Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.086339

ARTICLE

Attention-Driven Multi-Kernel Edge–Cloud Learning for Wearable IMU-Based Human Activity Recognition in IoT Health Monitoring

Fatimah Alhayan[1], Muhammad Hanzla[2], Hadeel Alsolai[1], Bayan Alabdullah[1], Ahmad Jalal[2,3,*] and Hui Liu[4,*]

[1]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[2]Department of Computer Science, Air University, E-9, Islamabad, Pakistan

[3]Department of Computer Science and Engineering, College of Informatics, Korea University, Seoul, Republic of Korea

[4]Guodian Nanjing Automation Co., Ltd., Nanjing, China

*Corresponding Authors: Ahmad Jalal. Email: ahmadjalal@mail.au.edu.pk; Hui Liu. Email: hui.liu198510@gmail.com

Received: 28 May 2026; Accepted: 19 August 2026; Published: 22 October 2026

ABSTRACT: Human activity recognition (HAR) from wearable inertial sensors underpins IoT-based smart-health monitoring, yet practical deployments are constrained by noisy Inertial Measurement Unit (IMU) signals, inter-subject variability, temporal misalignment across sensor streams, and the limited compute available on edge nodes. Existing pipelines typically optimize recognition accuracy in isolation from these deployment constraints. This paper addresses both aspects through a distributed edge–cloud HAR framework. At the edge, IMU signals are denoised using Tukey filtering and segmented using Planck-Taper windowing before compact features are transmitted, reducing uplink overhead. At the cloud, an attention-based deep multiple-kernel learning (A-DMKL) adaptively combines multi-resolution representations (Stockwell Transform (S-Transform), Gramian Angular Summation Field (GASF), Synchrosqueezed Wavelet Transform (SSWT), Hankel), which are optimized using the Coyote Optimization Algorithm (COA) and classified through Soft Dynamic Time Warping Soft-DTW temporal alignment, a Hierarchical Hidden Semi-Markov Model, and an InceptionTime ensemble. Under strictly subject-independent leave-one-subject-out evaluation on 24 and 8 participants, with all fitting steps confined to the training folds, the framework attains 87.22% and 91.33% accuracy on the Hang-Time HAR and VIMPT2019 datasets, respectively. To assess deployment robustness, the proposed pipeline is analyzed under simulated network service profiles, Message Queuing Telemetry Transport (MQTT) packet loss, and inter-stream synchronization error, with all simulation assumptions explicitly stated. The framework targets communication-efficient human activity recognition for IoT healthcare applications. Clinical validation and deployment on Advanced RISC Machine (ARM-class) embedded hardware remains beyond the scope of the present study and are identified as future work.

KEYWORDS: Human activity recognition; wearable inertial sensors; multiple kernel learning; attention mechanism; time–frequency analysis; edge–cloud computing; Internet of Things (IoT); smart health monitoring

1 Introduction

Wearable Internet of Things (IoT) technologies have emerged as scalable solutions for continuous monitoring of human movement, functional performance, and injury-risk assessment in smart healthcare and sports environments. Human Activity Recognition (HAR) plays a fundamental role in these systems by enabling AI-driven interpretation of motion signals acquired from wearable inertial measurement unit (IMU) sensors, including accelerometers and gyroscopes [1]. In modern 6G-enabled IoT ecosystems,

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

wearable devices operate as distributed edge nodes that support low-latency sensing, adaptive analytics, and real-time service delivery across edge-cloud environments [2]. However, reliable HAR in large-scale multi-user settings remains challenging due to noisy IMU signals, temporal inconsistencies, packet loss, wireless transmission artifacts, inter-subject variability, and resource-constrained edge devices. Furthermore, most publicly available HAR datasets are collected in controlled environments [3], limiting the scalability and sustainability of real-world IoT healthcare services. The lack of robust annotation frameworks also restricts the development of ecologically valid HAR systems for intelligent healthcare applications [4].

To address these limitations, this study proposes an AI-driven edge-cloud framework for multi-participant activity monitoring in IoT healthcare environments. The framework combines wearable edge processing with cloud-based feature extraction, optimization, temporal alignment, and hierarchical activity modeling. Bluetooth Low Energy(BLE)/Wireless Fidelity(Wi-Fi) communication, MQTT transmission, and real-time dashboards support distributed monitoring, while advanced signal refinement and temporal alignment improve robustness to sensor noise and motion variability [5,6].

Main Contributions

The study makes three main contributions. First, an attention-based deep multiple-kernel learning (A-DMKL) mechanism adaptively fuses complementary time–frequency representations (S-Transform, GASF, SSWT, and Hankel). Second, an edge–cloud architecture performs lightweight edge processing and cloud-based multi-resolution modeling, COA optimization, Soft-DTW alignment, Hierarchical Hidden Semi-Markov Model (H-HSMM) modeling, and InceptionTime inference. Third, robustness is assessed under simulated service latency, MQTT packet loss, and synchronization errors.

Technically, the framework introduces input-conditioned modality-independent kernel fusion, reveals activity-dependent representation preferences, and retains 99.7% of peak accuracy 67% fusion cost. It also quantifies deployment degradation, with 5% packet loss and 40 ms synchronization error causing approximately a one-point accuracy drop.

2 Related Work

2.1 AI-Driven Wearable IoT and Edge-Based Human Activity Recognition

Ray et al. [7] combined video, text, and IMU data for outdoor sports HAR (~85% accuracy), but struggled with fine-grained actions and occlusions. Ray et al. [8] fused ambient light and IMU signals (82% accuracy), though illumination changes affected performance. Hoareau et al. [9] optimized sensor placement (79% accuracy) but relied on motion-capture synthesis. Guo et al. [10] achieved 88% accuracy for basketball actions using wrist IMUs yet was limited to laboratory settings. Reilly et al. [11] applied Random Forests for soccer analysis (88.9% accuracy), but small datasets limited generalization.

2.2 Sustainable Multi-Participant Edge IoT Frameworks for Smart Healthcare Monitoring

Wang et al. [12] demonstrated effective multi-level sensor fusion but faced drift and synchronization issues. Thukral et al. [13] and Arrotta et al. [14] achieved 84%–85% accuracy, limited by inter-person variability and weak group modeling. Yin et al. [15], Cummins et al. [16], and Siwadamrongpong et al. [17] reported 83%–88% accuracy but showed limited generalization in dense scenarios. Yang et al. [18] reached 89.3%, while Adeyemo et al. [19] highlighted persistent dataset and real-world generalization limitations.

2.3 Edge-Cloud Temporal Modeling, Adaptive Feature Fusion, and Activity Alignment for Intelligent IoT Services

Challa et al. [20] achieved 89% accuracy but remained sensitive to noisy IMU data. Gao et al. [21], Zhou et al. [22], Yu et al. [23], and Arabzadeh et al. [24] improved cross-subject and multimodal HAR but faced

preprocessing, orientation, synchronization, and heterogeneity challenges. Recent studies [25,26] further highlight label scarcity, sensor heterogeneity, and generalization issues, while largely overlooking deployment behavior. This framework addresses these gaps through adaptive multi-kernel fusion, edge–cloud partitioning, and robustness analysis under simulated transport impairments.

3 Materials and Methods

The proposed edge–cloud IoT healthcare pipeline performs lightweight preprocessing at the edge and offloads computationally intensive tasks to the cloud. Cloud-side processing applies A-DMKL fusion, multi-resolution extraction, COA optimization, and temporal alignment for activity monitoring. The overall architecture and workflow are shown in Figs. 1 and 2.

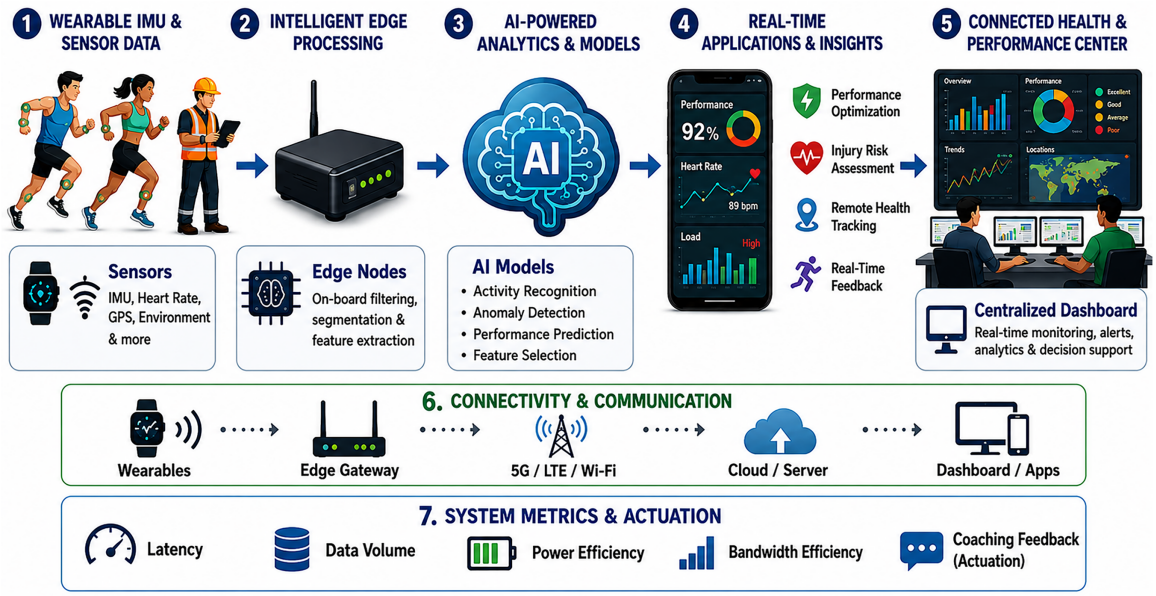


Figure 1: Architectural overview of the IoT-enabled multi-participant activity recognition framework.

