Array 15 (2022) 100190

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

Stochastic recognition of human daily activities via hybrid descriptors and random forest using wearable sensors

Nurkholish Halim

System & Technology Architect, PT. Bank OCBC-NISP, Tbk, Indonesia

A R T I C L E I N F O

INDEX TERMS:
Wearable sensors
Random forest (RF)
Human daily activity recognition (HDAR)
Stochastic gradient descent

A B S T R A C T

Human daily activity recognition (HDAR) using wearable sensors is an important task for researchers aiming to develop an effective and feasible model which is capable of accurately detecting human motion patterns. These applications provide elderly care, surveillance systems, and wellness tracking. Despite the pervasive use, recognition and monitoring of human physical activities remains inaccurate, which may contribute to negative reactions and feedback. This paper addresses a data-driven approach to recognizing human daily activities in an indoor-outdoor environment. To improve the classification and recognition of human life-log activities (for example, walking, drinking, and exercising), a model is introduced that incorporates pre-processing (such as denoising), hybrid features extraction from four domains, including time, frequency, wavelet, and time-frequency respectively. After that, stochastic gradient descent is exploited to optimize the selected features. The optimal extracted features are advanced to random forest classifiers in order to develop adaptive for human life-log activities. Additionally, the proposed HDAR model is experimentally evaluated on three benchmark datasets, namely, USC-HAD, which is comprised of 12 physical activities, IM-WSHA, which involves 11 life-log activities, and MOTIONSENSE which contains six static and dynamic activities, respectively. The experimental results show that the proposed HDAR method significantly achieves better results and outperforms others in terms of recognition rates of 91.08%, 91.45%, and 93.16% respectively, when the USC-HAD, IM-WSHA, and MOTIONSENE databases are applied.

1. Introduction

Automatic human daily activity recognition (HDAR) has become an active topic of body-worn sensor based behavioral study in the last decade due to its significance in many real-world applications, including ergonomics, fitness tracking, elderly care, surveillance, security, and sports, etc [1–5]. Additionally, the ubiquity and usability of wearable-based inertial sensors in smart watches, fitness monitoring bands, and smartphones embedded with fused sensors (e.g., accelerometers, gyroscopes, and magnetometers) open up new opportunities for real-time monitoring of human daily life activities. The utilization of these wearable sensors for monitoring studies has been in practice and demand, but there is still a need for a system capable of recognizing activities with limited contextual information [6]. In essence, activity data collected from multiple dimensions are analyzed in order to recognize and monitor various motion patterns and behaviors. As a result, recognition of daily living activities, including walking, standing, sleeping, and cooking, is particularly pivotal for smart homes and elderly care monitoring. However, the system can not always accurately reflect human activity detection, which gives rise to complex movements and understanding behaviors.

Wearable sensors have transformed every aspect of human daily life, from e-health care to personal living comfort. With the advancement of wearable-based inertial sensors, these sensors are playing a significant role in our daily lives by allowing us to access our environments such as temperature, humidity, and thermostat, etc. Fortunately, greater affordances, capabilities, and features are emerging with advancements in wearable sensors. Additionally, these wearable-based inertial sensors have resulted in significant demand in research for healthcare, security and surveillance-based systems, wellbeing, and biofeedback systems [7,8]. Furthermore, each of these real-world applications involves real-time and continuous monitoring. These applications enable vital access to information about the wellbeing of vulnerable elderly individuals and children by incorporating multiple wearable-based inertial sensors on different parts of the body [9,10]. Similarly, security and surveillance-based systems are capable of detecting potentially hazardous or odd events in their vicinity and notifying emergency response teams immediately. With regard to fitness tracking, wearable inertial

E-mail address: nurkholish.h@ocbcnisp.com.

https://doi.org/10.1016/j.array.2022.100190
Received 8 May 2022; Accepted 18 May 2022
Available online 23 May 2022

sensors can provide continuous monitoring for daily fitness in order to make training more effective and efficient. Intelligent home-based systems provide continual physical and cognitive stimulation that aids children's growth and enhances their learning capabilities. In the context of healthcare, biofeedback treatment works well with VR-based systems to measure and assess the functioning of the body and to monitor changes in physiological functions, including blood pressure and heart rate. Additionally, these systems can also be useful for anxiety and stress reduction techniques.

Recently, we have a variety of inertial sensors at our disposal, such as accelerometers, gyroscopes, and magnetometers, to facilitate us in recognizing environmental changes [11,12]. The fusion of these sensors enables these devices to extract vital information about human complex body patterns in three-dimensional space. For example, an accelerometer is a type of sensor generally used for HAR and can measure both static (such as gravity) and dynamic (vibrations) acceleration forces impacting on the sensors, providing relevant data for the detection of complex motion patterns. The gyroscope is another popular sensor that can measure angular velocity. In the context of wearable sensors, magnetometer sensors analyze the relative changes in existing magnetic field intensity and provide compass calibration information. In general, these three sensors are fused to enhance the capability of human activity monitoring. The augmentation of accelerometers, gyroscopes, and magnetometers into a single device is stated as an inertial sensor or inertial measurement unit (IMU) [13].

