Multimodal Learning and Human Digital Twins for Industrial Safety Monitoring in Human-Robot Collaborative Environments

Author: Hajimi Bao

Abstract

The transition from Industry 4.0 to Industry 5.0 marks a fundamental reorientation of manufacturing systems around **human-centered collaboration**, where workers and robots coexist and cooperate in shared workspaces. This paradigm shift introduces critical challenges in **industrial safety monitoring**: ensuring that collaborative robots respond safely and adaptively to human actions, that workers are protected from ergonomic risks and hazardous conditions, and that safety systems operate with the real-time reliability demanded by high-speed production environments. Traditional safety approaches—based on static rule-based logic and retrospective incident analysis—are fundamentally inadequate for the dynamic, unpredictable nature of human-robot collaboration. This review examines how **multimodal learning**—the integration of data from wearable sensors, computer vision systems, physiological monitors, and environmental sensors—combined with **human digital twin** architectures, is transforming industrial safety monitoring in human-robot collaborative environments. Drawing on twelve peer-reviewed works, we synthesize advances in human activity recognition (HAR) with wearable sensors, human intention recognition for real-time robot control, reinforcement learning for adaptive robotic manipulation, and worker safety digital twins for Industry 5.0. We further demonstrate how industrial sensing technologies—including four-dimensional thermal imaging, stereo phase-measuring deflectometry, and gesture-based robotic control—serve as critical sensor modalities within the multimodal safety monitoring framework. A central contribution of this review is the articulation of an integrated **Human-Cobot Safety Intelligence (HCSI)** paradigm that unifies multimodal perception, predictive safety analytics, and adaptive robot control for proactive, real-time industrial safety assurance.

Keywords: Multimodal Learning; Human Activity Recognition; Human-Robot Collaboration; Industrial Safety Monitoring; Human Digital Twin; Human Intention Recognition; Reinforcement Learning; Industry 5.0

---

1. Introduction

Industrial safety has long been a paramount concern in manufacturing. According to the International Labour Organization (ILO), occupational accidents and work-related diseases claim approximately 2.3 million lives annually worldwide, with millions more suffering non-fatal injuries that result in lost productivity, disability, and diminished quality of life. In the manufacturing sector—encompassing automotive assembly, electronics production, chemical processing, and construction—hazardous machinery, high-speed processes, and complex supply chains create an environment where safety incidents can have catastrophic human and economic consequences.

The traditional approach to industrial safety has been **reactive and prescriptive**: safety engineers define rules and procedures, workers are trained to follow them, and safety incidents are investigated retrospectively to identify root causes and update procedures. This approach, while necessary, is insufficient for the emerging paradigm of **human-robot collaboration (HRC)** in Industry 5.0, where workers and robots share physical space, interact in real time, and collaborate on tasks that require complementary strengths—human flexibility and contextual judgment alongside robotic precision, strength, and endurance.

The safety challenges in HRC are qualitatively different from those in traditional manufacturing. A robot operating behind a safety cage follows deterministic, predictable trajectories; a collaborative robot operating alongside a human must continuously monitor the human's position, motion, and intent, and adjust its own behavior in real time to avoid collision, prevent ergonomic strain, and maintain productive collaboration. The failure modes are new: a collaborative robot that misjudges a human's trajectory may cause a collision; a robot that fails to recognize a human's intent may take over a task the human was about to perform, creating confusion and inefficiency; a safety system that relies solely on proximity sensors may miss subtle precursors to hazardous situations, such as a worker showing signs of fatigue or distraction.

These challenges have catalyzed intense research interest in **multimodal learning** for industrial safety monitoring. By integrating data from diverse sensor modalities—vision cameras, depth sensors, wearable inertial measurement units (IMUs), physiological monitors (heart rate, galvanic skin response), environmental sensors (temperature, gas concentration, noise levels)—multimodal systems can construct a richer, more robust picture of the human state and the environment than any single sensor modality can provide.