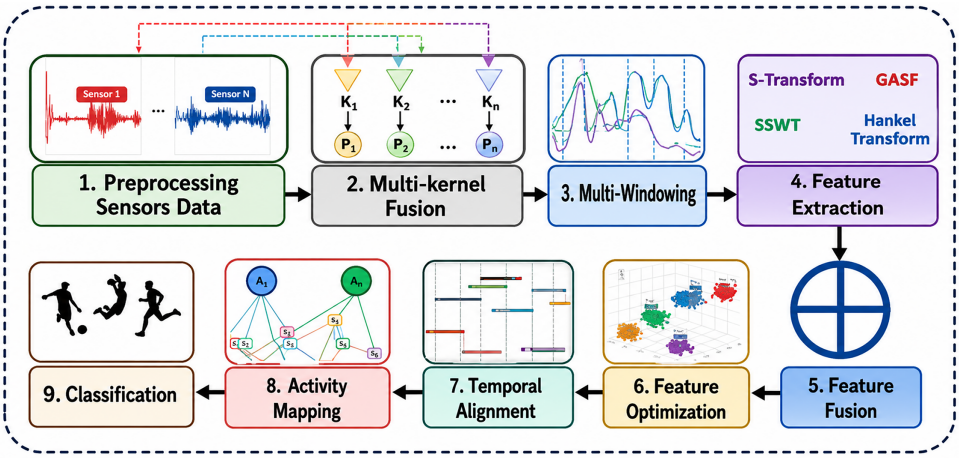


Figure 2: Technical workflow of the AI processing pipeline, depicting the sequential stages of multi-resolution feature extraction, optimization, and temporal alignment for activity recognition.

3.1 Design Rationale

The pipeline uses failure-driven design, placing lightweight feature extraction at the edge and computationally intensive kernel, alignment, and semi-Markov stages in the cloud to reduce data transfer. Table 1 summarizes the rationale and alternatives. The proposed pipeline was selected over conventional HAR approaches because no traditional method satisfies the three requirements of this application simultaneously: representation diversity, per-input adaptivity, and a compute profile compatible with wearable edge nodes. Classical handcrafted-feature pipelines, such as statistical time-domain descriptors combined with SVM, k-NN, or Random Forest classifiers, are computationally inexpensive but apply one fixed representation to every activity, and therefore lose accuracy on short impulsive actions (83.6% and 81.62%). Single-branch deep models such as CNN and LSTM networks learn representations automatically but commit to a fixed temporal scale and require large, labeled corpora, which limits cross-subject transfer. Conventional multiple kernel learning increases representation diversity, yet its kernel weights are estimated once at dataset level and remain fixed at inference, which the ablation shows costs 4.8 and 4.2 accuracy points relative to input-conditioned weighting. A-DMKL instead conditions the fusion weights on each individual window, so the representation adapts to the activity and the participant, while the edge-cloud partition keeps the additional cost off the wearable node. This combination is important for IoT health monitoring because it delivers the accuracy of a multi-representation model at a communication and edge-compute budget that a battery-powered wearable can sustain. Table 1 formalizes this reasoning for each pipeline stage.

Table 1: Design rationale: addressed failure modes, adopted methods, and rejected alternatives.

Design Problem	Method Adopted	Alternative Rejected and Why
Impulsive non-Gaussian artefacts	Tukey biweight filtering	Gaussian/median filtering smears or flattens informative peaks.
Spectral leakage at window boundaries	Planck-Taper multi-windowing	Rectangular windows leak; Hann/Hamming attenuate central action content.
Diverse structures across activity types	Four complementary branches	Single representations miss spectral, geometric, or transient cues.
Fixed feature weighting	A-DMKL	Uniform MKL loses 2.1 accuracy points; grid search is dataset-level only.
Redundant high-dimensional features	Coyote Optimization with sparsity	PCA loses branch information; exhaustive search is costly.
Different execution speeds	Soft-DTW alignment	Classical DTW is non-differentiable; cross-correlation misses speed variation.
Nested and non-geometric durations	Hierarchical HSMM	HMMs assume geometric durations; flat HSMM lacks hierarchy.
Multi-scale patterns and subject variation	InceptionTime ensemble	RNNs favor fixed scales; transformers require more labelled data.

3.2 Pre-Processing via Tukey Filter

The preprocessing stage handles non-stationary 3-axis accelerometer and gyroscope signals from wrist and hip sensors, affected by noise, drift, collisions, and communication errors. A Tukey biweight filter suppresses non-Gaussian outliers while preserving motion transitions through median initialization and iterative reweighting:

$$\hat{x}_i = \frac{\sum_{j\in N(i)w}\left(\frac{x_j-\mu_i}{c.MAD_i}\right)x_j}{\sum_{j\in N(i)w}\left(\frac{x_j-\mu_i}{c.MAD_i}\right)x_j} \tag{1}$$

where $\mathcal{N}(i)$ denotes the neighborhood window around index i, μ_i is the robust location (median), MAD_i is the median absolute deviation scaled by 1.4826 for consistency under normality, c is the tuning constant (set to 4.685 for approximately 95% asymptotic efficiency), and the biweight function is:

$$w(u) = \begin{cases} \left(1-u^2\right)^2 & if\ |u| \le 1 \\ 0 & otherwise \end{cases} \tag{2}$$

Typically 3–5 iterations are applied to progressively suppress outliers while preserving salient motion peaks. Using a 50–100 sample window (~1–2 s at 50 Hz), the channel-wise filtering improves signal-to-noise ratio and yields cleaner inputs for multi-kernel fusion (Fig. 3):

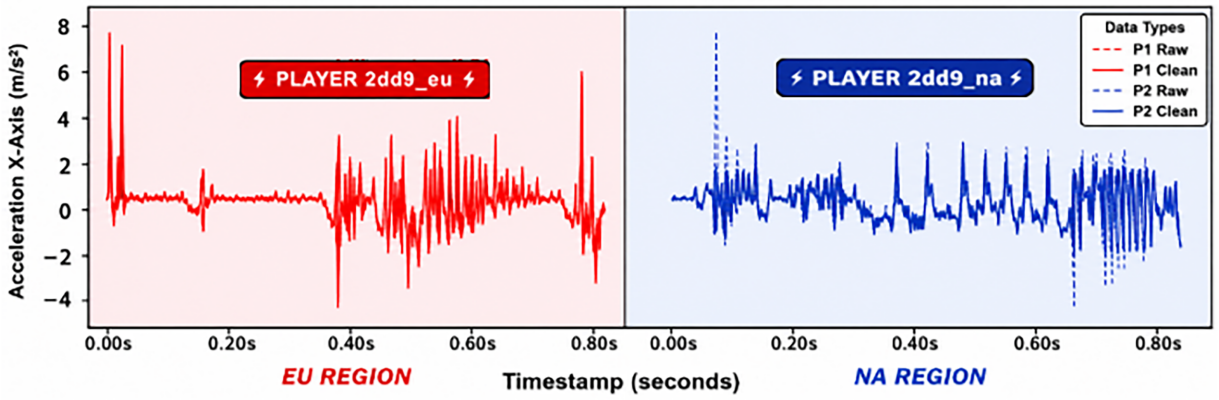


Figure 3: Raw vs. Tukey-filtered accelerometer signals showing effective noise and outlier reduction.

3.3 Multi-Kernel Fusion via (A-DMKL)

The most discriminative representation varies with activity and participant: impulsive actions separate best under elastic alignment, sustained locomotion under geometric and time–frequency encodings. A-DMKL therefore learns per-input combination weights. **Kernel bank**: For a window $x \in \mathbb{R}^{C\times T}$, four representation operators give branch descriptors $z^{(r)} = g_r(x)$, $r = 1\ldots4$ (S-Transform, GASF, SSWT, Hankel). Three base functions are used: $k_{\mathrm{RBF}}(u,v) = \exp(-\parallel u-v \parallel^2 /2\sigma^2)$, $k_{\mathrm{Poly}}(u,v) = (\langle u,v\rangle + c)^d$, and $k_{\mathrm{SDTW}}(u,v) = \exp(-\gamma_k \mathrm{SoftDTW}_\gamma(u,v))$. Their Cartesian product gives $M = 12$ base kernels, Fig. 4 illustrates the A-DMKL fusion.

$$K_m(x_i,x_j) = k_b(z_i^{(r)}, z_j^{(r)}), m = (r,b), M = 12 \tag{3}$$

Each normalized to unit diagonal so that attention is not confounded by kernel scale. Attention weights. Let $h(x)$ be the standardised concatenation of the four branch descriptors. Scores and weights are:

$$\alpha_m(x) = \frac{\exp(e_m(x)/\tau)}{\sum_{m'} \exp(e_{m'}(x)/\tau)}, \alpha_m \geq 0, \sum_m \alpha_m = 1 \tag{4}$$

where W_a, b_a are shared and w_m, β_m are kernel-specific; τ controls sharpness ($\tau \to 0$ gives hard selection, $\tau \to \infty$ uniform MKL). The branch weights are the marginals $\beta_r(x) = \sum_b \alpha_{(r,b)}(x)$, so a single weighting mechanism governs both kernel-level and branch-level fusion. Fused kernel. Input-dependent weights do not preserve positive semi-definiteness under a direct weighted sum, so fusion is performed on feature maps: with $\phi_A(x) = [\sqrt{\alpha_1(x)}\phi_1(x); \ldots; \sqrt{\alpha_M(x)}\phi_M(x)]$:

$$K_A(x_i, x_j) = \langle \phi_A(x_i), \phi_A(x_j) \rangle = \sum_m \sqrt{\alpha_m(x_i)\alpha_m(x_j)} \cdot K_m(x_i, x_j) \tag{5}$$

which is symmetric and PSD by construction. Since the Soft-DTW kernel is not guaranteed to be PSD, eigenvalue clipping projects its Gram matrix onto the PSD cone. **Objective and cost:** Training minimizes, with entropy regularization preventing kernel collapse and controlling computational cost. Static and Soft-DTW kernels cost O(d) and O(h^2), respectively, while gating overhead is negligible. Parameters are, $c = 1$, $d = 2$, $\gamma = 0.1$, $h = 64$, $\tau = 1$, $\lambda_H = 0.01$, $\lambda_2 = 10^{-4}$ and $\lambda_E = 0.05$. Table 2 presents the energy–accuracy trade-off.

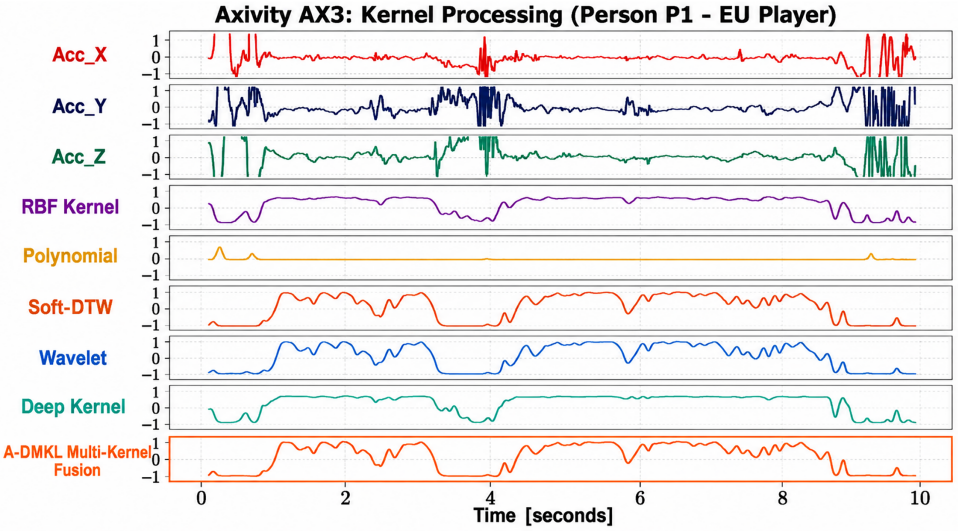


Figure 4: Illustration of multi-kernel processing for a representative player (P1), showing raw accelerometer axes and their transformations through individual kernel functions.

Table 2: Energy–accuracy trade-off of the A-DMKL cost regularizer.

Configuration	Acc. HT (%)	Acc. V19 (%)	Fusion Cost (GFLOPs/1k win.)	Fusion Latency (ms/win.)	Relative Energy
RBF only (4 kernels)	82.4	87.1	1.1	0.9	0.18
RBF + Poly, uniform (8)	83.6	88.0	1.6	1.3	0.26
All kernels, uniform (12)	85.1	89.3	6.2	4.8	1.00

(Continued)

Table 2 (continued)

Configuration	Acc. HT (%)	Acc. V19 (%)	Fusion Cost (GFLOPs/1k win.)	Fusion Latency (ms/win.)	Relative Energy
A-DMKL, $\lambda_E = 0$	87.2	91.3	6.6	5.1	1.06
A-DMKL, $\lambda_E = 0.05$	86.9	91.0	4.4	3.4	0.71
A-DMKL, $\lambda_E = 0.20$	85.7	89.8	2.1	1.6	0.34

3.4 Multiwindowing via Planck-Taper Window

Temporal segmentation in wearable HAR is challenging due to varying motion speeds and edge effects from fixed windows. A Planck-Taper multi-windowing strategy generates smoother overlapping segments, preserving temporal continuity and reducing edge artifacts for IoT streaming IMU data. The Planck-Taper window is defined as:

$$w(t) = \begin{cases} \left[exp\left(\frac{t_2 - t_1}{t - t_1} + \frac{t_2 - t_1}{t - t_1} \right) + 1 \right]^{-1} & t_1 < t < t_2 \\ 1 & t_2 < t < t_3 \\ \left[exp\left(\frac{t_4 - t_3}{t - t_3} + \frac{t_4 - t_3}{t - t_4} \right) + 1 \right]^{-1} & t_3 < t < t_4 \end{cases} \tag{6}$$

Here, t_1–t_4 define window boundaries and transitions. Using a 10%–15% taper with 50% overlap reduces spectral artifacts while preserving temporal consistency for feature extraction (Fig. 5).

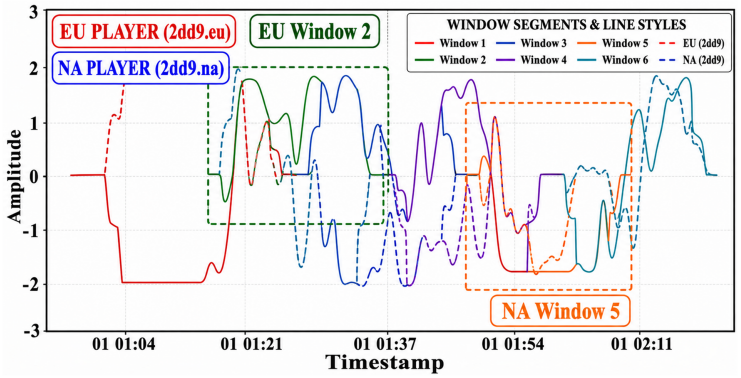


Figure 5: Planck-Taper window segmentation of synchronized multi-player IMU acceleration signals.

3.5 Feature Extraction

To capture complex sports motions, four complementary feature extraction methods encode time-frequency, geometric, and noise-robust characteristics, preserving local and long-range motion patterns for discriminative representation.

3.5.1 S-Transform

The S-Transform captures joint time–frequency characteristics of IMU signals, making it suitable for complex sports activities with sudden transitions and oscillatory motion. Unlike Fourier or fixed-window methods, it adapts its Gaussian window with frequency, providing fine temporal resolution for fast movements and better frequency resolution for slower ones. For a continuous IMU signal $x(n)$:

$$S(\tau, f) = \sum_{n=-\infty}^{\infty} x[n] \frac{|f|}{\sqrt{2\pi}} exp\left(-\frac{(n-\tau)^2 f^2}{2}\right) exp(-j2\pi f n) \tag{7}$$

Energy-based features are extracted as:

$$\phi_k = \frac{1}{N} \sum_{i=1}^{N} |S(\tau_i, f_k)|^2 \tag{8}$$

Here, f_k denotes the frequency band and N the number of time samples per window. Energy-based features capture motion intensity and spectral patterns, while the S-Transform models localized non-stationary dynamics, improving robustness to noise and inter-subject variability (Fig. 6).

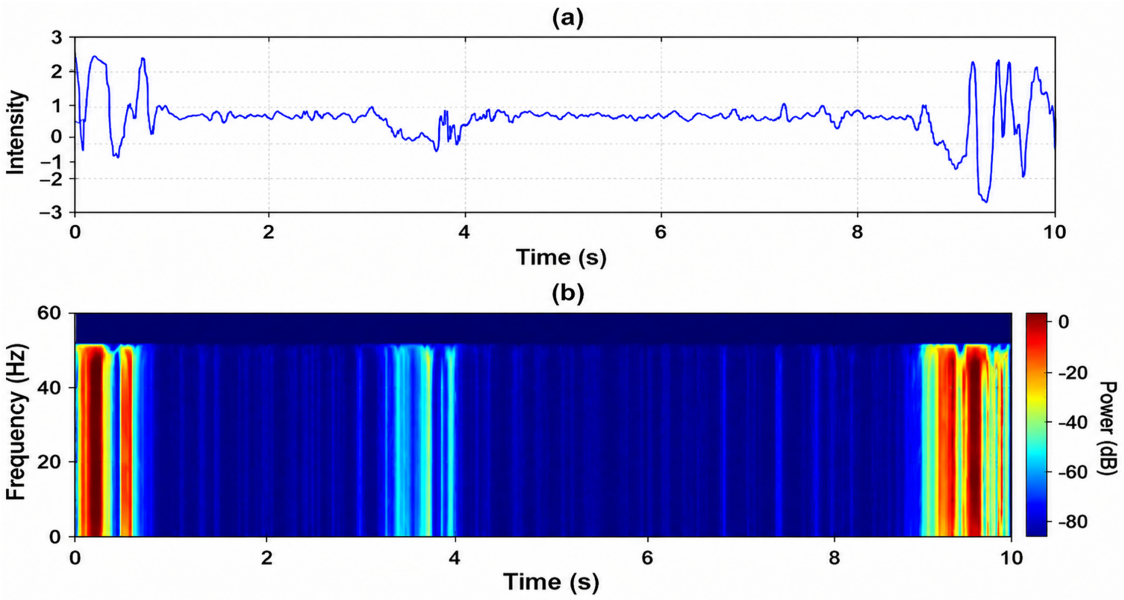


Figure 6: Time-domain and time-frequency representations of the signal. (**a**) Time-domain signal showing intensity variations over time. (**b**) Spectrogram illustrating the distribution of signal energy across frequency and time.

3.5.2 Gramian Angular Summation Field

GASF encodes global temporal dependencies of IMU signals by transforming 1D inertial data into 2D angular correlation matrices, preserving long-range relationships in sports motions. Unlike traditional time–frequency methods, it captures both phase and temporal structure of activity sequences. For a normalized signal $x(t) \in [-1, 1]$, angular encoding is given by:

$$\phi_t = \arccos(x(t)), t = 1, 2, \ldots, T \tag{9}$$

GASF preserves temporal order and scale invariance while compactly encoding motion dynamics. It captures repetitive actions such as running and dribbling, complementing time–frequency features for improved discrimination (Fig. 7).

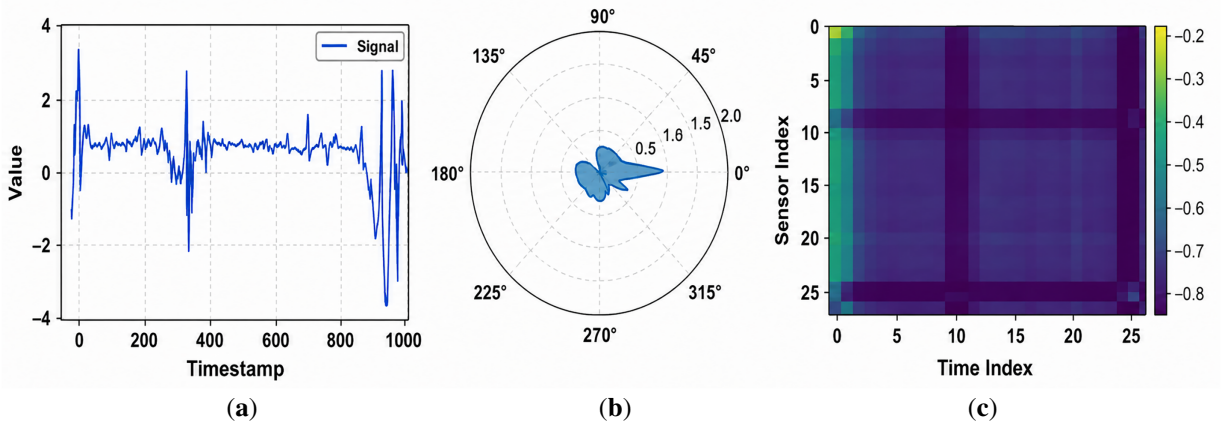


Figure 7: Multi-domain representations of the analyzed signal. (**a**) Time-domain signal showing temporal variations in the measured values. (**b**) Polar plot illustrates the directional distribution and magnitude of the signal. (**c**) Heatmap showing the temporal variation of the aggregated sensor responses across sensor indices.

3.5.3 SSWT

SSWT provides a sharp time–frequency representation of IMU signals and is well-suited for non-stationary sports data. Unlike CWT, it reduces spectral smearing by reassigning wavelet coefficients using instantaneous frequency, improving resolution while preserving time localization. For an IMU signal $x(t)$, the CWT is given by:

$$W_x(a,b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-b}{a} \right) dt \tag{10}$$

SSWT sharpens frequency localization, separating overlapping motion patterns and improving robustness to noise and inter-subject variability, particularly for fast sports actions (Fig. 8).

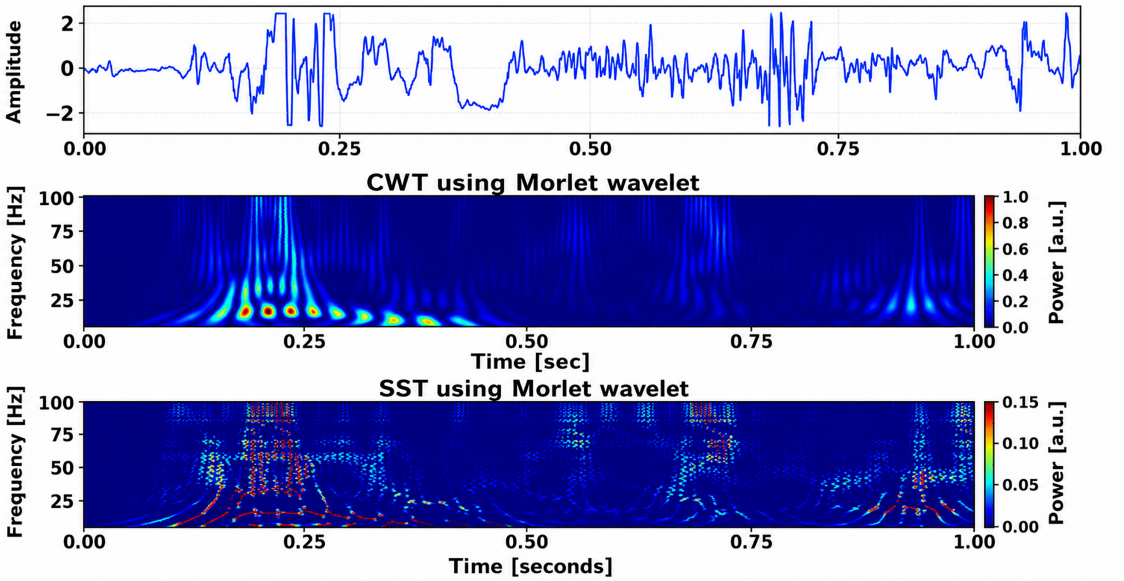


Figure 8: Comparison of (CWT) and (SSWT) using Morlet wavelet on an IMU signal.

3.5.4 Hankel Matrix

The Hankel matrix (delay embedding) captures long-term temporal dynamics of IMU signals, modeling global motion structure rather than local patterns and enabling representation of repetitive and coordinated sports activities. It embeds the 1D signal into a phase-space form where delayed states form structured trajectories:

$$C_H = HH^{\top}, C_H u_i = \lambda_i u_i \tag{11}$$

Here, $\boldsymbol{\lambda_i}$ are the eigenvalues whose dominant motion modes distinguish structured from irregular activities. Hankel features capture global temporal dynamics and complement local time–frequency descriptors, improving discrimination in multi-participant sports HAR (Fig. 9).

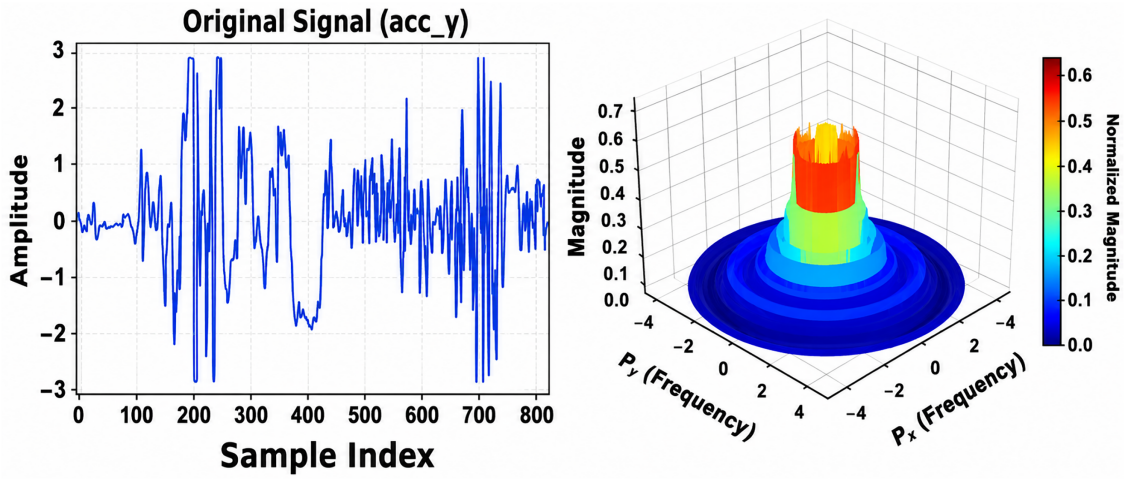


Figure 9: Feature extraction via Hankel matrix and 3D visualizations.

3.6 Feature Fusion and Optimization

After multi-domain feature extraction, fusion and optimization are applied to reduce redundancy and improve generalization, since direct concatenation of S-Transform, GASF, SSWT, and Hankel features leads to imbalance. First, features are normalized using z-score:

$$\tilde{F}^{(i)} = \frac{F^{(i)} - \mu^{(i)}}{\sigma^{(i)}} \tag{12}$$

The normalized representations are then adaptively fused:

$$F_{fusion} = \sum_{i=1}^{4} \alpha_i \tilde{F}^{(i)}, \sum \alpha_i = 1 \tag{13}$$

This weighted fusion adaptively assigns higher importance to more informative modalities while suppressing less relevant ones, allowing the model to dynamically emphasize the most discriminative feature type for each input rather than enforcing equal contribution. Finally, Coyote Optimization Algorithm (COA) [27] is applied to select the most informative subset by balancing classification performance and sparsity:

$$J(X_c) = \eta E_{cls} + (1-\eta)\frac{\| X_c \|_0}{D} \tag{14}$$

This integrated fusion–optimization pipeline produces a compact, discriminative feature representation suitable for robust multi-participant HAR. The optimized features result is shown in Fig. 10.