Therefore, the objective of this paper is to deal with the classification of human daily activities (walking, cooking, sleeping, drinking, and ironing, etc.) via wearable-based inertial sensors that detect complex motion patterns in nine degrees of freedom (9-DoF). The system detects changes in the position, rotation, and orientation of the body in three-dimensional space in order to determine static postures and dynamic motion patterns. The main aim of this work is to improve recognition and detection while lowering the complexity of the systems needed to recognize human daily activities. The proposed human daily activity recognition involves four main steps: data acquisition and signal denoising, hybrid feature extraction and selection, optimization, and classification. Initially, inertial data is acquired and denoised with a third-order median filter in order to reduce the noise ratio of the inertial data. Then, we adopted hybrid feature extraction techniques (i.e., time domain, frequency domain, wavelet features, and time-frequency domain features) from the denoised data. In the context of features, these hybrid descriptors are further optimized with stochastic gradient descent (SGD) in order to select the optimal descriptors for further classification. Finally, the model incorporates a random forest classification algorithm to recognize and preset parameters for classifying human daily activities from optimal feature vectors in order to attain a significant recognition rate. To evaluate the performance of our proposed HDAR system, we applied our proposed model to three public benchmark datasets: the IM-Wearable Smart Home Activities (IM-WSHA), the University of Southern California Human Activity Dataset (USC-HAD), and the MotionSense datasets. The major contributions and highlights of this study is summarized as follows:

- Hybrid descriptor techniques are adapted from diverse domains, including the time domain, time-frequency domain, frequency domain, and wavelet domain.
- In order to deal with the complex human activity patterns and to improve the recognition rate of all three datasets, we proposed an SGD feature selection-based random forest model that provided contextual information coupled with classifying activities.
- Additionally, a detailed analysis was carried out on three public benchmark datasets, namely, IM-WSHA, USC-HAD, and MotionSense datasets for human daily activities. The experimental results display a higher recognition rate for the proposed methods than for other state-of-the-art methods.
- Furthermore, we also compared the performance with other state-of-the-art classifiers for further HDAR analysis.
- We also provide baseline literature information that other studies can employ to analyze potential fusion techniques.
- The remainder of the article is structured as follows. In Section II, we briefly review the related work on inertial sensor fusion methods for HDAR and vision-based sensors. Section III addresses the details of the proposed architecture of our HDAR model. In Section IV, we evaluate the experimental setup for three public benchmark datasets and also present the empirical comparison of the various metrics for that purpose.
- Finally, Section V summarizes the findings of the study with a conclusion.

2. Related work

Various related works exploited machine learning methods for human daily activity recognition using a set of multiple sensors, including image and inertial sensors [14]. This section comprehensively summarizes previously and currently conducted researched HDAR analysis via vision sensors and wearable-based inertial sensors.

2.1. Image sensor-based HDAR analysis

In image-based HDAR systems, many researchers have exploited image and video sensing technologies from multiple sensors, including RGB and RGB-D sensors, which are primarily used in security and surveillance systems for the tracking and detection of 3D movements of humans. Espinosa et al. [15] proposed a fall detection system via a 2D convolution neural network (CNN) and multiple vision sensors. They demonstrated an approach for feature extraction via the fixed window method based on time and the optical flow method. Additionally, they tested the proposed approach only on a single public dataset. Experimental results have shown better performance compared with other state-of-the-art methods, but in order to evaluate the whole performance of the proposed approach, multiple datasets should be introduced. In Ref. [16], Anitha et al. presented image processing techniques for human action recognition. Initially, video sequences of individuals walking, jogging, and hand waving are converted into 2D frames. Then, these obtained frames are denoised followed by feature extraction via Laplace smoothing transform (LST). Finally, k-nearest neighbor (KNN) is utilized for classification tasks. The main limitation of this work is that the KNN classifier performs poorly with high-dimensional data. Additionally, it also takes an excessive amount of time to calculate the distance between datasets points, both of which significantly reduce the model's performance. Sharif et al. [17], proposed a model for a human action recognition system. Their proposed model comprises two phases. Initially, various human motion regions are detected via a fusion method which is based on uniform distributions and expectation maximization (EM) segmentation. Secondly, augmented features from the video segments from multiple databases involving histogram of gradients (HOG), local binary patterns (LBP), and Harlick features are extracted. Finally, modified joint entropy along with the Euclidean-based feature selection techniques are combined with a multi-class SVM algorithm in order to study human daily activities in real-time. The proposed model has produced better results in low-dimensionality datasets. However, to make the proposed model more diverse, the database dimension should be increased. Additionally, the proposed model is affected by varying lighting conditions, which results in segmentation accuracy. In Ref. [18], Hu et al. proposed a hierarchical model using Kinect sensors for human interaction recognition. Additionally, this hierarchical model is used to determine the most salient features of multiple-person interactions. Therefore, at the most upper level, an interaction is split into two types of atomic actions, including salient and non-salient actions. Furthermore, at the lowest level, a salient is presented to detect the joint that experiences the largest displacement. Finally, the hierarchical model has been evaluated against an SVM based multi-class classifier. Kong et al. [19], introduced a discriminative model for classifying actions in a partial video sequence. Their proposed model captured the whole evolution action through time and also involve the spatiotemporal nature of a video sequence. But the main drawback is that it requires all predefined rules for human daily interaction activities. In Ref. [20], Ince et al. designed a biometric-based system to monitor human daily activities in 3D space using multiple skeleton joint angle patterns. Additionally, this system utilizes an RGB-depth camera, which seems to be ideal for surveillance-based systems and in a healthcare environment. But the main limitation of the proposed model is not enough to deal with the false skeleton tracking induced incorrect angle calculations which lead to imprecise classification. Wang et al. [21], deals with a probabilistic based graphical model for real-time monitoring of human daily living activities. Initially, they split the model into two parts, including discriminative boundaries and subtractive transitions. Additionally, they address the segmentation problem for human activities. However, these methods function only in offline mode.

2.2. Wearable inertial sensor-based HDAR analysis

Advancements in wearable sensing technologies assist researchers in building different smart systems to recognize and monitor human daily life activities. Currently, wearable inertial sensors are capable of detecting abnormal and uncertain events and can provide support in times of need. