Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.086599

ARTICLE

Edge-AI Enabled Multi-Agent Intelligence Framework for Evolutionary Optimization and Mobility Localization in 6G-IoT Smart Sensing Systems

Muhammad Hanzla[1], Bayan Alabdullah[2], Mohammad Shorfuzzaman[3,*], Mohammed Alonazi[4], Jasem Almotiri[5] and Ahmad Jalal[1,6,*]

[1]Department of Computer Science, Air University, Islamabad, Pakistan
[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia
[3]Department of Software Engineering, College of Engineering and Advanced Computing, Alfaisal University, Riyadh, Saudi Arabia
[4]Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia
[5]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia
[6]Department of Computer Science and Engineering, College of Informatics, Korea University, Seoul, Republic of Korea

*Corresponding Authors: Mohammad Shorfuzzaman. Email: mshorfuzzaman@alfaisal.edu; Ahmad Jalal. Email: ahmadjalal@mail.au.edu.pk

Received: 02 June 2026; Accepted: 26 August 2026; Published: 22 October 2026

ABSTRACT: The increasing adoption of wearable Internet of Things (IoT) devices and wireless sensor networks has accelerated the demand for intelligent, low-latency human activity recognition and localization systems. However, dynamic environments, sensor noise, motion variability, and communication delays continue to hinder reliable real-time sensing. To address these challenges, this study presents an Edge-AI enabled multi-agent intelligence framework for human activity recognition and mobility localization within a 6G-enabled IoT edge-cloud continuum. The proposed framework employs Self-supervised Transformer-based Adaptive Sensor Encoding and learning-aware temporal windowing to generate robust representations from heterogeneous sensing modalities. Extracted locomotion and localization features are optimized using the Aquila Optimizer (AO) and classified through a deep neuro-fuzzy inference model. Furthermore, an edge-assisted monitoring platform, EvolveSense, is developed to support real-time visualization and low-latency intelligent analytics. Experimental evaluation on the ExtraSensory and Opportunity datasets achieved activity recognition accuracies of 93.63% and 95.57%, respectively, while localization accuracy reached 91.33% and 92.40%. The results demonstrate the effectiveness of Edge AI, intelligent sensing, and optimized multi-modal learning for next-generation IoT healthcare and smart environment applications.

KEYWORDS: Multi-agent intelligence; Internet of Things (IoT); human activity recognition; mobility localization; edge computing; wearable sensors; smart healthcare; edge AI

1 Introduction

The rapid growth of Wireless Sensor Networks (WSN), Internet of Things (IoT), and emerging 6G communication technologies have significantly advanced human activity recognition (HAR) and localization systems [1]. Wearable and mobile sensing devices, including accelerometers, gyroscopes, GPS, and acoustic sensors, enable continuous monitoring of human movement and positioning in intelligent environments [2]. These capabilities have enabled a wide range of applications such as smart healthcare, rehabilitation, indoor navigation, sports analytics, and intelligent IoT-based systems [3]. Despite these advancements,

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.

achieving reliable localization and activity recognition in real-world dynamic environments remains challenging. Sensor noise, motion variability, synchronization issues, and communication latency continue to degrade system performance [4]. In addition, existing approaches often exhibit limited adaptability and inefficient feature fusion when deployed in heterogeneous IoT environments [5]. Another important limitation is that many existing studies treat localization and activity recognition as separate problems, which reduces contextual understanding and limits real-time decision-making in dense sensing scenarios [6]. To overcome these limitations, there is a growing need for Edge AI-enabled intelligent systems that can process multi-modal sensor data closer to the data source while ensuring low latency, scalability, and improved context-awareness. Edge computing combined with IoT, and 6G connectivity provides a promising framework for distributed intelligence, enabling real-time analytics and reducing reliance on centralized cloud processing.

This paper proposes an Edge AI-enabled framework for human locomotion analysis and localization within a 6G IoT edge–cloud continuum. It leverages Self-supervised Transformer-based Adaptive Sensor Encoding (STASE) Learning-Aware Windowing for Inertial and Neural Sensing and LAWINS for efficient multimodal representation learning, with Aquila Optimizer for feature refinement and a neuro-fuzzy model for classification and localization. The system is deployed via EvolveSense for real-time, low-latency monitoring and visualization.

The main contributions of this work are summarized as follows:

- **Edge AI-Based Integrated Framework:** A unified 6G-enabled architecture for joint human activity recognition and localization in IoT and WSN environments.
- **Adaptive Multi-Modal Representation Learning:** A transformer-based encoding mechanism with learning-aware temporal windowing for robust and synchronized sensor feature extraction.
- **AI-Driven Feature Optimization:** An Aquila Optimizer-based strategy for selecting discriminative features for locomotion and localization tasks.
- **Edge-Enabled Intelligent Inference:** A deep neuro-fuzzy model designed for adaptive and efficient decision-making in edge-based IoT systems.
- **Real-Time Edge Monitoring Platform:** Development of the EvolveSense platform for visualization and low-latency monitoring in smart IoT environments.

The remainder of this paper is organized as follows. Section 2 reviews related work on HAR and localization in IoT and edge intelligence systems. Section 3 presents the proposed Edge AI framework and system architecture. Section 4 describes the experimental setup and evaluation methodology. Section 5 concludes the paper and outlines future research directions.

2 Related Work

Smartphone, wearable, and IoT sensing enable continuous HAR and localization in real-world settings. Recent studies use classical ML, deep learning, and Edge AI to handle noise, variability, and position estimation, but still face high computational cost, fixed temporal windows, and limited interpretability. This section briefly reviews recent five-year studies, focusing on methods, strengths, and limitations.

2.1 HAR and Localization with Smartphone, Wearable, and IoT Sensors

Recent studies explore smartphone, wearable, and IoT sensing for HAR and localization using classical, deep learning, and hybrid methods. Moreira et al. [7] propose a ConvLSTM-based HAR–PDR framework (~2.4 m error) but with limited generalization. Koşar and Barshan [8] use a CNN–LSTM model with high accuracy but increased computational cost due to spectrograms. Ciortuz et al. [9] show deep

models outperform classical ones, though SVM remains competitive for simpler tasks. Mekruksavanich and Jitpattanakul [10] use attention-based Bi-GRU, improving complex activity recognition but without real-time validation. Classical methods remain effective in constrained settings, with Silva Zendron et al. [11] achieving ~90% accuracy. Deep and multimodal approaches show limited robustness gains: Najeh et al. [12] and Yoon and Kim [13] combine deep regression with Kalman filtering (3–5 m error) but rely on GPS and show weak indoor robustness. Alazeb et al. [14] report ~88%–90% accuracy with weak generalization. Overall, Klein [15] and Borsuk et al. [16] highlight performance drops in unseen settings.

2.2 Motivation and Superiority of the Proposed Framework

Unlike prior HAR–localization pipelines that integrate off-the-shelf components, the methodological novelty of this work lies in four contributions and their coupling. First, STASE introduces a self-supervised, signal-corruption reconstruction objective tailored to noisy, heterogeneous inertial streams, learning noise-invariant representations without any activity labels in contrast to supervised transformer-HAR models that require large, labelled datasets. Second, LAWINS replaces fixed/heuristic windowing with a differentiable, variance-aware gating mechanism that segments signals using the temporal dynamics learned by STASE, directly coupling representation learning with temporal segmentation. A shared representation is used for both activity recognition and mobility localization, unlike prior work that treats them separately, and the Aquila Optimizer performs joint feature selection in this dual-domain space. DNFC adds interpretable fuzzy reasoning with uncertainty handling for reliable edge inference. Together, STASE improves signal quality, LAWINS preserves transitions, Aquila removes redundant features, and DNFC handles ambiguity. Ablation results confirm the full pipeline outperforms all reduced variants.

3 Materials and Methods

The proposed framework is a 6G-enabled Edge AI-based multi-modal sensing system for real-time HAR and localization built on a 6G-edge-cloud continuum (Fig. 1). It follows a layered architecture for scalable, low-latency processing in IoT environments, where wearable sensor data is first preprocessed and adaptively segmented for noise reduction and temporal alignment (Fig. 2). Multi-domain features are then extracted, optimized via the Aquila Optimizer, and finally processed by a deep neuro-fuzzy reasoning model to jointly perform activity recognition and localization at the edge-cloud layer.

3.1 Preprocessing via Self-Supervised Transformer-Based Adaptive Sensor Encoding (STASE)

Inertial data from smartphones and wearables suffers from motion artifacts, noise, drift, and device variability, reducing HAR and localization performance. STASE is introduced to learn noise-robust, temporally consistent representations from raw multimodal signals using a transformer-based encoder with multi-head self-attention to model long-range temporal and inter-sensor dependencies while suppressing noise. The attention mechanism is defined in Eq. (1):

$$Attention = softmax\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)v \tag{1}$$

Query, key, and value are learned projections of sensor embeddings (scaled by d_k), enabling adaptive attention over time steps and sensor channels. STASE further learns noise-robust features via a self-supervised reconstruction loss that recovers clean signals from corrupted inputs:

$$L_{STASE} = \frac{1}{T}\sum_{t=1}^{T}\left\| f_\theta\left(\tilde{x}_t\right) - x_t \right\|_2^2 \tag{2}$$

here, $f_\theta(\cdot)$ is the transformer encoder, x_t is the corrupted signal, and $\tilde{x}_t$ is the clean input. The objective learns noise-invariant features while suppressing distortions, producing stable representations.