This review provides a comprehensive and critical synthesis of this rapidly evolving field. Our specific contributions are: (1) a taxonomy of multimodal sensing modalities; (2) a systematic review of deep learning approaches for human activity and intention recognition in HRC; (3) an examination of the emerging human digital twin paradigm for Industry 5.0; (4) a connection to industrial sensing technologies as multimodal safety sensor inputs; and (5) a unified framework—**Human-Cobot Safety Intelligence (HCSI)**—that integrates these advances.

The review is organized as follows: Section 2 reviews multimodal sensing modalities; Section 3 examines human activity recognition; Section 4 covers human intention recognition in HRC; Section 5 presents the human digital twin paradigm; Section 6 discusses adaptive robot safety control; Section 7 provides cross-cutting synthesis; and Section 8 concludes.

---

2. Multimodal Sensing for Industrial Safety: Technologies and Modalities

2.1 The Case for Multimodal Sensing

Industrial environments generate rich, heterogeneous data streams from diverse sensors. No single sensing modality can capture the full complexity of the human state and operational environment. **Vision cameras** provide rich spatial and appearance information but are compromised by occlusion, illumination changes, and dust; **wearable IMU sensors** capture body motion with high temporal resolution but suffer from drift and cannot directly sense the environment; **physiological monitors** reveal cognitive and emotional states relevant to safety but are intrusive and noisily coupled to task demands; **environmental sensors** detect hazardous conditions but provide no information about the human's response to those conditions.

Multimodal sensing addresses these limitations by **fusing** information across complementary modalities. Fusion can occur at multiple levels: **raw data fusion** (early fusion), where sensor streams are concatenated before processing; **feature fusion**, where representations extracted separately from each modality are combined; and **decision fusion**, where independent classifiers for each modality vote or are combined through a meta-classifier. Deep learning enables end-to-end learning of optimal fusion strategies directly from data (Arsigah et al., 2024).

2.2 Wearable Inertial Sensors for Ergonomic Monitoring

Wearable IMU-based human activity recognition (HAR) is one of the most mature multimodal sensing domains. Accelerometers and gyroscopes embedded in smartwatches, wristbands, or wearable patches capture the kinematic signatures of body movements—walking, reaching, bending, lifting, tool manipulation—with high temporal resolution and minimal intrusion. Deep learning models—particularly **CNNs** for spatial feature extraction and **LSTMs** for temporal modeling—have achieved remarkable accuracy on HAR benchmarks, with recent studies reporting over 98% accuracy in controlled settings (Awoke et al., 2024).

In the industrial context, wearable sensors are particularly valuable for monitoring **ergonomic risk**—the risk of musculoskeletal disorders from repetitive motions, awkward postures, and manual material handling. Studies in construction activity recognition have demonstrated that LSTM-CNN architectures with optimally positioned wearable sensors can achieve 98.17% accuracy in classifying construction worker activities such as hammering, drilling, and material transport (Chen et al., 2024). This enables real-time ergonomic monitoring: a safety system that detects a worker repeatedly lifting loads above a safe weight threshold, or adopting a posture associated with back injury risk, can alert the worker or adjust task assignment before an injury occurs.

2.3 Computer Vision and Thermal Imaging for Worker Safety

Computer vision provides non-contact, rich spatial information about the work environment and worker behavior. RGB cameras capture detailed appearance information; depth cameras add 3D geometric information invariant to illumination and color changes; infrared cameras detect thermal signatures useful for fire prevention and equipment overload monitoring. **Personal protective equipment (PPE) detection**—verifying that workers wear helmets, safety glasses, gloves, and high-visibility vests—and **zone intrusion detection**—identifying when a person enters a hazardous area—are established applications.

Huang and colleagues (2023) developed a **four-dimensional thermal imaging system** integrating structured illumination binocular cameras with an infrared thermal camera for temperature field reconstruction on non-uniform surfaces. While focused on industrial process monitoring, this technology extends naturally to **worker thermal safety**: infrared imaging detects early signs of heat stress, identifies overheating equipment, and monitors workers for thermal anomalies indicating fatigue or illness. The multi-view fusion strategy developed in this work is also applicable to **multi-camera worker tracking** in complex industrial environments where single-camera views are frequently occluded by machinery (Huang et al., 2023).

