© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

From Inertial to Ambient: A Systematic Review of Sensor Modalities and Data Acquisition Strategies in Human Activity Recognition

OLUWOLE A. AYEGBUSI[1], MOSES O. ONYESOLU[2]
[1]*Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria*
[2]*Department of Software Engineering, Nnamdi Azikiwe University, Awka, Nigeria*

Abstract- Human Activity Recognition (HAR) is a rapidly advancing research domain with transformative applications across healthcare, smart homes, rehabilitation, elderly monitoring, sports analytics, and context-aware computing. The performance and deployability of HAR systems are fundamentally determined by the sensor modalities through which human motion and behavioral data are captured. This systematic review provides a comprehensive examination of the full spectrum of sensor modalities employed in HAR research spanning wearable inertial sensors, physiological sensors, vision-based sensors, depth cameras, ambient sensing systems, and emerging multimodal fusion frameworks tracing, technical characteristics, application domains, and comparative strengths and limitations. The review further examines data acquisition protocols, benchmark datasets, and the persistent challenges of inter-subject variability, sensor placement sensitivity, privacy constraints, and computational efficiency that continue to shape the field. Emerging directions including federated sensing, edge-optimized data acquisition, self-supervised learning from unlabeled sensor streams, and privacy-preserving modalities are critically evaluated as promising pathways toward robust, scalable, and ethically responsible HAR deployment. By synthesizing evidence across more than a decade of HAR sensor research, this review provides a structured reference for researchers and practitioners designing next-generation activity recognition systems, and identifies the most consequential open challenges and future directions for the field.

Keywords: human activity recognition; sensor modalities; wearable sensors; inertial measurement unit; ambient sensing; multimodal fusion; data acquisition; deep learning; privacy-preserving har

I. INTRODUCTION

Human Activity Recognition (HAR) is the automated computational inference of physical human actions and behavioral states from sensor-derived observations. This field has emerged as one of the most consequential and rapidly evolving research domains at the intersection of ubiquitous, artificial, and pervasive computing (Lara & Labrador, 2013; Bulling et al., 2014). By enabling machines to continuously and unobtrusively interpret human physical behavior, HAR systems provide foundational situational awareness for a remarkably broad range of real-world applications. In clinical and rehabilitative healthcare, HAR enables continuous remote patient monitoring, objective functional assessment, early detection of health deterioration, and automated rehabilitation progress tracking, reducing dependence on infrequent clinical observations (Bibbò & Vellasco, 2023 ; Bibbò & Serrano, 2025). In smart home and ambient assisted living environments, HAR underpins adaptive environmental automation, occupancy-sensitive energy management, and safety monitoring systems for fall detection and anomalous behavior recognition in elderly and vulnerable populations (Gorce & Jacquier-Bret, 2025; Gosavi et al., 2026). In sports science, fitness, and military applications, HAR enables objective performance monitoring, personalized training feedback, and physical readiness assessment at scale.

Sensor modalities fundamentally influence the accuracy, robustness, computational requirements, deployability, and privacy characteristics of Human Activity Recognition (HAR) systems. The choice of sensing technology determines the physical phenomena that can be observed, as well as important properties such as spatial and temporal resolution, environmental dependence, computational and energy requirements, user burden, and privacy implications (Bian et al., 2022). HAR research has consequently expanded across diverse sensing paradigms, including wearable inertial and physiological sensors, ambient and environmental sensors, computer vision, radio-frequency sensing, and multimodal systems, with each modality offering distinct advantages and limitations(Bian et al., 2022 ; Shin et al.,2025). Although several comprehensive surveys have examined HAR sensing modalities, datasets, algorithms, and challenges, their coverage and emphasis vary considerably, with some focusing primarily on wearable sensing, others on vision and sensor-based HAR, multimodal recognition, or particular sensing technologies(Dang et al., 2020; Yadav et al., 2021). Moreover, issues surrounding data acquisition, dataset diversity, evaluation standardization, cross-environment generalization, resource constraints, privacy, and real-world deployment remain important challenges in HAR research. These considerations motivate a systematic examination that jointly considers sensor modalities, data acquisition strategies, benchmark datasets and evaluation practices, multimodal fusion, and the practical challenges associated with reliable and responsible real-world HAR deployment.

Early HAR research relied predominantly on vision-based systems using fixed cameras, later supplemented by depth sensors and skeleton tracking technologies (Aggarwal & Ryoo, 2011). The democratization of miniaturized, low-cost microelectromechanical systems (MEMS) inertial sensors particularly triaxial accelerometers and gyroscopes embedded in smartphones and dedicated wearable platforms subsequently established inertial measurement as the dominant sensing paradigm for HAR, enabling large-scale, privacy-preserving data collection in naturalistic real-world environments (Lara & Labrador, 2013; Bulling et al., 2014). More recently, the integration of physiological sensors, ambient environmental monitors, radio-frequency sensing systems, and multimodal sensor arrays has substantially expanded the HAR sensing landscape, offering complementary perspectives on human physical behavior that no single modality can provide in isolation (Ma et al., 2019; Ramamurthy & Roy, 2018).

Concurrently, advances in machine learning have shifted Human Activity Recognition (HAR) from conventional pipelines based on handcrafted feature engineering and classifier design toward end-to-end deep learning architectures, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and, more recently, Transformer-based models. These approaches have reduced reliance on manually engineered features and enabled increasingly effective representation learning from sensor data (Wang et al., 2019; Vaswani et al., 2017). Nevertheless, HAR performance and generalization remain strongly influenced by the quality, diversity, representativeness, and acquisition conditions of the sensor data, alongside factors such as sensor placement, subject variability, and domain shift (Wang et al., 2019; Wang et al., 2019b). Understanding the sensor-modality landscape is therefore a central consideration in designing HAR systems capable of robust performance under diverse real-world conditions

This review makes the following contributions to the HAR sensor literature. First, it provides comprehensive systematic survey to date of HAR sensor modalities, spanning inertial, physiological, vision-based, depth, ambient, radio-frequency, and multimodal sensing technologies, with detailed technical characterization of each modality's physical measurement principles, signal characteristics, and HAR-relevant properties. Second, it critically examines data acquisition protocols, windowing strategies, annotation methodologies, and benchmark datasets, identifying best practices and persistent methodological limitations. Third, it synthesizes comparative performance evidence across sensing modalities and identifies the application domains, activity complexity levels, and deployment contexts for which each modality is most appropriate. Fourth, it comprehensively characterizes the open challenges including inter-subject variability, privacy constraints, computational efficiency, and ecological validity that continue to limit HAR sensor research. Fifth, it identifies and evaluates emerging sensing directions and data acquisition strategies that hold the greatest promise for advancing robust, scalable, and privacy-conscious next-generation HAR systems.

The remainder of this review is organized as follows: Section 2 presents an overview of the HAR sensing landscape. Section 3 provides a comprehensive review of wearable inertial and physiological sensor modalities. Section 4 examines vision-based and depth sensing modalities. Section 5 covers ambient and radio-frequency sensing approaches. Section 6 addresses multimodal sensor fusion strategies. Section 7 reviews data acquisition protocols and benchmark datasets. Section 8 identifies current challenges. Section 9 discusses emerging sensing directions and future research. Section 10 concludes the review.

II. OVERVIEW OF THE HAR SENSING LANDSCAPE

2.1 Taxonomy of HAR Sensor Modalities

The diversity of sensor modalities employed in HAR research reflects the multidimensional nature of human physical behavior, which manifests simultaneously through body kinematics, physiological state changes, environmental interactions, and optical appearance. For the purposes of this review, HAR sensor modalities are organized into five principal categories: (1) wearable inertial and motion sensors; (2) wearable physiological sensors; (3) vision-based optical sensors; (4) ambient and environmental sensors; and (5) radio-frequency and contactless sensing systems. Each category captures a distinct physical dimension of human activity, with characteristic measurement principles, signal properties, application domains, and deployment constraints.

Table 1 provides a high-level taxonomy of the principal HAR sensing modalities reviewed in this paper, highlighting their advantages, limitations and typical use.

Table 1: Comparison of HAR Sensor Modalities

Sensor/Modality	Typical Sensors	Information Captured	Major Advantages	Major Limitations	Key Applications
Inertial/Kinematic	Accelerometer, gyroscope, magnetometer, IMU	Acceleration, rotation, orientation, body movement	Low cost, compact, low power, wearable, strong temporal information	Sensor placement and orientation can affect performance	Activities of daily living(ADL)
Physiological	ECG, EMG, PPG, heart-rate, skin-temperature sensors	Cardiac, muscular and physiological responses	Provides information beyond physical movement; useful for health-related activities	Individual variability; signal noise; sensor placement and contact issues	Exercise, fatigue monitoring, rehabilitation, healthcare
Environmental/Ambient	PIR, temperature, humidity, light, pressure, CO_2	Environmental context and presence/movement	Can provide contextual information without body attachment	environment dependent	Smart homes, indoor activities, context recognition
Vision-based	RGB, infrared, depth cameras	movement, spatial appearance	Rich spatial information; can recognize	Privacy concerns, occlusion,	Gesture recognition, sports,

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

			complex movements	lighting and viewpoint sensitivity	surveillance, interaction
Radio/Wave-based	Wi-Fi CSI, RF, UWB, radar, mmWave	Motion, distance, position, body movement	Can support contactless/device-free recognition	Hardware and signal-processing complexity	Indoor monitoring

2.2 Evolution of HAR Sensing Paradigms

The evolution of HAR sensing can be broadly viewed through three overlapping paradigms, each characterized by different technological capabilities and dominant sensing approaches. Early vision-based HAR, particularly during the 1990s and early 2000s, relied heavily on fixed cameras and motion representations such as frame differencing and optical flow to recognize activities from video. Although these approaches established important foundations for vision-based activity recognition, they were sensitive to illumination and occlusion and raised significant privacy concerns (Aggarwal & Ryoo, 2011). Subsequently, advances in MEMS technology, smartphones, and wearable devices accelerated the adoption of inertial sensing through accelerometers and gyroscopes, enabling more unobtrusive and scalable activity-data collection and supporting increasingly practical HAR applications (Lara & Labrador, 2013; Mukhopadhyay, 2015). Inertial sensing consequently became one of the most widely adopted sensing paradigms in wearable and mobile HAR.

The contemporary multimodal era is characterized by the increasing integration of diverse sensing modalities, including inertial, physiological, vision, and ambient sensors, together with deep-learning methods that enable automatic representation and feature learning from heterogeneous sensor data. Such multimodal approaches can provide complementary information and richer contextual representations for human activity recognition, although their effectiveness depends on appropriate sensor selection, synchronization, and fusion strategies (Wang et al., 2019; Wang et al., 2019b).

2.3 Factors Governing Sensor Modality Selection

The selection of a sensor modality in HAR is influenced by the target activity, sensing environment, recognition requirements, sensor placement, cost, energy consumption, computational resources, robustness, privacy, and user comfort. The characteristics of the target activities influence the type and placement of sensors required, while deployment constraints such as battery capacity, device size, and processing capability can limit the practicality of particular sensing modalities. Inertial sensors are widely adopted in wearable and mobile HAR because they can provide relatively low-cost, compact, and computationally efficient motion sensing, whereas vision- and RF-based approaches can provide complementary contextual information but may introduce additional computational, environmental, or privacy considerations (Zhang et al., 2022; Wang et al., 2019). Sensor placement is also important because the location of a wearable sensor affects the motion information captured and can consequently influence recognition performance (Atallah et al., 2011). Energy efficiency is particularly important for continuously operating wearable systems, where sensing, wireless communication, and on-device processing can contribute substantially to energy consumption and reduced battery life (Contoli et al., 2024). Therefore, sensor selection generally involves balancing recognition capability, practicality, energy efficiency, privacy, robustness, and deployment requirements according to the intended HAR application.

III. WEARABLE INERTIAL AND PHYSIOLOGICAL SENSOR MODALITIES

3.1 Accelerometers

3.1.1 Operating Principles and Signal Characteristics

Accelerometers are among the most widely used sensors in wearable HAR because they provide direct measurements of body-motion-related acceleration and can be readily integrated into smartphones, smartwatches, and other compact wearable devices.

They measure acceleration along one or more orthogonal axes, typically producing uni-, bi-, or tri-axial time-series signals that characterize movement in different directions. Accelerometer measurements contain both dynamic components associated with body movement and a gravitational component that can provide information about sensor orientation and posture. Consequently, accelerometer signals can capture temporal patterns associated with activities such as walking, running, sitting, standing, and transitions between activities. Their small size, relatively low cost, and low power requirements have contributed to their widespread adoption in wearable HAR systems (Demrozi et al., 2020; Lara & Labrador, 2013).

For HAR, accelerometer data are commonly acquired as discrete time-series signals and subsequently subjected to preprocessing and segmentation, followed by handcrafted feature extraction or direct input into machine- and deep-learning models. Signal characteristics, including sampling frequency, measurement range, resolution, sensor placement, orientation, and noise, can influence the quality of the resulting representations and recognition performance. Tri-axial accelerometers are particularly useful because they capture multidirectional motion, although differences in sensor placement and orientation can introduce variability across recordings and users. Recent research has increasingly applied machine-learning methods to accelerometer time-series data for activity identification, although conventional signal-processing and feature-based approaches remain widely used (Letts et al., 2024). Deep-learning approaches can further enable direct learning of representations from raw sensor signals, reducing reliance on manually engineered features (Demrozi et al., 2020).

Accelerometers are among the most extensively studied sensing modalities in HAR research and are included in many widely used benchmark datasets. Their sensitivity to both dynamic body motion and gravitational orientation enables discrimination across a broad range of activities, from sedentary postures such as sitting, standing, and lying to dynamic activities such as walking, running, cycling, jumping, and stair climbing. Bao and Intille (2004) provided an influential early demonstration that body-worn accelerometers could distinguish a range of everyday activities, achieving 84% overall accuracy using decision-tree classification across 20 activities and 20 subjects.

Subsequent accelerometer-based studies have reported moderate to very high classification performance, although results vary considerably according to the activity set, sensor configuration, placement, evaluation protocol, and subject characteristics. Janidarmian et al. (2017) demonstrated that wearable acceleration-based HAR performance is influenced by factors including sensor placement, sensing configuration, and feature representation. Consequently, although accelerometers remain a practical and effective modality for HAR, performance reported on controlled laboratory datasets should be interpreted cautiously when assessing generalization to unseen users and real-world environments.

3.2 Gyroscopes

3.2.1 Operating Principles and Signal Characteristics

3.2.1 Operating Principles and Signal Characteristics

Gyroscopes are inertial sensors that measure angular velocity, typically expressed in degrees per second (°/s) or radians per second (rad/s), along one or more orthogonal axes. In wearable HAR, tri-axial gyroscopes are commonly integrated with accelerometers to capture rotational dynamics of body segments, providing information that complements linear acceleration measurements. This complementary information can aid the discrimination of activities involving rotational motion, changes in body orientation, and complex movement patterns (Demrozi et al., 2020; Bian et al., 2022). The resulting signals are continuous time-series measurements whose characteristics depend on factors such as sampling frequency, measurement range, resolution, sensor placement, noise, and orientation. Unlike accelerometers, gyroscopes primarily measure rotational motion; however, bias and drift can accumulate over time and affect orientation estimation. In HAR, gyroscope signals are therefore commonly segmented into temporal windows and processed using handcrafted time- and

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

frequency-domain features or directly with machine- and deep-learning models. Their complementary information, particularly when combined with accelerometer data, has made gyroscopes a common component of multimodal inertial sensing systems used in wearable HAR (Webber et al., 2024; Saha et al., 2024).

The signal characteristics of gyroscopes depend on the sensing device and operating configuration. Important parameters include full-scale measurement range, sensitivity, resolution, output data rate, bandwidth, noise characteristics, and power consumption. These parameters vary substantially across commercial MEMS gyroscopes, with some devices offering full-scale ranges extending to approximately ±4000°/s. Sampling rates used in wearable human-motion monitoring vary according to the target activities, sensor hardware, and application requirements (Zhang et al., 2023). Therefore, neither a particular measurement range nor sampling frequency should be regarded as universal; sensor configuration should instead be selected according to the motion dynamics of the target activities and the computational and energy constraints of the HAR system.

An important limitation of MEMS gyroscopes is bias instability and drift. Small measurement biases and other sensor errors can accumulate when angular-velocity measurements are integrated over time to estimate orientation or angular displacement, resulting in increasing estimation errors. Temperature variations can also contribute to bias and scale-factor variations in MEMS gyroscopes because changes in operating temperature affect their mechanical and electrical characteristics. These effects can reduce the reliability of orientation estimation, particularly during prolonged integration of angular-velocity measurements (Du et al., 2018). Consequently, calibration and compensation techniques, including temperature-based calibration, filtering, and model-based compensation, are commonly employed to reduce bias and thermal drift (Niu et al., 2013; Du et al., 2018).

3.2.2 HAR Applications and Complementarity with Accelerometers

Gyroscopes provide angular-velocity information that complements the specific-force measurements obtained from accelerometers, thereby providing a more comprehensive representation of human motion in wearable HAR. This complementary information is particularly useful for activities involving rotational motion, changes in body orientation, and repetitive or cyclical movements, where rotational patterns may provide discriminative information beyond acceleration alone. Consequently, accelerometer–gyroscope combinations are widely used in IMU-based HAR systems to characterize complementary linear and rotational aspects of human movement (Demrozi et al., 2020; Webber et al., 2024).

The usefulness of multisensor inertial representation is illustrated by the PAMAP2 Physical Activity Monitoring dataset introduced by Reiss and Stricker (2012). PAMAP2 employs three wearable IMUs positioned on the dominant wrist, chest, and dominant ankle, with each IMU incorporating triaxial accelerometer, gyroscope, and magnetometer sensors. The dataset includes activities such as walking, running, cycling, Nordic walking, ascending and descending stairs, rope jumping, and soccer, with the inertial sensors sampled at 100 Hz (Reiss & Stricker, 2012).

From a sensing perspective, accelerometers and gyroscopes should therefore be regarded as complementary rather than redundant modalities. Accelerometers measure specific force, comprising dynamic components associated with body motion together with the gravitational component, whereas gyroscopes measure angular velocity. Their combination can provide richer kinematic information and may improve discrimination of activities involving complex rotational, dynamic, or transitional movements, although the benefits depend on activity characteristics, sensor placement, signal quality, and the recognition or fusion method.

