ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 12 December 2025

A Novel Approach for Feature Extraction to Efficiently Determine the Unique and Distinctive Characteristics of Human Activity

[1] Suraj Deb Barma*, [2] Dr. Abhijit Biswas

[1] PhD Scholar, Department of Computer Science & Engineering, The ICFAI University Tripura, Kamalghat, Mohanpur, West Tripura, India

[2] Associate Professor, Department of Computer Science & Engineering, The ICFAI University Tripura, Kamalghat, Mohanpur, West Tripura, India

Email: [1] surajdebbarma.cse@gmail.com

Abstract— ***Human Activity Recognition (HAR) using either wearable or ambient sensors has become a key component in healthcare monitoring, rehabilitation, and smart environments. Ever since HAR was introduced, the extraction of discriminative features capable of capturing temporal and spatial characteristics representative of different human activities, while standing robust against user variation and sensor noise, has been one of the most challenging open issues. This paper addresses these issues with an innovative hybrid feature extraction approach, called Hybrid Multi-Domain Feature Extraction (HMDFE), which hinges upon time–frequency transformations, morphological analysis, and correlation-based descriptors to detect pattern and micro-movements within human activity signals. This letter transforms raw accelerometer and gyroscope signals using Continuous Wavelet Transform (CWT) and Short-Time Fourier Transform (STFT); extracts multi-scale statistical and morphological features; and encodes inter-axis correlations via Gramian Angular Fields (GAF). Experiments on benchmark datasets like UCI HAR, WISDM, and PAMAP2 show that the proposed approach outperforms existing classical handcrafted and deep learning-based feature extractors regarding classification accuracy and F1-score.***

Index Terms— ***Human Activity Recognition (HAR), Feature Extraction, Multi-domain Features, Wavelet Transform, Morphological Analysis, Machine Learning, Sensor Data***

I. INTRODUCTION

Human Activity Recognition has emerged as one of the prime factors in context-aware computing and intelligent health monitoring systems. The main objective of HAR is to identify the activity which has been performed by a human in a video that was captured from an uncontrolled background environment. Research on HAR has widely been accomplished due to its wide applicability in various applications involving computer gaming, virtual reality, security surveillance, human machine interaction, video understanding, and assistive robots [13-16]. The broad range of applications it covers includes elderly care, gesture tracking, fitness analytics, and workplace safety. It depends upon the correct feature extraction regarding discriminating between dynamic and static activities of humans such as walking, sitting, climbing stairs, or running [4, 10].

One of the most intriguing uses of HAR is the ubiquitous detection of human activity. It is well known that the primary cause of human obesity is a habit of physical inactivity. People can keep a much more accurate track of the calories they consume or gain throughout the day by tracking their physical activity. In addition to preventing obesity and other negative health effects, this will motivate them to continue exercising. Another delightful example is Home Care Monitoring, which allows disabled and elderly patients a continuous health and well-being supervision while they perform Activities of Daily Living (ADL) at home. [9]

These traditional methods for feature extraction cannot generalize well across users or devices and include statistical moments, energy measures, and frequency-domain coefficients. Deep learning models, on the other hand, automatically learn their representations with CNNs, LSTMs, and Transformers, require a large amount of labeled data, and are not interpretable. Hence, the requirement to develop more hybrid-type techniques that can integrate the interpretability of handcrafted techniques along with the robustness of deep feature learning is increasing.

This paper introduces a Hybrid Multi-Domain Feature Extraction framework, HMDFE, which represents an effective means of identifying unique patterns of human activities by integrating complementary cues from time, frequency, and morphological domains.

The rest of this paper is organized as follows: Section II presents other related works; Section III describes the approach used in our work to extract features based on human range of movements; Section IV describes the experimental evaluation and the tests carried out; Section V discusses the results obtained from the investigation; finally, Section VI summarizes the findings from the experiment.

