Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.064601

ARTICLE

A Hybrid Deep Learning Pipeline for Wearable Sensors-Based Human Activity Recognition

Asaad Algarni[1], Iqra Aijaz Abro[2], Mohammed Alshehri[3], Yahya AlQahtani[4], Abdulmonem Alshahrani[4] and Hui Liu[5,*]

[1]Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha, 91911, Saudi Arabia
[2]Faculty of Computing and AI, Air University, Islamabad, 44000, Pakistan
[3]Department of Computer Science, King Khalid University, Abha, 61421, Saudi Arabia
[4]Department of Informatics and Computer Systems, King Khalid University, Abha, 61421, Saudi Arabia
[5]Cognitive Systems Lab, University of Bremen, Bremen, 28359, Germany

*Corresponding Author: Hui Liu. Email: hui.liu@uni-bremen.de

Received: 19 February 2025; Accepted: 26 May 2025; Published: 30 July 2025

ABSTRACT: Inertial Sensor-based Daily Activity Recognition (IS-DAR) requires adaptable, data-efficient methods for effective multi-sensor use. This study presents an advanced detection system using body-worn sensors to accurately recognize activities. A structured pipeline enhances IS-DAR by applying signal preprocessing, feature extraction and optimization, followed by classification. Before segmentation, a Chebyshev filter removes noise, and Blackman windowing improves signal representation. Discriminative features—Gaussian Mixture Model (GMM) with Mel-Frequency Cepstral Coefficients (MFCC), spectral entropy, quaternion-based features, and Gammatone Cepstral Coefficients (GCC)—are fused to expand the feature space. Unlike existing approaches, the proposed IS-DAR system uniquely integrates diverse handcrafted features using a novel fusion strategy combined with Bayesian-based optimization, enabling a more accurate and generalized activity recognition. The key contribution lies in the joint optimization and fusion of features via Bayesian-based subset selection, resulting in a compact and highly discriminative feature representation. These features are then fed into a Convolutional Neural Network (CNN) to effectively detect spatial-temporal patterns in activity signals. Testing on two public datasets—IM-WSHA and ENABL3S—achieved accuracy levels of 93.0% and 92.0%, respectively. The integration of advanced feature extraction methods with fusion and optimization techniques significantly enhanced detection performance, surpassing traditional methods. The obtained results establish the effectiveness of the proposed IS-DAR system for deployment in real-world activity recognition applications.

KEYWORDS: Wearable sensors; deep learning; pattern recognition; feature extraction

1 Introduction

Precise tracking of daily human movement is vital for improving healthcare, disability treatment, safety, and exercise monitoring [1]. Wearable sensor advances enable IS-DAR systems to monitor motion in real time using accelerometers, gyroscopes, and magnetometers. These sensors provide detailed motion data but still face challenges in accurately classifying complex activities like walking, jogging, jumping, and falling.

Existing methods have drawbacks: video-based systems struggle with lighting and occlusion; marker-based systems require controlled environments; and inertial systems, though portable, often produce noisy data and generalize poorly. Many studies also rely on limited sensor fusion and basic feature selection, underusing available data.

This study proposes a robust IS-DAR system using body-mounted inertial sensors and a noise-resistant signal processing pipeline. Signals are denoised with a sixth-order Chebyshev filter and segmented via Blackman windows. It extracts and fuses a rich set of features—GMM, MFCC, spectral entropy, quaternion-based, and GCC—then applies Bayesian feature selection. A CNN classifies the fused input by learning spatial-temporal patterns.

Tested on the IM-WSHA and ENABL3S datasets, the system achieves 93.0% and 92.0% accuracy, outperforming traditional methods. Its integration of advanced filtering, feature fusion, and deep learning enhances recognition accuracy and efficiency. This research delivers six key innovations:

- We present a systematic deep learning framework integrating advanced feature extraction, fusion, and Bayesian optimization to enhance IS-DAR robustness. This structured approach leverages both domain-specific handcrafted features and deep learning-based representations for superior classification performance.
- While the individual features (MFCC, GCC, spectral entropy, quaternion orientation, and GMM) are well-known in Human Activity Recognition (HAR) literature, the novelty of this work stems from the systematic integration and cross-domain fusion of these complementary modalities into a single feature space, optimized through Bayesian-driven selection. This process ensures that only the most informative and noise-robust features contribute to classification, mitigating redundancy and enhancing model generalization.
- Although all utilized features have precedent in prior works, our contribution lies in the cross-modal feature synergy achieved via Bayesian Optimization-driven selection and fusion, offering a compact and high-utility representation that directly enhances CNN-based classification.
- A comparative analysis compares Bayesian Optimization for discriminative feature selection and hyper-parameter tuning to traditional methods such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD). The results show Bayesian Optimization's better ability to choose discriminatory features and optimize classification performance.
- Our research enhances deep learning-based IS-DAR by maximizing the collaboration of feature fusion and Bayesian Optimization in a CNN classifier. Model robustness to sensor variability and noise is greatly enhanced by this systematic approach.
- Intensive experimentation over the IM-WSHA and ENABL3S datasets establishes the efficacy of our system through state-of-the-art recognition accuracy. Comparative studies on newer HAR models affirm the superiority of our approach, establishing it as more practically viable and computationally efficient.

This paper is structured as follows: Section 2 presents the datasets; Section 3 discusses related work; Section 4 describes the planned methodology; Section 5 provides details of the experimental setup; Section 6 presents experimental results; and Section 7 concludes with future directions.

2 Datasets Description

2.1 IM-WSHA

Real-time daily activities were recorded using three inertial measurement units (IMUs) which were positioned on the chest the thigh along with the wrist for data collection on the IM-WSHA dataset [2]. This dataset contains data about the kinematic and static movement activities of ten subjects who performed their actions in a smart home setting. Roughly two hundred everyday activities were recorded which included phone conversation, vacuum cleaning, watching TV, using computer, reading books, ironing, walking, exercise, cooking, drinking, and brushing hair.

2.2 ENABL3S

The ENABL3S dataset involved motion capture sensors that included four IMUs on both wrists and shanks as well as four Electromyography (EMG) sensors attached to biceps and thighs. Fifteen participants carried out repeated five daily activities as part of the collection process. The recorded activities include standing still, squatting and standing up, jumping, raising right hand, and jogging.

3 Literature Review

Different methods exist for recognizing daily activities which utilize video-based sensors, body-worn markers and inertial measurement units (IMUs). This section reviews current methodologies from the three categories by analyzing their advantages and disadvantages for activity recognition:

3.1 Video-Based Activity Recognition

Ko et al. [3] built a video-based Human Activity Recognition (HAR) system that utilizes angle inclination and a keypoint descriptor network to identify movement direction. It performed well on lightweight devices, while issues such as pose variability, motion complexity, and occlusion persist. Kang and Wildes [4] came up with a strategy to identify biological activity by processing both positional and oscillatory motion in videos. Its system separated humans from non-human movements well in controlled environments, though dynamic backgrounds caused it to fail and demand high-quality video feed. Hassan et al. [5] developed a HAR system that integrates DenseNet121 for feature extraction and optimized Long Short-Term Memory (LSTM) to identify patterns over time. Although it surpassed current models in benchmark databases, it underscored the challenge of modeling dependencies on time and emphasized the importance of generic HAR systems. Jatesiktat et al. [6] applied deep learning to classify walking and running through postural features improved by using temporal filtering to deal with transitioning motion. Although promising in sport and rehabilitation, its application in complicated scenarios requires further examination. Finally, Kamble and Bichkar [7] built a two-tier HAR system involving the Hidden Markov Model (HMM) for rough modeling of activities and the Support Vector Machine (SVM) for detailed identification. They enhanced similar action identification but still had problems like noise, occlusion, and interference between different subjects.