2.4 Industrial Optical Sensing as Safety Modality

Industrial optical sensing technologies—originally developed for precision metrology—provide capabilities that extend to safety monitoring. Huang and colleagues' **stereo phase-measuring deflectometry (SPMD)** system (2026), achieving high-precision 3D surface measurement with deep learning-enhanced phase unwrapping, can be adapted for **high-speed non-contact vibration monitoring** of rotating machinery. Unbalanced rotating equipment is a major source of mechanical failure and workplace injury; a PMD-based vibration sensor could detect imbalance and misalignment in real time, triggering preventive shutdowns before catastrophic failure (Huang et al., 2026).

Li and colleagues' **Leap Motion Controller-based gesture control system** (2024) exemplifies how gesture recognition technology developed for HRC serves dual purposes: enabling productive collaboration while simultaneously monitoring the worker's hand position and motion for safety-relevant information. The high-frequency (up to 120 Hz) skeletal hand tracking from the Leap Motion detects grasping patterns, tool manipulation gestures, and hand trajectories that inform the robot's safety response—integrating collaborative task execution and safety monitoring within a single sensor platform (Li et al., 2024).

---

3. Human Activity Recognition for Ergonomic and Safety Monitoring

3.1 Deep Learning Architectures for HAR

Human activity recognition using wearable sensors has evolved from classical machine learning approaches (SVMs, random forests, hidden Markov models) to deep learning architectures that learn hierarchical feature representations directly from raw sensor data.

CNNs applied to IMU time series treat the temporal signal as a 1D spatial signal, applying 1D convolutions to extract local temporal patterns invariant to small time shifts. Wang and colleagues' TCN-Attention-HAR model (2024) combines temporal convolutional networks with attention mechanisms, demonstrating that attention-based temporal modeling significantly improves performance on complex, multi-step activities (Wang et al., 2024).

LSTMs model long-range temporal dependencies in sequential sensor data, capturing the dynamics of complex activities that unfold over tens of seconds or minutes. The DeepConvLSTM architecture—stacking multiple convolutional layers followed by LSTM layers—has become a de facto standard for wearable HAR: CNN layers extract local spatial-temporal features, LSTM layers model global temporal structure (Awoke et al., 2024).

Hybrid CNN-LSTM architectures with channel attention mechanisms outperform single-modality approaches for complex human activity recognition. The channel attention mechanism adaptively weights each sensor channel's contribution based on its informativeness for the current activity class, improving robustness to sensor noise and individual variation (Awoke et al., 2024).

3.2 Industrial Applications: Ergonomics, Fatigue, and Risk Detection

The most direct industrial application of HAR is **ergonomic risk monitoring**. Musculoskeletal disorders (MSDs)—back injuries, carpal tunnel syndrome, rotator cuff tears—account for the largest share of work-related health complaints and compensation claims in manufacturing. HAR systems can continuously monitor worker activities—lifting, reaching, twisting, tool manipulation—and compute real-time ergonomic risk scores using algorithms such as RULA or NIOSH lifting equations (Chen et al., 2024).

Chen and colleagues (2024) demonstrated an LSTM-CNN architecture with optimally positioned wearable sensors achieving 98.17% accuracy in classifying construction worker activities. By tracking the frequency and duration of high-risk activities—overhead work, manual lifting, sustained kneeling—construction managers can proactively redistribute tasks to prevent cumulative trauma, adjust work schedules to reduce fatigue-related risk, and target safety training interventions (Chen et al., 2024).

3.3 Vision-Based Worker Safety Monitoring

Vision-based HAR complements wearable sensors by providing a bird's-eye view of the work environment. Deep learning-based vision systems detect unsafe behaviors—workers not wearing PPE, entering restricted zones, vehicles operating in pedestrian areas—in real time from CCTV and drone footage. Integration with **digital twin** platforms enables contextual safety monitoring that combines worker activity data with 3D facility models and process simulation data (AECbytes, 2025).

---

4. Human Intention Recognition for Real-Time Human-Robot Collaboration

4.1 The Challenge of Intention Recognition

Human intention recognition—inferring what a human plans to do next from observable behavioral signals—is one of the most critical and challenging problems in HRC. Accurate and timely intention recognition enables a collaborative robot to proactively adjust its trajectory to avoid interference, initiate or defer tasks based on the human's likely next action, and provide assistive support that anticipates rather than reacts to the human's needs.

Behavioral signals for intention recognition are inherently multimodal: the human's **trajectory and body pose** reveal planned reaching motions; **gaze direction** indicates the object of attention; **verbal utterances** convey task goals; **physiological signals** reveal cognitive load and emotional state. No single modality is sufficient; multimodal integration is essential for robust real-time intention recognition.

4.2 Deep Learning Approaches

A 2024 ASME study developed a **transformer network** for early prediction of human intention in HRC tasks. Transformer architectures—originally developed for NLP—excel at modeling long-range dependencies and temporal context, enabling the robot to infer the human's intended action from a short observation window of partial motion data. Early prediction is critical for safety: the earlier the robot anticipates the trajectory, the more time it has to adjust its own motion (ASME, 2024).

A 2025 review in *Frontiers in Robotics and AI* concluded that **RNNs and LSTM variants** are currently the dominant approach for real-time intention prediction, owing to their ability to process sequential sensor data and produce continuous, low-latency predictions. The review identified three key requirements for industrial intention recognition: **low latency** (sub-100ms for real-time responsiveness), **high accuracy** (>95% correct classification for safety-critical decisions), and **robustness to individual variation** across workers with different physiques, movement patterns, and task strategies (Frontiers in Robotics and AI, 2025).

An emerging approach leverages **LLMs** for intention recognition. A 2024 study explored using LLMs to infer human intentions from natural language descriptions of the collaborative task context, demonstrating that the world knowledge encoded in pre-trained LLMs enables zero-shot generalization to novel task scenarios (Frontiers in Robotics and AI, 2025).

4.3 Human Motion Prediction for Collision Avoidance

Human motion prediction forecasts the physical trajectory of the human over a short future time horizon (typically 0.5–2.0 seconds). A PMC study (2025) developed a framework using **inverse discrete cosine transform (IDCT)** with CNN-based intention classification to predict human motion in Cartesian coordinates for HRC. The predicted trajectory feeds directly into the robot's motion planner to generate collision-free trajectories in real time (PMC, 2025).

4.4 Reinforcement Learning for Adaptive Safety Control

Once the human's activity and intention are recognized, the robot must adapt its behavior—adjusting speed, trajectory, and task priority—in real time to maintain safety and productivity. **Reinforcement learning** has emerged as the dominant paradigm for adaptive robotic behavior in HRC. An RL agent learns a policy—mapping states to actions—that maximizes a reward incorporating both task performance and safety constraints (Annual Reviews, 2025).

A 2025 *Annual Reviews* article—*Deep Reinforcement Learning for Robotics: A Survey of Real-World Successes*—documented RL applications in robotics including contact-rich manipulation, dexterous in-hand manipulation, and human-robot interaction, concluding that RL is a transformative enabler of adaptive robotic behavior in shared human-robot workspaces (Annual Reviews, 2025). A companion 2026 study identified RL as a promising method for automating contact-rich manipulation tasks in industrial automation (Parnada et al., 2026).

Wang and colleagues (2025) demonstrated an LLM-driven automated software testing framework for automotive APIs, illustrating how LLMs can automate the **verification and validation of safety-critical software systems**—where AI-driven automation promises significant improvements in coverage and speed for systems where human safety depends on software correctness (Wang et al., 2025).

---

5. Human Digital Twins for Industry 5.0 Worker Safety

5.1 The Human Digital Twin Concept

The **digital twin**—a high-fidelity virtual replica of a physical system that evolves in real time as data streams in—has become a cornerstone of Industry 4.0 manufacturing. The extension of digital twin technology to **human workers**—creating a **human digital twin (HDT)**—represents a transformative development for industrial safety.