II. RELATED WORK

Time-domain and frequency-domain feature extraction methods have been investigated by using the traditional and deep learning approaches for identifying the human activities in their everyday life. During earlier researches, the HAR mainly concentrated on handcrafted time-domain parameters like mean, zero-crossing rate, variance, and signal magnitude area (SMA). Periodicities and energy distributions were recorded using Fourier and wavelet-based techniques (Anguita *et al.*, 2013; Banos *et al.*, 2014). However, due to changes in sensor orientation or inter-subject variability, these parameters often tend to fail.

Recognition accuracy has greatly increased recently by the establishment of Deep learning models such as DeepConvLSTM (Ordóñez & Roggen, 2016), Inception Time (Fawaz *et al.*, 2019), and Transformer-based HAR (Xu *et al.*, 2021). However, those systems have limitation due to large amount of computational costs and huge dependency on data.

The feature extraction from RGBD sensor data based on human range of movement indicated better performance in Wee-Hong et al. (2013). However, this method failed to identify some activities thereby creating a sense of confusion.

In this work, the various feature extraction strategies surveyed were time-based and frequency-based; the results of time-based features were superior compared to the frequency-based ones. Besides, combining information from all sensors significantly enhanced it. Zhu et al. 2017 presented a competitive result with another public dataset by using the proposed model in this study.

The proposed TCN-Bi-LSTM HAR system, which combines the benefits of Temporal Convolutional Networks with Bidirectional Long Short-Term Memory, postulates impressive performance in the tasks of classification and knowledge recognition by demonstrating experiments on three public datasets: namely, UCI-HAR, PAMAP2, and WISDM (Sadhna Bijrothiya et al., 2025) 805–813 813 Although this system is effective for scaled and pre-defined datasets, there are some limitations related to computational demands due to its complexity and generalization challenges over unseen data.

The developed integrated feature extraction model and GRRF feature selection for smartphone-based HAR utilizes the RF-based feature selection method to overcome the issues of overfitting and the curse of dimensionality efficiently, which improves the generalization performance of the classifiers in efficiently identifying different human physical activities.

Multilevel Features Cascade Fusion (MFCF) network integrates different temporal and spatial feature maps obtained through the residual blocks of the backbone network that enhances the accuracy of behavior recognition. Inspite of the enhancement in results, certain shortcomings are noticed. For example, there is scope to further improve the accuracy in recognition and to devise worthy algorithms whenever there is complex dataset for better feature extraction and fusion when dealing with complex datasets (H. Han *et al.*, 2025).

Recent hybrid approaches, such as Zeng et al. (2022) and Liu et al. (2023), combine multi-domain or handcrafted-deep features. However, they often lack explicit mechanisms for modeling morphological signal structures-local extrema, envelope shapes, and inter-axis dependencies.

Our approach addresses this by designing a multi-representation feature extraction pipeline that models the local morphology, time–frequency energy, and spatial correlations of multi-sensor signals explicitly.

III. METHODOLOGY

A. Overview

The proposed HMDFE framework basically consists of three core modules:

- **Time–Frequency Decomposition**—Wavelet and STFT-based spectral features.
- **Morphological Profile Analysis**—shape-based features via dilation/erosion.
- **Correlation and Phase Mapping**—Gramian Angular and Recurrence Encodings.

All the features are to be combined together and normalized so as to form a high-level descriptor, that could be fed into any machine learning or deep learning classifier.

B. Time–Frequency Features

The time-frequency features are explored such that each accelerometer and gyroscope axis signal $x(t)$ are decomposed using Continuous Wavelet Transform (CWT):

$$S(u,s) = \int x(t)\psi^*\left(\frac{t-u}{s}\right)dt$$

where "ψ" is the mother wavelet, "s" is the scale, and "u" is the translation.

The Spectral energy features are computed as:

$$E_s = \frac{1}{T}\sum_t \mid S(u,s) \mid^2$$

An average of these is then derived across scales so as to form a compact spectral signature. Using overlapping windows for capturing the steady-state frequencies, CWT is complemented by the Short-Time Fourier Transform (STFT). Thus, the Dominant frequency bands, an spectral entropy, and the bandwidth are extracted.