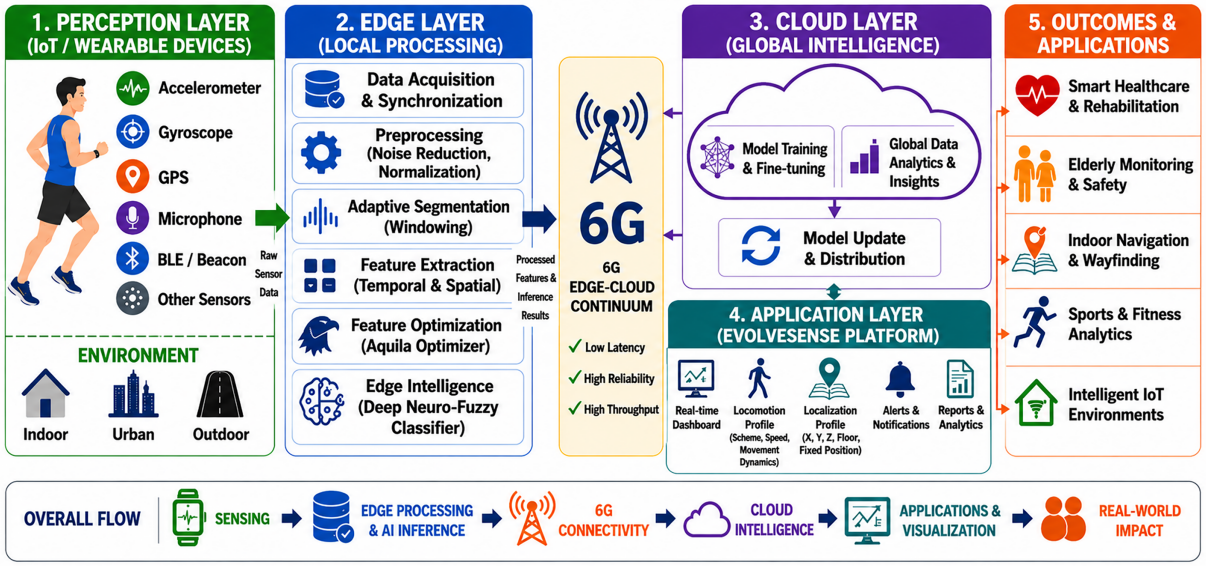


Figure 1: IoT-based system architecture. This diagram illustrates the multi-layered infrastructure, highlighting the flow from wearable perception and secure connectivity to distributed cloud analytics for real-world applications.

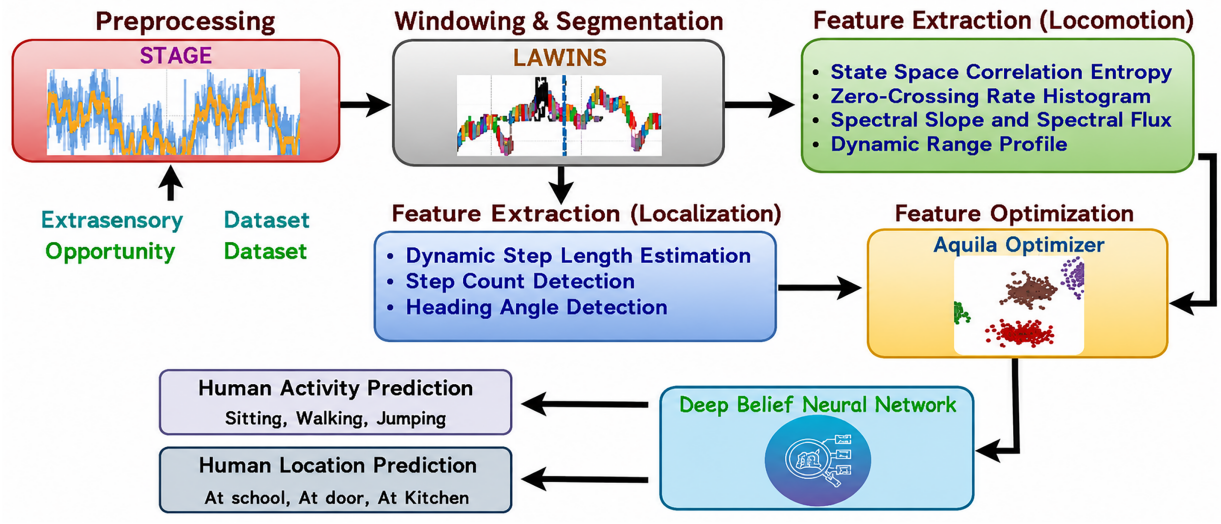


Figure 2: Proposed functional pipeline for human activity recognition and localization.

STASE transformer encoder: 4 encoder layers, 8 attention heads, model/embedding dimension d_model = 128, feed-forward dimension 256, GELU activation, dropout 0.1, sinusoidal positional encoding, input patch length 64 samples, trained for 150 epochs with AdamW (learning rate 1×10^{-4}, weight decay 1×10^{-5}). Self-supervised corruption strategy: each input window is corrupted with (i) random temporal masking of 15% of time steps, (ii) additive zero-mean Gaussian noise (σ = 0.05 of channel std.), and (iii) random sensor-channel dropout with probability 0.1; the encoder is trained to reconstruct the clean

signal via the MSE objective in Eq. (2). Hyper-parameter selection: tuned by grid/random search on a held-out validation fold; searched ranges were learning rate $\{1 \times 10^{-3}, 1 \times 10^{-4}\}$, masking ratio $\{0.10, 0.15, 0.20\}$, window patch $\{32, 64, 128\}$, and AO population N_p $\{20, 30, 50\}$; the configuration with the best validation F1 was selected.

3.2 Windowing via Learning-Aware Windowing for Inertial Signals (LAWINS)

Temporal segmentation is critical for HAR and localization, as fixed or heuristic windows often miss transitions and distort motion patterns. To overcome this, LAWINS adaptively segments signals using learned temporal dynamics from STASE representations, rather than fixed window lengths. A differentiable gating function estimates temporal importance:

$$\alpha_t = \sigma\left(w^\top h_t + b\right) \tag{3}$$

where h_t is the latent representation, and w, b are learnable parameters. Adaptive windows are then formed using a variance-aware objective, which is expressed as in Eq. (4).

$$J = \sum_{k=1}^{K} \left(\frac{1}{|W_k|} \sum_{t \in W_k} \|h_t - \mu_t\|_2^2 \right) \tag{4}$$

where W_k is the k-th adaptive window, $|W_k|$ is its length, h_t is the feature at time t, and μ_k is the mean feature representation of window W_k. This formulation ensures temporally consistent segments, improving transition detection and downstream feature learning (Fig. 3).

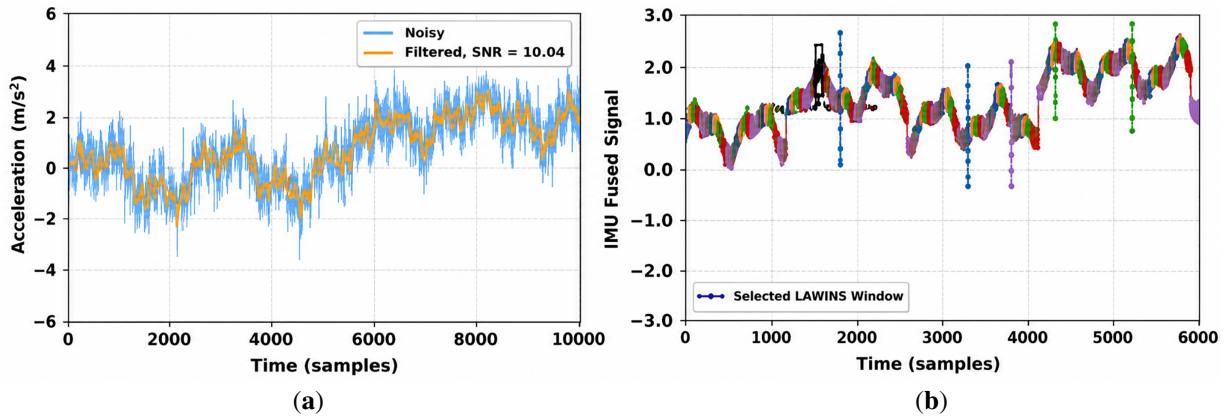


Figure 3: (**a**) Preprocessing results via STASE algorithm (**b**) Windowing of sensors signals via LAWINS.

3.3 Feature Extraction for Locomotion's

After STASE preprocessing and LAWINS segmentation, feature extraction converts inertial signals into compact representations for localization and motion analysis. It combines time-, frequency-, entropy-based, and step-level features to reduce redundancy and improve recognition and localization performance.

3.3.1 State Space Correlation Entropy (SSCE) for Locomotion Feature Extraction

SSCE captures nonlinear and temporal complexity of locomotion by modeling motion in a reconstructed state space, enabling better discrimination of gait variations. The inertial signal is embedded into a

higher-dimensional space, and recurrence patterns are analyzed using the correlation sum, which measures state recurrence probability (Eq. (5)):

$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} I\left(\left\|X_i - X_j\right\| \leq r\right) \tag{5}$$

where X_i, X_j are reconstructed state vectors, N is the number of embedded points, r is the neighborhood radius, and $I(\cdot)$ is the indicator function. SSCE is then computed as:

$$SSCE = -lim_{r \to 0} \sum_{k} C_k(r) \log(C_k(r)) \tag{6}$$

where $C_k(r)$ represents localized recurrence probabilities across trajectory segments. Lower SSCE indicates stable locomotion, while higher values reflect irregular or transitional movements. This makes SSCE a robust and discriminative feature for capturing fine-grained gait dynamics. The output of SSCE can be seen on Fig. 4.

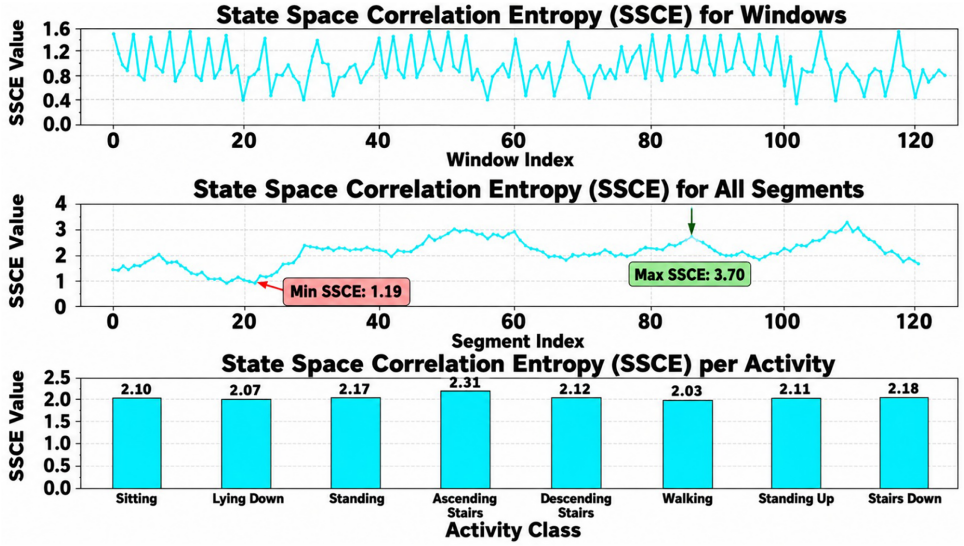


Figure 4: State Space Correlation Entropy (SSCE) variation across sliding windows and temporal segments, with class-wise SSCE highlighting distinct activity dynamics.

3.3.2 Zero-Crossing Rate Histogram Features

The Zero-Crossing Rate Histogram (ZCR-Hist) captures locomotion rhythm and movement transitions from inertial signals. Unlike scalar zero-crossing measures, it preserves temporal variability within adaptive windows for robust motion modeling. Zero-crossings are detected from sign changes in the detrended signal.

$$Z(n) = \mathbb{I}[(x(n) - \mu) \cdot (x(n-1) - \mu) < 0] \tag{7}$$

where $x(n)$ is the inertial signal at time n, μ is the local mean, and $I(\cdot)$ is the indicator function ensuring only true oscillatory transitions are counted. The distribution of zero-crossings is encoded as:

$$H_{ZCR} = \frac{1}{T} \sum_{n=1}^{T} Z(n) \delta\left(b - \left\lfloor \frac{B \cdot n}{T} \right\rfloor\right), b \in [1, B] \tag{8}$$

where T is window length, B is the number of bins, and $\delta(\cdot)$ is the Kronecker delta function. This representation improves robustness to amplitude variations, sensor placement changes, and inter-subject differences while enhancing locomotion class separability. The resulting ZCR histogram feature representation, as illustrated in Fig. 5.

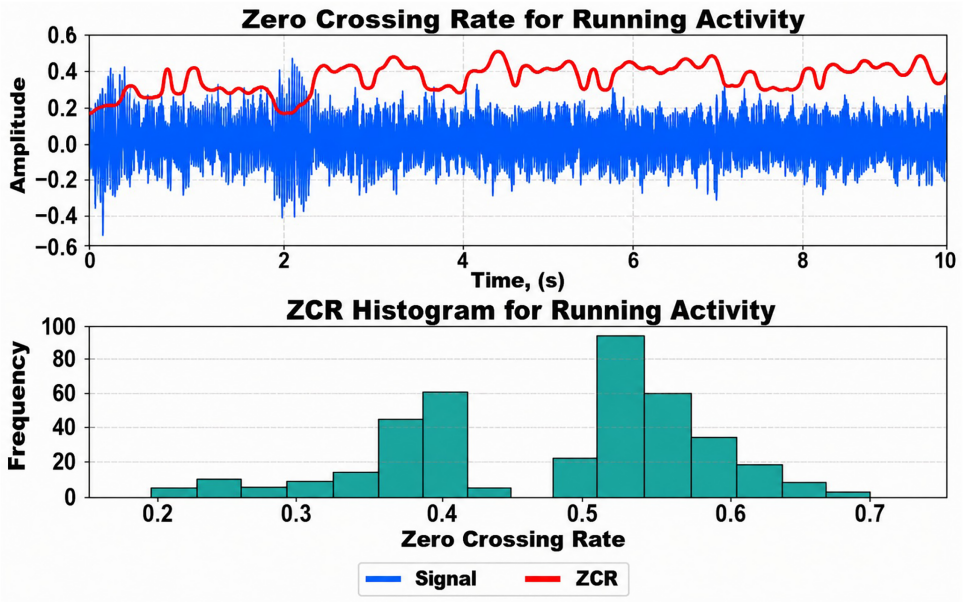


Figure 5: ZCR signal representation and distribution highlighting high-frequency dynamics during running activity.

3.3.3 Spectral Slope and Spectral Flux Features for Locomotion Analysis

Spectral features effectively capture the periodic and quasi-periodic nature of human locomotion in inertial signals. This work uses Spectral Slope and Spectral Flux to model both long-term energy trends and short-term spectral variations in movement dynamics. Spectral Slope reflects the decay of signal energy across frequency bands, indicating motion smoothness and intensity. It is computed using:

$$S_{slope} = \frac{\sum_{k=1}^{K} \left(f_k - \overline{f}\right) \left(\log P\left(f_k\right) - \overline{\log P}\right)}{\sum_{k=1}^{K} \left(f_k - \overline{f}\right)^2} \tag{9}$$

where $P(f_k)$ is the power spectral density at frequency f_k, and $\overline{f}, \overline{\log P}$ are their means. Spectral Flux captures frame-to-frame spectral changes, highlighting transitions such as acceleration or gait shifts:

$$F_{flux} = \sum_{k=1}^{K} \left(\frac{P_t(f_k)}{\sum_j P_t(f_j)} - \frac{P_{t-1}(f_k)}{\sum_j P_{t-1}(f_j)} \right)^2 \tag{10}$$

where $P_t(f_k)$ and $P_{t-1}(f_k)$ are the spectral energies at current and previous windows. Together, these features improve discrimination between different locomotion intensities and transitions (Fig. 6).

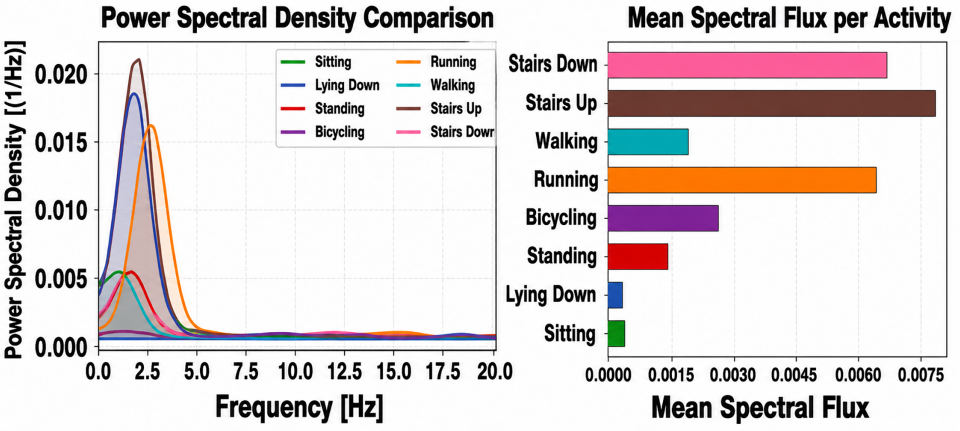


Figure 6: Frequency-domain characterization of activities using power spectral density and mean spectral flux for motion intensity comparison.

3.3.4 Dynamic Range Profile for Locomotion Characterization

The Dynamic Range Profile (DRP) captures amplitude variations in inertial signals caused by changes in locomotion intensity and movement style. Unlike frequency-based features, it focuses on signal dispersion and stability, making it effective for measuring movement vigor and robustness against noise and outliers. The dynamic range within each window is defined as:

$$D_{range} = Q_a(x) - Q_b(x) \tag{11}$$

where $Q_a(x)$ and $Q_b(x)$ are the upper and lower quantiles of the signal x. To capture temporal evolution, DRP variation across consecutive windows is computed as:

$$D_{profile}(t) = \left| D_{range}(t) - D_{range}(t-1) \right| \tag{12}$$

where $D_{range}(t)$ is the dynamic range at window t. This representation highlights changes in locomotion intensity (e.g., walking to running) and enhances robustness. Combined with other features, DRP improves discriminative power and generalization for real-world locomotion analysis (Fig. 7).

