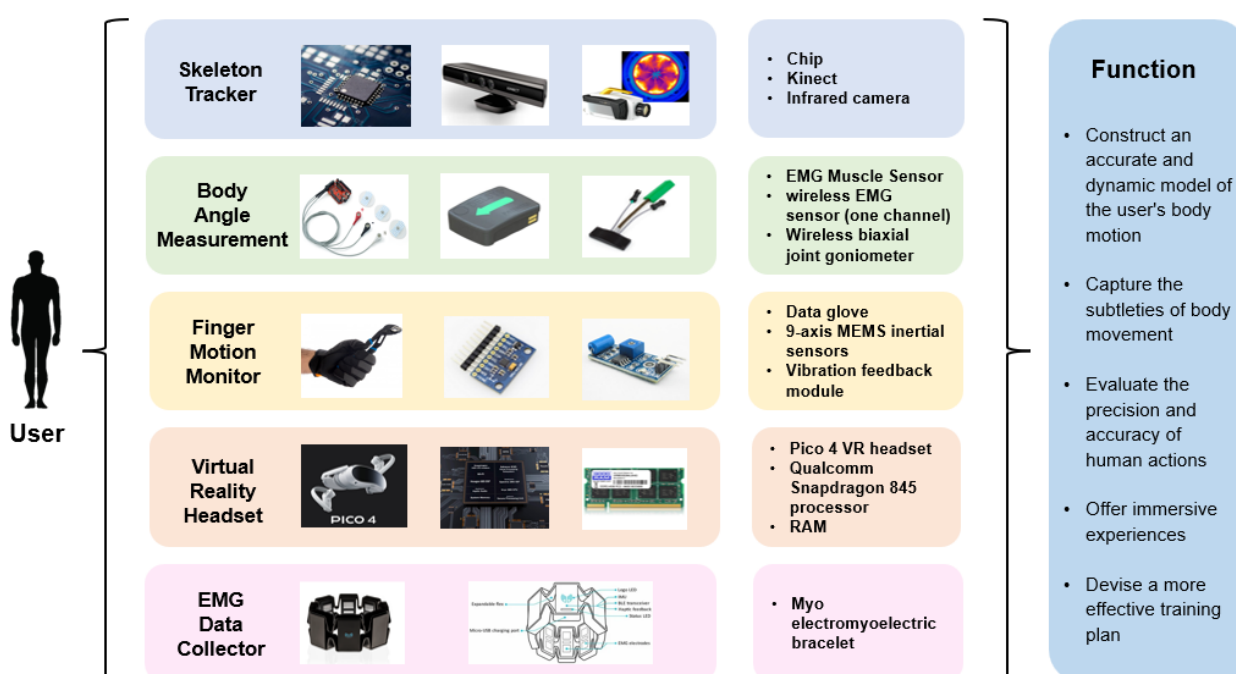

Figure 3. Multi-modal sensor fusion system in physical training.

3.1. Skeleton tracker

The skeleton tracker's primary function in this research is to construct an accurate and dynamic model of the user's body motion during physical training. It uses multiple cameras and chips to detect depth, track bones and recognize faces and voices. Meanwhile, it allows for high-resolution output (1920×1080, 30 fps in color; 512×424, 30 fps in depth) with remarkable accuracy. Depth detection relies on the projection of infrared rays through the infrared camera to generate reflected light, which facilitates the determination of an object's position and the creation of a depth image based on spare

flight time. Skeleton tracking was primarily designed to construct a bone map for up to six players and accurately track the corresponding bone nodes throughout the physical training session.

3.2. Body angle measurement

The body angle measurement apparatus was designed to capture the subtleties of body movement, focusing primarily on the arm segment, by using EMG signals. It gives real-time data on muscle activity during training. The EMG signal collector primarily comprises a wireless collector, a wireless EMG sensor (one channel) and a wireless biaxial joint goniometer (two tracks). In our system, the Biometrics GZ12 EMG signal collector records electrical muscle activity in response to different physical actions. This information is then exported for in-depth analysis, allowing us to examine the efficiency of the training, identify potential strain or injury risks and suggest improvements in a training method.