C. Morphological Features

The system emphasizes that the **shape characteristics** of the signal envelope revealed the morphological operations as per the study.

For each of the time series $x(t)$:

1. $\text{Opening}_r(x) = \text{Dilation}_r(\text{Erosion}_r(x))$
2. $\text{Closing}_r(x) = \text{Erosion}_r(\text{Dilation}_r(x))$

where "r" denoting the structuring element scale.

The features such as area under profile, morphological gradient, and scale-wise mean amplitude difference is to record local dynamics and transitions.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 12 December 2025

D. Angular Encodings and Correlation

Two nos. of 2-Dimensional image-like encodings are engaged to simulate the inter-axis and temporal dependencies:

- Firstly, normalized signals are being mapped into a cosine-sum picture that reflects phase correlations by using the Gramian Angular Field (GAF).
- Next the State-space recurrences are being captured by the Recurrence Plot (RP) by using:

$$R_{i,j} = \mathbb{I}\{\| x_i - x_j \| < \epsilon\}$$

These periodic and quasi-periodic structures indicative to each activity are then highlighted in these depictions.

E. Fusion of Feature and Classification

Feature vectors of every domain are combined and normalized:

$$\mathbf{F} = [\mathbf{f}_{\text{time}}, \mathbf{f}_{\text{freq}}, \mathbf{f}_{\text{morph}}, \mathbf{f}_{\text{corr}}]$$

Using Principal Component Analysis (PCA) the dimensionality has to be reduced. Then the resulting representation are fed into neural networks (MLP, CNN) and machine learning classifiers (Random Forest, SVM).

IV. EXPERIMENTAL SETUP

Three common datasets—UCI HAR, WISDM, and PAMAP2—were used in the experiments. The sensors, namely, accelerometer and gyroscope values that are collected at 50–100 Hz, are included in each dataset. The data were divided into 50% by overlapping on window size of 2.56-second segments. In this study,10-fold cross-subject validation to assess accuracy, precision, recall, and F1-score are being used. As baselines are referred, the examples of baselines are CNN-LSTM, Transformer-based models, and Statistical Feature + SVM.

Table I: Parameter Description

Parameters	Descriptions
Sampling Rate	50–100 Hz
Datasets	UCI HAR, WISDM, and PAMAP2
Window Size	2.56 s with 50% overlap segments
Metrics	F1-score Validation, Accuracy, Precision, and Recall

Figure 1. Depiction of the evaluation parameters.

The following baseline models are used to extract features by evaluating their performance:

- Inception Time 1-D;
- CNN-LSTM Hybrid;
- Statistical feature + SVM;
- Transformer-based HAR

A. Datasets and Tasks

Three sets of benchmark datasets and controlled synthetic sets are used to assess the suggested novel Hybrid Multidomain Feature Extraction (HMDFE).

Table II. Dataset Description

Dataset	No. of Subjects	Sampling Rate (in Hz)	No. of Activities	Duration (in hours)
UCI-HAR	30	50	6	10
WISDM	36	20–200	18	20
PAMAP2	9	100	12	8

Figure 2. An outline of the dataset description for the suggested framework for Hybrid Multi-Domain Feature Extraction (HMDFE).

The table figure illustrates a data flow from raw sensor signals through wavelet and morphological transforms to final feature fusion and classification.

1. UCI HAR: six activities-walking, walking upstairs, walking downstairs, sitting, standing, and lying, 30 subjects, and a 50 Hz smartphone accelerometer and gyroscope.

2.WISDM: Smartphone accelerometer data at frequencies of 20–200 Hz combined to 50 Hz; six activities: walking, jogging, sitting, standing, upstairs, downstairs.

3. PAMAP2: wearable IMU (100 Hz) + heart rate; 12 activities including lying, sitting, standing, walking, running, cycling, and housework.

4. Synthetic-HAR: multichannel accelerometer + gyroscope signals with activity-specific parametric profiles, colored Gaussian noise, and subject variability for method stress testing. The main task is single-label activity recognition. Among the metrics are macro-precision/recall/F1, accuracy, and per-class F1.

B. Data Collection and Preprocessing

1. **Sensor Units & Axes**: The metrics of the sensors employed for this study are — Gyroscope (G_1, G_2, G_3) in °/s; accelerometer (A_1, A_2, A_3) in m/s².

For testing robustness of cross-dataset, solely PAMAP2 heart rate (HR) is used.

2. Synchronization and Resampling Time: In the resampling stage, all data streams are resampled to 50 Hz followed by alignment for each participant.

3. Filtration: In order to remove the undesired elements from the data samples for achieving the targeted results, the data samples need to undergo filtration.

- *Fourth-order Butterworth zero-phase filter*: This filter is the best one for removing frequency elements that are undesired, for example, senor noise or drift. To eliminate high-frequency noise and gravitational drift for achieving best result, the band-pass of Accelerometer must be from 0.3–20 Hz; While the band-pass for Gyroscope, must be from the range between 0.3–25 Hz.
- *Gravity separation (acceleration)*: This preprocessing step measures gravity using low-pass at 0.3 Hz so as to get dynamic acceleration by means of subtract dynamic motion components.

4. Segmentation: The feature set was formed using a 2.56-

second sliding window with 50% overlap (128 samples). All windows having more than 15% missing data are excluded; otherwise, the vacant regions are filled using local linear interpolation.

5. Synthetic-HAR generation:

- Base motion profiles: Profiles of human motion or activity are indicated via piecewise sinusoids with amplitude jitter (±15%), random phase, and activity-specific dominating frequencies (for example, walking 1.8–2.2 Hz, running 2.6–3.2 Hz, cycling 1.0–1.6 Hz).
- Noise: Signal to Noise Ratio (SNR) is adopted in the categories of {15, 20, 25 dB} with colored Gaussian noise (β=1.5).
- Subject variability: For variability of the datasets, orientation disturbance is identified via random rotation matrix (≤15°) and axis-wise scale drift (±10%).
- Class balance: for every activity => ≈2,800 windows, 2 hours is allotted.

6. Normalization: z-score for each channel based solely on data from the training split (used for validation/test).

C. Feature Extraction (HMDFE)

Four groups of features are visualized in each pane. All features are computed on a per-channel basis and, where applicable, aggregated with pairwise statistics and mean/variance across axes.

1. *Temporal (T)*

Hjorth parameters: activity, mobility, complexity; peak count; mean; variance; RMS; median absolute deviation: MAD; interquartile range: IQR; zero-crossing rate: ZCR; signal magnitude area: SMA; slope of linear trend; autoregressive coefficients: AR (order 4).

2. *Morphological (M)*

Sample entropy, the permutation entropy, duty cycle, the rise/fall time, the peak/valley amplitude and width distributions (mean/var), and a local waveform symmetry (peak vs. valley duration).

3. *Correlation/Spatial (C)*

The Pairwise Pearson correlations in between the axes, tilt-invariant of magnitude features (a, “g”), accelerometer–gyroscope couplings (corr(a, “g”), the axis energy ratios, and cross-covariance lag of the maximum correlation.

4. *Time–Frequency (TF)*

In this case, we have Spectral centroid, bandwidth, spectral entropy & spectral flatness, dominant frequency & amplitude, harmonic ratio (first/second peak), the short-time DFT energy decay, Welch PSD (Hann windows, 50% overlap), and the band energies: 0.5–3 Hz, 3–6 Hz, 6–12 Hz, and >12 Hz. Depending upon the dataset and availability of axis, each of the window may have between 180 and 220 features. While Implementation, NumPy/SciPy feature functions with vectorized operations and float32 representations is used.