3.2 Body-Worn Marker-Based Activity Recognition

Khan et al. [8] developed a marker-based motion tracking system for home rehabilitation, enabling real-time joint movement analysis. It improved therapeutic decisions by precisely monitoring joint location and rotation, though it required accurate marker placement and was limited to rehab contexts. Wickramarachchi [9] proposed a system with a 6D skeletal model to detect abnormal gait using raw marker data and multilayer perceptron classifiers. It achieved high accuracy but focused solely on pathological gait detection. Mekruksavanich et al. [10] combined CNN and LSTM models to classify exercise activities using data from both

IMUs and MOCAP markers. Both sensors delivered similar accuracy, supporting their interchangeability, though the system required high-quality input and was specific to exercise recognition. Wang et al. [11] validated the Opti_Knee system for tracking knee kinematics during walking and flexion-extension tasks. Its output closely matched biplanar fluoroscopy, making it useful for clinical gait analysis, despite some accuracy loss from skin movement artifacts. Niemann et al. [12] introduced a marker-based HAR system for logistics, integrating motion data with environmental context for better classification. While effective, it depended on precise marker placement and controlled conditions. Lastly, Lee and Park [13] offered a low-cost alternative using wearable skin markers (WSMs) detectable by standard webcams. Ideal for home use, the system reduced equipment needs but was sensitive to lighting and camera setup.

3.3 Inertial Sensor-Based Activity Recognition

Khan et al. [14] developed a wearable inertial sensor system using accelerometers and gyroscopes for activity and localization recognition. Their LSTM-based model, supported by noise filtering and feature extraction, outperformed existing methods across multiple datasets. Du et al. [15], a fuzzy-logic-based HAR system was introduced for hip exoskeletons, using onboard inertial sensors to detect hip joint angle extremes. It achieved 92.46% accuracy in intrasubject tests and 93.16% in intersubject validation, all without needing extra sensors. Sarcevic [16] used inertial data with a feature aggregation and genetic algorithm optimization approach, enhancing recognition speed and accuracy via dimensionality reduction. Shin et al. [17] designed a GMM-based locomotion recognition system using IMUs to classify five terrain types. It accurately captured walking conditions and linked performance to gait speed and rhythm, supporting robotic exoskeleton intent detection. Nouriani et al. [18] built a real-time HAR system with a chest-mounted IMU to monitor postural transitions, especially useful for Parkinson's patients. However, its effectiveness was limited by rigid sensor placement and user-specific motion transfer requirements. Celik et al. [19] proposed a CNN model that turned time-series inertial data into images to improve recognition in neurological populations. Although it handled class imbalance and dataset size issues, it was subject to lengthy training periods and had difficulty in generalizing across a limited pool of participants. Finally, Trabelsi et al. [20] compared deep learning models on six public databases. They obtained the best performance in a wavelet transforms–2D CNN combination, although performance was reduced on datasets such as WISDM and PAMAP2, highlighting the difficulty in generalizing HAR systems across varying setups.

4 Proposed System Methodology

The system that is proposed employs a hybrid method that blends hand-designed feature extraction and deep learning-based categorization by six consecutive stages. Preprocessing raw data from the IM-WSHA and ENABL3S databases is carried out using a Chebyshev filter followed by segmentation using a Blackman window. Features such as GMM, MFCC, GCC, spectral entropy, and quaternion-oriented features are independently extracted, then fused into a single representation. Bayesian Optimization selects relevant features and tunes hyperparameters. A Convolutional Neural Network (CNN) processes optimized features to capture spatial-temporal dependencies. This hybrid framework, integrating both handcrafted and learned features, improves generalization and robustness. Classification results are evaluated to verify performance, with all features computed in parallel and fused before classification, as shown in Fig. 1.

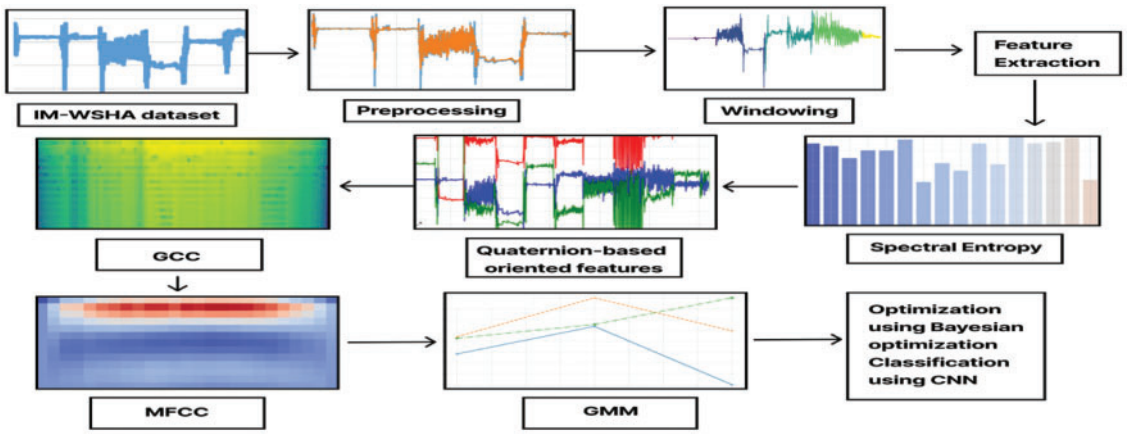

Figure 1: A comprehensive overview of the proposed IS-DAR system

4.1 Preprocessing

Before feature extraction, sixth-order Type I Chebyshev filtering methods are traditionally used to preprocess sensor data to increase the quality of the signals. These filters were selected based on their steep roll-off of the gains that efficiently reject high-frequency noise and maintain the important signal components. They also ensure that there is a predictable gain level in the passband, ensuring that there is little distortion in the passband to preserve the quality of the extracted features [14]. An imposed passband ripple of 0.5 dB safeguarded the biomedical signals without squandering too much power on high-frequency noise. Independent filtering was used for individual sensor channels to optimize the quality of the signals before further processing. Filtering by this method also reduces the effects of environmental interference and motion artifacts, ensuring that important motion characteristics are preserved in their true form for successful classification [21]. Sixth-order Type I Chebyshev filtering frequency response can be expressed by this mathematical equation:

$$|H(f)| = \frac{1}{\sqrt{1+ \epsilon^2 \; T_6^2\left(\frac{f}{f_c}\right)}} \tag{1}$$

where $H(f)$ is the filter gain at frequency f, ϵ represents the ripple factor, T_6 is the sixth order chebyshev polynomial, and f_c shows the cutoff frequency. Fig. 2 demonstrates the effects of the preprocessing procedure by contrasting raw and filtered signals. Filtering greatly enhances the quality of signals, making it possible for follow-up processing stages to use them as clean and stable data. Through this, the system attains improved classification performance and consistency in recognizing activities.