3.3 Inertial Measurement Units (IMUs) and Sensor Fusion

Inertial Measurement Units (IMUs) integrate multiple inertial sensors, most commonly triaxial accelerometers and gyroscopes, and may additionally include a triaxial magnetometer. In wearable HAR, this combination provides complementary information about dynamic motion, rotational movement, and, when available, magnetic-field-based orientation. The integration of these measurements can provide a richer representation of human movement than a single inertial sensor and is therefore widely used in wearable activity-recognition systems (Demrozi et al., 2020; Gomaa & Khamis, 2023).

Sensor fusion combines measurements from multiple sensors to exploit their complementary characteristics while mitigating the limitations of individual sensors. In HAR, fusion may be performed at the data, feature, or decision level, depending on where information from different sensors is combined. Combining accelerometer and gyroscope measurements can provide complementary motion information and may improve recognition robustness, while magnetometer integration can provide additional magnetic-field information useful for orientation estimation. However, fusion also introduces challenges involving synchronization, differences in sensor characteristics, computational cost, noise, and interference (Aguileta et al., 2019; Demrozi et al., 2020). The effectiveness of sensor fusion in wearable HAR therefore depends on factors such as sensor placement, sampling rate, signal quality, synchronization, and the selected fusion strategy.

Chung et al. (2019) systematically investigated on-body sensor positioning and data-acquisition configurations using multiple wearable IMUs and demonstrated that, for selected activities of daily living, effective HAR could be achieved using a reduced sensor configuration and sampling rates as low as 10 Hz. Their study also employed classifier-level multimodal sensor fusion and showed that combining class-probability outputs from multiple sensor sources could improve classification performance. These findings highlight the importance of jointly considering sensor configuration, data-acquisition parameters, and fusion strategy when designing wearable HAR systems.

3.4 Magnetometers

Magnetometers measure the magnitude and directional components of a magnetic field and, in wearable systems, are commonly used to sense the Earth's geomagnetic field. Within an Inertial Measurement Unit (IMU), triaxial magnetometers provide magnetic-field information that can support heading and orientation estimation when combined with accelerometer and gyroscope measurements (Demrozi et al., 2020; Zhang et al., 2022). Because the geomagnetic field provides a relatively stable local reference, magnetometers can provide additional orientation-related information, particularly for heading estimation, and may therefore contribute discriminative cues for activities involving changes in body orientation.

Magnetometer signals are typically represented as three-axis time-series measurements corresponding to the magnetic-field components along the sensor's orthogonal axes. Signal quality depends on factors such as sensor sensitivity, sampling rate, measurement range, noise, calibration, sensor orientation, and the surrounding magnetic environment. A major limitation of magnetometers is their susceptibility to magnetic disturbances caused by ferromagnetic materials, hard- and soft-iron effects, and magnetic or electromagnetic fields generated by nearby electronic devices (Soken, 2017). Consequently, careful calibration and appropriate sensor-fusion techniques are important for reliable orientation estimation. In wearable HAR, magnetometers therefore generally serve as a complementary sensing modality rather than a replacement for accelerometer- and gyroscope-based measurements (Demrozi et al., 2020; Zhang et al., 2022).

3.5 Barometric Pressure Sensors

Barometric pressure sensors measure atmospheric pressure and, in HAR, are primarily used to estimate relative changes in elevation or vertical displacement. With the integration of compact MEMS barometers into smartphones and wearable devices, these sensors have become useful for recognizing activities involving changes in elevation, such as ascending and descending stairs, riding elevators or escalators, and moving between floors. Temporal patterns and rates of pressure change can provide complementary information for distinguishing such activities, making barometers particularly useful when combined with inertial sensors (Manivannan et al., 2020).

In wearable HAR, barometric signals are generally sampled at lower frequencies than inertial signals because atmospheric-pressure variations associated with changes in elevation occur relatively slowly. Lower sampling requirements can reduce sensing and processing demands and may therefore support energy-efficient operation. Lower sampling requirements can reduce sensing and processing demands and may therefore support energy-efficient operation. Barometric measurements are also generally less sensitive to device orientation than inertial measurements, although sensor placement and local environmental conditions can still influence the recorded pressure. Atmospheric pressure measurements can be affected by weather, temperature, indoor environmental conditions, local pressure variations, sensor accuracy, and differences in reference pressure between locations or devices (Manivannan et al., 2020). Consequently, barometric sensors are generally most valuable as a complementary modality, particularly for recognizing vertical activities and elevation changes, rather than as a standalone solution for general-purpose HAR.

3.6 Physiological Wearable Sensors

3.6.1 Photoplethysmography (PPG) and Heart Rate

Photoplethysmography (PPG) is a non-invasive optical sensing technique that detects variations in light intensity resulting from changes in optical absorption associated with pulsatile blood-volume changes in tissue. A typical PPG sensor consists of one or more light-emitting diodes (LEDs) and a photodetector, with the resulting waveform providing a basis for estimating pulse rate and, under appropriate conditions, heart rate. Wearable PPG systems commonly employ green, red, or near-infrared wavelengths and are widely integrated into wrist-worn devices such as smartwatches and fitness trackers (Tamura et al., 2014; Charlton et al., 2022).

In HAR, PPG provides physiological information that complements motion-related measurements from accelerometers and gyroscopes. Changes in heart rate associated with physical exertion can provide additional information about physiological response and, indirectly, activity intensity. However, PPG signals are susceptible to motion artifacts, particularly during dynamic activities, while sensor placement, skin contact, tissue characteristics, and variations in peripheral blood flow can also affect signal quality. Signal-processing and artifact-reduction techniques, together with sensor fusion with inertial measurements, can therefore be employed to improve the reliability of PPG-derived physiological features in wearable HAR systems (Charlton et al., 2022).

A major limitation of PPG-based sensing in HAR is its susceptibility to motion artifacts. Movement of the sensor relative to the skin, changes in contact pressure, tissue deformation, and variations in the optical path can contaminate the PPG waveform, particularly during dynamic and vigorous activities. These artifacts can reduce the accuracy and reliability of heart-rate estimation and consequently affect the quality of physiological features used for activity classification (Tamura et al., 2014; Charlton et al., 2022). In addition, cardiovascular responses to changes in physical activity generally exhibit different temporal dynamics from the rapid kinematic changes captured by accelerometers and gyroscopes. Consequently, PPG-derived heart rate generally changes more gradually than inertial measurements and can provide complementary information about the physiological response to activity. PPG is therefore best regarded as a complementary physiological modality that can enrich inertial sensing rather than replace it in wearable HAR systems.

3.6.2 Electromyography (EMG) sensors

Electromyography (EMG) sensors measure electrical activity associated with muscle activation during voluntary movement. Surface electromyography (sEMG) uses electrodes positioned on the skin over selected muscles to detect electrical potentials generated by the activity of underlying motor units. The resulting signals provide high temporal-

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

resolution information about neuromuscular activity and can be characterized using time-, amplitude-, and frequency-domain features, as well as spatial and inter-channel activation patterns. These characteristics provide discriminative information about muscle recruitment and movement-related activity patterns (Phinyomark et al., 2012).

In human activity recognition (HAR), EMG is particularly relevant to gesture recognition, hand and forearm movement classification, prosthetic control, rehabilitation, and muscle-activity assessment. Unlike inertial sensors, which primarily characterize external body kinematics, EMG provides information about the neuromuscular activity underlying movement. This complementary information can be valuable for fine-grained hand and finger movements whose kinematic signatures may be limited or difficult to distinguish using inertial measurements alone. EMG-derived features can also provide information about changes in muscle activation and physiological states such as muscle fatigue, thereby complementing motion-based sensing (Phinyomark et al., 2012).

Despite these advantages, sEMG presents practical challenges for continuous wearable HAR. Signal quality can be affected by electrode placement, skin–electrode impedance, electrode displacement, anatomical differences between users, muscle cross-talk, and motion artifacts. Reliable measurements require appropriate electrode positioning and adequate skin–electrode contact, which can make long-term deployment more demanding than inertial sensing (Ghapanchizadeh et al., 2017; Guo et al., 2020).. Conventional wet electrodes may require skin preparation to achieve good electrical contact and can present adhesion and comfort challenges during prolonged use. Although dry and textile electrodes offer alternatives for wearable applications, they may introduce their own challenges related to contact stability and motion artifacts. These factors may limit the practicality of EMG for unobtrusive, continuous HAR in everyday environments, despite its high information content for muscle-related activity recognition(Guo et al., 2020).

IV. VISION-BASED AND DEPTH SENSING MODALITIES

4.1 RGB Cameras

4.1.1 Operating Principles and Signal Characteristics

RGB cameras are vision-based sensing devices that capture visible-light information through red, green, and blue color channels to produce two-dimensional color images. In Human Activity Recognition (HAR), sequences of RGB frames provide rich spatial information within individual frames and temporal information across successive frames, enabling the analysis of human appearance, body posture, movement, and interactions with surrounding objects. Unlike wearable inertial sensors, which directly measure motion-related physical quantities such as acceleration and angular velocity, RGB cameras enable non-contact observation of human activities and can capture holistic whole-body movements without requiring users to wear dedicated sensing hardware (Aggarwal & Xia, 2014; Beddiar et al., 2020). RGB video streams are represented as sequential image frames characterized by acquisition parameters such as spatial resolution, frame rate, field of view, exposure settings, and camera placement. These parameters influence the spatial and temporal information available for activity classification, while RGB-based systems remain sensitive to environmental factors such as illumination variation, scene occlusion, viewpoint changes, and background clutter, in addition to significant privacy concerns (Beddiar et al., 2020).

4.2 Depth Sensors

4.2.1 Operating Principles

Depth sensors capture information about the distance between the sensing device and objects in a scene, commonly producing depth maps in which each pixel represents an estimated distance to the corresponding scene point. Common depth-sensing technologies used in HAR include structured light, time-of-flight (ToF), and stereo vision. Structured-light systems project a known pattern onto a scene and estimate depth from its observed deformation, whereas ToF sensors estimate distance from the travel time of emitted light and its return. Stereo systems estimate depth from the disparity between corresponding image points captured by spatially separated cameras

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

(Han et al., 2013; Aggarwal & Xia, 2014). These approaches provide geometric information that can be used to characterize body shape, posture, joint positions, and movement while reducing some of the appearance-related ambiguity associated with conventional RGB images.

In HAR, depth information can be represented directly as depth maps or reconstructed three-dimensional point clouds, or transformed into representations such as depth-based silhouettes and skeletal joint coordinates. These representations can be processed using spatiotemporal features, convolutional neural networks, and other deep-learning architectures. A major advantage of depth sensing is its reduced dependence on RGB color and surface texture, making it useful for posture and motion analysis. However, depth measurements can be affected by occlusion, limited sensing range, reflective or transparent surfaces, sensor viewpoint, missing or noisy measurements, and environmental or illumination conditions depending on the sensing technology. Depth sensing has therefore been extensively investigated in indoor and controlled HAR settings, although its performance and practicality depend strongly on the underlying sensing technology and deployment environment (Han et al., 2013; Aggarwal & Xia, 2014).

4.3 Infrared and Thermal Imaging Sensors

Infrared-based sensing provides non-contact information about human presence, movement, and thermal patterns and has been investigated as an alternative to conventional RGB vision for HAR. Infrared sensing encompasses technologies ranging from passive infrared (PIR) detectors to thermographic cameras, which differ substantially in sensing principle, spatial resolution, and information content. PIR sensors primarily detect changes in incident infrared radiation associated with movement, whereas thermal cameras capture spatial distributions of long-wave infrared radiation emitted by objects and can represent variations in apparent surface temperature as thermal images. Unlike RGB cameras, thermal imaging does not rely on visible illumination and can therefore support human detection and activity monitoring in low-light or dark environments (Bian et al., 2022).

In HAR, infrared and thermal sensing have been applied to human presence detection, posture and movement recognition, fall detection, and activity monitoring, particularly in smart-home, healthcare, and assisted-living environments. Thermal data can be represented as sequences of thermal images or low-resolution infrared arrays and processed using image-based features, silhouettes, CNNs, and other spatiotemporal learning methods. Low-resolution infrared sensing can provide a potential privacy advantage by reducing identifiable visual detail compared with conventional RGB imagery, while retaining information about human movement and activity (Karayaneva et al., 2023; Newaz & Hanada, 2024).

However, infrared and thermal sensing have limitations, including sensor cost, resolution constraints, occlusion, viewpoint and sensor-placement effects, environmental-temperature dependence, and reduced availability of detailed texture and appearance information compared with RGB imagery. Thermal measurements can also vary with ambient temperature, emissivity, reflections, distance, and sensor characteristics, while the effectiveness of low-resolution infrared sensing depends on sensor layout and viewing perspective. Consequently, infrared and thermal sensors are particularly attractive for privacy-conscious, low-light, and indoor HAR applications, although their suitability depends on the required spatial detail, environmental conditions, sensing range, and recognition task (Fernández-Cuevas et al., 2024).

V. AMBIENT AND RADIO-FREQUENCY SENSING MODALITIES

5.1 Passive Infrared (PIR) Motion Sensors

Passive Infrared (PIR) sensors are ambient, non-contact sensing devices that detect temporal changes in incident infrared radiation, typically associated with the movement of warm objects such as humans within their field of view. Conventional PIR sensors use pyroelectric elements to detect variations in infrared radiation and do not actively illuminate or transmit radiation. In HAR, PIR sensors are primarily used for human presence and movement detection and have been investigated in ambient and smart-

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

environment applications (Demrozi et al., 2020; Bian et al., 2022).

PIR measurements generally provide low-dimensional temporal information about changes in detected infrared radiation rather than detailed representations of body posture or limb motion. Consequently, multiple PIR sensors can be deployed at different locations to infer activity patterns from their sequences of activation. Their advantages include low cost, low power consumption, unobtrusive operation, and a potential visual-privacy advantage over conventional RGB cameras because conventional PIR sensors do not normally acquire identifiable images. However, their limited spatial and semantic information makes them generally less suitable for fine-grained activities requiring detailed body-motion or posture information. Recognition performance can also depend on sensor placement, field of view, environmental conditions, and the number and spatial arrangement of sensors (Bian et al., 2022).

5.2 Pressure and Force Sensing

Pressure and force sensors measure mechanical loading, including contact force and pressure distribution, and can provide information complementary to motion-based sensing in HAR. Depending on the sensing configuration, they can be deployed as wearable sensors, such as pressure-sensitive insoles, or as ambient systems, such as pressure mats and instrumented surfaces. Common technologies include force-sensitive resistors (FSRs), resistive textile sensors, capacitive pressure sensors, and pressure-sensitive mats. In HAR, these sensors are particularly useful for detecting load distribution, foot–ground interaction, gait patterns, posture, and transitions between activities (Bian et al., 2022).

Plantar-pressure sensing is an important application in which pressure-sensitive insoles or instrumented footwear capture the spatial and temporal distribution of loading beneath the feet. These measurements can provide information about gait characteristics, plantar loading, weight-transfer patterns, and posture that complements inertial measurements. For example, Sazonov et al. (2011) demonstrated that a shoe-based system combining plantar-pressure and acceleration measurements could recognize several common postures and locomotor activities.

However, pressure-based HAR systems are influenced by sensor placement, calibration, mechanical loading, contact conditions, and individual differences in gait and body mechanics. Instrumented floors and pressure mats also constrain the sensing area, whereas sensorized footwear or insoles require the user to wear the sensing system. Consequently, pressure and force sensing is particularly valuable for gait, posture, load-transfer, and transition recognition, while its suitability for continuous HAR depends strongly on the sensing configuration and deployment environment (Bian et al., 2022; Sazonov et al., 2011)

5.3 Smart Home Object and Appliance Interaction Sensors

Smart-home object and appliance sensing captures events or state changes associated with household objects, fixtures, and appliances, providing contextual information for ambient HAR. These sensors can include door and cabinet contact sensors, appliance-state sensors, smart switches, electrical-usage sensors, RFID-based object-identification systems, and sensors embedded in household objects or furniture. Rather than directly measuring body motion, such sensors provide information about changes in the state or use of objects for example, a refrigerator door opening, a cupboard being accessed, a light being switched on, or an appliance being activated. These object-related events can provide contextual evidence for recognizing activities of daily living in smart environments (Bouchabou et al., 2021; Sanchez-Comas et al., 2020).

In HAR, object- and appliance-related sensors are particularly useful for recognizing activities of daily living and household routines, where interactions with specific objects provide contextual information about the activity being performed. Sequences of sensor activations can be combined to infer higher-level activities because the temporal order and semantic meaning of individual sensor events provide useful contextual information (Bouchabou et al., 2021). For example, combinations of cupboard, refrigerator, and appliance events may provide evidence of food-preparation activities. Many such sensors generate binary or low-dimensional event streams, although some appliance and energy sensors provide continuous measurements. Their observations depend strongly on the objects that are instrumented, sensor placement, household layout, and individual interaction patterns. Consequently, models developed in one home may face generalization challenges when deployed in homes with different layouts, objects, appliances, or routines (Bouchabou et al., 2021).

A major advantage of object- and appliance-based sensing is that it provides contextual information that can complement motion-oriented sensing modalities such as PIR, pressure, environmental, and wearable sensors. Combining different ambient sensing sources can provide a richer representation of complex activities of daily living than relying on a single sensor type. However, extensive instrumentation can increase installation and maintenance requirements, while object-related sensors generally provide indirect evidence of a person's activity rather than direct measurements of body movement. Long-term smart-home HAR systems must therefore also consider changes in household environments, sensor availability, and resident behavior (Hiremath & Plötz, 2023).

5.4 WiFi Channel State Information (CSI) Sensing

Wi-Fi Channel State Information (CSI) sensing is a device-free radio-frequency sensing modality that exploits changes in wireless signal propagation caused by human movement. CSI characterizes the wireless channel between transmitting and receiving antennas across multiple subcarriers and contains information about signal amplitude and phase. When a person moves within or near the propagation environment, changes in multipath propagation and signal reflections modify the measured CSI, which can be analyzed to infer human activities without requiring the person to carry a sensing device (Ma et al., 2019). This makes CSI attractive for unobtrusive HAR in indoor environments and smart-home applications.

In HAR, CSI-based systems typically analyze temporal variations in CSI amplitude and/or phase using signal-processing techniques and machine-learning or deep-learning models to recognize activities such as walking, sitting, standing, falling, and other movements. Because commodity Wi-Fi infrastructure can potentially be used for sensing, CSI provides an opportunity for device-free activity recognition without cameras or dedicated wearable sensors. CSI-based sensing can operate in some non-line-of-sight configurations and has also been demonstrated for through-wall activity recognition, although performance depends strongly on propagation conditions, wall characteristics, and the sensing configuration.