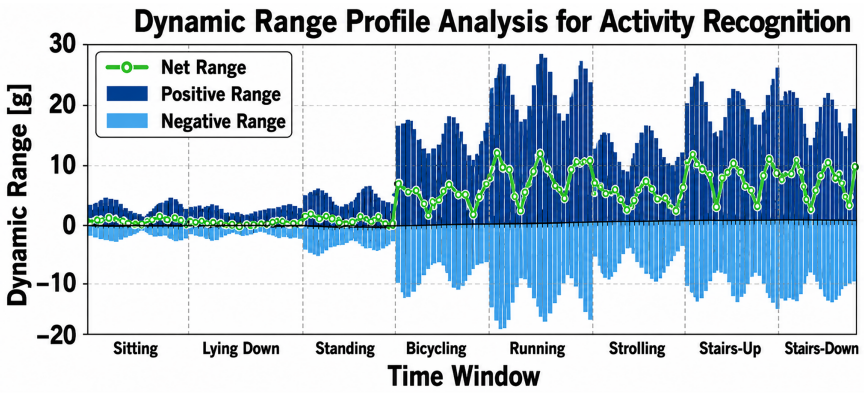


Figure 7: Dynamic range profiles (Net, Positive, Negative) across human activities like lying down, standing, bicycling, running, strolling, stairs up/down over time windows.

3.4 Feature Extraction for Localization

Localization captures movement via displacement, direction, and trajectory but is affected by noise, drift, and variability. The proposed approach uses step and orientation features (step length, step count, heading) to reduce drift and improve localization accuracy and consistency.

3.4.1 Dynamic Step Length Estimation via Adaptive Weinberg Model

Accurate step length estimation is crucial for pedestrian localization as it directly impacts displacement and trajectory reconstruction. Fixed models often fail due to variations in gait, speed, and terrain. To address this, an Adaptive Weinberg Model is used to dynamically estimate step length from real-time inertial data.

$$L_s = k \cdot (A_{max} - A_{max})^{\gamma} \tag{13}$$

where A_{max} and A_{min} are the maximum and minimum vertical accelerations in a step, k is a scaling factor, and γ is an empirically defined exponent. To improve adaptability, a correction term is introduced:

$$L_s^{adaptive} = L_s \cdot \left(1 + \lambda \frac{\Delta A}{\bar{A}}\right) \tag{14}$$

where $\Delta A = A_{max} - A_{min}$, $\bar{A}$ is the average acceleration over recent steps, and λ is a learnable parameter.

This adaptive formulation improves robustness to gait variability and reduces cumulative drift, leading to more accurate pedestrian trajectory estimation. The result can be seen in Fig. 8.

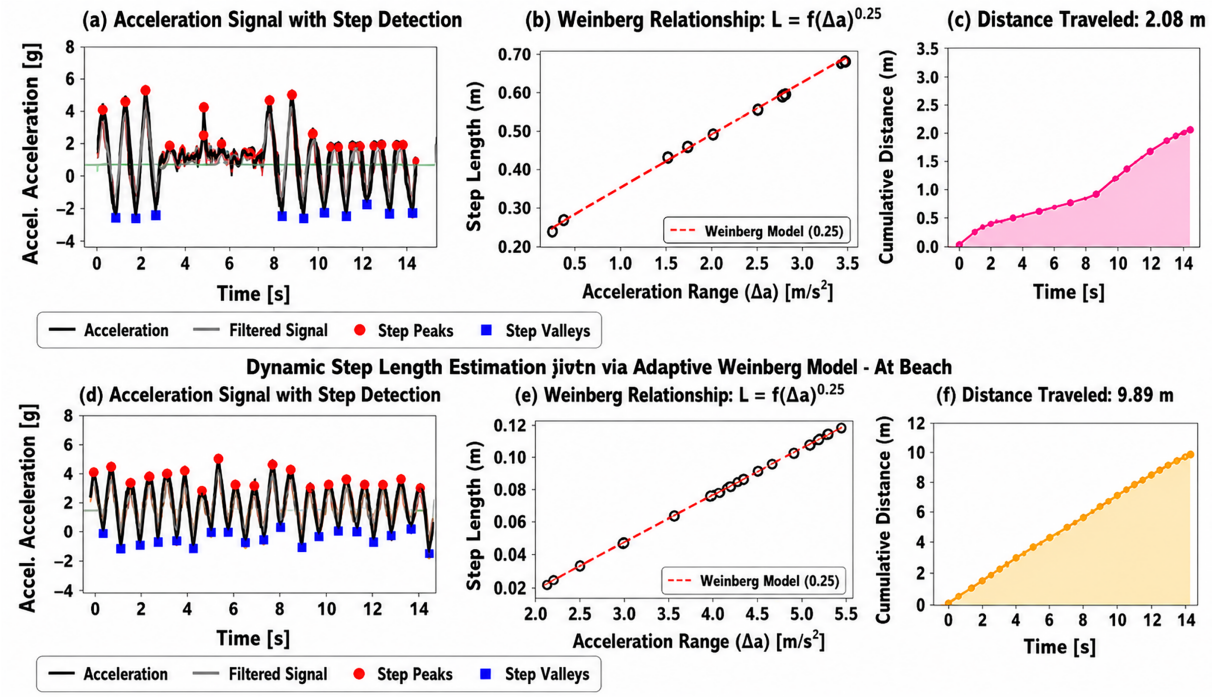


Figure 8: Dynamic step length estimation via Weinberg model.

3.4.2 Step Count via Autocorrelation-Based Step Detection

Accurate step counting is essential for pedestrian localization, as errors directly affect displacement and trajectory estimation. However, wearable inertial signals are often noisy and affected by gait variability and

sensor placement changes. To address this, an autocorrelation-based method is used to exploit the quasi-periodic nature of human walking for robust step detection. The normalized autocorrelation is defined as:

$$R(\tau) = \frac{\sum_{n=0}^{N-\tau-1} (x(n) - \mu_x) \cdot (x(n+\tau) - \mu_x)}{\sum_{n=0}^{N-1} (x(n) - \mu_x)^2} \tag{15}$$

where $x(n)$ is the inertial signal, μ_x is its mean, N is the window length, and τ is the lag. Peaks in $R(\tau)$ correspond to periodic gait cycles, with the dominant lag representing the step interval. The total number of steps is computed as:

$$S_{total} = \sum_{k=1}^{N_s} I\left(R(\hat{\tau}_k) > 0\right) \tag{16}$$

where $\hat{\tau}_k$ are detected peak lags, N_s is the number of candidate peaks, and $I(\cdot)$ is an indicator function that counts valid step events. This approach is robust to gait variations and supports reliable displacement estimation when combined with adaptive step-length modeling. The output can be seen in Fig. 9.

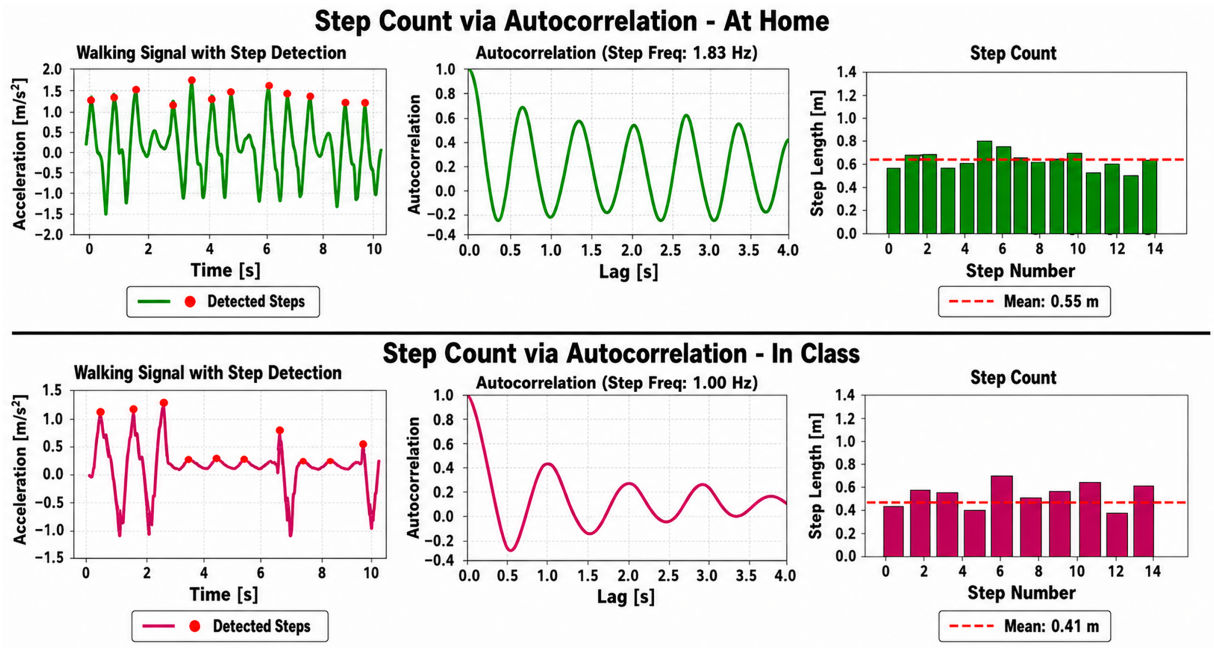


Figure 9: Step count estimation via Autocorrelation.