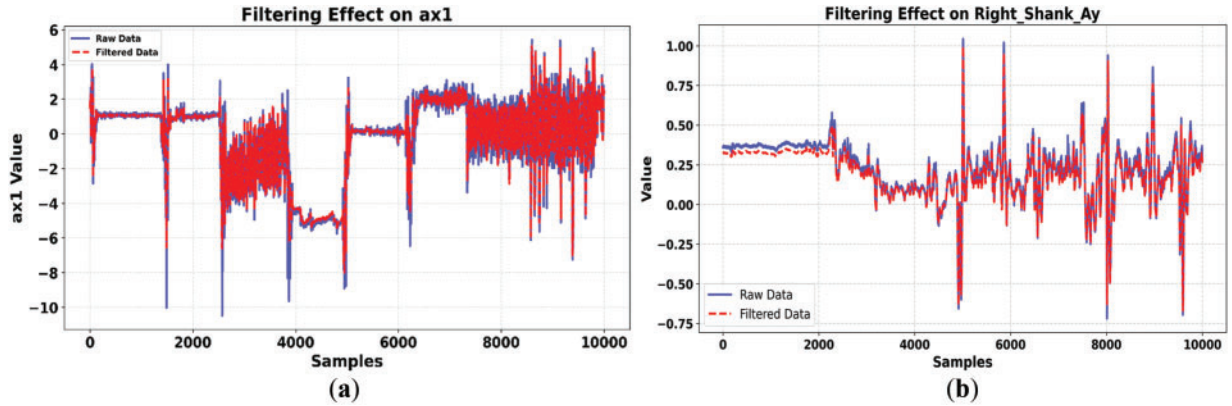

Figure 2: Comparison of raw and filtered signals using a sixth-order Chebyshev Type I filter for noise suppression in (**a**) IM-WSHA and (**b**) ENABL3S datasets

4.2 Windowing

The Blackman windowing technique was applied to segment sensor data by smoothing transitions at segment edges, reducing spectral leakage and improving frequency-domain precision. This function was applied across all sensor channels before feature extraction to preserve essential motion characteristics. Mathematically, the Blackman window is defined as:

$$W_k = 0.42 - 0.5\cos\left(\frac{2\pi k}{M-1}\right) + 0.08\cos\left(\frac{4\pi k}{M-1}\right) \tag{2}$$

where W_k is the window function at sample k, M is the total number of samples in a window, and k ranges from 0 to $M - 1$. A fixed window size, determined by activity duration, was used with overlap to maintain continuity between segments. Each data point was weighted using the corresponding window coefficient, producing a smoothed signal. As shown in Fig. 3, distinct window segments are visualized with different colors.

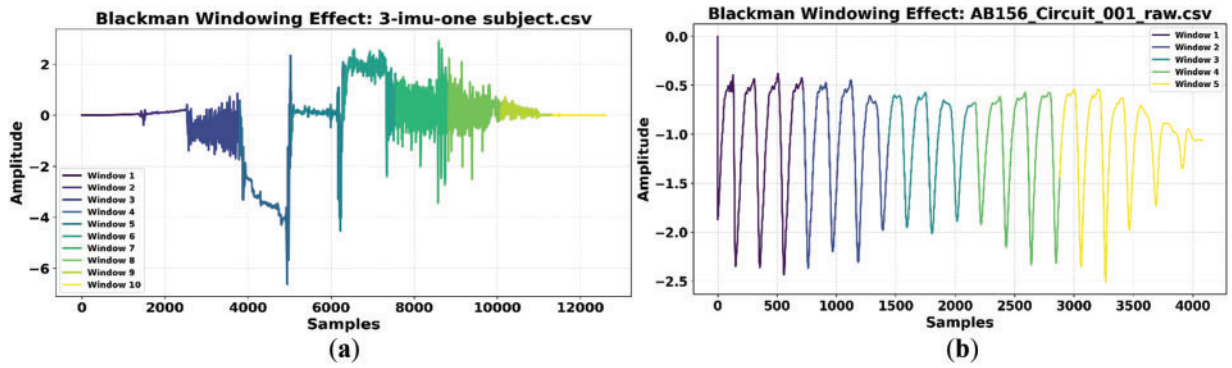

Figure 3: Windowed signal segments using Blackman windows to enhance feature extraction for (**a**) IM-WSHA and (**b**) ENABL3S datasets

4.3 Feature Extraction

The proposed system extracts several key features from inertial sensor data following the Blackman windowing process. The extracted feature set includes GMM coefficients, MFCC, spectral entropy, quaternion-oriented features, and GCC. These qualities encompass both the time-frequency domain features and the motion patterns in space that are critical for the correct identification of activities. These features were chosen based on the following rationale: GMM coefficients describe motion distribution, enabling the system to recognize variability among activities that involve similar motion. MFCC extracts frequency-domain features that are critical in recognizing activities by spectral patterns. Spectral entropy describes the complexity of signals that is useful in differentiating between dynamic and static activities. Quaternion-based features assist in retaining 3D orientation and rotational dynamics, which are important in correctly identifying activities where complex body motions exist. Lastly, GCC improves the frequency description to make it less susceptible to activities that involve coordinated motion across several sensor channels. The following subsections describe the extraction of the features and evaluation of these in the proposed system.

4.3.1 Gaussian Mixture Model Coefficients

The Gaussian Mixture Model (GMM) was applied to extract statistical features of preprocessed inertial sensor data, i.e., cluster means, weight vectors, and covariance matrices (N = 3). These parameters represent distributions of activity signals and depict changes in important movements. Mean vectors are estimated through Maximum Likelihood Estimation, weight vectors are optimized in terms of maximizing probability, and covariance is used to reflect dispersion of signals in order to differentiate activities. The GMM formulas are:

$$\mu_i = \frac{1}{M}\sum_{j=1}^{M} xj, w_i = \frac{1}{N}\sum_{j=1}^{M} P\left(x_j|i\right), \sum_i \frac{1}{M}\sum_{j=1}^{M}(x_j - \mu_i)^2 \tag{3}$$

where x_j represents the sensor signal sensor, $P(x_j|i)$ represents the posterior probability for cluster i, M is the number of samples, and N is the number of GMM components. GMM was applied to accelerometer, gyroscope, and magnetometer signals (X, Y, Z axes) from all placements. Extracted values were normalized for consistency across datasets. As shown in Fig. 4, GMM components visualize the statistical distribution of activity features, improving classification reliability by capturing signal mean and variance.

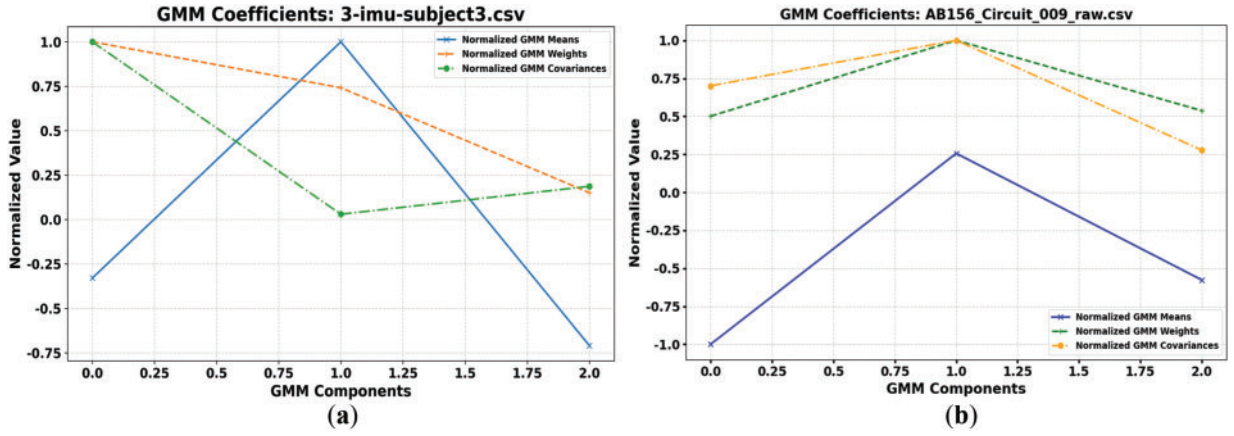

Figure 4: Visualization of GMM coefficients—means, weights, and covariances—illustrating distributional variations in (**a**) IM-WSHA and (**b**) ENABL3S datasets

4.3.2 Mel Frequency Cepstral Coefficients

Mel-Frequency Cepstral Coefficients (MFCCs) were extracted from inertial sensor data (accelerometer, gyroscope, magnetometer) to capture motion-relevant frequency features while minimizing signal noise. MFCCs transform raw signals into compact, discriminative features suitable for activity recognition.

Two extraction methods were applied: the first computed MFCCs individually across general IMU sensor channels, saving results as CSV files and visualizing them using heatmaps. The second processed structured IMU data by aggregating MFCCs across predefined channels into a unified format for classification. Both followed a common pipeline: Blackman windowing → FFT → Mel-filtering → log scaling → DCT. The kth MFCC is computed as:

$$C_k = \sum_{n=1}^{N} \log(E_n) \cos\left(\frac{\pi k (n - 0.5)}{N}\right) \tag{4}$$

here E_n represents the Mel-filtered energy of the nth frequency bin, and N is the total number of Mel-filters applied. Thirteen MFCCs were extracted from each sensor axis (*ax*, *ay*, *az*, *gx*, *gy*, *gz*, *mx*, *my*, *mz*) and used for classification.

To extract MFCC features, two implementations were applied to separate IMU datasets. The first processed individual sensor channels (e.g., accelerometer, gyroscope) independently, saving and visualizing MFCCs as heatmaps. The second aggregated MFCCs from structured acceleration, gyroscope, and magnetometer channels into a unified format. Both used a common pipeline: Blackman windowing, FFT, Mel-filtering, and DCT. Fig. 5 shows MFCC heatmaps from both datasets. Fig. 5a represents the Walking activity, with a consistent spectral pattern across time frames from a single IMU channel. Fig. 5b corresponds to the Jump activity, showing short bursts of high spectral energy across multiple channels. The x-axis indicates time frames, and the y-axis represents MFCC coefficients. Axis titles have been standardized. Color intensity reflects spectral energy—higher values indicate stronger frequency components, often linked to physical motion. These variations reveal patterns useful for classifying activities based on their frequency characteristics.

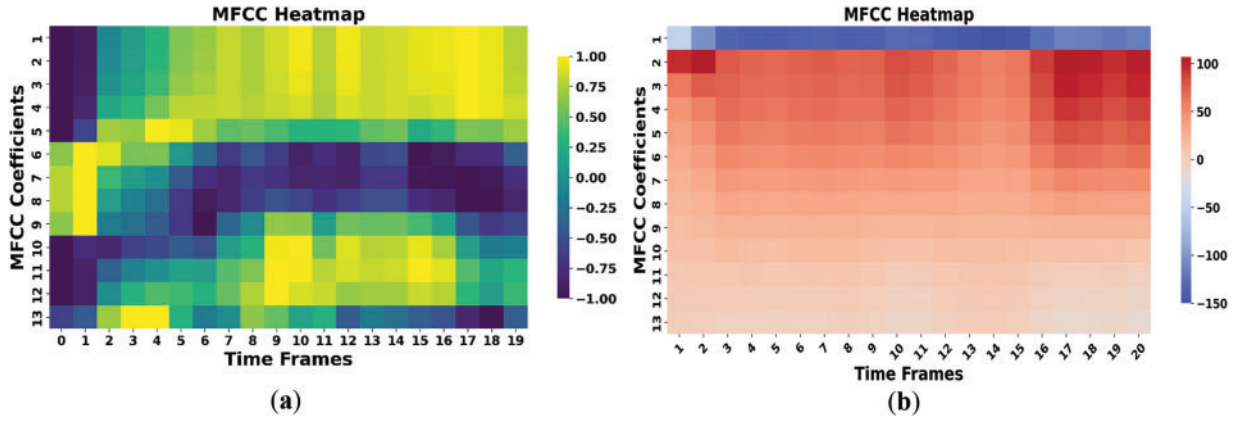

Figure 5: Temporal evolution of MFCCs highlighting spectral characteristics in (**a**) IM-WSHA and (**b**) ENABL3S datasets

4.3.3 Spectral Entropy

Spectral entropy measures the complexity of sensor signals by quantifying how power is distributed across frequencies. Periodic motions yield low entropy, while irregular movements result in higher values due to increased randomness. It is computed by estimating the Power Spectral Density (PSD) via the Welch method, normalizing it to form a probability distribution, and applying Shannon entropy:

$$H_s = -\sum_{j=1}^{M} P\left(f_j\right) \log P(f_j) \tag{5}$$

here $P(f_j)$ represents the normalized power at frequency f_j, and M is the total number of frequency bins. Entropy was extracted from all inertial channels (ax, ay, az, gx, gy, gz, mx, my, mz) to capture motion variability, aiding in distinguishing structured from unstructured activities. A visualization of Spectral Entropy values is displayed in Fig. 6, where different sensor channels show distinct movement complexity patterns. Specifically, Fig. 6a represents the IM-WSHA dataset, while Fig. 6b corresponds to the ENABL3S dataset. This comparative representation highlights the variability in spectral entropy across different sensors in two independent datasets, emphasizing how sensor placement and dataset-specific factors influence entropy characteristics. The intent behind Fig. 6 is to analyze inter-sensor variability rather than temporal entropy fluctuations. While spectral entropy over time could reveal phase transitions in movement, this visualization focuses on sensor-wise entropy distribution, aiding in feature selection and sensor placement analysis. By integrating Spectral Entropy into the feature set, the system enhances its ability to classify different activity patterns effectively.

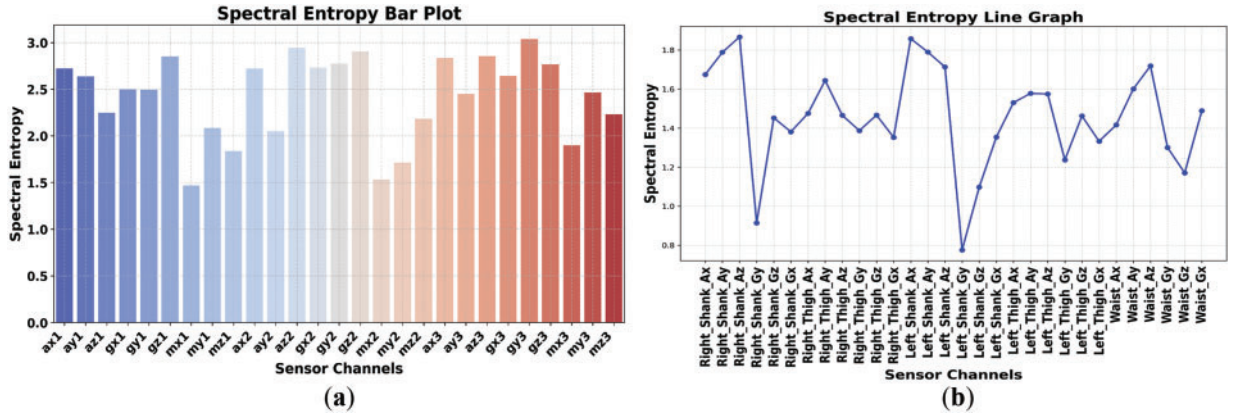

Figure 6: Spectral entropy computed for each sensor in (**a**) IM-WSHA and (**b**) ENABL3S datasets, illustrating variability across sensor channels