3.3. Finger motion monitor

Accurate measurement and analysis of hand movements necessitates sophisticated algorithms and advanced sensors. The finger motion monitor, represented here by our data glove, is a critical component that captures nuanced data about hand joint movements during physical training. Equipped with 9-axis MEMS inertial sensors and a vibration feedback module, the glove captures and communicates detailed hand motion data. By employing inverse dynamics, the glove accurately reconstructs bone motion, enabling the faithful reproduction of natural movement in virtual environments. Furthermore, the palm of the glove incorporates a built-in vibration feedback module, which triggers vibration effects corresponding to different scenarios, thereby enhancing the immersive experience in computer designs. To ensure seamless communication, the gloves utilize 24-GHz wireless transmission, achieving a high frame rate of 120 Hz for one hand and 240 Hz for both hands. This high frame rate, coupled with a low delay transmission effect within 10 ms, ensures smooth and responsive interactions. In this research context, we used it to understand the involvement of hand and finger movements during the exercise process. These data are particularly valuable for the evaluation of the precision and accuracy of hand movements during training, allowing us to devise more effective training regimens or corrective measures as required.

3.4. Virtual reality headset

The Pico 4 VR headset is our system's main user interface. The headset provides a fully immersive VR environment, allowing the user to perform physical training actions in a controlled and responsive setting. The Pico 4 is a standalone headset with six-degree-of-freedom inside-out tracking, the weighing 450 g. It prioritizes comfort with an adjustable head strap and face cushion, offering a smooth and immersive experience with a 101-degree FOV and 75 Hz refresh rate. Powered by a Qualcomm Snapdragon 845 processor, it has 4 GB of RAM and 128 GB of storage. The Pico 4 includes built-in stereo speakers, a 3.5 mm audio jack, two six-degree-of-freedom controllers for intuitive interaction, WiFi 6, Bluetooth 5.0 and hand-tracking capabilities, and it runs on the proprietary PICOS. In summary, the Pico 4 VR headset offers immersive experiences without external sensors or cameras. It is well suited for enhancing physical training and facilitating engaging virtual environments.

3.5. System overview

Our hand-tracking system represents a significant advancement in the field of the precise tracking of complex hand movements (Figure 4). Unlike traditional systems that are confined to a fixed workspace, ours allows users to move freely while wearing the equipment, thus introducing a more natural interactive user experience.
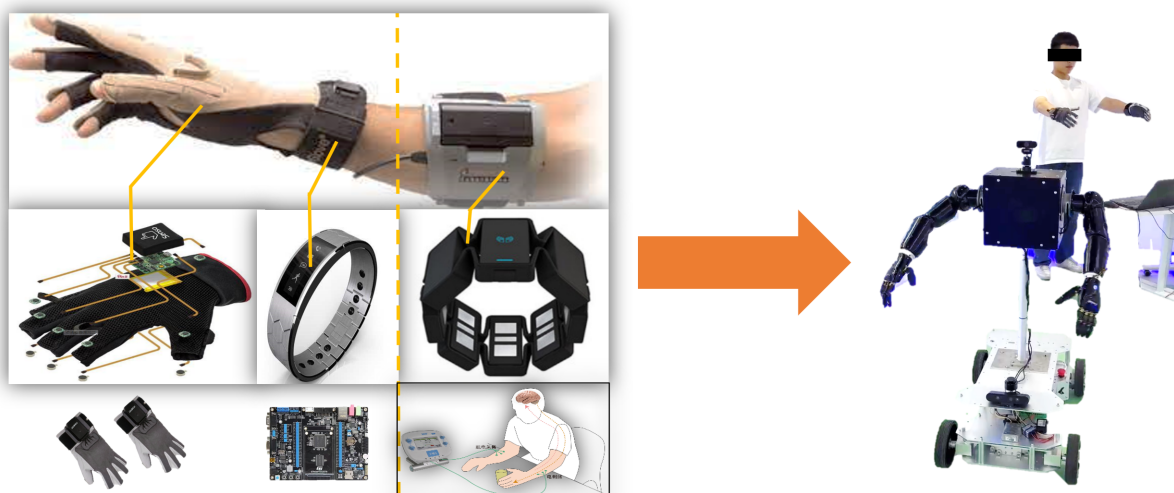

Figure 4. The system overview. On the left, a detailed schematic showcases the specific sensors and equipment related to hand tracking, meticulously designed for accurate data acquisition of hand movements. On the right, an illustrative depiction of a user engaged in physical training, demonstrating the freedom and flexibility afforded by the wearable equipment in real-world applications.

The system comprises two individual subsystems for each hand, each utilizing a glove equipped with a constellation of sensors. Each glove houses six nine-axis IMUs and five flex sensors, providing a total of 12 IMUs and 10 flex sensors for both hands. The IMUs incorporate a gyroscope, accelerometer and magnetometer with measurement ranges of ± 2000 dps, ± 16 g and ± 8.1 Ga, respectively. These sensors work in unison to provide accurate tracking of the hand's spatial orientation and acceleration, as well as the earth's magnetic field, to determine directional heading.