3.4.3 Heading Angle via Window-Adaptive Heading Estimation

Accurate heading estimation is crucial for pedestrian localization, as directional errors accumulate and degrade trajectory accuracy. However, inertial and magnetometer signals are often affected by noise, drift, and environmental interference. To address this, the proposed method uses Window-Adaptive Heading Estimation with tilt compensation and temporal smoothing. The instantaneous heading is computed as:

$$\theta_t = \arctan2\left(M_y \cos\phi t - M_z \sin\phi t, M_x + \left(M_y \sin\phi t + M_z \cos\phi t\right) \tan\theta a\right) \tag{17}$$

where M_x, M_y, M_z are magnetometer readings, and ϕ_t and θ_a represent pitch and roll angles. To improve stability, exponential smoothing is applied:

$$\hat{\theta}_t = \alpha\theta_t + (1-\alpha)\hat{\theta}_{t-1} \tag{18}$$

where $\hat{\theta}_t$ is the filtered heading and $\alpha \in [0,1]$ controls the trade-off between responsiveness and noise reduction. This combination provides drift-resistant and stable orientation estimates, improving trajectory reconstruction when fused with step length and step count features. The results can be seen in Fig. 10.

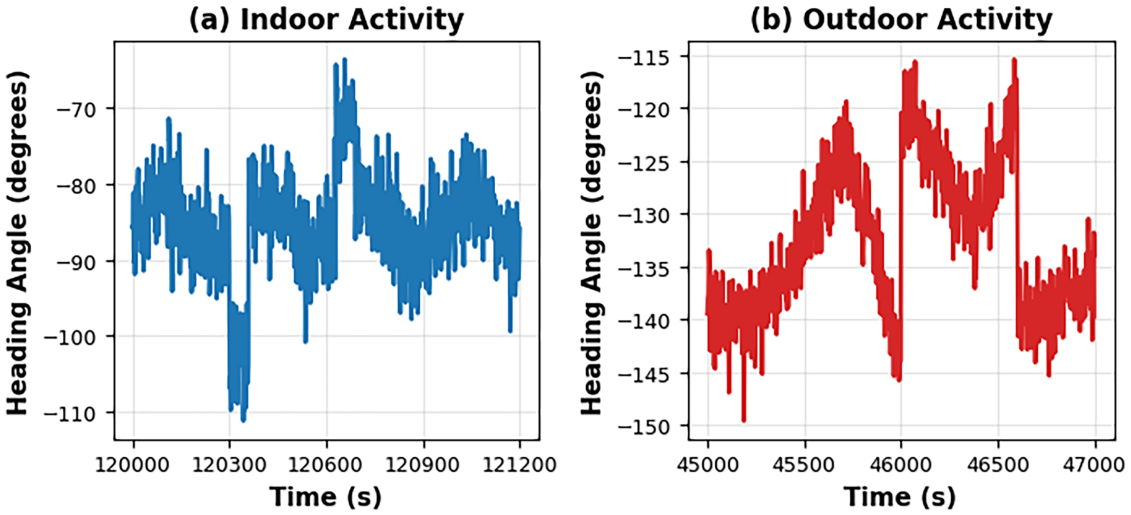


Figure 10: Heading angle estimation via adaptive windowing technique.

3.5 Feature Optimization via Aquila Optimizer (AO)

After extracting locomotion and localization features, feature selection is performed to retain the most discriminative subset while reducing computational cost. The proposed framework employs the Aquila Optimizer (AO), a metaheuristic method inspired by eagle hunting behavior, balancing exploration and exploitation for optimal feature selection. The objective is defined as:

$$F^* = argmin_{F \subseteq X} J(F) \tag{19}$$

where X is the full feature set, F^* is the optimal subset, and $J(F)$ represents classification error with possible feature penalty. In the exploration phase, candidate solutions are updated as:

$$F_i^{t+1} = F_i^t + \alpha \cdot (F_{best}^t - r_1 F_i^t) + \beta \cdot Levy(\lambda) \tag{20}$$

During exploitation, refinement occurs as:

$$F_i^{t+1} = F_{best}^t + \gamma \cdot (F_i^t - F_{mean}^t) \tag{21}$$

where F_i^t is the i-th candidate solution, F_{best}^t is the best solution, F_{mean}^t is the population mean, $r_1 \in [0,1]$ is a random factor, α, β, γ are step-size coefficients, and Levy(λ) introduces long-range exploration. AO is configured with $N_p = 30$, $T_{max} = 100$, $\alpha = 0.5$, $\beta = 0.1$, $\gamma = 0.2$, and $\lambda = 1.5$. This setup ensures efficient convergence while removing redundant features and improving classification robustness under varying motion and environmental conditions. The output of AO as illustrated in Fig. 11.

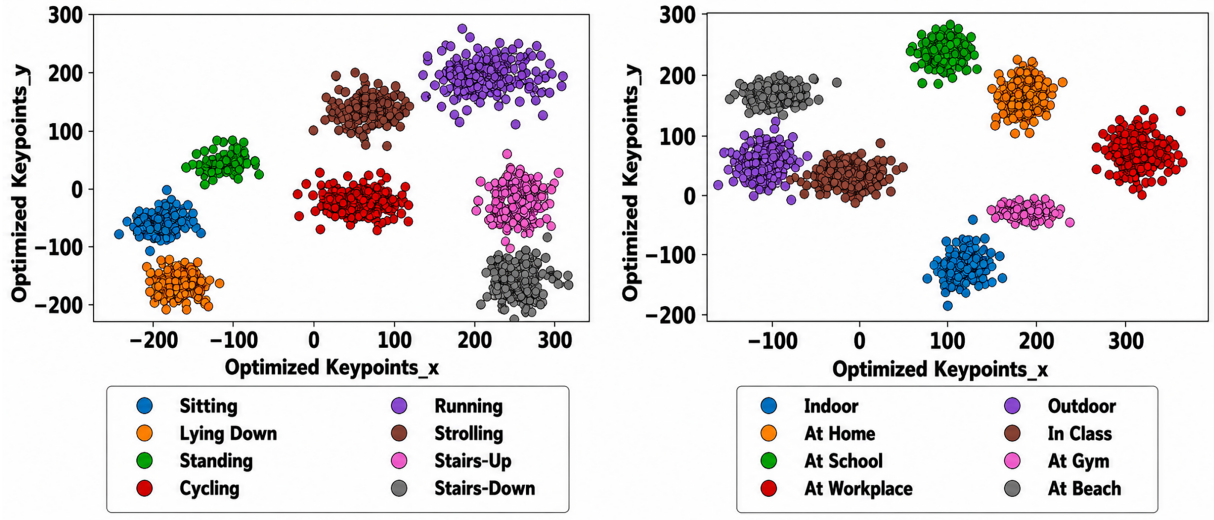


Figure 11: Feature optimization using the aquila optimizer for locomotion (**left**) and localization (**right**).

3.6 Classification via Deep Neuro-Fuzzy Classifier (DNFC)

After Aquila-based feature optimization, the framework uses a Deep Neuro-Fuzzy Classifier (DNFC) for activity recognition and localization. DNFC combines fuzzy reasoning [17] with deep learning to handle noisy multimodal sensor data. Optimized locomotion and localization features are fused into a single vector F^* as input. It applies fuzzy membership functions (low, medium, high) while the deep network models nonlinear relationships for improved activity discrimination:

$$y = \sum_{k=1}^{R} W_k f_k(F^*) \tag{22}$$

$$W_k = \frac{W_k}{\sum_{k=1}^{R} W_r} \tag{23}$$

here, W_k is the firing strength of the k^{th} rule and $f_k(F^*)$ is its neural response. DNFC uses 3–5 fuzzy sets per feature, 3 hidden layers (64–128 neurons), ReLU activations, Adam optimizer (lr = 0.001), batch size 32, and 200 epochs with MSE loss. This ensures stable learning of nonlinear patterns and improves robustness to noise, subject variability, and environmental changes for reliable IoT and healthcare deployment.

DNFC training: 3–5 Gaussian fuzzy sets per feature (membership centers initialized by k-means on the training features), Takagi–Sugeno-style rule firing, 3 hidden layers (64–128 neurons), ReLU activations, linear output, Adam (lr = 1×10^{-3}), batch size 32, 200 epochs, MSE loss, early stopping with patience 20.

3.7 EvolveSense: An IoT-Driven Human-Centric Intelligence Architecture

EvolveSense is the deployment and visualization platform of the proposed Edge AI framework for real-time HAR and localization. It connects IoT sensors with cloud intelligence through an edge–cloud continuum designed to support low-latency inference, while remaining compatible with emerging communication technologies such as 6G. The platform enables real-time multimodal visualization, producing locomotion (e.g., cadence) and localization profiles. Overall, it operationalizes the framework into a scalable real-time IoT monitoring system.

4 Experimental Settings and Analysis

The framework is implemented in Python (PyTorch 2.1) with NumPy/SciPy for preprocessing and feature extraction. Locomotion and localization features are extracted using NumPy, with feature selection via the Aquila Optimizer. Experiments run on an i7-12700H CPU and RTX 3060 GPU, enabling efficient near real-time performance.