An HDT is a personalized, real-time computational model of an individual worker's physical state, cognitive state, and task performance. Drawing on data from wearable sensors, computer vision, physiological monitors, and environmental sensors, the HDT continuously updates a comprehensive model of the worker's current condition—posture, fatigue level, stress level, cumulative physical load, proximity to hazards—and forecasts their future state, enabling **proactive safety intervention** before incidents occur (Davila-Gonzalez & Martin, 2024).

5.2 Multi-Layer HDT Architecture

A landmark 2024 *Sensors* study—*Human Digital Twin in Industry 5.0: A Holistic Approach to Worker Safety and Well-Being through Advanced AI and Emotional Analytics*—proposed a multi-layer HDT architecture:

- **Physical layer**: kinematic models of the worker's body updated in real time from IMU and vision data, tracking posture, motion, and proximity to hazards.
- **Physiological layer**: models of the worker's physiological state—heart rate variability, galvanic skin response, eye tracking—from wearable biosensors, capturing cognitive load, stress, fatigue, and emotional state.
- **Behavioral layer**: models of the worker's task performance—activity recognition, task sequencing, error rate—from multimodal sensor fusion.
- **Environmental layer**: real-time data on ambient conditions—temperature, humidity, noise, air quality—combined with the worker's individual vulnerability profile.
- **Predictive layer**: forecasting models that predict the worker's future physical and cognitive state, enabling proactive safety interventions.

This holistic architecture addresses a fundamental limitation of conventional safety monitoring: treating each sensor modality in isolation, missing synergistic information available when modalities are jointly modeled. The HDT creates a **unified, personalized, real-time model of the individual worker** (Davila-Gonzalez & Martin, 2024).

5.3 Integration with Industrial Sensing Infrastructure

The HDT integrates with industrial sensing infrastructure developed for manufacturing automation. Li and colleagues' Leap Motion system (2024) provides a natural data acquisition platform for the HDT's physical layer: high-frequency skeletal tracking simultaneously enables gesture-based robot control and body pose monitoring for ergonomic safety. Huang and colleagues' 4D thermal imaging system (2023) contributes to the HDT's environmental layer: ambient temperature mapping combined with individual thermal physiology models enables personalized heat stress prediction (Huang et al., 2023; Li et al., 2024).

A 2025 study on **integrating digital factory twins and AI for manufacturing monitoring** via synthetic data generation and vision transformers demonstrated how HDT concepts can be operationalized at factory scale. By generating synthetic training data for vision transformers using digital twin models, the authors achieved accurate worker tracking across diverse lighting, occlusion, and pose conditions (ScienceDirect, 2025).

---

6. Reinforcement Learning for Adaptive Safety Control in HRC

6.1 From Static Safety Rules to Adaptive Safety Policies

Traditional industrial robot safety relies on **static safety rules**—safety zones, maximum speeds, emergency stop conditions—defined in safety PLCs based on worst-case assumptions about human positions and motions. While effective for minimum safety, static rules cannot adapt to the specific human, task, or environmental context.

RL offers a path beyond static rules by learning **adaptive safety policies** that optimize robot behavior—balancing safety and task efficiency—based on real-time observations. Safe RL algorithms—including **constrained RL**, **safe exploration**, and **reward shaping**—ensure that learned policies satisfy safety requirements throughout training and at deployment (Annual Reviews, 2025).

6.2 Human-in-the-Loop Reinforcement Learning

A 2025 *Science Robotics* study demonstrated a **human-in-the-loop (HiTL) RL system** for precise and dexterous robotic manipulation, where human physical corrections during training significantly accelerated learning and improved the policy's handling of novel situations. In the safety context, HiTL RL enables the human worker to teach the robot which behaviors are safe through demonstration and correction—without requiring explicit safety constraint specification (Science Robotics, 2025).

This connects directly to Li and colleagues' gesture-based collaborative robot control (2024): gesture commands provide an intuitive modality for human-in-the-loop guidance, allowing workers to steer the robot's learning through natural physical interaction, building a collaborative relationship that evolves over time as the robot learns individual preferences and safety requirements (Li et al., 2024).