The flex sensors, strategically positioned, monitor the bending of the fingers, thereby capturing even the subtlest of hand movements. These sensors are calibrated to deliver a static accuracy of $0.2°$ RMS for roll/pitch and a dynamic accuracy of $1.0°$ RMS. The angular measurement resolution stands at an impressive $0.02°$.

The gloves employ dual-band (2.4 GHz/5.8 GHz) wireless communication technology, achieving a synchronization accuracy of 10 µs, an industry-leading figure. Furthermore, the system's data communication supports manual and automatic channel switching, thereby mitigating the influence of potential wireless interference in the surroundings.

The system's hardware components are engineered for compatibility with 100BASE-T interfaces and the IEEE 802.3 af/at standards for Power over Ethernet, making it a versatile solution for various user requirements.

Further enhancing the tracking capabilities of our system is a Microsoft Kinect V2, incorporated for the capture of body skeleton data and recognition of overall body movements. The amalgamation of hand movement data from our custom gloves and body movement data from Kinect V2 results in a comprehensive and seamless body tracking experience for the user.

Augmenting the sensor suite of each glove is a wireless EMG data collector from Biometrics. This collector gathers critical data pertaining to forearm muscle activity, providing another data source for enhanced movement recognition. Our approach of extensive sensor fusion, utilizing data from multiple sensor types, allows the system to achieve an unprecedented level of precision in hand movement recognition.

Powering our tracking system is the Robot Operating System (ROS), which was run on an Ubuntu 20 machine. This computer communicates with the gloves via a high-speed wireless router, ensuring minimal latency and maximum throughput in data transmission. Furthermore, MATLAB was employed to process the collected data, which were subject to interaction in real-time within a VR scene in the PICO 4 system. This VR scene, developed in Unity, maintains a real-time communication link with MATLAB, thus enabling fluid interaction between the user and the virtual environment.

Given its precision, our system is optimal for intricate physical training regimens, including those that emphasize precise hand movements such as rehabilitation exercises. It is invaluable for trainers seeking detailed feedback on a trainee's form, grip and finger positioning.

Table 2. Device and software parameters.

Device/Software	Version/Parameters
Inertial Sensor	Gyroscope: ±2000 dps, Accelerometer: ±16 g, Magnetometer: ±8.1 Ga
Flex Sensor	10 (5 per hand)
Kinect V2	Body skeleton tracking
EMG Data Collector	Biometrics
Communication Technology	Dual-band (2.4 GHz/5.8 GHz), Sync accuracy: $10\,\mu$s
ROS	Melodic Morenia (1.14.11)
Ubuntu	20.04 LTS
MATLAB	R2023a
PICO 4 VR System	Latest firmware as of June 2023
Unity	2023.1.2

In conclusion, our system stands out due to its detailed hand tracking, full-body motion capture and virtual environment integration. It provides a flexible solution that fits various use cases.

4. Proposed sensor fusion solution

First, we capture the basic posture of the human body through the skeleton tracking feature of the Kinect V2. For each joint, Kinect V2 provides its position in 3D space. To compute the angle between two joints, we can use the dot product of vectors formed by the joints, which is calculated as follows:

$$\theta = \cos^{-1}\left(\frac{\vec{a} \cdot \vec{b}}{|\vec{a}||\vec{b}|}\right) \tag{4.1}$$

where $\vec{a}$ and $\vec{b}$ are the vectors formed by the joints, and θ is the angle between these vectors.

4.1. KINECT skeleton tracking

Kinect provides the 3D positions of various joints in the human body. Each joint J_i is defined in a 3D Cartesian space, and thus has a position vector in this space denoted by $\mathbf{P}_i = (X_i, Y_i, Z_i)$. The data acquisition rate of Kinect V2 is 30 frames per second.