4.3.4 Quaternion-Oriented Features

Quaternion-based features provide rotational invariance, making them ideal for 3D inertial sensor data analysis. Unlike Euler angles, quaternions avoid issues like gimbal lock, offering stable orientation tracking. The process begins by normalizing accelerometers, gyroscope, and magnetometer readings. Accelerometer data estimates gravity, while the magnetometer refines orientation using Earth's magnetic field. These references compute a quaternion:

$$Q = w + xi + yj + zk \tag{6}$$

where w is the scalar part, and x, y, z are the vector components representing the axis direction of rotation movement. The sensor unit obtains normalized quaternions through this equation:

$$Q_i = \frac{1}{||a_i||}(0, a_{ix}, a_{iy}, a_{iz}) \tag{7}$$

where a_{ix}, a_{iy}, a_{iz} represent the normalized components of acceleration. Different sensor positions yield quaternion values that can be compounded together in an extensive descriptor of orientation. Each sensor offers an independent quaternion that describes local changes in orientation. To form a coherent presentation, quaternions of various sensors are concatenated without compromising their rotational integrity. This assembly ensures that motion patterns are well captured, resulting in a better descriptor for activity recognition. Implementation derives quaternion-based features from sensors located at three distinct points on the body to provide a better description of movement dynamics. Integration allows the system to record full-body orientation variations, which minimizes interference or noise and enhances the accuracy of classification. The operation aligns all quaternion readings in the same reference frame through sensor fusion strategies to ensure that the resulting orientation descriptor is robust across various activities.

Fig. 7 shows quaternion components derived from two various datasets: Fig. 7a is for the IM-WSHA dataset, and Fig. 7b is for quaternion values in the ENABL3S dataset. Both subfigures show time-varying quaternion components obtained for various activities, displaying how orientation changes for differing movement sequences. Having consistent visualization across datasets makes it easier to interpret, allowing rotational motion trends to be compared directly.

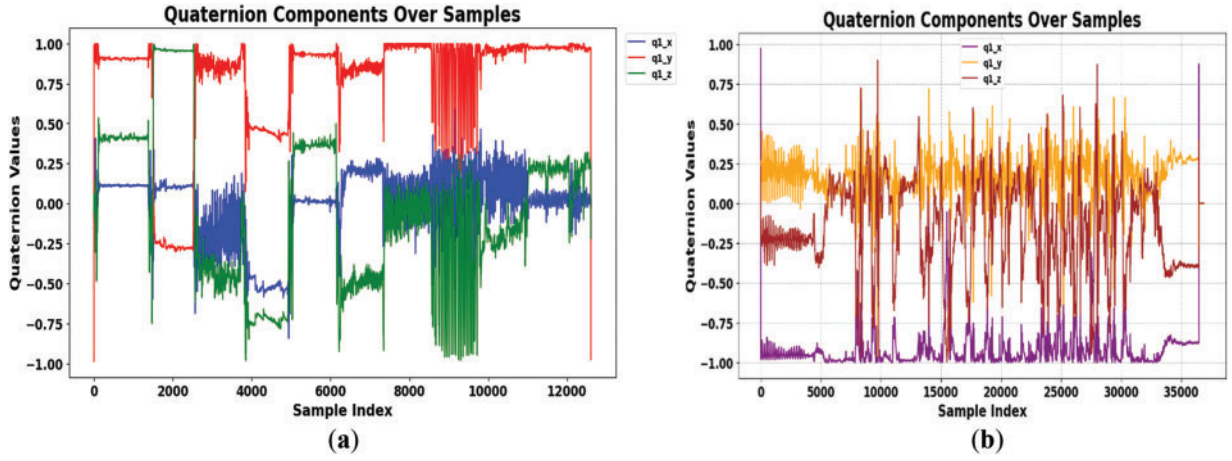

Figure 7: Quaternion components over samples for (**a**) IM-WSHA and (**b**) ENABL3S datasets, illustrating orientation variations during activities over time

4.3.5 Gammatone Cepstral Coefficients

Gammatone Cepstral Coefficients (GCCs), an advanced variant of MFCCs, improve inertial signal processing and robustness in dynamic environments. The extraction process starts by segmenting the preprocessed signal into overlapping frames to preserve temporal continuity. FFT is applied to obtain frequency representations, followed by spectral energy mapping through 26 Gammatone filters—replacing traditional triangular filters. A cubic transformation enhances sensitivity to high-frequency components. Then, Discrete

Cosine Transform (DCT) extracts decorrelated cepstral coefficients as GCC features, computed using:

$$C_m = \sum_{n=0}^{N-1} E_c[n] \cos\left[\frac{\pi m}{N}\left(n + \frac{1}{2}\right)\right] \tag{8}$$

$$C_m^{'} = \log(1 + |C_m|) \tag{9}$$

where $E_c = |H(f)|^3$ captures the cubic energy from the filter output. GCCs were extracted from all IMU channels (accelerometer, gyroscope, magnetometer), normalized, and visualized as heatmaps to highlight spectral activity patterns (see Fig. 8a,b). These features enhance class separability and improve recognition performance, especially during high-frequency, dynamic motions.

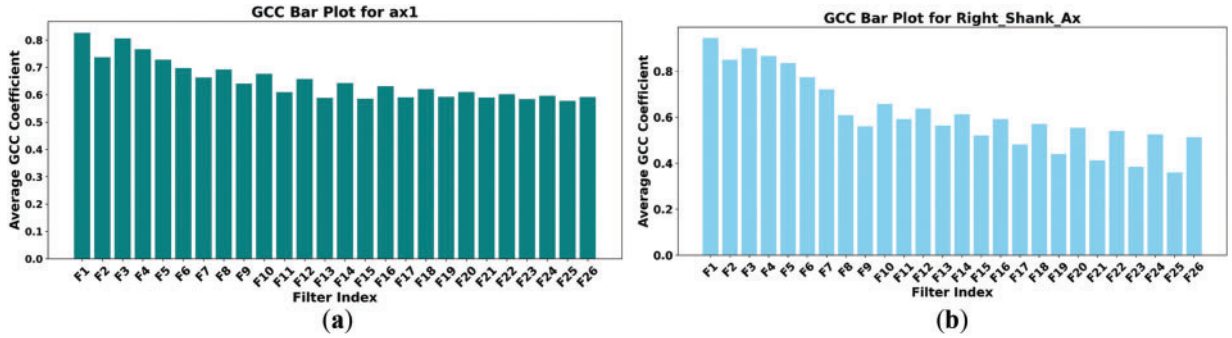

Figure 8: GCC bar plots for (**a**) IM-WSHA and (**b**) ENABL3S datasets, illustrating frequency band variations over time frames

4.4 Feature Fusion

Features extracted from inertial sensor data undergo a feature-level fusion using a common column before optimization for information integration. Different characteristics of human movement become observable through the combination of extracted features that include MFCC, GCC, spectral entropy, quaternion-based motion descriptors, and GMM-based features. The combined multicomponent feature vector maintains spectral and statistical data elements through sequential addition. The combined features yield better separability through this approach, which benefits the next Bayesian optimization step, which is to select optimal features and eliminate redundancies from the representation. The proposed method fuses different features to boost model generalization capabilities which results in better classification performance.