D. Selection of Features and Reduction of Dimensionality

- Variance threshold: Features that are nearly constant ($\sigma2 < 1e\text{-}6$) are typically excluded by this threshold.
- Mutual Information (MI) ranking: the top k is maintained with 120 attributes.
- Optional PCA: Unless specified, this method maintains 95% of the variance; however, it is not used in HMDFE unless it is necessary.
- Collinearity control: To achieve collinearity control, features with |ρ| > 0.98 relative to higher-ranked features should be eliminated.

E. Classifiers and Baselines

1. **Primary classifier (HMDFE)**: **SVM (RBF kernel)**
   Class-balanced weights; probability estimates enabled; C ∈ {1, 3, 10}, γ ∈ {1e-3, 3e-3, 1e-2}.
2. **Baselines**

- **DeepConvLSTM**: LSTM (128), dropout 0.5, Adam 1e-3, batch 128, 50 epochs, early halting (patience 7), 4 Conv blocks (filters=64, kernel=5).
- **Statistical+SVM**: The SVM (RBF) is configured as previously, with only temporal statistics that is subset of T.
- **PCA+CNN** (previously noted): 1D-CNN (64-64-128) with GAP and softmax → PCA (95%).

F. Protocols for Training and Validation

1. **Within-subject**: The window is split into 70/15/15 that is organized as per activity.
2. **Subject-independent** (primary): For UCI HAR Leave-One-Subject-Out (LOSO) protocol is desired and the method of k-fold by means of subject (where k=5) is for WISDM/PAMAP2.
3. **Cross-dataset**: In this study, the dataset is trained on UCI HAR, and tests on WISDM and PAMAP2.
4. **Hyper-parameter selection**: The embedded CV is used for parameter selection on training folds only while optimal configuration is achieved by Macro-F1.
5. **Random seeds**: For the following random seeds {13, 21, 42, 77, 99}, the mean of ± standard deviation is used.

G. Evaluation Metrics & Statistical Testing

In this study, following metrics and standard testing are adopted.

- Macro-F1 and accuracy are used for the primary; per-class F1, confusion matrix, ROC-AUC (one-vs-rest), and macro-precision/recall are for the secondary.
- Significance: Holm-Bonferroni correction method is employed in the comparisons of techniques, utilizing paired Wilcoxon signed-rank (per-fold Macro-F1), α=0.05.
- Cliff's delta is the effect size.

V. RESULTS AND DISCUSSION

In this research, the suggested HMDFE framework surpassed both traditional and deep feature-based models, with 96.8% accuracy on UCI HAR, 94.5% on WISDM, and 93.7% on PAMAP2. Morphological and correlation factors made the biggest difference in finding transitional behaviors (such going from sitting to standing) more accurately. Feature fusion made robustness better when sensors were noisy.

Table III. Performance Comparison (Accuracy / F1-score)

Method	UCI HAR (%)	WISDM (%)	PAMAP2 (%)
Proposed HMDFE	96.8 / 96.2	94.5 / 94.1	93.7 / 93.4
DeepConvLSTM	92.4 / 91.8	90.2 / 89.5	88.9 / 88.2
Statistical + SVM	88.9 / 88.2	85.7 / 85.0	83.6 / 82.9

Figure 3. Sample morphological profiles and spectrograms for particular activities, such as walking vs sitting.

These pictures show the unique patterns in time-frequency and shape that the proposed features were able to find. The table above shows how different methods did on HAR Benchmarks when tested across multiple datasets. The proposed Hybrid Multidomain Feature Extraction (HMDFE) method always beats the current baselines on all of the test datasets. The UCI HAR dataset shows that HMDFE is more than 4% better than DeepConvLSTM, with an accuracy/F1-score of 96.8 / 96.2. HMDFE shows good generalization on smartphone-based motion data on WISDM and stays very stable with 94.5 / 94.1.

The model gets 93.7 / 93.4 on PAMAP2, which has hard tasks and sensors that work in more than one way. This shows that it is strong in cross-domain recognition.

The findings validate that the integration of temporal, statistical, and spectral features enhances the discriminative representation of human activity signals, facilitating superior generalization across diverse sensor modalities.