Now, consider three joints J_1, J_2 and J_3 in sequence, where J_2 is the joint of interest and J_1 and J_3 are adjacent joints. The vectors formed from J_2 to J_1 and J_2 to J_3 are respectively denoted by $\mathbf{A}$ and $\mathbf{B}$, which can be calculated as follows:

$$\mathbf{A} = \mathbf{P}_1 - \mathbf{P}_2 = (X_1 - X_2, Y_1 - Y_2, Z_1 - Z_2) \tag{4.2}$$

$$\mathbf{B} = \mathbf{P}_3 - \mathbf{P}_2 = (X_3 - X_2, Y_3 - Y_2, Z_3 - Z_2) \tag{4.3}$$

The angle θ between these two vectors, which corresponds to the angle of the joint J_2, can be calculated by using the dot product of $\mathbf{A}$ and $\mathbf{B}$:

$$\theta = \cos^{-1}\left(\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}\right) = \cos^{-1}\left(\frac{A_x B_x + A_y B_y + A_z B_z}{\sqrt{(A_x^2 + A_y^2 + A_z^2)(B_x^2 + B_y^2 + B_z^2)}}\right) \tag{4.4}$$

where $\|\mathbf{A}\|$ and $\|\mathbf{B}\|$ are the magnitudes of vectors $\mathbf{A}$ and $\mathbf{B}$ respectively.

4.2. Glove hand tracking

The glove is furnished with inertial sensors and the flex sensors, capturing the movements of fingers. The fusion of data from these sensors allows us to achieve accurate finger movement estimation. The IMU sensors operate at a sampling rate of 200 Hz, and flex sensors operate at a rate of 100 Hz. All sensors are synchronized via a central microcontroller that timestamps the data before passing it on for processing.

We leverage an adaptive neural network algorithm to process the sensor data and predict hand movements. It helps to overcome the limitations of traditional methods, like sensitivity to the initial conditions and the constraints of linear systems, thus providing robust and precise hand tracking.

The flex sensors in the glove measure the amount of bend in each finger joint. For each joint F_i we have a corresponding bending measurement $\mathcal{B}_i$.

The IMU provides data for the orientation of the glove (and, consequently, the hand and the fingers) in the form of quaternion values $Q = (q_w, q_x, q_y, q_z)$. The Euler angles can be calculated from the quaternion as follows:

$$
\begin{aligned}
\text{roll} &= \text{atan2}(2(q_y q_z + q_w q_x), q_w^2 - q_x^2 - q_y^2 + q_z^2) \\
\text{pitch} &= \text{atan2}(2(q_x q_z + q_w q_y), q_w^2 + q_x^2 - q_y^2 - q_z^2) \\
\text{yaw} &= \text{asin}(-2(q_x q_y - q_w q_z))
\end{aligned}
\tag{4.5}
$$

For precise measurement of the finger joint angle θ_{F_i}, sensor fusion is implemented by using the Kalman filter. The state estimate $\hat{x}_{k|k}$ and estimate uncertainty $P_{k|k}$ are calculated as follows:

$$
\begin{aligned}
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1}) \\
P_{k|k} &= (I - K_k H)P_{k|k-1}
\end{aligned}
\tag{4.6}
$$

Where K_k is the Kalman gain, z_k is the actual sensor reading, H is the transformation matrix, I is the identity matrix, $\hat{x}_{k|k-1}$ is the a priori estimate, and $P_{k|k-1}$ is the a priori estimate uncertainty.

The bending measurement $\mathcal{B}_i$ and IMU orientation information are used as inputs to the Kalman filter to obtain the accurate estimate of the joint angle θ_{F_i}.

4.3. Sensor fusion optimization

In the pursuit of optimal sensor fusion, algorithm selection is crucial. For this research, the gated recurrent unit-RNN has been employed to predict the finger joint angles based on the sequential sensor readings. The primary reason for adopting the GRU-RNN model over other deep learning or traditional algorithms lies in its inherent capabilities that are tailored for sequential data. The GRU makes use of two gating mechanisms, namely the reset gate and the update gate, which control the flow of information inside of the unit. The sensors' raw data feed into the GRU model at their respective sampling rates, with missing data points interpolated to ensure continuous input.

For a given sequence of sensor readings $\{s_1, s_2, \ldots, s_t\}$, at each time step t, the update gate z_t and reset gate r_t are computed as follows:

$$
\begin{aligned}
z_t &= \sigma(W_z[s_{t-1}, s_t] + b_z) \\
r_t &= \sigma(W_r[s_{t-1}, s_t] + b_r)
\end{aligned}
\tag{4.7}
$$

where W_z, W_r are weight matrices, b_z, b_r are bias terms and σ is the sigmoid function.
The candidate activation $\tilde{h}_t$ is then calculated as

$$
\tilde{h}_t = \tanh(W[s_{t-1} \odot r_t, s_t] + b)
\tag{4.8}
$$

Finally, the hidden state h_t is updated as follows:

$$
h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t
\tag{4.9}
$$

This hidden state h_t represents the predicted joint angle at time step t.

A key property of GRUs (and RNNs in general) is that they have the ability to handle sequences of data and model temporal dependencies, which make them particularly suited for our application.

However, the learning of GRUs involves the challenging task of learning long-term dependencies due to the vanishing gradients problem and requires careful initialization and potentially regularization methods to ensure stable convergence.

4.4. Virtual reality interaction

To interact with the PICO 4 VR system, we used Unity3D to develop a VR scene and establish communication with MATLAB. The human movement data processed by MATLAB are sent to Unity in real-time, which then updates the VR scene accordingly. This provides the user with an immersive experience as their physical movements are accurately reflected in the virtual environment, which is especially beneficial for physical exercise applications.

For future work, we intend to enhance the system by incorporating more advanced sensor fusion algorithms and neural networks, further improving the robustness and accuracy of human movement tracking in VR-based human-computer interaction. We also plan to expand its application areas to other domains, such as medical rehabilitation and professional sports training.

5. Experimental demonstration

In order to evaluate the performance of the proposed GRU method and compare it with other algorithms, a comprehensive set of experiments were conducted. A dataset comprising time-series data from the motion-tracking system was prepared. This dataset consisted of sensor readings from the hand-tracking gloves, Kinect V2 and EMG data collectors.

Each time step in the data corresponds to a specific joint angle configuration for the human subject. The goal of the algorithms is to predict the joint angles at the next time step given the current and previous sensor readings. The ground truth for these predictions comes from the actual joint angles measured by the motion-tracking system.

The experiment involved two main steps: feature extraction and model training. The data from the Kinect and glove sensors were pre-processed and segmented into windows of 2 second with 50% overlap. For each window, a set of features were extracted, which included statistical features (mean, variance, skewness and kurtosis), frequency-domain features (FFT coefficients) and time-domain features (zero-crossings, peak values). These features were then used as input to the machine learning models.

The models were trained on 80% of the data, with the remaining 20% used for testing. The performance of each model was evaluated by using five metrics: accuracy, precision, recall, F1 score and average prediction time. The accuracy was calculated as the proportion of correctly predicted samples. The precision was calculated as the ratio of correctly predicted positive observations to the total predicted positive observations. The recall was calculated as the ratio of correctly predicted positive observations to all observations in the actual class. The F1 score was calculated as 2(Recall * Precision) / (Recall + Precision). The average prediction time was measured by averaging the time taken by the model to predict the joint angles for all samples in the test set.

The accuracy, precision, recall, and F1 score are popular metrics used in machine learning, and they were computed as follows:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}}$$

The comparison of the GRU-RNN algorithm with other machine learning methods is summed up in Table 3. It is clear from the results that the GRU- RNN algorithm outperforms the other algorithms in terms of accuracy, precision, recall and F1 score. This implies that the GRU-RNN algorithm is more successful in terms of correctly predicting the joint angles from the sensor data. Furthermore, the GRU-RNN algorithm also demonstrates superiority in terms of the average prediction time, further enhancing its suitability for real-time applications. The superior performance of the GRU-RNN algorithm can be attributed to its ability to effectively capture the temporal dependencies in the sensor data, which is crucial for accurate joint angle prediction. This makes the GRU-RNN algorithm a robust and efficient choice for the motion-tracking system.

Table 3. Comparison of different algorithms on joint angle prediction task.

Algorithm	Accuracy	Precision	Recall	F1 Score	Average Prediction Time (ms)
GRU-RNN	0.97	0.98	0.96	0.97	4.5
Feedforward Neural Network	0.85	0.87	0.83	0.85	5.2
Support Vector Machine	0.81	0.83	0.79	0.81	6.3
Random Forest	0.76	0.77	0.75	0.76	7.8

As observed in Table 3, our proposed method outperforms the other techniques in terms of accuracy, precision, recall and F1 score. Furthermore, it provides the best real-time processing capabilities, with a shorter computational time than other methods. Thus, our system's application of networks demonstrates superior performance, further bolstering its potential for precise, robust and real-time hand motion tracking in dynamic environments.

The bar chart provided in Figure 5 presents a comparison of the GRU-RNN algorithm with other common machine learning algorithms in terms of accuracy.
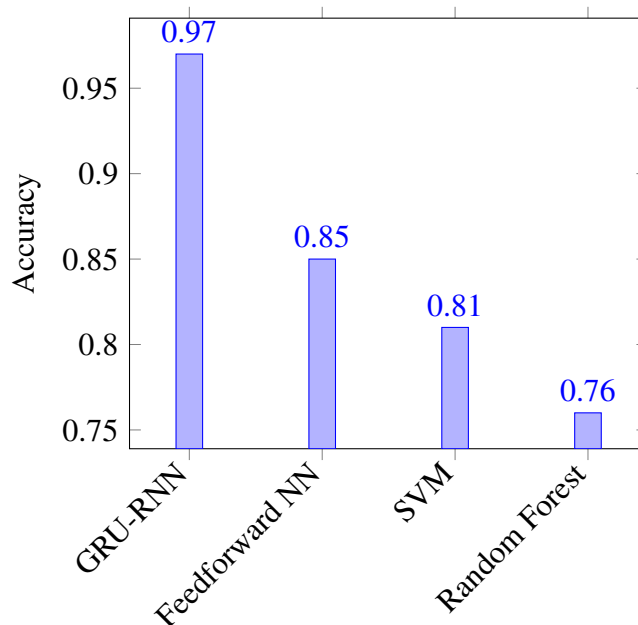
Figure 5. Comparison of the accuracy of different algorithms.

Evidently, the GRU-RNN outperforms the other algorithms in terms of accuracy. It is important to note that the superiority of the GRU-RNN is attributable to its inherent ability to capture temporal dependencies in sequential data. In the case of joint angle prediction, the current joint angles are closely related to the previous states due to the physical constraints of human body motion. This context is something that the GRU-RNN captures well.

The feedforward neural network, support vector machine (SVM) and random forest methods, though being competent machine learning models, are unable to handle sequential data as effectively. This is due to their inability to maintain a "memory" of previous inputs in the sequence.

The feedforward neural network, in this case, fails to consider the temporal sequence of data. On the other hand, the SVM and random forest work on the assumption that the data samples are independent, which is not the case with sequential sensor data. Hence, their performance falls short as compared to the GRU-RNN.

In conclusion, for tasks that involve sequential data, such as the joint angle prediction in our motion-capture system, the use of the GRU-RNN provides significant advantages over other methods in terms of accuracy. This finding is vital for the enhancement of the effectiveness and reliability of our system in order to deliver a more immersive VR experience for users.

6. Conclusion and future work

We have successfully developed a multi-modal sensor fusion system that can enhance physical training by delivering detailed, real-time feedback on the user's performance. Through the use of diverse sensor sources including IMUs, flex sensors, EMG sensors and Microsoft's Kinect V2, our system can capture comprehensive body and hand motion data. Our results demonstrate that the GRU-RNN algorithm, as employed in our system, outperforms traditional machine learning

algorithms in terms of accuracy, precision, recall and F1 score. Furthermore, the integration of our system with the PICO 4 VR environment offers an immersive and interactive experience, fostering user engagement in physical training. Unlike many traditional systems, ours overcomes the constraint of a fixed workspace, allowing users the flexibility to train in a free-form environment. Future work will focus on further improving the accuracy of motion estimation and enhancing the immersive experience of the VR environment.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Joubert C, Chainay H (2018) Aging brain: the effect of combined cognitive and physical training on cognition as compared to cognitive and physical training alone–a systematic review. *Clin Interv Aging* 13: 1267–1301. https://doi.org/10.2147/CIA.S165399

2. Zouhal H, Ben Abderrahman A, Khodamoradi A, et al. (2020) Effects of physical training on anthropometrics, physical and physiological capacities in individuals with obesity: a systematic review. *Obes Rev* 21: e13039. https://doi.org/10.1111/obr.13039

3. Pedersen BK (2006) The anti-inflammatory effect of exercise: its role in diabetes and cardiovascular disease control. *Essays Biochem* 42: 105–117. https://doi.org/10.1042/bse0420105

4. Gilani SRM, Feizabad AK (2019) The effects of aerobic exercise training on mental health and self-esteem of type 2 diabetes mellitus patients. *Health Psychol Res* 7: 6576. https://doi.org/10.4081/hpr.2019.6576

5. Shepherd HA, Evans T, Gupta S, et al. (2021) The impact of COVID-19 on high school student-athlete experiences with physical activity, mental health, and social connection. *Int J Environ Res Public Health* 18: 3515. https://doi.org/10.3390/ijerph18073515

6. Sjøgaard G, Christensen JR, Justesen JB, et al. (2016) Exercise is more than medicine: The working age population's well-being and productivity. *J Sport Health Sci* 5: 159–165 https://doi.org/10.1016/j.jshs.2016.04.004

7. McCrone K (2014) *Sport and the physical emancipation of English women (RLE Sports Studies): 1870-1914*, London: Routledge. https://doi.org/10.4324/9781315772844