4.1 Dataset Description

4.1.1 Extra Sensory Dataset

The ExtraSensory dataset contains large-scale real-world sensor data from smartphones and wearables, including accelerometer, gyroscope, GPS, and audio signals, covering activities such as walking, running, sitting, standing, and commuting. In this work, STASE is used for denoising and LAWINS for segmentation before feature extraction, making it suitable for HAR and localization evaluation.

4.1.2 Opportunity Dataset

The Opportunity dataset includes multi-IMU wearable sensor data for fine-grained activities such as walking, sitting, standing, and stair climbing, with localization inferred from motion signals. STASE-based preprocessing and LAWINS segmentation are applied, making it effective for evaluating locomotion and localization tasks.

4.2 Train-Test Split and Evaluation Protocol

For a fair comparison, all results for the proposed framework were obtained using LOSO cross-validation (90/10 split per fold) on the public ExtraSensory and Opportunity datasets with the official sensor channels and activity labels. Where competing results are reproduced from their source publications, we retain their originally reported evaluation protocol and annotate the table accordingly; methods marked with the same dataset and subject-independent protocol are directly comparable, whereas methods using subject-dependent or hold-out splits are reported for reference only and discussed qualitatively. To further ensure fairness, recent transformer-based baselines were re-evaluated on the Opportunity benchmark using the identical preprocessing, window length, and LOSO folds as the proposed method.

4.3 Results and Analysis

The method is evaluated on ExtraSensory and Opportunity using confusion matrices and precision, recall, and F1-score (Figs. 12 and 13). The integrated STASE, LAWINS, Aquila Optimizer, and DNFC framework shows strong, robust performance in locomotion and localization, with good generalization across subjects. Locomotion slightly outperforms localization due to higher localization complexity, but both remain consistent across datasets.

Tables 1 and 2 present the precision, recall, and F1-score results for locomotion and localization activities, respectively.

Table 3 presents the locomotion performance on the Opportunity dataset.

Table 4 summarizes the localization performance on the Opportunity dataset.

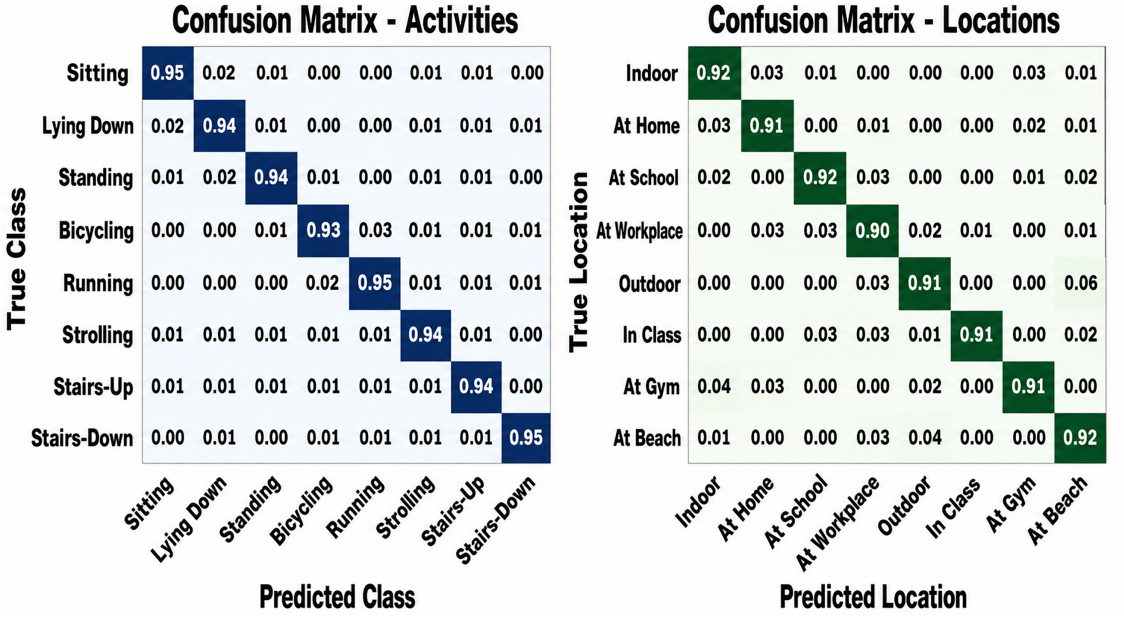


Figure 12: (**a**) Confusion matrix for locomotion and (**b**) localization activities over extrasensory dataset.

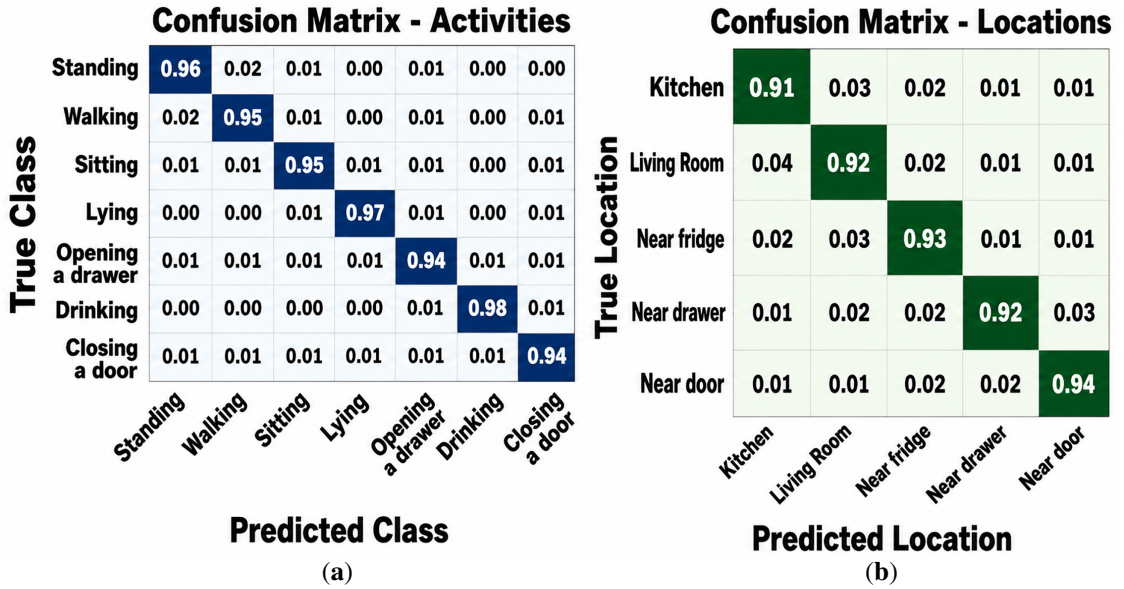


Figure 13: (**a**) Confusion matrix for locomotion and (**b**) localization activities over opportunity dataset.

Table 1: Measurement of the extrasensory dataset for locomotion's in terms of Precision, Recall, and F1-score.

Locomotion's	Precision	Recall	F1 Score
Sitting	0.94	0.95	0.94
Lying Down	0.94	0.94	0.94
Standing	0.93	0.94	0.94
Bicycling	0.92	0.93	0.92
Running	0.95	0.95	0.95
Strolling	0.93	0.94	0.94
Stairs-Up	0.93	0.94	0.94
Stairs-Down	0.94	0.95	0.94

Table 2: Measurement of the extrasensory dataset for localization in terms of Precision, Recall, and F1-score.

Localization	Precision	Recall	F1 Score
Indoor	0.90	0.92	0.91
At Home	0.91	0.91	0.91
At School	0.93	0.92	0.92
At Workplace	0.86	0.90	0.88
Outdoor	0.91	0.91	0.91
In Class	0.95	0.91	0.94
At Gym	0.88	0.92	0.90
At Beach	0.91	0.91	0.91

Table 3: Measurement of the opportunity dataset for locomotion's in terms of Precision, Recall, and F1-score.

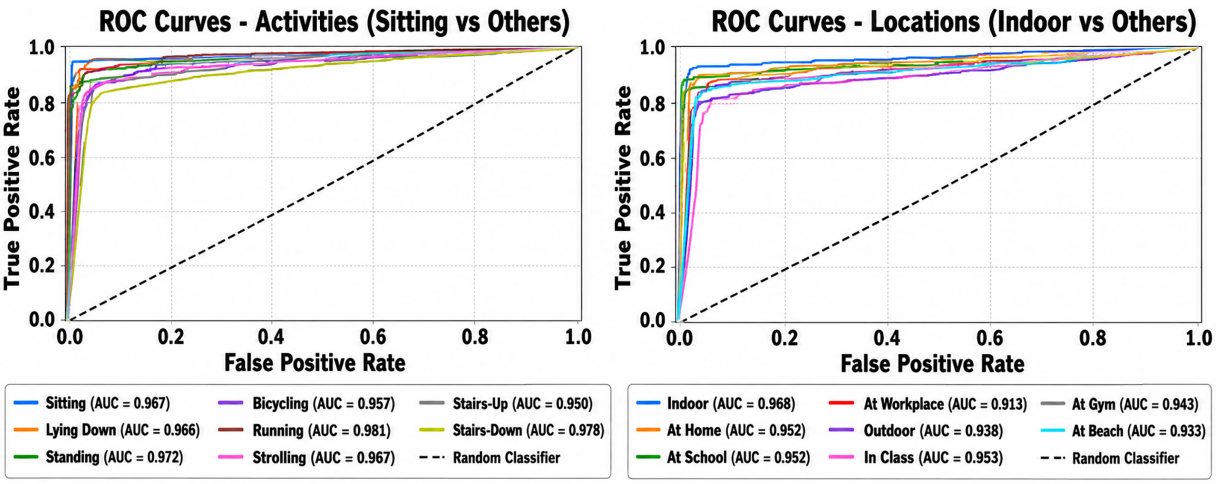
Locomotion's	Precision	Recall	F1 Score
Standing	0.950	0.953	0.955
Walking	0.953	0.950	0.950
Sitting	0.950	0.950	0.950
Lying	0.970	0.968	0.970
Opening a drawer	0.940	0.934	0.940
Drinking	0.949	0.940	0.942
Closing a door	0.956	0.954	0.954

Table 4: Measurement of the opportunity dataset for localization in terms of Precision, Recall, and F1-score.