Despite these advantages, CSI sensing is sensitive to multipath propagation, environmental changes, transmitter–receiver configuration, antenna arrangement, human position, and interference. Changes in furniture, room layout, or sensing-device placement can alter the underlying channel characteristics and affect recognition performance. CSI phase measurements may also contain hardware-related distortions and therefore often require appropriate calibration or preprocessing. Consequently, CSI-based HAR offers a potentially privacy-preserving and unobtrusive alternative to wearable and vision-based sensing, but robustness and generalization across different environments remain important challenges (Ma et al., 2019).

5.5 Ultra-Wideband (UWB) and Other Radar Sensing

Ultra-Wideband (UWB) radar and other radar-based sensing systems provide non-contact measurements of human presence, position, and movement by analyzing electromagnetic signals transmitted toward and reflected or scattered by targets. UWB radar employs wide-bandwidth signals that can provide information about target range and motion, while Doppler and micro-Doppler characteristics can characterize gross body movement and finer motion components associated with moving body segments. These properties make radar-based sensing attractive for HAR in situations where wearable or vision-based sensing is undesirable, including low-light environments and, depending on the sensing configuration and propagation conditions, some visually occluded or through-wall scenarios (Li et al., 2019).

Radar-based HAR can employ representations such as range profiles, range-Doppler maps, Doppler spectrograms, and micro-Doppler signatures, which can subsequently be analyzed using signal-processing techniques, machine-learning algorithms, or deep-learning models. These representations have been used for recognizing activities such as walking, sitting, standing, falling, and other human movements. The wide bandwidth of UWB radar can provide high range resolution, while radar sensing does not rely on visible illumination and can therefore operate in darkness and other challenging lighting conditions (Li et al., 2019).

Despite these advantages, radar-based HAR is affected by multipath propagation, environmental changes, sensor placement, target position, interference, and the presence of multiple people. Changes in the sensing environment or configuration can alter the received radar signatures and consequently affect recognition performance and model generalization. Through-wall operation is also dependent on radar frequency, wall properties, target position, and system configuration rather than being a universal capability of all radar systems (Savvidou et al., 2024).

Radar sensing nevertheless offers important advantages for ambient HAR, including contactless operation, independence from visible illumination, and potentially lower visual privacy intrusion than conventional cameras because radar does not normally generate recognizable visual images. These characteristics make radar particularly promising for smart-home, healthcare, assisted-living, and other privacy-sensitive monitoring applications (Savvidou et al., 2024; Li et al., 2019).

5.6 Acoustic Sensing

Acoustic sensing uses microphones or other acoustic transducers to capture sounds generated by human activities or interactions with the surrounding environment. In HAR, acoustic signals can contain information about speech, footsteps, clapping, coughing, object manipulation, and other activity-related sounds. Acoustic sensing can provide non-contact monitoring without requiring the person to wear a dedicated sensing device. Captured signals are typically represented as temporal waveforms and may be transformed into frequency- or time–frequency representations, such as spectrograms and Mel-frequency cepstral coefficients (MFCCs), for subsequent sound-event or activity classification (Bian et al., 2022; Ntalampiras et al., 2011).

Acoustic sensing has been investigated for recognizing activities such as walking, running, falling, coughing, clapping, and interactions with household objects, particularly in smart-home and ambient-assisted-living environments. Because acoustic signals can capture events that may not be readily observable through motion-oriented sensors, they can complement modalities such as PIR, inertial, pressure, and object-interaction sensors. Deep-learning approaches, including CNN-based models applied to spectrograms and other learned acoustic representations, can automatically extract discriminative features from activity-related sounds. However, acoustic sensing is strongly affected by background noise, room acoustics, source-to-microphone distance, reverberation, competing sound sources, and microphone placement, which can reduce recognition performance in uncontrolled environments (Ntalampiras et al., 2011; Bian et al., 2022).

VI. MULTIMODAL SENSOR FUSION STRATEGIES

Multimodal sensor fusion combines information from two or more sensing modalities to obtain a more comprehensive representation of human activities. In HAR, different sensors provide complementary information: accelerometers and gyroscopes capture complementary aspects of body kinematics, including linear acceleration and angular velocity; physiological sensors provide information about bodily responses; while PIR, pressure, object-interaction, Wi-Fi CSI, and radar sensors provide environmental, contextual, or device-free information. Combining these modalities can potentially improve recognition robustness and mitigate limitations associated with reliance on a single sensing source (Atrey et al., 2010; Bian et al., 2022). Multimodal fusion can generally be categorized according to the stage at which information from different sensors is integrated, including data-level, feature-level, and decision-level fusion (Atrey et al., 2010; Ramachandram & Taylor, 2017).

6.1 Fusion Levels

Multimodal HAR systems can perform fusion at different stages of the recognition pipeline. Data-level fusion, often referred to as early fusion, combines raw or minimally processed signals from multiple sensors before feature extraction or model learning. This approach preserves substantial information from the original signals but requires appropriate synchronization, alignment, and compatibility between sensors with potentially different sampling rates, units, and signal characteristics. Feature-level fusion independently extracts features or learned representations from each modality and then combines the resulting representations before classification. This approach allows modality-specific characteristics to be modeled while potentially reducing the dimensionality and heterogeneity of the raw sensor data. Decision-level fusion, also known as late fusion, processes each modality independently and combines the outputs of individual classifiers or decision models to obtain a final prediction (Atrey et al., 2010; Ramachandram & Taylor, 2017). The appropriate fusion level depends on the characteristics and quality of the sensing modalities, computational and communication requirements, and the desired balance between information preservation, robustness, and system complexity..

6.2 Feature-Level Fusion

Feature-level fusion combines representations extracted from different sensor modalities into a joint representation that is subsequently provided to a classifier or downstream recognition model. For example, accelerometer and gyroscope features can be combined to represent complementary aspects of body motion, while inertial features can be integrated with physiological or ambient representations to incorporate additional information about physiological state or environmental context. By integrating complementary information, feature-level fusion may provide greater discriminative capability than relying on individual modalities alone (Atrey et al., 2010; Bian et al., 2022).

A major advantage of feature-level fusion is its ability to integrate modality-specific representations before the final decision stage. However, the resulting representation can become high-dimensional when multiple sensors or large learned embeddings are combined. Differences in feature scales, temporal resolution, sampling characteristics, and statistical distributions can also complicate multimodal integration. Consequently, normalization, dimensionality reduction, feature selection, and temporal alignment may be employed to construct an effective multimodal representation, depending on the characteristics of the sensing modalities and fusion architecture (Ramachandram & Taylor, 2017).

6.3 Decision-Level Fusion

Decision-level fusion, also referred to as late fusion, combines the individual predictions generated by separate modality-specific classifiers or recognition models. Each sensor modality is processed independently, allowing the corresponding recognition model to be optimized according to the characteristics of that modality. The resulting decisions or class-probability estimates can then be combined using methods such as majority voting, weighted voting, probability averaging, Bayesian approaches, or learned decision-fusion mechanisms (Atrey et al., 2010).

An important advantage of decision-level fusion is its modularity and flexibility. Individual sensing systems can be developed and optimized independently, and modalities can potentially be added, removed, or replaced with less disruption to the modality-specific processing pipelines, although the fusion mechanism may require modification or retraining. This is particularly useful when sensor modalities have substantially different data structures or require different processing techniques. However, because decision-level fusion operates after modality-specific analysis, fine-grained information contained in the original signals or intermediate feature representations may no longer be directly available to

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

the fusion stage. Its effectiveness consequently depends on the quality, reliability, calibration, and complementarity of the individual classifiers, as well as the strategy used to combine their outputs (Atrey et al., 2010; Ramachandram & Taylor, 2017).

6.4 Deep Learning and Attention-Based Fusion

Recent HAR research increasingly employs deep-learning architectures to learn multimodal representations automatically, reducing reliance on manually engineered features while still allowing modality-specific preprocessing where necessary. Modality-specific CNNs, recurrent networks, and Transformer-based encoders can extract representations from individual sensor streams, after which these representations can be integrated through concatenation, learned projection, cross-modal interaction, attention mechanisms, or dedicated fusion networks. Such approaches are particularly useful for heterogeneous sensor streams because the model can learn relationships between modalities during training rather than relying entirely on manually defined fusion rules (Ramachandram & Taylor, 2017, Sun et al., 2023).

Attention mechanisms and Transformer architectures are increasingly attractive for multimodal HAR because they can learn to assign different importance to sensor observations or feature representations according to their relevance to the recognition task (Ramachandram & Taylor, 2017, Sun et al., 2023).

For example, within a multimodal HAR system, inertial measurements may provide detailed information about body motion, while physiological or ambient sensors can provide complementary information about physiological state or environmental context. Cross-modal attention can therefore enable models to learn interactions between different sensing streams rather than simply concatenating their features. This is particularly relevant when sensor modalities contain complementary temporal and contextual information.
However, attention-based and Transformer-based fusion can introduce additional computational, memory, and training-data requirements, depending on model size and architecture. These requirements are particularly important for wearable and edge-based HAR systems, where processing capability, energy consumption, memory, and inference latency may be constrained. Consequently, multimodal deep-learning architectures should balance the potential recognition benefits of richer cross-modal representations against the computational and deployment costs associated with additional sensing and model complexity.