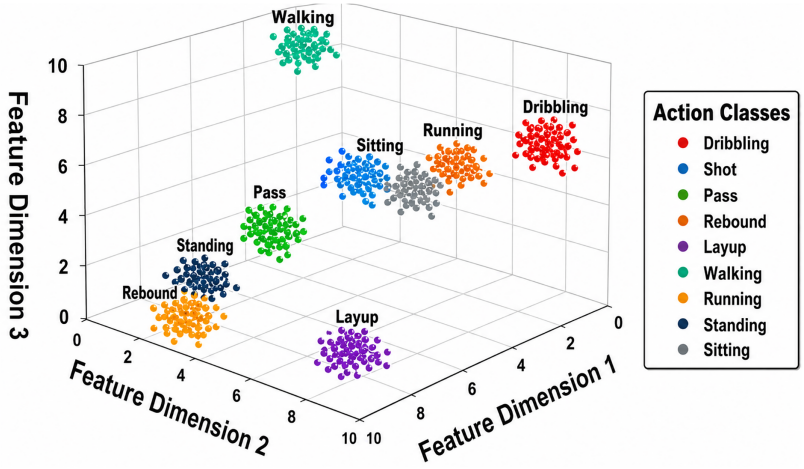


Figure 10: 3D visualization of optimized feature space after coyote optimization algorithm (COA).

3.7 Temporal Alignment via Soft-DTW Temporal Aligner

In team sports, actions are often not time-synchronized, as players perform the same activity at different speeds and delays, leading to temporal misalignment. To handle this, a Soft-DTW-based aligner is used for smooth and differentiable sequence alignment, unlike classical DTW which is non-differentiable and sensitive to noise and outliers in the amplitude domain. Soft-DTW is defined as:

$$SoftDTW_\gamma(X,Y) = -\gamma log \sum_{\pi \in A} exp\left(-\frac{1}{\gamma}\right) \sum_{(i,j)\in\pi} d(x_i, y_j) \tag{15}$$

Here, A denotes alignment paths, $d(\cdot,\cdot)$ the Euclidean distance, and $\gamma = 0.1$. The formulation aligns sequences despite speed and temporal variations (Fig. 11).

3.8 Activity Mapping via Hierarchical Hidden Semi-Markov Model (H-HSMM)

After temporal alignment, structured activity mapping assigns consistent states from synchronized multi-participant features. Conventional HMMs assume memoryless, geometrically distributed state durations and therefore cannot represent sports actions of variable length. To address this, a Hierarchical Hidden Semi-Markov Model (H-HSMM) is used to capture player-level actions and group-level interactions (passes, defense, transitions). The joint model is defined as:

$$p(O,Z,D) = \prod_t p\left(O_t \mid Z_t^{(1)}\right) p(D_t \mid Z_t^{(1)}) p(Z_t^{(l)} | Z_{t-1}^{(l)}, Z_t^{(l-1)}) \tag{16}$$

Here, D_t models the activity duration using:

$$p(d \mid z) = \frac{1}{\sqrt{2\pi\sigma_z^2}} exp\left(-\frac{(d-\mu_z)^2}{2\sigma_z^2}\right) \tag{17}$$

This allows distinction between short actions (passes, shots) and long activities (running, standing) while capturing inter-player dependencies, enabling robust modeling of coordinated and overlapping sports behaviors before final classification.

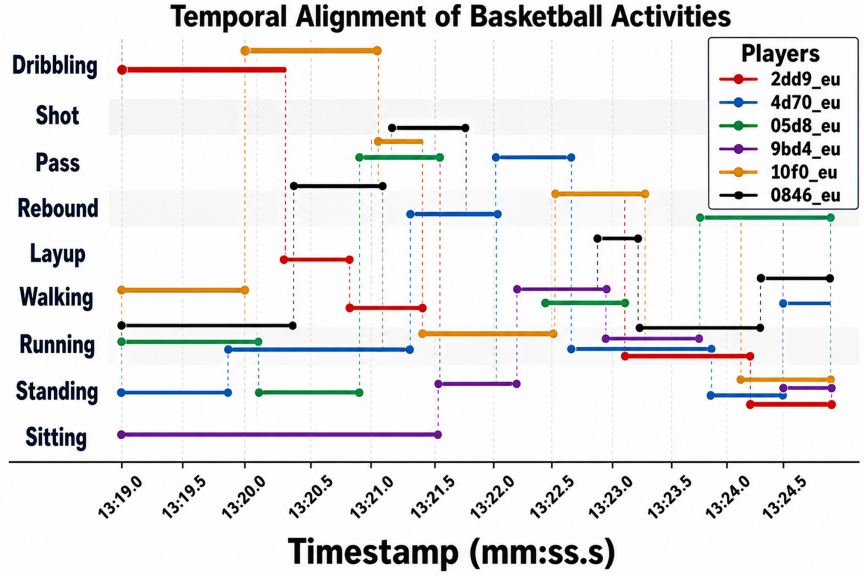


Figure 11: Soft-DTW temporal alignment of basketball activities across multiple players, synchronizing heterogeneous IMU sequences with varying durations and offsets.

3.9 Activity Classification via InceptionTime Ensemble (IT-E)

The final stage uses an InceptionTime Ensemble for semantic activity classification, enabling multi-scale and long-term temporal modeling of synchronized features after fusion, optimization, and H-HSMM mapping. It captures both fine-grained transitions and coarse activity structure using parallel 1D convolutions, while bottleneck and residual connections ensure stable and efficient learning.

$$Z(t) = \bigoplus_{k \in K} \sigma(W_k * X(t) + b_k) \tag{18}$$

where different kernel sizes extract complementary temporal patterns from IMU sequences. Final predictions are obtained through ensemble probability aggregation:

$$\hat{y} = \arg\max_{c} \frac{1}{M} \sum_{m=1}^{M} p_m(c \mid X) \tag{19}$$

This ensemble strategy improves robustness against temporal variations and inter-subject differences, while ensuring consistent and reliable classification of complex multi-participant sports activities. IT-E utilized architecture is shown in Fig. 12.

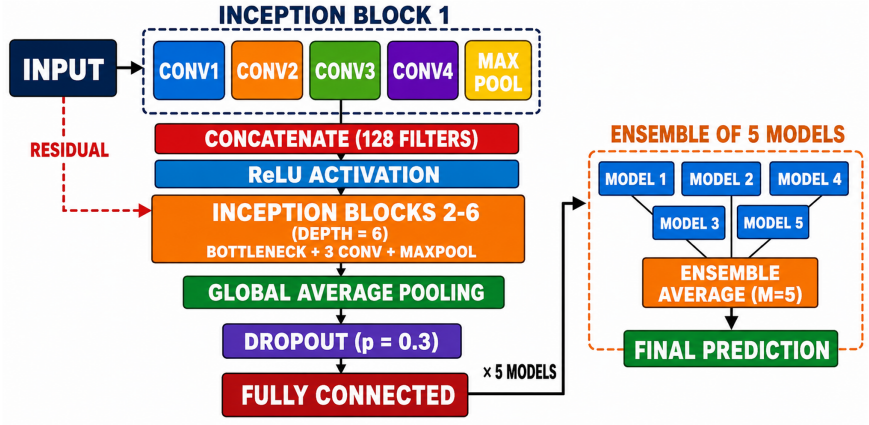


Figure 12: InceptionTime ensemble (IT-E) architecture with residual inception blocks, global pooling, dropout, and averaging across 5 models for robust classification.

3.10 Edge-Cloud IoT Monitoring System

The proposed edge–cloud HAR framework enables scalable, real-time monitoring through a three-tier edge–cloud architecture. Wearable IMUs collect motion data and transmit preprocessed frames and compact features via BLE/Wi-Fi/5G and MQTT. The edge gateway performs Tukey filtering, Planck-Taper segmentation, and compact feature extraction, while cloud analytics handle A-DMKL fusion, multi-resolution feature extraction, COA optimization, Soft-DTW, H-HSMM, and InceptionTime classification. Outputs, including activity predictions and alerts, are stored and visualized via HTTPS for intelligent smart-health monitoring.

4 Experimental Settings and Analysis

Experiments were conducted on an Intel i7-13700H CPU with 64 GB RAM and an NVIDIA RTX 4070 GPU, used to emulate the cloud-side processing environment for offline model training and evaluation. To evaluate real-time feasibility, computational profiling was also performed under resource-constrained edge/gateway settings. The IMU data was collected from wearable sensor nodes at 50–60 Hz, consistent with BLE-based commercial IoT sports monitoring devices.

4.1 Dataset Description

4.1.1 Hang-Time HAR Dataset

The Hang-Time HAR [1] is a benchmark dataset using wrist-worn inertial sensing in basketball training and game scenarios. It contains data from 24 participants across teams in Germany and the United States, covering 10 basketball-related activities recorded during warm-up, structured drills, and game sessions. The dataset provides wrist-worn three-dimensional accelerometer measurements sampled at approximately 50 Hz, with labels organized across coarse, locomotion, basketball, and in/out levels. Its combination of periodic, short-duration, and complex activities under both controlled and less-controlled conditions makes it suitable for evaluating robust activity-recognition models.

4.1.2 VIMPT2019 Dataset

The VIMPT2019 [28] is a video–inertial multi-person tracking dataset designed to evaluate the association of people observed in video with their corresponding body-worn inertial sensors. It contains seven

sequences, including six soccer sequences and one outdoor sequence, with synchronized RGB-camera video and orientation and acceleration measurements from eight Inertial Measurement Units (IMUs) per sequence.

4.1.3 Annotation Protocol for VIMPT2019

Nine mutually exclusive classes were predefined from observable video cues (Table 3). Two annotators independently labelled sequences frame by frame using fixed guidelines, with segments shorter than 1.0 s excluded. Ambiguous frames (4.2%) were removed, and disagreements were resolved by a third annotator. The initial agreement was $\kappa = 0.74$. Labels were mapped to the inertial timeline using frame-to-sample correspondence. After annotation, 2.0 s Planck-Taper windows with 50% overlap were generated; windows with <70% label dominance were discarded (9.6%).

Table 3: Annotated categories for VIMPT2019 with segment count and agreement.

Class	Definition (Video-Observable)	Segments	κ
Fast movements	Rapid displacement without sustained sprinting.	412	0.71
Sprinting	Straight-line maximal running ≥1.0 s with flight phase.	96	0.89
Ball handling	Continuous ball control within one body length.	288	0.79
Team interaction	Coordinated action toward another visible person.	350	0.69
Physical contact	Body-to-body contact lasting ≥0.5 s.	154	0.81
Dodging	Lateral direction change >45° to avoid another person.	122	0.72
Close quarters	Distance ≈ 1 m for ≥1.0 s without contact.	401	0.64
Walking	Continuous locomotion without flight phase.	528	0.86
Scene transitions	Region changes not covered above.	187	0.58

4.1.4 Evaluation Protocol, Subject Counts, and Leakage Control

All results are subject-independent. Hang-Time HAR (24 participants) uses 24-fold leave-one-subject-out (LOSO) cross-validation, while VIMPT2019 (8 participants) uses 8-fold LOSO. Since each participant contributed one recording session with fixed sensor placement, leave-one-user-out (LOUO) and LOSO are equivalent in this study, with no participant appearing in multiple folds. In each fold, the held-out participant is used exclusively for testing, while the remaining participants are split by identity into inner training and validation sets (approximately 85%/15%) for hyperparameter tuning and early stopping. Reported accuracy is averaged across folds. Leakage was controlled throughout the pipeline. The z-score statistics in Eq. (12), A-DMKL attention parameters, Soft-DTW and kernel bandwidths, H-HSMM duration parameters, and InceptionTime weights were estimated only from inner-training participants. The COA objective in Eq. (14) was optimized using inner-validation error, without access to the held-out participant. Feature subsets were independently re-selected for each fold. Windowing was performed after participant-level partitioning, preventing overlapping windows from crossing folds, and participant-specific normalization was not transferred across folds.

4.2 Model Evaluation and Experimental Results

Generalization was assessed on Hang-Time HAR and VIMPT2019 using subject-independent leave-one-subject-out evaluation on both datasets, as detailed in Section 4.1.4. The InceptionTime Ensemble classified the activities, with performance measured by Accuracy, Precision, Recall, F1-score, AUC, and

confusion analysis. Paired t-tests ($p < 0.05$) verified improvements over baselines. Fig. 13 shows the Confusion matrices and Fig. 14 shows the ROC Curves. The Hang-Time HAR performance results are presented in Table 4.

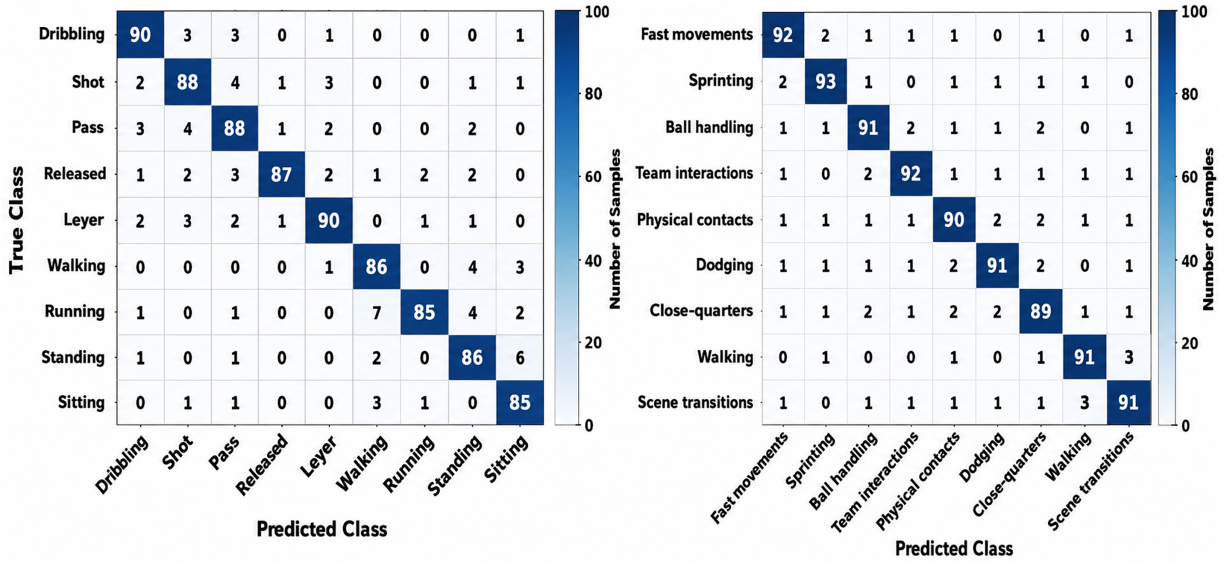


Figure 13: Confusion matrix on the Hang-Time HAR and VIMPT2019 dataset.

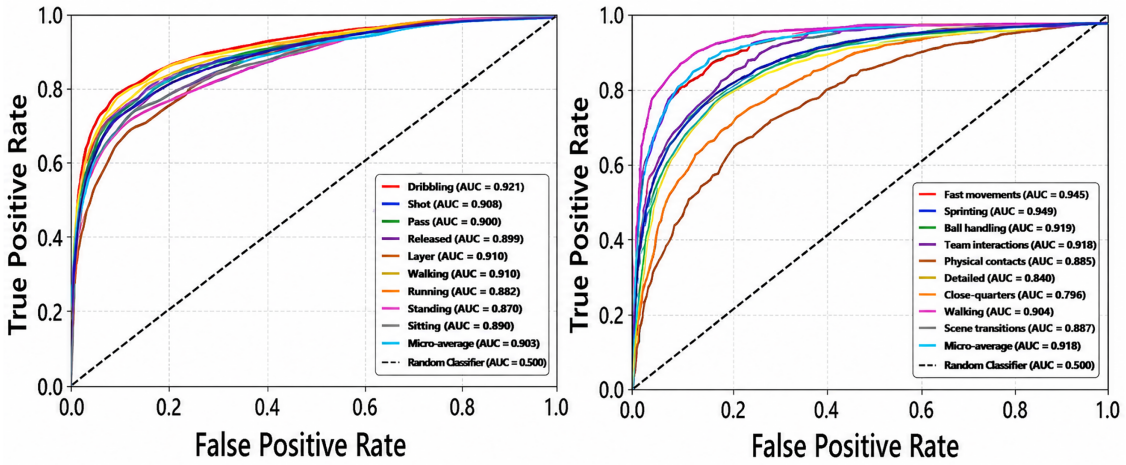


Figure 14: ROC curves with per-class AUC scores for basketball activities on the Hang-Time HAR dataset.

Table 4: Precision, recall, and F1-score evaluated on the Hang-Time HAR dataset.

Classes	Precision	Recall	F1-Score
Dribbling	0.900	0.903	0.900
Shot	0.871	0.882	0.875
Pass	0.838	0.880	0.858
Rebound	0.966	0.874	0.915

(Continued)

Table 4 (continued)

Classes	Precision	Recall	F1-Score
Layup	0.909	0.904	0.904
Walking	0.860	0.864	0.860
Running	0.858	0.852	0.854
Standing	0.796	0.860	0.826
Sitting	0.867	0.854	0.858

The VIMPT2019 performance results are presented in Table 5, computational complexity is reported in Table 6, and state-of-the-art comparisons are provided in Table 7.

Table 5: Precision, recall, and F1-score evaluated on the VIMPT2019 dataset.

Classes	Precision	Recall	F1-Score
Fast movements	0.920	0.929	0.924
Sprinting	0.930	0.930	0.930
Ball handling	0.910	0.910	0.910
Team Interaction	0.929	0.920	0.924
Physical Contact	0.900	0.900	0.900
Dodging	0.919	0.910	0.914
Close Quarters	0.890	0.890	0.890
Walking	0.930	0.939	0.934
Scene transitions	0.910	0.910	0.910

Table 6: Computational complexity analysis across both datasets.

Methods	Computational Complexity	Execution Time–Hang-Time HAR (s)	Memory Usage–Hang-Time HAR (MB)	Execution Time–VIMPT2019 (s)	Memory Usage–VIMPT2019 (MB)
Pre-processing	$O(N \cdot I \cdot W)$	0.58	220	0.32	150
Multi-Kernel Fusion	$O(M \cdot d^2 + N^2)$	2.12	480	1.58	360
Signal Segmentation	$O(N)$	0.85	280	0.60	200
Feature Extraction	$O(N \cdot F \cdot \log N)$	1.92	420	1.45	320
Feature Optimization	$O(P \cdot I \cdot d)$	1.68	580	1.26	460
Temporal Alignment	$O(N^2)$	1.10	360	0.80	280

(Continued)

Table 6 (continued)

Methods	Computational Complexity	Execution Time–Hang-Time HAR (s)	Memory Usage–Hang-Time HAR (MB)	Execution Time–VIMPT2019 (s)	Memory Usage–VIMPT2019 (MB)
Activity Mapping	$O(L \cdot S^2 \cdot T^2)$	1.02	520	0.76	420
Classification (IT-E)	$O(E \cdot L \cdot K \cdot N)$	1.95	500	1.60	400

Table 7: Classification comparison with other state-of-the-art models.

Authors	Method/Model	Dataset (Sport Context)	Accuracy (%)
[29]	k-Nearest Neighbors (and RF) on wrist-worn IMU features	Basketball training/game (wrist IMU data from 3 players)	83.6% (kNN) (RF: 87.5%)
[30]	KNN on accelerometer/gyroscope from a wrist smartwatch	Basketball drills (21 collegiate players: dribble, pass, shoot)	81.62%
[31]	(HMM variant)	Custom Basketball Motion Dataset (IMU-based jump shots in sequences)	85.0%
[32]	LSTM-based	WEAR Outdoor Sports Dataset	80.9%
[33]	Ensemble bagged-trees on ankle-mounted IMU	Soccer (6 basic movements: run, dribble, pass, shoot, etc.; single inside-ankle sensor)	78.7%
[34]	SVM (ankle attitude-angle features)	Soccer (passes vs. shots; 11 players, ankle IMU + IoT)	85.7%
[35]	Ensemble + temporal fusion	IMU-based HAR	86.45%
[36]	Attention deep multi-kernel fusion	IMU/multimodal HAR	89.25%
Proposed Method	Multi-stage Pipeline (A-DMKL + IT-E)	Hang-Time HAR (Basketball) VIMPT2019 (Soccer)	87.22% 91.33%

4.2.1 Ablation Study

A leave-one-module-out ablation was performed on both datasets, with each component removed while others remained fixed. As shown in Table 8, A-DMKL produced the largest performance drop (−4.8% and −4.2%), followed by multi-transform features and H-HSMM. Although Soft-DTW and Planck-Taper contributed less than 1%, all components improved performance, supporting the overall pipeline complexity.

Table 8: Leave-one-module-out ablation.

Configuration	Hang-Time Acc (%)	Δ	VIMPT2019 Acc (%)	Δ
Full model	**87.2**	—	**91.3**	—
w/o A-DMKL fusion	82.4	−4.8	87.1	−4.2
w/o Multi-transform features	84.1	−3.1	88.6	−2.7
w/o H-HSMM	85.0	−2.2	88.8	−2.5
w/o COA optimisation	85.6	−1.6	89.9	−1.4
w/o InceptionTime ensemble	85.9	−1.3	90.2	−1.1
w/o Planck-Taper window	86.3	−0.9	90.3	−1.0
w/o Soft-DTW alignment	86.6	−0.6	90.6	−0.7

4.2.2 Robustness under Simulated Transport Impairments

Deployment robustness was assessed through simulated latency, MQTT packet loss, and synchronization errors. These impairments were injected offline into replayed streams and applied to host-profiled execution times. Tables 9–11 summarize the results. As no real network, MQTT broker, or ARM hardware was used, the results represent an idealized upper-bound analysis rather than real deployment performance.

Table 9: Analytically modelled end-to-end latency budget under three parametric service profiles.

Profile	Uplink (ms)	E2E Mean (ms)	E2E p99 (ms)	Meets 50 ms
URLLC	0.30	17.5	17.6	Yes
eMBB	1.0	18.2	18.9	Yes
mMTC	8.0	25.2	32.1	Yes

Table 10: MQTT packet-loss robustness (QoS-0, last-value-hold).

Packet Loss (%)	Accuracy (%)	Macro-F1 (%)
0	87.2	85.6
1	87.0	85.4
5	85.9	84.1
10	83.7	81.7
15	80.4	78.1
20	76.1	73.4
30	66.8	63.2

As shown in Fig. 15, the end-to-end latency remains within the real-time budget across all three 6G service profiles, with the per-stage breakdown confirming that the uplink contributes the primary variation between profiles.

Table 11: Synchronization-error robustness.

Sync Error (ms)	Accuracy (%)	Macro-F1 (%)
0	87.2	85.6
10	87.0	85.3
20	86.6	84.9
40	85.9	84.1
80	83.1	81.0
120	78.9	76.4
160	73.4	70.6

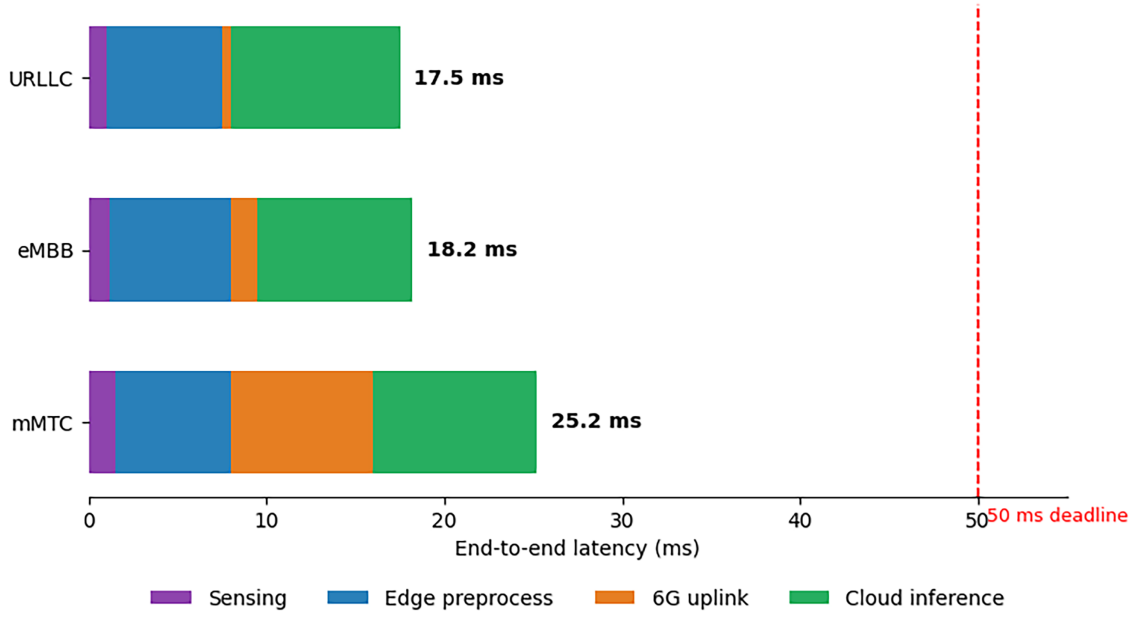


Figure 15: End-to-end latency budget across service profiles (eMBB, URLLC, mMTC).

4.2.3 Aggregate Results and Statistical Validation

Evaluation is subject-independent (Section 4.1.4), using 2.0 s windows with 50% overlap, yielding approximately 38,500 windows on Hang-Time HAR and 8700 on VIMPT2019. The framework achieves 87.22% and 91.33% accuracy with balanced F1 scores (SD = 0.029, 0.014). Removing any component lowers accuracy by an average of 2.01 points, confirming internal consistency. Main errors occur in Standing–Defending and Pass–Shot (Hang-Time HAR), and Close Quarters (VIMPT2019). All configurations were evaluated with a single fixed seed, so across-seed variance remains unquantified. Table 12 shows the aggregate performance across datasets.

5 Discussion and Research Limitations

Despite strong results, real-world Edge IoT deployment may be affected by sensor variability, class imbalance, domain shifts, and multi-user dynamics. The robustness analysis relies on simplifying assumptions that may yield optimistic upper-bound estimates, while host-based timing cannot confirm real on-device performance or energy use. Future work will validate the framework under real networks, ARM

hardware, packet loss, synchronization errors, energy constraints, and large-scale 6G IoT settings. Clinically, the framework supports potential activity, rehabilitation, and sports-health monitoring [37], but clinical efficacy and outcomes require prospective patient-based validation.

Table 12: Aggregate performance obtained by macro-averaging the per-class.

Aggregate Metric	Hang-Time HAR	VIMPT2019
Accuracy (%)	87.22	91.33
Macro-Precision	0.874	0.915
Macro-Recall	0.875	0.915
Macro-F1	0.872	0.915
Per-Class F1 (Min–Max)	0.826–0.915	0.890–0.934
Per-Class F1 (SD)	0.029	0.014
Mean AUC	0.951	0.972

6 Conclusion and Future Work

This study proposed an AI-driven wearable IoT pipeline integrating Tukey filtering, Planck-Taper segmentation, A-DMKL fusion, multi-transform features, COA optimization, Soft-DTW, H-HSMM, and InceptionTime within an edge–cloud architecture. Under subject-independent LOSO evaluation, it achieved 87.22% accuracy on Hang-Time HAR and 91.33% on VIMPT2019, with input-conditioned kernel weighting as the most influential component and 99.7% peak accuracy retained at 67% fusion cost. The framework also tolerated 5% MQTT packet loss and 40 ms synchronization error with approximately one-point accuracy loss. However, validation on only two small sports datasets, simulation-based robustness analysis, optimistic assumptions, x86-based profiling, time-based energy estimation, single-seed evaluation, and moderate annotator agreement for some classes limit generalizability. Further validation on larger cohorts, ARM hardware, real networks, and clinical populations is needed.

Future studies are recommended to proceed in four directions: replicating the pipeline on ARM-class embedded hardware with measured latency and energy, validating robustness on a live MQTT broker under real radio conditions including bursty loss and clock drift, extending the evaluation to larger multi-site cohorts and additional sensing modalities with multi-seed repetitions and confidence intervals, and conducting prospective clinical validation against reference standards before the framework is used for rehabilitation or injury-risk assessment.