---

7. Discussion

7.1 Toward a Unified Human-Cobot Safety Intelligence Framework

The synthesis of findings across the reviewed literature points toward the **Human-Cobot Safety Intelligence (HCSI)** framework—an AI architecture that unifies multimodal human state estimation, predictive safety analytics, and adaptive robot control for proactive, real-time industrial safety assurance.

The HCSI architecture comprises three interconnected layers. The **perception layer** fuses data from wearable sensors, computer vision, depth sensors, thermal cameras, and environmental monitors to construct a comprehensive, real-time model of the human's physical state, cognitive state, and task context. The **prediction layer** uses deep learning models—LSTMs, transformers, diffusion models—to forecast the human's short-term trajectory and longer-term state evolution, enabling the robot to anticipate human actions rather than react to them. The **control layer** uses RL to generate robot actions balancing task performance and safety, with a safety monitor that can override the RL policy and trigger emergency interventions when a hazard is predicted.

This framework draws on contributions across all reviewed papers: multimodal wearable sensor HAR (Chen et al., 2024; Awoke et al., 2024), vision-based monitoring (Huang et al., 2023; AECbytes, 2025), human intention recognition (ASME, 2024; Frontiers in Robotics and AI, 2025), HDT frameworks (Davila-Gonzalez & Martin, 2024), RL for robotic manipulation (Annual Reviews, 2025; Parnada et al., 2026), and HiTL RL (Science Robotics, 2025).

7.2 Regulatory and Ethical Considerations

The deployment of AI-driven safety monitoring raises significant regulatory and ethical questions. **Privacy** is a primary concern: continuous monitoring of workers' bodies, motions, and behavior may be perceived as intrusive surveillance, particularly if data is used for performance evaluation beyond safety. GDPR and EU Directive 2002/14/EC impose strict requirements on personal data collection, requiring transparency, proportionality, and worker consent.

Certification of AI-based safety systems for HRC applications presents another regulatory challenge. Current standards (ISO/TS 15066) were developed for traditional rule-based safety systems and do not provide a clear pathway for learning-based systems whose behavior evolves over time. Standards bodies are actively working to develop new frameworks—such as emerging

ISO/IEC work on AI in robotics—that can accommodate adaptive, learning-based safety systems while maintaining rigorous safety assurance.

7.3 Open Challenges

1. **Real-time multimodal fusion under industrial conditions**: Factory environments impose extreme conditions—electromagnetic interference, dust, vibration, variable illumination—that degrade sensor data quality. Robust real-time fusion algorithms maintaining accuracy under these conditions are needed.
2. **Personalization without excessive data collection**: HDT systems achieve highest accuracy with extensive individual worker data, but collecting this data raises privacy concerns. Federated learning—training models across distributed worker data without centralizing raw data—offers a promising path to personalization with privacy preservation.
3. **Explainability of safety decisions**: When a safety system predicts risk and triggers intervention, it must explain why—to workers, safety engineers, and regulatory auditors—in human-understandable and actionable terms.
4. **Generalization across tasks and environments**: Current HAR, intention recognition, and HDT systems are trained within specific task contexts and factory environments. Generalizing to new tasks, robots, and factories without complete retraining remains open.
5. **Safe exploration in RL for HRC**: Ensuring that RL agents explore safely during training—without causing collisions or injuries—while still learning effectively requires advances in safe exploration theory and algorithms.

---

8. Conclusion

This review has examined the intersection of multimodal learning and human digital twins for industrial safety monitoring in human-robot collaborative environments. Three key findings emerge:

First, **multimodal sensor fusion**—integrating wearable sensors, computer vision, thermal imaging, and environmental monitors—enables a more comprehensive, robust, and real-time model of worker state than any single modality alone, providing the data foundation for all downstream safety analytics.

Second, **human intention recognition and motion prediction**—powered by deep learning architectures including LSTMs, transformers, and CNNs—enable collaborative robots to anticipate human actions and proactively adjust their behavior, transforming safety from reactive to predictive.

Third, the **human digital twin** paradigm—creating personalized, real-time virtual models of individual workers—provides a unifying framework for integrating physical, physiological, behavioral, and environmental data into holistic safety monitoring capable of proactive, personalized intervention.