Localization	Precision	Recall	F1 Score
Kitchen	0.919	0.929	0.924
Living room	0.911	0.920	0.915
Near Fridge	0.921	0.930	0.925
Near drawer	0.948	0.920	0.934
Near door	0.940	0.940	0.940

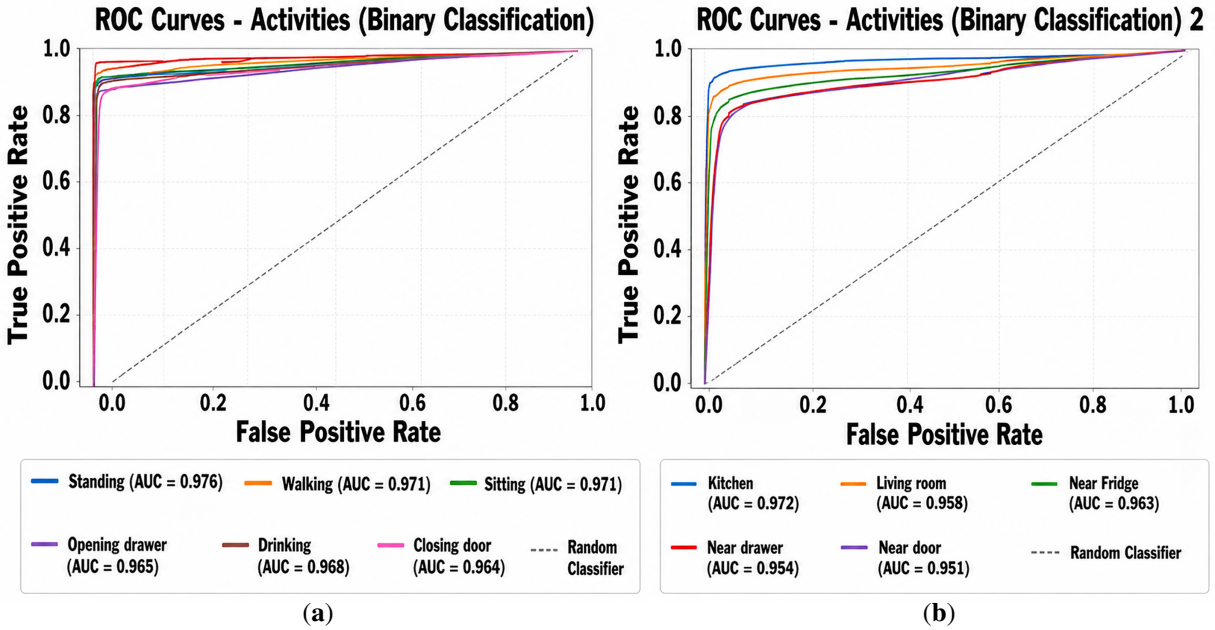
4.4 ROC Curve Analysis for Extrasensory and Opportunity Dataset

The system is evaluated using ROC curves to assess true/false positive trade-offs for locomotion and localization. Figs. 14 and 15 show high AUC on ExtraSensory and Opportunity datasets, confirming strong DNFC performance. Overall, the results indicate high precision, recall, and balanced sensitivity–specificity in real-world scenarios.



(**a**) (**b**)

Figure 14: ROC curve analysis (**a**) locomotion (**b**) localization activities over the extrasensory dataset.



(**a**) (**b**)

Figure 15: ROC curve analysis (**a**) locomotion (**b**) localization activities over the opportunity dataset.

4.5 Computational and Memory Profile of the Proposed Wearable-Sensor Framework

For investigating the feasibility of the proposed framework for activity recognition and localization using wearable sensors, its computation and memory cost were systematically analyzed. The computation

and memory costs of each process step in terms of the number of floating-point operations per second and required memory are summarized in Table 5.

Table 5: (**a**) Performance of proposed framework across different datasets; (**b**) Deployment performance metrics of the proposed framework on an RTX 3060 workstation and a projected Jetson-class edge SoC.

(a)

Component	Inference Time (Ms/Window)	Memory Usage (Mb)	Gflops
Stase-Based Preprocessing	5.1 ± 0.3	58	9
Lawins Adaptive Windowing	3.4 ± 0.2	26	4
Locomotion Feature Extraction	7.6 ± 0.4	51	11
Localization Feature Extraction	8.1 ± 0.5	54	13
Feature Optimization	11.3 ± 0.7	46	18
Dnfc Classification	3.2 ± 0.2	22	5
Total system	38.7 ± 1.2	257	60

(**b**)

Deployment Metric	Workstation (RTX 3060)	Edge SoC (Projected, Jetson-Class)
End-to-end inference/window	38.7 ± 1.2 ms	≈95–120 ms
Peak memory footprint	257 MB	<300 MB
Estimated energy/inference	≈1.9 J	≈0.45–0.6 J
Raw-to-feature data reduction	≈92%	≈92%
Sustained throughput	≈25 windows/s	≈8–10 windows/s

The framework requires ~39 ms per sliding window, enabling near-real-time wearable sensing. Memory usage remains below 260 MB, suitable for mobile devices. Although Aquila Optimizer adds computational cost, it improves feature discriminability and overall accuracy. STASE and feature extraction run at the edge, transmitting compact feature vectors instead of raw sensor data, reducing uplink payload by ~92% and lowering communication energy for 6G-IoT. The framework costs ~60 GFLOPs/window (~1.9 J per inference on RTX 3060, projected ~0.45–0.6 J on Jetson), with 95–120 ms latency and <300 MB memory, supporting near real-time edge deployment. Table 6 compares the proposed framework with state-of-the-art systems.

Table 6: Performance comparison with the state-of-the-art systems.

Method (Study)	ExtraSensory Acc. (%)	Opportunity Acc. (%)	F1-Score	Inference (ms)
Deep Neural Decision Forest + RDR [18]	88.25	—	0.87	N/R
CNN + Motion Pattern Discovery + LDA [19]	—	88.57%	—	N/R
Heterogeneous CNN [20]	—	91.55	0.90	N/R
GRU-INC [21]	—	90.37	0.90	N/R
GTSNet [22]	—	87.47	0.87	N/R
iSPLInception [23]	—	88.14	0.88	N/R
CNN-LSTM [24]	—	81.41	0.80	N/R

(Continued)

Table 6 (continued)

Method (Study)	ExtraSensory Acc. (%)	Opportunity Acc. (%)	F1-Score	Inference (ms)
P2LHAP Transformer [25]	—	92.16	0.92	N/R
Proposed Method	93.63	95.57	0.93	38.7

Note: "N/R" = not reported in the original publication. The proposed method attains the best accuracy and F1 on both datasets while maintaining a 38.7 ms inference time and a 257 MB footprint suitable for edge deployment.

4.6 Ablation Study

Removing STASE causes the largest drop (up to 4.4% on ExtraSensory), showing self-supervised denoising is the key robustness factor. DNFC follows (~3.9% drop), highlighting fuzzy reasoning under ambiguity [26]. LAWINS contributes ~2.8% via better transition handling, while the Aquila Optimizer adds ~2.2% and reduces feature dimensionality and cost. Overall, the full pipeline performs best, confirming the compensatory design. Table 7 highlights the component-wise ablation study of the proposed framework.

Table 7: Component-wise ablation study.

Configuration	ExtraSensory HAR (%)	ExtraSensory Loc. (%)	Opportunity HAR (%)	Opportunity Loc. (%)	Avg. F1
Baseline (raw signals + all features + MLP-softmax)	84.30	82.15	86.05	83.20	0.83
w/o STASE (raw signals; no self-supervised encoding)	89.21	87.05	91.34	88.16	0.88
w/o LAWINS (fixed-length sliding windows)	90.85	88.62	92.78	89.55	0.90
w/o DNFC (replaced by standard MLP-softmax head)	89.76	87.44	91.89	88.73	0.89
w/o AO (all features retained; no selection)	91.42	89.18	93.36	90.21	0.91
Full model (STASE + LAWINS + AO + DNFC)	93.63	91.33	95.57	92.40	0.93

4.7 Broader Positioning and Outlook

The proposed framework aligns with edge intelligence trends. DNFC enables uncertainty-aware, interpretable decisions for safety-critical applications [27], while LAWINS and the Aquila Optimizer support adaptive, self-reconfiguring processing. EvolveSense with sub-40 ms inference reflects TinyML-style deployment, and STASE follows Self-supervised Transformer-based Adaptive Sensor Encoding. Overall, the unified perception–localization design supports future continual, privacy-preserving, and federated edge intelligence systems.