6.5 Challenges of Multimodal Fusion

Despite its potential benefits, multimodal sensor fusion introduces several challenges. Temporal synchronization is important because different sensors may operate at different sampling frequencies and experience different acquisition delays. For example, inertial sensors commonly operate at tens or hundreds of hertz, whereas ambient and physiological sensors may operate at substantially different rates depending on the sensing modality and application. Missing, corrupted, or intermittently unavailable sensor data can also affect multimodal models, particularly when the architecture assumes that all modalities are continuously available. Other challenges include differences in signal representation, sensor placement, calibration requirements, communication overhead, computational complexity, energy consumption, and privacy considerations (Bian et al., 2022; Ramachandram & Taylor, 2017).

Another important challenge is generalization across users and environments. Sensor placement, body morphology, movement style, environmental layout, and device characteristics can introduce substantial variability between subjects and deployment settings. Adding more sensors does not necessarily guarantee improved HAR performance; an additional modality should provide sufficiently complementary information to justify its sensing, processing, communication, and energy costs. Robust multimodal HAR systems should consequently consider sensor reliability, modality redundancy, temporal synchronization, computational efficiency, and the possibility of missing or degraded modalities during both model development and deployment.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

VII. DATA ACQUISITION PROTOCOLS AND BENCHMARK DATASETS

7.1 Data Acquisition Considerations

7.1.1 Controlled vs. Real-World Collection

HAR datasets span a continuum from highly controlled laboratory protocols, in which participants perform predefined activities under supervised conditions, to free-living or in-the-wild recordings in which participants engage in their normal daily routines. Controlled protocols provide reproducible experimental conditions and generally facilitate precise activity labeling, but may reduce ecological validity because participants perform activities according to predefined instructions. Naturalistic collection can better capture the variability, duration, context, and transitions encountered during everyday behavior, but can introduce greater challenges in continuous annotation, label uncertainty, class imbalance, and behavioral variability.

Many established HAR benchmarks, including UCI HAR and PAMAP2, rely substantially on predefined activity protocols. UCI HAR was collected from 30 participants who performed six predefined activities while carrying a waist-mounted smartphone equipped with accelerometer and gyroscope sensors at 50 Hz; the recordings were video-recorded to support manual labeling (Anguita et al., 2013). PAMAP2 similarly used a structured protocol comprising multiple physical activities recorded with wearable sensors (Reiss & Stricker, 2012). The OPPORTUNITY dataset occupies an intermediate or mixed position, combining predefined drill sessions with more naturalistic Activities of Daily Living (ADL) recordings, thereby providing both controlled and less-scripted activity sequences (Roggen et al., 2010). More recent datasets such as HARTH and ExtraSensory emphasize free-living or in-the-wild behavior. HARTH recorded 22 participants for approximately 90–120 minutes during their regular working hours and used a chest-mounted camera for frame-by-frame expert annotation (Logacjov et al., 2021). ExtraSensory collected multimodal smartphone and smartwatch data from 60 users during their regular natural behavior and incorporated user-provided context labels, providing an example of in-the-wild multimodal sensing (Vaizman et al., 2017).

7.1.2 Annotation Strategies

Accurate activity annotation is fundamental to the development and evaluation of supervised HAR systems. Annotation strategies differ in temporal precision, scalability, participant burden, and susceptibility to labeling errors. Manual event logging through dedicated buttons or mobile applications can provide direct activity labels but may introduce timing errors, missed events, and additional participant burden. Video-assisted annotation provides a means of reconstructing activity boundaries retrospectively and can achieve high temporal precision when sensor and video streams are synchronized, although it requires substantial annotation effort and may introduce privacy considerations (Logacjov et al., 2021). The HARTH dataset, for example, used a chest-mounted camera together with expert frame-by-frame annotation to support detailed labeling of activities (Logacjov et al., 2021). Semi-automatic approaches can additionally use computational models to generate provisional labels for subsequent human verification, potentially reducing the manual effort required for large-scale annotation.

7.1.3 Windowing and Segmentation

Continuous sensor streams are commonly segmented into finite temporal windows before conventional feature extraction or window-based deep-learning classification. The sliding-window approach is one of the most widely adopted segmentation strategies in sensor-based HAR, with window length and overlap selected according to activity characteristics, sampling frequency, computational requirements, and the target application. Banos et al. (2014) systematically evaluated window sizes ranging from 0.25 to 7 s and demonstrated that window length can substantially affect recognition performance, with the most suitable duration varying according to the activity and recognition task. Subsequent studies have commonly employed windows on the order of approximately 1–2 s, particularly where a balance between temporal context and recognition latency is required, although longer windows may provide additional temporal context for activities whose discriminative patterns extend over longer durations.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722556

Window overlap provides another important design parameter. A 50% overlap is commonly adopted because it increases the temporal density of training and prediction windows while generally requiring less additional computation than very high-overlap configurations. However, overlap is not universally required, and non-overlapping and higher-overlap configurations have also been reported. Window configuration should therefore be treated as a task-dependent design decision rather than a universal preprocessing standard.

VIII. CURRENT CHALLENGES IN HAR SENSOR MODALITIES AND DATA ACQUISITION

Despite substantial progress in sensor technologies and deep-learning-based HAR, practical deployment remains constrained by challenges associated with sensing variability, data quality, annotation, computational resources, and generalization. These challenges are particularly evident when systems move from controlled benchmarks to heterogeneous and free-living environments.

8.1 Sensor Heterogeneity, Placement, and Orientation

HAR systems increasingly combine accelerometers, gyroscopes, magnetometers, physiological sensors, cameras, microphones, and ambient sensors. These modalities differ in sampling frequency, signal characteristics, noise characteristics, and spatial and temporal resolution, making multimodal synchronization and fusion challenging. In wearable and smartphone-based HAR, sensor position and orientation are particularly influential because the same activity can generate substantially different signals at different body locations or device orientations (Barua et al., 2024). Consequently, models developed under fixed-placement conditions may not generalize reliably to arbitrary device locations or orientations encountered during everyday use (Barua et al., 2024).

8.2 Sampling Rate, Noise, and Data Quality

Sampling frequency involves a trade-off between temporal information, measurement accuracy, computational requirements, data volume, and battery consumption. Reviews of smartphone-based HAR report substantial variation in sampling rates according to the sensing modality and application requirements, with higher sampling rates generally increasing data-acquisition and energy demands (Straczkiewicz et al., 2021). Sensor noise, bias, motion artifacts, poor sensor contact, and environmental interference can further degrade signal quality and consequently affect recognition performance. Robust preprocessing, sensor fusion, and signal-quality assessment are therefore important considerations for reliable HAR.

8.3 Temporal Segmentation and Activity Transitions

Segmentation remains an important design decision because window length can directly affect recognition accuracy, detection latency, and computational requirements. Banos et al. (2014) demonstrated that there is no universally optimal window size and that recognition performance varies according to the activity and recognition task. A systematic review of smartphone-based HAR found that window lengths commonly ranged from approximately 1–5 s, with shorter windows of 1–2 s generally sufficient for posture and mobility recognition, whereas somewhat longer windows of 4–5 s provided better classification performance for some activities (Straczkiewicz et al., 2021). This issue becomes particularly challenging at activity boundaries, where a fixed window may contain portions of two consecutive activities, potentially producing ambiguous or mixed activity labels.

8.4 Annotation, Label Noise, and Class Imbalance

Reliable annotation is a major challenge, particularly for continuous free-living recordings. Manual and self-reported labels may contain temporal inaccuracies, missing information, or reporting errors, whereas video-assisted annotation can provide more precise temporal labels but requires considerable human effort. Naturalistic datasets can additionally exhibit substantial class imbalance because common activities occur much more frequently than rare activities. The ExtraSensory dataset, for example, was collected from users during their normal daily activities and contains multimodal smartphone and smartwatch data together with rich contextual labels, illustrating both the value and complexity of in-the-wild HAR data (Vaizman et al., 2017). Consequently, accuracy alone may be insufficient for evaluating imbalanced HAR datasets, making class-wise precision, recall, F1-score, and balanced accuracy useful complementary metrics (Sokolova & Lapalme, 2009; Chicco & Jurman, 2020).

8.5 Inter-Subject Variability and Cross-Domain Generalization

Human activities vary across individuals and even across repeated performances by the same individual. Differences in body characteristics, movement patterns, sensor placement, device configuration, and environmental conditions can introduce substantial distribution shifts between users and datasets (Straczkiewicz et al., 2021; Barua et al., 2024). Consequently, high performance obtained from a random train–test split does not necessarily demonstrate robustness to unseen users, sensor configurations, or deployment environments. Subject-independent validation and cross-dataset evaluation are therefore important for assessing generalization, while transfer learning, domain adaptation, and domain-generalization methods provide potential approaches for addressing distribution shifts between training and deployment conditions.

8.6 Missing Data and Sensor Failure

Real-world wearable systems are susceptible to missing observations caused by communication interruptions, temporary sensor detachment, device removal, battery limitations, or individual sensor failure. These problems become particularly important in multimodal systems because some fusion architectures may assume that all modalities are continuously available and synchronized. Missing or corrupted sensor streams can consequently degrade recognition performance and system reliability. Developing models capable of operating under missing or degraded sensor conditions is therefore an important requirement for robust long-term HAR (Atrey et al., 2010; Ramachandram & Taylor, 2017)

8.7 Energy and Computational Constraints

Continuous sensing and inference can impose considerable energy and computational demands on smartphones, smartwatches, and other wearable devices. High-frequency sensing, wireless communication, and computationally intensive deep-learning models can increase energy consumption and consequently limit continuous operation. Energy efficiency is therefore an important consideration in the deployment of wearable HAR systems (Contoli et al., 2024). Lightweight architectures, adaptive sampling, model compression, quantization, pruning, and efficient on-device inference provide potential approaches for balancing recognition performance with computational and energy constraints.