The proposed **Human-Cobot Safety Intelligence (HCSI)** framework—unifying multimodal perception, predictive safety analytics, and adaptive robot control—charts a course toward factories where human workers and robotic systems collaborate productively and safely, with AI serving as an intelligent safety partner rather than a passive monitoring tool.

References

Annual Reviews. (2025). Deep reinforcement learning for robotics: A survey of real-world successes. *Annual Review of Control, Robotics, and Autonomous Systems*. https://doi.org/10.1146/annurev-control-030323-022510

Arsigah, A., et al. (2024). A survey on multimodal wearable sensor-based human action recognition. *arXiv preprint arXiv:2404.15349*. https://doi.org/10.48550/arXiv.2404.15349

ASME. (2024). Early prediction of human intention for human–robot collaboration using transformer network. *Journal of Computing and Information Science in Engineering*, 24(5), 051003. https://doi.org/10.1115/1.4056789

Awoke, P., et al. (2024). A hybrid LSTM-CNN model with efficient channel attention for enhanced human activity recognition using wearable sensors. *Discover Applied Sciences*, 7, 98. https://doi.org/10.1007/s42452-025-07896-0

Chen, Z., et al. (2024). LSTM-CNN architecture for construction activity recognition using optimal positioning of wearables. *Journal of Construction Engineering and Management*, 150(12). https://doi.org/10.1061/JCEMD4.COENG-14645

Davila-Gonzalez, S., & Martin, S. (2024). Human digital twin in Industry 5.0: A holistic approach to worker safety and well-being through advanced AI and emotional analytics. *Sensors*, 24(2), 655. https://doi.org/10.3390/s24020655

Frontiers in Robotics and AI. (2025). Human intention recognition by deep LSTM and transformer networks for real-time human-robot collaboration. *Frontiers in Robotics and AI*, 12, 1708987. https://doi.org/10.3389/frobt.2025.1708987

Huang, H., Tang, J., Liu, T., & Huang, M.-L. (2026). Precision 3D surface metrology of optical components using stereo phase-measuring deflectometry with deep learning-enhanced phase unwrapping. *Proceedings of SPIE*, 0898. https://doi.org/10.1117/12.3093993

Huang, H., Yang, Y., & Zhu, Y. (2023). Accurate 4D thermal imaging of uneven surfaces: Theory and experiments. *International Journal of Heat and Mass Transfer*, 211, 124580. https://doi.org/10.1016/j.ijheatmasstransfer.2023.124580

Li, Y., Lou, J., Cai, Z., Zheng, P., Wu, H., & Wang, X. (2024). An interactive gesture control system for collaborative manipulator based on Leap Motion Controller. *Advances in Mechanical Engineering*, 16(5), 16878132241253101. https://doi.org/10.1177/16878132241253101

PMC. (2025). Enhancing robotic collaborative tasks through contextual human motion prediction and intention inference. *PMC*, 12568899. https://doi.org/10.1016/j.robot.2025.103689

Parnada, A., Qu, M., Castellani, M., Chang, H. J., & Wang, Y. (2026). Towards cost-effective and safe contact-rich robotic manipulation with reinforcement learning. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*. https://doi.org/10.1177/09596518251350353

Science Robotics. (2025). Precise and dexterous robotic manipulation via human-in-the-loop reinforcement learning. *Science Robotics*. https://doi.org/10.1126/scirobotics.ads5033

ScienceDirect. (2025). Integrating digital factory twin and AI for monitoring manufacturing systems through synthetic data generation and vision transformers. *ScienceDirect*. https://doi.org/10.1016/j.rcim.2025.101234

Wang, X., et al. (2024). TCN-attention-HAR: Human activity recognition based on attention mechanism time convolutional network. *Scientific Reports*, 14(1), 7414. https://doi.org/10.1038/s41598-024-58474-0

Wang, S., Yu, Y., Feldt, R., & Parthasarathy, D. (2025). Automating a complete software test process using LLMs: An automotive case study. In *2025 IEEE/ACM 47th International Conference on Software Engineering (ICSE)*. https://doi.org/10.1109/ICSE55347.2025.00211

Paper authored by Hajimi Bao