Acknowledgement: We thanks to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R719), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R719), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Muhammad Hanzla and Fatimah Alhayan contributed to the conceptualization of the study and development of the proposed methodology. Hadeel Alsolai and Bayan Alabdullah were involved in system design, supervision, and critical revision of the manuscript. Ahmad Jalal and Hui Liu contributed to experimental validation, result analysis, and manuscript review and editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets analyzed in this study are publicly available. The Hang-Time HAR dataset is available via Zenodo at https://zenodo.org/records/7920485, with the corresponding dataset description published in *Sensors* (https://doi.org/10.3390/s23135879). The VIMPT2019 dataset is available from the Institut für Informationsverarbeitung (TNT), Leibniz University Hannover, at https://www.tnt.uni-hannover.de/en/project/VIMPT2019/.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hoelzemann A, Romero JL, Bock M, Laerhoven KV, Lv Q. Hang-time HAR: a benchmark dataset for basketball activity recognition using wrist-worn inertial sensors. Sensors. 2023;23(13):5879. doi:10.3390/s23135879.
2. Cai Y, Guo B, Salim FD, Hong Z. Towards generalizable human activity recognition: a survey. arXiv:2508.12213. 2025. doi:10.48550/arXiv.2508.12213.
3. Gomaa W, Khamis MA. A perspective on human activity recognition from inertial motion data. Neural Comput Appl. 2023;35(28):20463–568. doi:10.1007/s00521-023-08863-9.
4. Murphy SJ, Vitali RV. Assessment of IMU configurations for human activity recognition using ensemble learning. IEEE Access. 2024;12:111433–42. doi:10.1109/access.2024.3440215.
5. Baca A, Dabnichki P, Hu CW, Kornfeind P, Exel J. Ubiquitous computing in sports and physical activity—recent trends and developments. Sensors. 2022;22(21):8370. doi:10.3390/s22218370.
6. Friemert D, Schnur D, Runkel S, Borsch J, Karamanidis K, Dellen B, et al. Limitations of public biomechanical movement datasets for deep learning: issues of metadata, standardization, and variety in motion types. medRxiv. 2025. doi:10.1101/2025.05.29.25328474.
7. Ray LSS, Xia Q, Rey VF, Wu K, Lukowicz P. Improving IMU based human activity recognition using simulated multimodal representations and a MoE classifier. Front Comput Sci. 2025;7:1569205. doi:10.3389/fcomp.2025.1569205.
8. Ray LSS, Geißler D, Liu M, Zhou B, Suh S, Lukowicz P. ALS-HAR: harnessing wearable ambient light sensors to enhance IMU-based human activity recognition. In: Pattern recognition. Cham, Switzerland: Springer Nature; 2025. p. 133–47. doi:10.1007/978-3-031-78110-0_9.
9. Hoareau D, Jodin G, Chantal PA, Bretin S, Prioux J, Razan F. Synthetized inertial measurement units (IMUs) to evaluate the placement of wearable sensors on human body for motion recognition. J Eng. 2022;2022(5):536–43. doi:10.1049/tje2.12137.
10. Guo X, Brown E, Chan PPK, Chan RHM, Cheung RTH. Skill level classification in basketball free-throws using a single inertial sensor. Appl Sci. 2023;13(9):5401. doi:10.3390/app13095401.
11. Reilly B, Morgan O, Czanner G, Robinson MA. Automated classification of changes of direction in soccer using inertial measurement units. Sensors. 2021;21(14):4625. doi:10.3390/s21144625.
12. Wang ZZ, Xia X, Chen Q. Multi-level data fusion enables collaborative dynamics analysis in team sports using wearable sensor networks. Sci Rep. 2025;15(1):28210. doi:10.1038/s41598-025-12920-9.
13. Thukral M, Haresamudram H, Plötz T. Cross-domain HAR: few-shot transfer learning for human activity recognition. ACM Trans Intell Syst Technol. 2025;16(1):1–35. doi:10.1145/3704921.
14. Arrotta L, Civitarese G, Chen X, Cumin J, Bettini C. Multi-subject human activities: a survey of recognition and evaluation methods based on a formal framework. Expert Syst Appl. 2025;267(12):126178. doi:10.1016/j.eswa.2024.126178.
15. Yin H, Sinnott RO, Jayaputera GT. A survey of video-based human action recognition in team sports. Artif Intell Rev. 2024;57(11):293. doi:10.1007/s10462-024-10934-9.
16. Cummins C, Orr R, O'Connor H, West C. Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. Phys Med. 2013;43(10):1025–42. doi:10.1007/s40279-013-0069-2.

17. Siwadamrongpong W, Chinrungrueng J, Hasegawa S, Nantajeewarawat E. Fall detection and prediction based on IMU and EMG sensors for elders. In: Proceedings of the 2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE); 2022 Jun 22–25; Bangkok, Thailand.
18. Yang Y, Mohammadzadeh F, Khishe M, Ahmed AN, Abualhaj MM, Ghazal TM. Deep learning for sports motion recognition with a high-precision framework for performance enhancement. Sci Rep. 2025;15(1):38861. doi:10.1038/s41598-025-22701-z.
19. Adeyemo VE, Palczewska A, Jones B, Weaving D, Whitehead S. Optimising classification in sport: a replication study using physical and technical-tactical indicators in rugby league. Sci Med Footb. 2024;8(1):68–75.
20. Challa SK, Kumar A, Semwal VB, Dua N. An optimized deep learning model for human activity recognition using inertial measurement units. Expert Syst. 2023;40(10):e13457. doi:10.1111/exsy.13457.
21. Gao Z, Wang Y, Chen J, Xing J, Patel S, Liu X, et al. MMTSA: multi-modal temporal segment attention network for efficient human activity recognition. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2023;7(3):1–26.
22. Zhou H, Arakawa R, Agarwal Y, Goel M. IMUCoCo: enabling flexible on-body IMU placement for human pose estimation and activity recognition. In: Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology; 2025 Sep 28–Oct 1; Busan, Republic of Korea. p. 1–16. doi:10.1145/3746059.3747695.
23. Yu Y, Li S, Jiang Y, Xiao H, Long J, Tang H, et al. CMD-HAR: cross-modal disentanglement for wearable human activity recognition. arXiv:2503.21843. 2025.
24. Arabzadeh S, Almasganj F, Ahmadi MM. CNN autoencoders for hierarchical feature extraction and fusion in multi-sensor human activity recognition. arXiv:2502.04489. 2025.
25. Xie D, Zhang X, Gao X, Zhao H, Du D. MAF-Net: a multimodal data fusion approach for human action recognition. PLoS One. 2025;20(4):e0319656. doi:10.1371/journal.pone.0319656.
26. Bian S, Liu M, Ray LSS, Zhou B, Guo B, Yu Z, et al. Foundation models defining a new era in sensor-based human activity recognition: a survey and outlook. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2026;10(2):1–38. doi:10.1145/3810230.
27. Pierezan J, Coelho LDS. Coyote optimization algorithm: a new metaheuristic for global optimization problems. In: Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC); 2018 Jul 8–13; Rio de Janeiro, Brazil. New York, NY, USA: IEEE; 2018. p. 1–8. doi:10.1109/CEC.2018.8477769.
28. Henschel R, von Marcard T, Rosenhahn B. Accurate long-term multiple people tracking using video and body-worn IMUs. IEEE Trans Image Process. 2020;29:8476–89. doi:10.1109/TIP.2020.3013801.
29. Hasegawa T. Channel-free human activity recognition via inductive-bias-aware fusion design for heterogeneous IoT sensor environments. arXiv:2604.21369. 2026.
30. Hölzemann A, Van Laerhoven K. Using wrist-worn activity recognition for basketball game analysis. In: Proceedings of the 5th International Workshop on Sensor-Based Activity Recognition and Interaction; 2018 Sep 20–21; Berlin, Germany. p. 1–6. doi:10.1145/3266157.3266217.
31. Asmara RA, Hendrawan ND, Handayani AN, Arai K. Basketball activity recognition using supervised machine learning implemented on tizen OS smartwatch. J Ilm Tek Elektro Komput Dan Inform. 2022;8(3):447. doi:10.26555/jiteki.v8i3.23668.
32. Cuperman R, Jansen KMB, Ciszewski MG. An end-to-end deep learning pipeline for football activity recognition based on wearable acceleration sensors. Sensors. 2022;22(4):1347. doi:10.3390/s22041347.
33. Eggert B, Mundt M, Markert B. IMU-based activity recognition of the basketball jump shot. ISBS Proc Arch. 2020;38(1):344–7.
34. Kondo Y, Ishii S, Aoyagi H, Hossain T, Yokokubo A, Lopez G. FootbSense: soccer moves identification using a single IMU. In: Sensor- and video-based activity and behavior computing. Singapore: Springer; 2022. p. 115–31. doi:10.1007/978-981-19-0361-8_7.
35. Lu S, Zhang X, Wang J, Wang Y, Fan M, Zhou Y. An IoT-based motion tracking system for next-generation foot-related sports training and talent selection. J Healthc Eng. 2021;2021:9958256. doi:10.1155/2021/9958256.

36. Dahou A, Al-qaness MA, Abd Elaziz M, Helmi AM, Trouba NT, Ewess A, et al. Ensemble learning with temporal fusion for human activity recognition using model-based augmentation. Biomed Signal Process Control. 2025;110(1):108168. doi:10.1016/j.bspc.2025.108168.
37. Zam A, Bohlooli A, Jamshidi K. Unsupervised deep domain adaptation algorithm for video based human activity recognition via recurrent neural networks. Eng Appl Artif Intell. 2024;136(Part A):108922. doi:10.1016/j.engappai.2024.108922.