8.8 Privacy, Reproducibility, and Standardization

The increasing use of cameras, microphones, physiological sensors, and continuous wearable monitoring raises important privacy and ethical concerns, particularly in uncontrolled or everyday environments where sensing can capture sensitive behavioral, physiological, or contextual information. At the same time, heterogeneity in sensor hardware, placement, sampling rates, preprocessing, segmentation, annotation, and evaluation protocols can limit reproducibility and complicate quantitative comparisons across HAR studies. Recent reviews therefore emphasize the need for more standardized evaluation practices and datasets that better reflect real-world conditions (Arshad et al., 2022; Qureshi et al., 2025). Privacy-enhancing approaches such as on-device processing and federated learning can reduce the need to transfer or centralize raw sensor data, although they introduce additional challenges related to computational and communication resources, data heterogeneity, personalization, and privacy (Aouedi et al., 2024; Grataloup & Kurpicz-Briki, 2024).

8.9 Summary of the Challenges

Overall, a principal challenge in HAR extends beyond achieving high classification accuracy on benchmark datasets. The central challenge is developing recognition systems that remain accurate, efficient, and reliable when sensing conditions are heterogeneous, users are unseen, activities are continuous, labels are imperfect, and devices operate under real-world resource constraints. These limitations motivate greater emphasis on naturalistic data collection, robust multimodal fusion, transition-aware temporal modeling, missing-sensor tolerance, energy-efficient architectures, and cross-domain evaluation.

IX. EMERGING DIRECTIONS AND FUTURE RESEARCH

Future HAR research is moving toward adaptive, multimodal, efficient, and generalizable systems capable of operating under real-world conditions.

9.1 Multimodal and Adaptive Sensor Fusion

Future systems will increasingly integrate inertial, physiological, ambient, audio, and vision sensors to exploit complementary information. Attention- and Transformer-based fusion can learn cross-modal relationships, while adaptive sensor selection may reduce unnecessary sensing and energy consumption.

9.2 Self-Supervised and Foundation Models

Self-supervised learning can exploit large quantities of unlabeled sensor data to reduce dependence on manual annotation. Foundation and pretrained models also offer potential for cross-dataset and cross-device transfer, although their generalization to heterogeneous sensor-based HAR remains an open challenge

9.3 Efficient Transformer-Based HAR

Future research should develop lightweight and efficient Transformers capable of modeling long-term and transitional activities while reducing memory, computational cost, and inference latency for wearable and edge devices.

9.4 Continuous, Personalized, and Context-Aware HAR

HAR should increasingly address continuous activity and transition recognition rather than isolated activities. Adaptive segmentation, personalization, and contextual information may improve recognition of variable-duration and ambiguous activities in free-living environments.

9.5 Robustness, Privacy, and Edge Intelligence

Practical HAR systems must tolerate sensor displacement, missing modalities, noise, unseen users, and domain shifts. Domain generalization, missing-sensor learning, privacy-enhancing methods, model compression, quantization, adaptive sampling, and efficient edge inference are therefore important directions

9.6 Naturalistic Datasets and Standardized Evaluation

Future datasets should better represent free-living behavior, diverse users, natural transitions, sensor displacement, and missing data. Evaluation should extend beyond random train–test splits to subject-independent, cross-device, cross-dataset, and real-time testing, while reporting computational cost, latency, memory, and energy consumption alongside accuracy and F1-score.

REFERENCES

[1] Aggarwal, J.K., & Ryoo, M.S. (2011). Human activity analysis. ACM Computing Surveys (CSUR), 43, 1 - 43.

[2] Aggarwal, J. K., & Xia, L. (2014). Human activity recognition from 3D data: A review. Pattern Recognition Letters, 48, 70–80. https://doi.org/10.1016/j.patrec.2014.04.011

[3] Aguileta, A. A., Brena, R. F., Mayora, O., Molino-Minero-Re, E., & Trejo, L. A. (2019). Multi-sensor fusion for activity recognition—A survey. Sensors, 19(17), 3808. https://doi.org/10.3390/s19173808

[4] Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). A public domain dataset for human activity recognition using smartphones. In Proceedings of the 21st International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013), 437–442.

[5] Aouedi, O., Sacco, A., Khan, L. U., Nguyen, D. C., & Guizani, M. (2024). Federated learning for human activity recognition: Overview, advances, and challenges. IEEE Open Journal of the Communications Society, 5, 7341–7367. https://doi.org/10.1109/OJCOMS.2024.3484228

[6] Arshad, M. H., Bilal, M., & Gani, A. (2022). Human Activity Recognition: Review, Taxonomy and Open Challenges. Sensors

(Basel, Switzerland), 22(17), 6463. https://doi.org/10.3390/s22176463

[7] Atallah, L., Lo, B., King, R., & Yang, G.-Z. (2011). Sensor positioning for activity recognition using wearable accelerometers. IEEE Transactions on Biomedical Circuits and Systems, 5(4), 320–329. https://doi.org/10.1109/TBCAS.2011.2160540

[8] Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Systems, 16(6), 345–379.

[9] Banos, O., Galvez, J.-M., Damas, M., Pomares, H., & Rojas, I. (2014). Window Size Impact in Human Activity Recognition. Sensors, 14(4), 6474-6499. https://doi.org/10.3390/s140406474

[10] Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In International Conference on Pervasive Computing (pp. 1–17). Springer.

[11] Barua, A., Jiang, X., & Fuller, D. (2024). The effectiveness of simple heuristic features in sensor orientation and placement problems in human activity recognition using a single smartphone accelerometer. BioMedical Engineering OnLine, 23, 21. https://doi.org/10.1186/s12938-024-01213-3.

[12] Beddiar, D. R., Nini, B., Sabokrou, M., & Hadid, A. (2020). Vision-based human activity recognition: A survey. Multimedia Tools and Applications, 79, 30509–30555. https://doi.org/10.1007/s11042-020-09004-3

[13] Bian, S., Liu, M., Zhou, B., & Lukowicz, P. (2022). The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey. Sensors (Basel, Switzerland), 22(12), 4596. https://doi.org/10.3390/s22124596

[14] Bibbò, L., & Vellasco, M. M. B. R. (2023). Human Activity Recognition (HAR) in Healthcare. Applied Sciences, 13(24), 13009. https://doi.org/10.3390/app132413009

[15] Bibbò, L., & Serrano, A. (2025). Human Activity Recognition (HAR) in Healthcare, 2nd Edition. Applied Sciences, 15(10), 5762. https://doi.org/10.3390/app15105762

[16] Bouchabou, D., Nguyen, S. M., Lohr, C., Leduc, B., & Kanellos, I. (2021). A survey of human activity recognition in smart homes based on IoT sensors algorithms: Taxonomies, challenges, and opportunities with deep learning. Sensors, 21(18), 6037. https://doi.org/10.3390/s21186037

[17] Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. ACM Computing Surveys, 46(3), Article 33. https://doi.org/10.1145/2499621

[18] Charlton, P. H., Kotzen, K., Mejía-Mejía, E., Aston, P. J., Budidha, K., Mant, J., Pettit, C., Behar, J. A., & Kyriacou, P. A. (2022). Detecting beats in the photoplethysmogram: benchmarking open-source algorithms. Physiological measurement, 43(8), 085007. https://doi.org/10.1088/1361-6579/ac826d

[19] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21, 6. https://doi.org/10.1186/s12864-019-6413-7

[20] Chung, S., Lim, J., Noh, K. J., Kim, G., & Jeong, H. (2019). Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. Sensors, 19(7), 1716. https://doi.org/10.3390/s19071716

[21] Contoli, C., Freschi, V., & Lattanzi, E. (2024). Energy-aware human activity recognition for wearable devices: A comprehensive review. Pervasive and Mobile Computing, 104, 101976. https://doi.org/10.1016/j.pmcj.2024.101976

[22] Dang, L. M., Min, K., Wang, H., Piran, M. J., Lee, C. H., & Moon, H. (2020). Sensor-based and vision-based human activity recognition: A comprehensive survey. Pattern Recognition, 108, Article 107561. https://doi.org/10.1016/j.patcog.2020.107561.

[23] Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive

survey. IEEE Access, 8, 210816–210836. https://doi.org/10.1109/ACCESS.2020.3037715

[24] Du, J., Gerdtman, C., & Lindén, M. (2018). Signal quality improvement algorithms for MEMS gyroscope-based human motion analysis systems: A systematic review. Sensors, 18(4), 1123. https://doi.org/10.3390/s18041123

[25] Fernández-Cuevas, I., Bouzas Marins, J. C., Arnáiz Lastras, J., Gómez Carmona, P. M., Piñonosa Cano, S., García-Concepción, M. Á., & Sillero-Quintana, M. (2015). Classification of factors influencing the use of infrared thermography in humans: A review. Infrared Physics & Technology, 71, 28–55. https://doi.org/10.1016/j.infrared.2015.02.007

[26] Ghapanchizadeh, H., Ahmad, S. A., Ishak, A. J., & Al-Quraishi, M. S. (2017). Review of surface electrode placement for recording electromyography signals. Biomedical Research, 2017(Special Issue: ComplexWorldofNeuroscience), S1–S7.