8. Latey P (2001) The pilates method: history and philosophy. *J Bodyw Mov Ther* 5: 275–282. https://doi.org/10.1054/jbmt.2001.0237

9. Hauguel-Moreau M, Naudin C, N'Guyen L, et al. (2020) Smart bracelet to assess physical activity after cardiac surgery: a prospective study. *PloS one* 15: e0241368. https://doi.org/10.1371/journal.pone.0241368

10. Javaloyes A, Sarabia JM, Lamberts RP, et al. (2019) Training prescription guided by heart-rate variability in cycling. *Int J Sports Physiol Perform* 14: 23–32. https://doi.org/10.1123/ijspp.2018-0122

11. Shi Y, Li L, Yang J, et al. (2023) Center-based transfer feature learning with classifier adaptation for surface defect recognition. *Mech Syst Signal Process* 188: 110001. https://doi.org/10.1016/j.ymssp.2022.110001

12. Shi Y, Li H, Fu X, et al. (2023) Self-powered difunctional sensors based on sliding contact-electrification and tribovoltaic effects for pneumatic monitoring and controlling. *Nano Energy* 110: 108339. https://doi.org/10.1016/j.nanoen.2023.108339

13. Qi W, Aliverti A (2019) A multimodal wearable system for continuous and real-time breathing pattern monitoring during daily activity. *IEEE J Biomed Health Inform* 24: 2199–2207. https://doi.org/10.1109/JBHI.2019.2963048

14. De Pasquale G, Mastrototaro L, Pia L, et al. (2018) Wearable system with embedded force sensors for neurologic rehabilitation trainings. *2018 Symposium on Design, Test, Integration & Packaging of MEMS and MOEMS (DTIP)*, Rome, IEEE: 1–4. https://doi.org/10.1109/DTIP.2018.8394187

15. Mendes AS, Silva LA, Blas HSS, et al. (2021) Physical movement helps learning: teaching using tracking objects with depth camera. *Trends and Applications in Information Systems and Technologies: Volume 4 9,* Springer International Publishing, 2021: 183–193. https://doi.org/10.1007/978-3-030-72654-6_18

16. Egle F, Kluge F, Schoene D, et al. (2022) Development of an inertial sensor-based exergame for combined cognitive and physical training. *2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN),* IEEE, 2022: 1–4. https://doi.org/10.1109/BSN56160.2022.9928474

17. Batalik L, Dosbaba F, Hartman M, et al. (2020) Benefits and effectiveness of using a wrist heart rate monitor as a telerehabilitation device in cardiac patients: a randomized controlled trial. *Medicine* 99: e19556. https://doi.org/10.1097/MD.0000000000019556

18. Bo H, Sun Z (2023) Construction of heart rate monitoring platform for college physical training based on wireless network. *Wireless Netw* 29: 3005—3016. https://doi.org/10.1007/s11276-022-03226-z

19. Su H, Qi W, Schmirander Y, et al. (2022) A human activity-aware shared control solution for medical human–robot interaction. *Assem Autom* 42: 388–394. https://doi.org/10.1108/aa-12-2021-0174

20. Qi W, Ovur SE, Li Z, et al. (2021) Multi-sensor guided hand gesture recognition for a teleoperated robot using a recurrent neural network. *IEEE Robot Autom Lett* 6: 6039–6045. https://doi.org/10.1109/LRA.2021.3089999

21. Patel AN, Jung TP, Sejnowski TJ (2018) A wearable multi-modal bio-sensing system towards real-world applications. *IEEE Trans Biomed Eng* 66: 1137–1147. https://doi.org/10.1109/TBME.2018.2868759

22. Weiner P, Neef C, Shibata Y, et al. (2019) An embedded, multi-modal sensor system for scalable robotic and prosthetic hand fingers. *Sensors* 20: 101. https://doi.org/10.3390/s20010101

23. Marco VR, Kalkkuhl J, Raisch J, et al. (2020) Multi-modal sensor fusion for highly accurate vehicle motion state estimation. *Control Eng Pract* 100: 104409. https://doi.org/10.1016/j.conengprac.2020.104409

24. Ma R, Hu F (2016) An intelligent thermal sensing system for automatic, quantitative assessment of motion training in lower-limb rehabilitation. *IEEE Trans Syst Man Cybern Syst* 48: 661–669. https://doi.org/10.1109/TSMC.2016.2636660

25. Torreño N, Munguía-Izquierdo D, Coutts A, et al. (2016) Relationship between external and internal loads of professional soccer players during full matches in official games using global positioning systems and heart-rate technology. *Int J Sports Physiol Perform* 11: 940–946. https://doi.org/10.1123/ijspp.2015-0252