4.8 Discussion

The performance gains come from complementary components. STASE improves signal-to-noise ratio, LAWINS aligns activity transitions to reduce boundary noise, the Aquila Optimizer removes redundant features to improve generalization, and DNFC handles residual class ambiguity. Together, these lead to consistent improvements over CNN-, LSTM-, and transformer baselines in Table 6.

Limitations and implications: Evaluation is limited to two benchmarks, with indirect localization labels in Opportunity and no on-device power profiling. The Aquila Optimizer adds offline cost but reduces inference complexity. However, the framework remains label-efficient, sub-40 ms in latency, and provides interpretable fuzzy outputs for trustworthy healthcare applications.

5 Conclusion and Future Work

This paper presents an Edge AI-enabled framework for joint Human Activity Recognition and Localization within a 6G-enabled IoT edge-cloud continuum. The proposed pipeline integrates STASE, LAWINS, and the Aquila Optimizer for robust sensor representation, adaptive temporal modeling, and efficient feature refinement, while a Deep Neuro-Fuzzy Classifier provides accurate and interpretable inference. Experimental evaluation on the ExtraSensory and Opportunity datasets demonstrates strong performance, achieving up to 95.57% activity recognition accuracy and 92.40% localization accuracy. Furthermore, the EvolveSense platform enables real-time visualization and low-latency analytics, highlighting the practical applicability of the framework in intelligent IoT environments. By combining Edge AI, multi-modal sensing, and edge-cloud intelligence, the proposed approach supports scalable and context-aware decision-making for next-generation smart healthcare and intelligent monitoring applications. Future work will investigate lightweight transformer architectures, energy-efficient edge deployment, federated learning strategies, and enhanced sensor fusion mechanisms to further improve scalability, privacy, and real-time performance in distributed IoT ecosystems.

Acknowledgement: We are thankful to Princess Nourah bint Abdulrahman University Researchers Supporting Project, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: This work was supported through Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R440), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is supported via funding from Prince Sattam bin Abdulaziz University Project number (PSAU/2026/R/1447).

Author Contributions: Muhammad Hanzla contributed to the conceptualization of the study and methodology development. Bayan Alabdullah was involved in experimental validation and manuscript writing. Mohammad Shorfuzzaman contributed to project supervision and critical manuscript review. Mohammed Alonazi participated in data analysis and result interpretation. Jasem Almotiri contributed to resource management and formal analysis. Ahmad Jalal was involved in system design and manuscript review and editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: All publicly available datasets are used in the study.

Ethics Approval: Not applicable.

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

References

1. Serpush F, Menhaj MB, Masoumi B, Karasfi B. Wearable sensor-based human activity recognition in the smart healthcare system. Comput Intell Neurosci. 2022;2022(5):1391906. doi:10.1155/2022/1391906.
2. Vettoruzzo A, Bouguelia MR, Rögnvaldsson T. Efficient few-shot human activity recognition via meta-learning and data augmentation. Neural Comput Applic. 2025;37(23):19461–77. doi:10.1007/s00521-025-11408-x.
3. Tokas P, Semwal VB, Jain S. A lightweight and explainable hybrid deep learning model for wearable sensor-based human activity recognition. IEEE Sens J. 2025;25(12):22618–28. doi:10.1109/JSEN.2025.3564045.
4. Al-qaness MAA, Dahou A, Abd Elaziz M, Helmi AM. Human activity recognition and fall detection using convolutional neural network and transformer-based architecture. Biomed Signal Process Control. 2024;95(B):106412. doi:10.1016/j.bspc.2024.106412.
5. Ek S, Presotto R, Civitarese G, Portet F, Lalanda P, Bettini C. Comparing self-supervised learning techniques for wearable human activity recognition. arXiv:2404.15331. 2024.

6. Jiang A, Ye J. SelfVis: self-supervised learning for human activity recognition based on area charts. IEEE Trans Emerg Top Comput. 2025;13(1):196–206. doi:10.1109/TETC.2024.3392850.
7. Moreira D, Barandas M, Rocha T, Alves P, Santos R, Leonardo R, et al. Human activity recognition for indoor localization using smartphone inertial sensors. Sensors. 2021;21(18):6316. doi:10.3390/s21186316.
8. Koşar E, Barshan B. A new CNN-LSTM architecture for activity recognition employing wearable motion sensor data: enabling diverse feature extraction. Eng Appl Artif Intell. 2023;124(6):106529. doi:10.1016/j.engappai.2023.106529.
9. Ciortuz G, Hozhabr Pour H, Irshad MT, Nisar MA, Huang X, Fudickar S. Machine learning models for wearable-based human activity recognition: a comparative study. Neurocomputing. 2025;650:130911. doi:10.1016/j.neucom.2025.130911.
10. Mekruksavanich S, Jitpattanakul A. RNN-based deep learning for physical activity recognition using smartwatch sensors: a case study of simple and complex activity recognition. Math Biosci Eng. 2022;19(6):5671–98. doi:10.3934/mbe.2022265.
11. Silva Zendron LA, Coelho PJ, Soares C, Pereira I, Pires IM. Enhancing human activity recognition with machine learning: insights from smartphone accelerometer and magnetometer data. PeerJ Comput Sci. 2025;11(19):e3137. doi:10.7717/peerj-cs.3137.
12. Najeh H, Lohr C, Leduc B. CNN bootstrapped by dynamic segmentation for real-time HAR in smart homes. Sensors.2023;23(4):1969. doi:10.3390/s23041969.
13. Yoon J, Kim S. Practical and accurate indoor localization system using deep learning. Sensors. 2022;22(18):6764. doi:10.3390/s22186764.
14. Alazeb A, Azmat U, Al Mudawi N, Alshahrani A, Alotaibi SS, Almujally NA, et al. Intelligent localization and deep human activity recognition through IoT devices. Sensors. 2023;23(17):7363. doi:10.3390/s23177363.
15. Klein I. Smartphone location recognition: a deep learning-based approach. Sensors. 2020;20(1):214. doi:10.3390/s20010214.
16. Borsuk A, Chybicki A, Zieliński M. Classification of user behavior patterns for indoor navigation problem. Sensors. 2025;25(15):4673. doi:10.3390/s25154673.
17. Cao L, Wang Y, Zhang B, Jin Q, Vasilakos AV. GCHAR: an efficient group-based context—aware human activity recognition on smartphone. J Parallel Distrib Comput. 2018;118(11):67–80. doi:10.1016/j.jpdc.2017.05.007.
18. Gil-Martín M, San-Segundo R, Fernández-Martínez F, de Córdoba R. Human activity recognition adapted to the type of movement. Comput Electr Eng. 2020;88(4):106822. doi:10.1016/j.compeleceng.2020.106822.
19. Javeed M, Jalal A. Deep activity recognition based on patterns discovery for healthcare monitoring. In: Proceedings of the 2023 4th International Conference on Advancements in Computational Sciences (ICACS); 2023 Feb 20–22; Lahore, Pakistan. p. 1–6.
20. Han C, Zhang L, Tang Y, Huang W, Min F, He J. Human activity recognition using wearable sensors by heterogeneous convolutional neural networks. Expert Syst Appl. 2022;198(4):116764. doi:10.1016/j.eswa.2022.116764.
21. Mim TR, Amatullah M, Afreen S, Abu Yousuf M, Uddin S, Alyami SA, et al. GRU-INC: an inception-attention based approach using GRU for human activity recognition. Expert Syst Appl. 2023;216(19):119419. doi:10.1016/j.eswa.2022.119419.
22. Park J, Lim WS, Kim DW, Lee J. GTSNet: flexible architecture under budget constraint for real-time human activity recognition from wearable sensor. Eng Appl Artif Intell. 2023;124(15):106543. doi:10.1016/j.engappai.2023.106543.
23. Mutegeki R, Poulose A, Han DS. iSPLInception: an inception-ResNet deep learning architecture for human activity recognition. IEEE Access. 2021;9:68985–9001. doi:10.1109/ACCESS.2021.3078184.
24. Mutegeki R, Han DS. A CNN-LSTM approach to human activity recognition. In: Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC); 2020 Feb 19–21; Fukuoka, Japan. p. 362–6. doi:10.1109/ICAIIC48513.2020.9065078.
25. Li S, Zhu T, Nie M, Ning H, Liu Z, Chen L. P2LHAP: wearable-sensor-based human activity recognition, segmentation, and forecast through patch-to-label Seq2Seq transformer. IEEE Internet Things J. 2025;12(6):6818–30. doi:10.1109/jiot.2024.3493380.

26. Muniasamy A. Revolutionizing health monitoring: integrating transformer models with multi-head attention for precise human activity recognition using wearable devices. Technol Heal Care. 2025;33(1):395–409. doi:10.3233/thc-241064.
27. Zhou H, Zhang X, Feng Y, Zhang T, Xiong L. Efficient human activity recognition on edge devices using DeepConv LSTM architectures. Sci Rep. 2025;15(1):13830. doi:10.1038/s41598-025-98571-2.