[27] Gomaa, W., & Khamis, M. A. (2023). A perspective on human activity recognition from inertial motion data. Neural Computing and Applications, 35, 20463–20568. https://doi.org/10.1007/s00521-023-08863-9

[28] Gorce, P., & Jacquier-Bret, J. (2025). Fall Detection in Elderly People: A Systematic Review of Ambient Assisted Living and Smart Home-Related Technology Performance. Sensors (Basel, Switzerland), 25(21), 6540. https://doi.org/10.3390/s25216540

[29] Gosavi, D., Chavan, D., Kaware, A., Bhaemal, S., & [Remaining Author Name(s)]. (2026). Smart elderly activity recognition and fall detection using machine learning. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 14(5), 1–5.//doi: 10.22214/ijraset.2026.82943

[30] Grataloup A and Kurpicz-Briki M (2024) A systematic survey on the application of federated learning in mental state detection and human activity recognition. Front. Digit. Health 6:1495999. doi: 10.3389/fdgth.2024.1495999

[31] Guo, L., Sandsjö, L., Ortiz-Catalan, M., & Skrifvars, M. (2020). Systematic review of textile-based electrodes for long-term and continuous surface electromyography recording. Textile Research Journal, 90(2), 227–244. https://doi.org/10.1177/0040517519858768

[32] Han, J., Shao, L., Xu, D., & Shotton, J. (2013). Enhanced computer vision with Microsoft Kinect sensor: A review. IEEE Transactions on Cybernetics, 43(5), 1314–1334. https://doi.org/10.1109/TCYB.2013.2265378

[33] Hiremath, S. K., & Plötz, T. (2023). The lifespan of human activity recognition systems for smart homes. Sensors, 23(18), 7729. https://doi.org/10.3390/s23187729

[34] Janidarmian, M., Fekr, A. R., Radecka, K., & Zilic, Z. (2017). A comprehensive analysis on wearable acceleration sensors in human activity recognition. Sensors, 17(3), 529. https://doi.org/10.3390/s17030529

[35] Karayaneva, Y., Sharifzadeh, S., Jing, Y., & Tan, B. (2023). Human activity recognition for AI-enabled healthcare using low-resolution infrared sensor data. Sensors, 23(1), 478. https://doi.org/10.3390/s23010478

[36] Lara, Ó. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. IEEE Communications Surveys and Tutorials, 15(3), 1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192

[37] Letts, E., Jakubowski, J. S., King-Dowling, S., Clevenger, K. A., Kobsar, D., & Obeid, J. (2024). Accelerometer techniques for capturing human movement validated against direct observation: A scoping review. Physiological Measurement, 45(7). https://doi.org/10.1088/1361-6579/ad45aa

[38] Li, X., He, Y., & Jing, X. (2019). A survey of deep learning-based human activity recognition in radar. Remote Sensing, 11(9), 1068. https://doi.org/10.3390/rs11091068

[39] Logacjov, A., Bach, K., Kongsvold, A., Bårdstu, H. B., & Mork, P. J. (2021). HARTH: A human activity recognition dataset for machine learning. Sensors, 21(23), 7853. https://doi.org/10.3390/s21237853

[40] Ma, Y., Zhou, G., & Wang, S. (2019). WiFi sensing with channel state information: A survey. ACM Computing Surveys (CSUR). https://doi.org/10.1145/3310194

[41] Manivannan, A., Chin, W. C. B., Barrat, A., & Bouffanais, R. (2020). On the challenges and potential of using barometric sensors to track human activity. Sensors, 20(23), 6786. https://doi.org/10.3390/s20236786

[42] Mukhopadhyay, S. C. (2015). Wearable sensors for human activity monitoring: A review. IEEE Sensors Journal, 15(3), 1321-1330. Article 6974987. https://doi.org/10.1109/JSEN.2014.2370945

[43] Ntalampiras, S., Potamitis, I., & Fakotakis, N. (2011). Probabilistic novelty detection for acoustic surveillance under real-world conditions. IEEE Transactions on Multimedia, 13(4), 713–719. https://doi.org/10.1109/TMM.2011.2122247

[44] Newaz, N. T., & Hanada, E. (2024). A low-resolution infrared array for unobtrusive human activity recognition that preserves privacy. Sensors, 24(3), 926. https://doi.org/10.3390/s24030926.

[45] Niu, X., Li, Y., Zhang, H., Wang, Q., & Ban, Y. (2013). Fast thermal calibration of low-grade inertial sensors and inertial measurement units. Sensors, 13(9), 12192–12217. https://doi.org/10.3390/s130912192

[46] Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. Expert Systems with Applications, 39(8), 7420–7431. https://doi.org/10.1016/j.eswa.2012.01.102

[47] Qureshi, T. S., Shahid, M. H., Farhan, A. A., & Alamri, S. (2025). A systematic literature review on human activity recognition using smart devices: Advances, challenges, and future directions. Artificial Intelligence Review, 58, 276. https://doi.org/10.1007/s10462-025-11275-x

[48] Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. IEEE Signal Processing Magazine, 34(6), 96–108. https://doi.org/10.1109/MSP.2017.2738401

[49] Ramamurthy, S. R., & Roy, N. (2018). Recent trends in machine learning based human activity recognition—A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), Article e1254. doi.org

[50] Reiss, A., & Stricker, D. (2012). Introducing a new benchmarked dataset for activity monitoring. Proceedings of the 2012 16th Annual International Symposium on Wearable Computers (ISWC), 108–109. IEEE. https://doi.org/10.1109/ISWC.2012.13

[51] Roggen, D., Calatroni, A., Nguyen-Dinh, L., Chavarriaga, R., & Sagha, H. (2010). OPPORTUNITY Activity Recognition [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5M027

[52] Saha, U., Saha, S., Kabir, M. T., Fattah, S. A., & Saquib, M. (2024). Decoding human activities: Analyzing wearable accelerometer and gyroscope data for activity recognition. IEEE Sensors Letters, 8(8), 1–4. https://doi.org/10.1109/LSENS.2024.3423340

[53] Sanchez-Comas, A., Synnes, K., & Hallberg, J. (2020). Hardware for recognition of human activities: A review of smart home and AAL related technologies. Sensors, 20(14), 4035. https://doi.org/10.3390/s20144035

[54] Savvidou, F., Tegos, S. A., Diamantoulakis, P. D., & Karagiannidis, G. K. (2024). Passive radar sensing for human activity recognition: A survey. IEEE Open Journal of Engineering in Medicine and Biology, 5, 700–706. https://doi.org/10.1109/OJEMB.2024.3420747

[55] Sazonov, E. S., Fulk, G. D., Hill, J., Schutz, Y., & Browning, R. (2011). Monitoring of posture allocations and activities by a shoe-based wearable sensor. IEEE Transactions on

Biomedical Engineering, 58(4), 983–990. https://doi.org/10.1109/TBME.2010.2046738

[56] Shin, J., Hassan, N., Miah, A. S. M., & Nishimura, S. (2025). A Comprehensive Methodological Survey of Human Activity Recognition Across Diverse Data Modalities. Sensors, 25(13), 4028. https://doi.org/10.3390/s25134028

[57] Soken, H. E. (2017). A survey of calibration algorithms for small satellite magnetometers. In 2017 IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace) (pp. 62–67). IEEE. https://doi.org/10.1109/MetroAeroSpace.2017.7999

[58] Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

[59] Straczkiewicz, M., James, P., & Onnela, J.-P. (2021). A systematic review of smartphone-based human activity recognition methods for health research. npj Digital Medicine, 4, 148. https://doi.org/10.1038/s41746-021-00514-4

[60] Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2023). Human action recognition from various data modalities: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(3), 3200–3225. https://doi.org/10.1109/TPAMI.2022.3183112

[61] Tamura, T., Maeda, Y., Sekine, M., & Yoshida, M. (2014). Wearable Photoplethysmographic Sensors—Past and Present. Electronics, 3(2), 282-302. https://doi.org/10.3390/electronics3020282

[62] Vaizman, Y., Ellis, K., & Lanckriet, G. R. (2017). Recognizing detailed human context in-the-wild from smartphones and smartwatches. IEEE Pervasive Computing, 16(4), 62–74. https://doi.org/10.1109/MPRV.2017.3971131

[63] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (pp. 5998–6008).

[64] Wang, Y., Cang, S., & Yu, H. (2019). A survey on wearable sensor modality centred human activity recognition in health care. Expert Systems with Applications, 137, 167–190. https://doi.org/10.1016/j.eswa.2019.04.057

[65] Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 119, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

[66] Webber, M., Rojas, R. F.(2024). Human Activity Recognition With Accelerometer and Gyroscope: A Data Fusion Approach. IEEE Sensors Journal. https://doi.org/10.1109/JSEN.2021.3079883

[67] Yadav, S. K., Tiwari, K., Pandey, H. M., & Akbar, S. A. (2021). A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. Knowledge-Based Systems, 223, Article 106970. https://doi.org/10.1016/j.knosys.2021.106970.

[68] Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. Sensors, 22(4), 1476. https://doi.org/10.3390/s22041476