26. Nur L, Suherman A, Subarjah H (2019) The use of global positioning system (GPS) polars to determine motion intensity. *J Eng Sci Technol* 14: 2132–2139. https://doi.org/10.1123/ijspp.2015-0252

27. Drobnjakovic F, Douangpaseuth JB, Gadea C, et al. (2018) Fusing data from inertial measurement units and a 3D camera for body tracking. *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC),* IEEE, 2018: 1–6. https://doi.org/10.1109/I2MTC.2018.8409754

28. Zhang X, Elnady AM, Randhawa BK, et al. (2018) Combining mental training and physical training with goal-oriented protocols in stroke rehabilitation: a feasibility case study. *Front Hum Neurosci* 12: 125. https://doi.org/10.3389/fnhum.2018.00125

29. Ryselis K, Petkus T, Blažauskas T, et al. (2020) Multiple Kinect based system to monitor and analyze key performance indicators of physical training. *Hum-Centric Comput Inf Sci* 10: 1–22. https://doi.org/10.1186/s13673-020-00256-4

30. Su H, Qi W, Hu Y, et al. (2020) An incremental learning framework for human-like redundancy optimization of anthropomorphic manipulators. *IEEE Trans Ind Inform* 18: 1864–1872. https://doi.org/10.1109/TII.2020.3036693

31. Qi W, Su H (2022) A cybertwin based multimodal network for ecg patterns monitoring using deep learning. *IEEE Trans Ind Inform* 18: 6663–6670. https://doi.org/10.1109/TII.2022.3159583

32. Yuan L, Andrews J, Mu H, et al. (2022) Interpretable passive multi-modal sensor fusion for human identification and activity recognition. *Sensors* 22: 5787. https://doi.org/10.3390/s22155787

33. Oskiper T, Chiu HP, Zhu Z, et al. (2010) Multi-modal sensor fusion algorithm for ubiquitous infrastructure-free localization in vision-impaired environments. *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems,* IEEE, 2010: 1513–1519. https://doi.org/10.1109/IROS.2010.5649562

34. Calvo AF, Holguin GA, Medeiros H (2018) Human activity recognition using multi-modal data fusion. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 23rd Iberoamerican Congress, CIARP 2018,* Spain, Springer International Publishing: 946–953. https://doi.org/10.1007/978-3-030-13469-3_109

35. Tian C, Xu Z, Wang L (2023) Arc fault detection using artificial intelligence: challenges and benefits. *Math Biosci Eng* 20: 12404–12432. https://doi.org/10.3934/mbe.2023552

36. Lei Y, Su Z, He X, et al. (2023) Immersive virtual reality application for intelligent manufacturing: Applications and art design. *Math Biosci Eng* 20: 4353–4387. https://doi.org/10.3934/mbe.2023202

37. Qi W, Fan H, Karimi HR, et al. (2023) An adaptive reinforcement learning-based multimodal data fusion framework for human–robot confrontation gaming. *Neural Netw* 164: 489–496. https://doi.org/10.1016/j.neunet.2023.04.043

38. Xu L, Zhang JQ, Yan Y (2004) A wavelet-based multisensor data fusion algorithm. *IEEE Trans Instrum Meas* 53: 1539–1545. https://doi.org/10.1109/TIM.2004.834066

39. Wang H, Song L, Liu J, et al. (2021) An efficient intelligent data fusion algorithm for wireless sensor network. *Procedia Comput Sci* 183: 418–424. https://doi.org/10.1016/j.procs.2021.02.079

40. Zhao J, Lv Y (2023) Output-feedback robust tracking control of uncertain systems via adaptive learning. *Int J Control Autom Syst* 21: 1108–1118. https://doi.org/10.1007/s12555-021-0882-6

41. Liu Z, Yang D, Wang Y, et al. (2023) EGNN: Graph structure learning based on evolutionary computation helps more in graph neural networks. *Appl Soft Comput* 135: 110040. https://doi.org/10.1016/j.asoc.2023.110040

42. Wang Y, Liu Z, Xu J, et al. (2022) Heterogeneous network representation learning approach for ethereum identity identification. *IEEE Trans Comput Soc Syst* 10: 890–899 https://doi.org/10.1109/TCSS.2022.3164719

43. Lei Y, Su Z, Cheng C (2023) Virtual reality in human-robot interaction: Challenges and benefits. *ERA* 31: 2374–2408. https://doi.org/10.3934/era.2023121