4.5 Bayesian Optimization

The system fuses features and applies Bayesian Optimization for feature selection, enhancing classification accuracy. This optimization balances exploration and exploitation, iteratively refining feature subsets to identify high-performing candidates. By selecting the most discriminative and noise-resilient features, Bayesian Optimization further reduces the impact of sensor drift and transient motion artifacts, ensuring that classification performance remains stable even in varying environmental conditions [22]. Bayesian Optimization was employed to identify an optimal subset of fused features, balancing classification accuracy with computational efficiency through sequential model-guided evaluations. The optimization goal took the following form:

$$\theta^* = arg\max_{\theta \in \Theta} f(\theta) + \lambda u(\theta) \tag{10}$$

where $f(\theta)$ represents the predictive mean of the objective function, $u(\theta)$ is the uncertainty estimate (modeled using Gaussian Processes), and λ is a trade-off parameter between exploration and exploitation. The objective function was designed to maximize classification performance by selecting the most informative, non-redundant feature subsets. Bayesian Optimization iteratively evaluates feature combinations, guided by an acquisition function that prioritizes improvements in recognition accuracy. A batch strategy was used to assess multiple feature subsets per iteration, enhancing efficiency. The process began with 10 randomly sampled subsets to explore the search space, followed by 50 iterations using the Expected Improvement (EI) function with $\xi = 0.01$. Each iteration evaluated five subsets in parallel to reduce convergence time.

Simultaneously, CNN hyperparameters were optimized: learning rate was drawn from a log-uniform range [0.0001, 0.01], while batch size, filter count, and kernel size were selected from {32, 64, 128} and {3, 5, 7}, respectively. This joint optimization produced robust feature-CNN configurations. Fig. 9 shows a 3D visualization of feature subset trajectories, with color-coded activity classes highlighting the model's discriminative performance. The approach effectively balances exploration and exploitation, resulting in improved classification accuracy and system reliability.

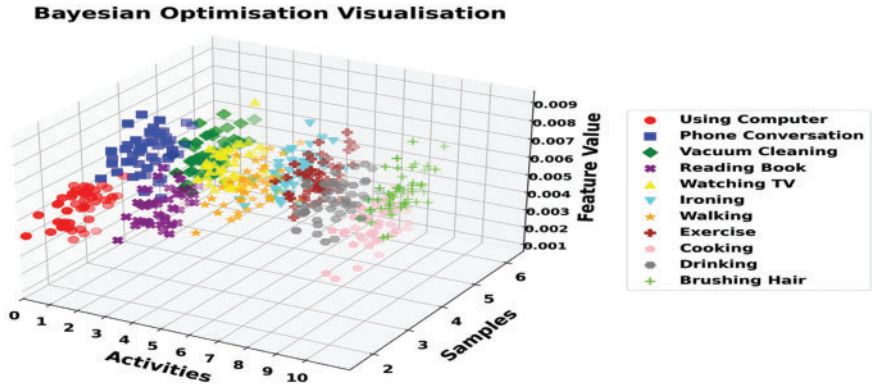

Figure 9: Optimization results for the IM-WSHA dataset, highlighting performance improvements

4.6 Classification

A CNN was used to classify daily life activities by capturing spatial relationships among extracted features.

Given an input feature map X of size $H_{in} \times W_{in} \times C_{in}$, convolution is performed with filters W of size $K \times K \times C_{in} \times F$ and bias b of size F. The output feature map Y is computed as:

$$O_{uvw} = \sum_{p=0}^{K-1} \sum_{q=0}^{K-1} \sum_{r=0}^{C_f-1} GpqrwF_{u+p,v+q,r} + \beta_w \tag{11}$$

The network uses successive convolutional layers with Rectified Linear Unit (ReLU) activations for non-linear mapping and batch normalization to accelerate training. Max pooling reduces spatial dimensions while preserving key features. A fully connected layer follows, with a softmax function converting outputs to a probability distribution:

$$\sigma(y)_m = \frac{e^{y_m}}{\sum_{n=1}^{D} e^{y_m}} \tag{12}$$

where y_m is the output from the previous layer, and e is Euler's constant (2.7183). CNN received its optimization through Bayesian Optimization which adjusted hyper-parameters like learning rate and batch size together with the number of filters and kernel sizes to achieve better results.

Bayesian Optimization systematically explored the hyper-parameter space by constructing a probabilistic model of the objective function, which aimed to maximize classification accuracy. The optimization process iteratively refined the selection of hyper-parameters—such as learning rate, batch size, number of filters, and kernel sizes—by evaluating their impact on model performance. An acquisition function guided this process, balancing exploration of new hyper-parameter configurations and exploitation of promising values. K-fold cross-validation was employed during training, and Bayesian Optimization dynamically adjusted hyper-parameters based on validation performance, leading to improved accuracy and reduced overfitting. Table 1 summarizes the CNN architecture and training configuration, with hyperparameters reflecting the optimal values identified via Bayesian Optimization.

Table 1: CNN architecture and training hyperparameters

Layer	Type	Parameters	Value
Input layer	Input features	Dimension $H_{in} \times W_{in} \times C_{in}$	$64 \times 64 \times 1$
Conv layer 1	Conv2D + ReLU	kernal size, filters, stride, padding	3×3, 64 filters, stride 1, same
Batch normalization	Normalization	–	Applied after Conv Layer 1
Max pooling 1	Pooling	Pool size, stride	2×2, stride 2
Conv layer 2	Conv2D + ReLU	kernal size, filters, stride, padding	3×3, 128 filters, stride 1, same
Batch normalization	Normalization	–	Applied after Conv Layer 2
Max pooling 2	Pooling	Pool size, stride	2×2, stride 2
Flatten layer	Flatten	–	Converts to 1D
Dense	Fully connected	Number of units	256 units
Dropout	Regularization	Dropout rate	0.5
Output layer	Softmax	Number of output classes	11
Training epochs	Hypermeters	–	50
Learning rate	Hyperparameter	(log-uniform: [0.0001, 0.01])	0.001
Batch size	Hyperparameter	From discrete set {32, 64, 128}	64
Optimizer	Training parameter	Optimization algorithm	Adam
Loss function	Training parameter	Multi-class classification loss	Categorical cross entropy

5 Experimental Setup

Experiments were conducted on a Windows 11 Pro system with an Intel Core i7 (2.40 GHz), using Python and libraries such as TensorFlow, NumPy, Pandas, and Scikit-learn. An n-fold cross-validation was applied to evaluate model performance, achieving high accuracy on the IM-WSHA and ENABL3S datasets. Data acquisition and evaluation methods are detailed in later sections.

6 Experimental Results

We evaluated the proposed IS-DAR using the IM-WSHA and ENABL3S datasets, employing N-fold cross-validation where each data instance contributed to both training and validation. The confusion matrices in Fig. 10a,b show the model's performance, with a mean accuracy of 93.0% on IM-WSHA and 92.0% on ENABL3S, reflecting the model's generalization ability.

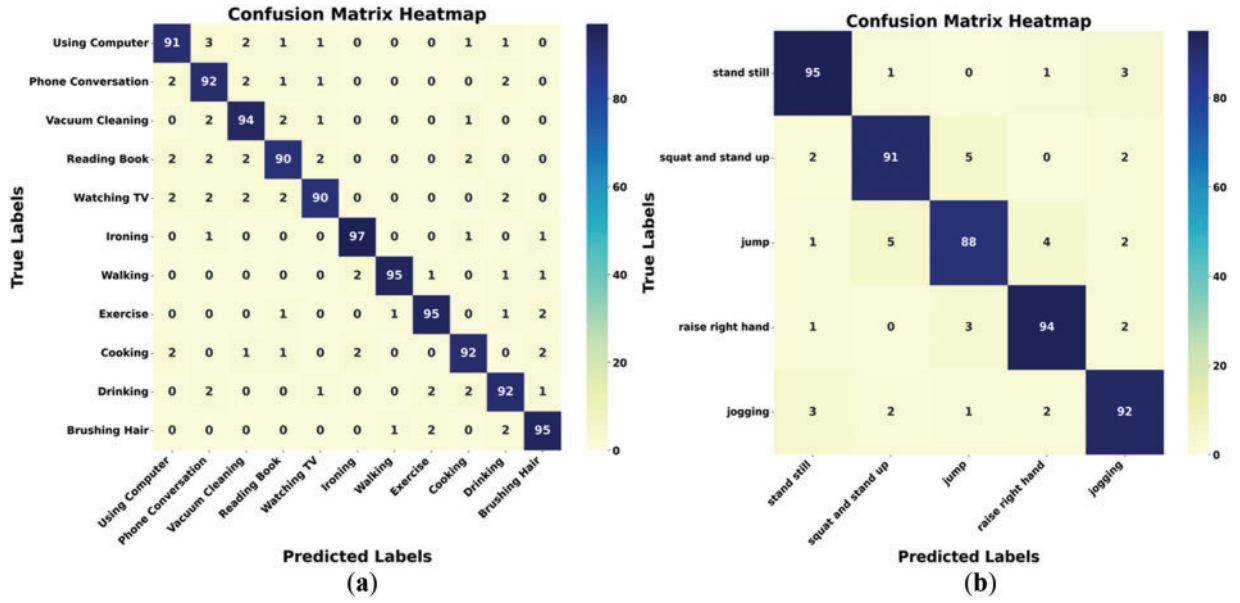

Figure 10: Confusion matrices showing accuracy for the (**a**) IM-WSHA (**b**) and ENABL3S datasets

To further validate the effectiveness of our proposed hybrid pipeline, we compare its classification accuracy with recent state-of-the-art HAR models. Our method outperforms existing approaches in both datasets. As shown in Table 4, our method outperforms existing approaches in both datasets. Specifically, on the ENABL3S dataset, our system achieves 92.0% accuracy, surpassing the CNN-based classification system (90.93%) and the LDA-based classification system (85.78%) reported in [23]. Similarly, for the IM-WSHA dataset, our approach achieves 93.0% accuracy, outperforming the RPLB + Stochastic Gradient Descent system (83.18%) and the RPLB + Adagrad system (81.73%) reported in [24].

6.1 Precision, Recall, and F1 Score for IS-DAR

This study has included precision, recall, and F1-score as evaluation metrics. The IM-WSHA and ENABL3S datasets' precision, recall, and F1 score are displayed in Tables 2 and 3, respectively. Finally, Table 4 shows the comparison of proposed system with other state-of-the-art systems over the IM-WSHA and ENABL3S datasets. The bold text in Tables 2 and Tables 3 indicates the dataset classes, while the bold numerical values at the end represent the average precision, recall, and F1 score.

Table 2: Classification report on the IM-WSHA dataset

Classes	Precision	Recall	F1-score
Using computer	0.92	0.91	0.92
Phone conversation	0.90	0.92	0.91
Vacuum cleaning	0.91	0.94	0.93
Reading book	0.92	0.90	0.91
Watching TV	0.94	0.90	0.92
Ironing	0.96	0.97	0.97
Walking	0.98	0.95	0.96

(Continued)

Table 2 (continued)

Classes	Precision	Recall	F1-score
Exercise	0.95	0.95	0.95
Cooking	0.93	0.92	0.93
Drinking	0.91	0.92	0.92
Brushing	0.92	0.95	0.94
Average	**0.93**	**0.93**	**0.93**

Table 3: Classification report on the ENABL3S dataset

Classes	Precision	Recall	F1-Score
Stand still	0.93	0.95	0.94
Squat and stand up	0.92	0.91	0.91
Jump	0.91	0.88	0.89
Raise right hand	0.93	0.94	0.94
Jogging	0.91	0.92	0.92
Average	**0.92**	**0.92**	**0.92**

Table 4: Comparisons of the proposed system with other systems

Methods	IM-WSHA (%)	ENABL3S (%)
CNN [23]	–	90.93
LDA [23]	–	85.78
RPLB + SGD [24]	83.18	–
RPLB + Adagrad system [24]	81.73	–
Standard HMM [2]	87.15	88.35
Modified HMM-based [2]	92.65	92.50
Proposed system mean accuracy (%)	93.00	92.00

6.2 Computational Complexity and Feature Selection Comparison

Despite its multi-stage design, the proposed method remains efficient—Bayesian Optimization achieves effective feature selection with low execution time (4.2 s) and memory usage (75 MB), as shown in Table 5. Compared to PCA, SVD, and IDA, it offers better accuracy while remaining practical for real-world use.

Table 5: Computational complexity comparison of feature selection methods

Feature selection method	Computational complexity	Execution time (s)	Memory usage	Accuracy
Bayesian optimization	O (n log n)	4.2	75	93.0
PCA	O ($n^2d + d^3$)	5.8	82	89.7
SVD	O (nd^2)	6.5	95	88.9
IDA	O (n^3)	8.1	110	87.3

7 Conclusion and Future Works

The proposed methodology efficiently detects locomotor movements through inertial sensors with a robust feature extraction, processing, and classification flow. It uses Chebyshev filtering, Blackman windowing, and extracts GMM, MFCC, spectral entropy, quaternion, and GCC features. The best features are shortlisted through Bayesian Optimization followed by classification using a CNN. The system outperforms the existing approaches on IM-WSHA and ENABL3S datasets, and it is extremely promising for use in real-time movement monitoring, medical care, and sports.

Although it is true that the process involves costly computations such as feature extraction and classification, its appropriateness relies on application needs in the real world. When high accuracy and robustness are more important than low-latency processing—such as diagnostics for medicine, cybersecurity intrusion detection, or financial fraud detection—the cost of processing is reasonable. Sensor drift is a potential issue on long-term deployment, although low degradation across a period of time indicates that the process is stable for extended use.

Although we prioritize performance, interpretability is especially important for applications in healthcare, security, and behavior monitoring. To facilitate this, we use hand-crafted features—spectral entropy, quaternion descriptors, and GMM statistics—that possess well-understood physical meaning with respect to body movement. These transparently inform activity signals increase user trust. For greater model explainability, we will incorporate post-hoc explanation mechanisms such as SHAP and LIME, particularly to elucidate CNN predictions in uncertain or high-consequence situations.

Even if the model is robust on IM-WSHA and ENABL3S, generalizability across unobserved datasets as well as different sensor setups is crucial for practical applications. Cross-dataset evaluation, domain adaptation, and tests with different sensor placements will be part of the next work to evaluate and extend the portability and flexibility of the model.

Further research will similarly investigate dataset fusion across different sensor placements for greater system robustness. Transformer-based models integrated with hybrid deep pipelines of models are predicted to increase accuracy and robustness. Exploration of light-weight architectures such as MobileNetV3, EfficientNet-Lite, or SqueezeNet should lower computing requirements for embedded deployment. Attention-based temporal encoders embedded within CNN or LSTM architectures can better capture long-range dependency modeling. Real-time monitoring of human activities is even more viable by optimizing the system for smartphone, wearable, or microcontroller edge deployment through the use of model quantization and hardware-aware neural architecture search. Lastly, integration with self-supervised learning or semi-supervised learning using multimodal inputs—even physiological signals—can increase flexibility under limited-label-availability resource-constrained environments.