Table IV. Impact of Each Feature Domain on Overall Activity Recognition Performance

Configuration	Accuracy (%)	Macro-F1 (%)
Full HMDFE	96.8	96.2
Without Morphological	94.1	93.6
Without Correlation	93.7	93.2
Time-Frequency Only	91.5	90.8

Fig. 4. The effect of feature domains for chosen activities, for e.g.walking versus sitting, on recognition accuracy.

This analysis illustrates how each feature subset contributes to the proposed Hybrid Multidomain Feature Extraction (HMDFE) framework.

The model’s performance drops by about 2–3% when morphological and correlation-based descriptors are taken out. In order to understand the spatial-temporal context, both of these traits are important. With an accuracy of only 91.5%, the time-frequency-only configuration is the least accurate. This shows that combining features from different domains makes recognition much better. The HMDFE's strength and even accuracy across different activity classes are shown by its 96.8% accuracy and 96.2% macro-F1.

Observations:

- HMDFE frequently outperforms conventional and deep feature methodologies.
- Morphological traits, especially, made it easier to switch between sitting and standing.
- Adding feature fusion made it stronger when sensors were noisy.

VI. CONCLUSION

For reliable Human Activity Recognition, we introduced a novel multi-domain feature extraction framework (HMDFE). The technique captures the distinct temporal-spatial patterns of human motion by integrating spectral, morphological, and correlation-based information. The method maintains interpretability and minimal computing cost while achieving state-of-the-art accuracy on several benchmark datasets. This work shows that the distinctive aspects of human activity signals can be efficiently captured by multi-domain hybrid features.

Subsequent research will investigate including self-supervised learning for unlabeled data, expanding to real-time mobile deployment, and using attention-based weighting to optimize feature selection.

REFERENCES

[1] D. Anguita et al., "A Public Domain Dataset for Human Activity Recognition Using Smartphones," ESANN, 2013.

[2] F. J. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," Sensors, 2016.

[3] H.I. Fawaz et. al., "Deep Learning for Time Series Classification: A Review," Data Mining and Knowledge Discovery, 2019.

[4] O. Banos et. al., "Window Size Impact in Human Activity Recognition," Sensors, 2014.

[5] Y. Xu et al., "Transformer-Based Sensor Fusion for," HAR, IEEE Access, 2021.

[6] M. Zeng et al., "Multi-domain Feature Learning for Robust Activity Recognition," Pattern Recognition Letters, 2022.

[7] Z. Liu et al., "Contrastive Multi-Representation Learning for Wearable HAR," IEEE IoT Journal, 2023.

[8] Wee-Hong et al., "Investigation of Feature Extraction for Unsupervised Learning in Human Activity Detection," Bulletin of Networking Computing, Systems, and Software, 2013.

[9] Zhu et al., "Feature extraction for robust physical activity recognition," Human-centric Computing and Information Sciences, 2017.

[10] Sadhna et al., "An Architecture for Human Activity Recognition Using TCN-Bi-LSTM HAR based on Wearable Sensor," Procedia Computer Science 260, 2025.

[11] Thakur et al., "An Integration of feature extraction and Guided Regularized Random Forest feature selection for Smartphone based Human Activity Recognition," Journal of Network and Computer Applications 204 , 2022.
[12] H. Han et al., "Multilevel features cascade fusion network for infrared video human behavior recognition," Displays 87 , 2025.
[13] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," IEEE Trans. Pattern Anal. Mach. Intell. , 2016.
[14] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," In: Proc. of IEEE, 2016.
[15] Z. Wang, S. Liu, J. Zhang, S. Chen, and Q. Guan, "A Spatio-temporal CRF for human interaction understanding," IEEE Trans. Circuits Syst. Video Technology, 2017.
[16] J. Liu, G. Wang, L. Y. Duan, K. Abdiyeva, and A. C. Kot, "Skeleton based human action recognition with global context-aware attention LSTM networks," IEEE Trans. Image Process., 2018.