Acknowledgement: The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Group Project.

Funding Statement: The APC was funded by the Open Access Initiative of the University of Bremen and the DFG via SuUB Bremen. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Group Project under grant number (RGP. 2/568/45). The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the Project Number "NBU-FFR-2025-231-04".

Author Contributions: Study conception and design: Iqra Aijaz Abro; data collection: Mohammed Alshehri and Yahya AlQahtani; analysis and interpretation of results: Abdulmonem Alshahrani and Asaad Algarni; draft manuscript preparation: Hui Liu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used in this study are publicly available and can be accessed via the following links: IM-WSHA Dataset: https://www.kaggle.com/datasets/sbtahir/im-wsha-dataset and ENABL3S Dataset: https://figshare.com/articles/dataset/Benchmark_datasets_for_bilateral_lower_limb_neuromechanical_signals_from_wearable_sensors_during_unassisted_locomotion_in_able-bodied_individuals/5362627 (accessed on 25 May 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Arvindhan A, Sharma A, Kumar A. Human routine analyzer using machine learning. In: Proceedings of the 2023 5th International Conference on Advances in Computing, Communication Control and Networking; 2023 Dec 15–16; Uttar Pradesh, India. p. 695–99. doi:10.1109/ICAC3N60023.2023.10541574.
2. Ghadi YY, Javeed M, Alarfaj M, Al Shloul T, Alsuhibany SA, Jalal A, et al. MS-DLD: multi-sensors based daily locomotion detection via kinematic-static energy and body-specific HMMs. IEEE Access. 2022;10:23964–79. doi:10.1109/access.2022.3154775.
3. Ko MP, Su C, Shie H. Human activity recognition system using angle inclination method and keypoints descriptor network. In: Proceedings of the 2024 Conference of Young Researchers in Electrical and Electronic Engineering (ElCon); 2024 Jan 29–31; Saint Petersburg, Russian Federation. p. 235–9. doi:10.1109/ElCon61730.2024.10468160.
4. Kang S-M, Wildes RP. Detecting biological locomotion in video: a computational approach. arXiv:2105.12661. 2021.
5. Hassan N, Miah ASM, Shin J. Enhancing human action recognition in videos through dense-level features extraction and optimized long short-term memory. In: Proceedings of the 2024 7th International Conference on Electronics, Communications, and Control Engineering (ICECC); 2024 Mar 22–24; Kuala Lumpur, Malaysia. p. 19–23. doi:10.1109/ICECC63398.2024.00011.
6. Jatesiktat P, Lim GM, Lim WS, Ang WT. Anatomical-marker-driven 3D markerless human motion capture. IEEE J Biomed Health Inform. 2024;1–14. doi:10.1109/jbhi.2024.3424869.
7. Kamble M, Bichkar RS. A hierarchical framework for video-based human activity recognition using body part interactions. Int J Electr Comput Eng Syst. 2023;14(8):881–91. doi:10.32985/ijeces.14.8.6.
8. Khan MH, Zöller M, Farid MS, Grzegorzek M. Marker-based movement analysis of human body parts in therapeutic procedure. Sensors. 2020;20(11):e3312. doi:10.3390/s20113312.
9. Wickramarachchi DN. Automated marker-based abnormal gait pattern detection using novel 6-dimensional skeleton. In: Proceedings of the 2024 18th International Symposium on Medical Information and Communication Technology (ISMICT); 2024 May 15–17; London, UK. p. 89–94. doi:10.1109/ISMICT61996.2024.10738177.
10. Mekruksavanich S, Jantawong P, Jitpattanakul A. A hybrid deep learning neural network for recognizing exercise activity using inertial sensor and motion capture system. In: Proceedings of the 2023 4th International Conference on Big Data Analytics and Practices (IBDAP); 2023 Aug 25–27; Bangkok, Thailand. p. 1–5. doi:10.1109/IBDAP58581.2023.10271955.
11. Wang S, Zeng X, Huangfu L, Xie Z, Ma L, Huang W, et al. Validation of a portable marker-based motion analysis system. J Orthop Surg Res. 2021;16(1):425. doi:10.1186/s13018-021-02576-2.
12. Niemann F. Context-aware activity recognition in logistics (CAARL)—An optical marker-based motion capture dataset Zenodo [Internet]. 2021. [cited 2025 May 25]. Available from: https://doi.org/10.5281/zenodo.5680951.
13. Lee KD, Park HS. Real-time motion analysis system using low-cost web cameras and wearable skin markers. Front Bioeng Biotechnol. 2022;9:790764. doi:10.3389/fbioe.2021.790764.
14. Khan D, Al Mudawi N, Abdelhaq M, Alazeb A, Alotaibi SS, Algarni A, et al. A wearable inertial sensor approach for locomotion and localization recognition on physical activity. Sensors. 2024;24(3):735. doi:10.3390/s24030735.
15. Du G, Zeng J, Gong C, Zheng E. Locomotion mode recognition with inertial signals for hip joint exoskeleton. Appl Bionics Biomech. 2021;2021(2):6673018. doi:10.1155/2021/6673018.

16. Sarcevic P. Inertial sensor-based movement classification with dimension reduction based on feature ag-gregation. In: Proceedings of the 2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo); 2022 Nov 21–22; Budapest, Hungary. p. 113–8. doi:10.1109/CINTI-MACRo57952.2022.10029519.
17. Shin D, Lee S, Hwang S. Locomotion mode recognition algorithm based on Gaussian mixture model using IMU sensors. Sensors. 2021;21(8):2785. doi:10.3390/s21082785.
18. Nouriani A, Jonason A, Jean J, McGovern R, Rajamani R. System-identification-based activity recognition algorithms with inertial sensors. IEEE J Biomed Health Inform. 2023;27(7):3119–28. doi:10.1109/jbhi.2023.3265856.
19. Celik Y, Aslan MF, Sabanci K, Stuart S, Woo WL, Godfrey A. Improving inertial sensor-based activity recognition in neurological populations. Sensors. 2022;22(24):9891. doi:10.3390/s22249891.
20. Trabelsi I, Francoise J, Bellik Y. Sensor-based activity recognition using deep learning: a comparative study. In: Proceedings of the 8th International Conference on Movement and Computing; 2022 Jun 22–24; Chicago, IL, USA. p. 1–8. doi:10.1145/3537972.3537996.
21. Al Mudawi N, Azmat U, Alazeb A, Alhasson HF, Alabdullah B, Rahman H, et al. IoT powered RNN for improved human activity recognition with enhanced localization and classification. Sci Rep. 2025;15(1):10328. doi:10.1038/s41598-025-94689-5.
22. Sultana S. Bayesian and genetic optimization for human activity recognition with CNN and LSTM. In: Proceedings of the 2024 International Conference on Intelligent & Innovative Practices in Engineering & Management (IIPEM); 2024 Nov 25; Singapore. p. 1–6. doi:10.1109/IIPEM62726.2024.10925732.
23. Zhang K, Wang J, de Silva CW, Fu C. Unsupervised cross-subject adaptation for predicting human locomotion intent. IEEE Trans Neural Syst Rehabil Eng. 2020;28(3):646–57. doi:10.1109/TNSRE.2020.2966749.
24. Gochoo M, Tahir SBUD, Jalal A, Kim K. Monitoring real-time personal locomotion behaviors over smart indoor-outdoor environments via body-worn sensors. IEEE Access. 2021;9:70556–70. doi:10.1109/access.2021.3078513.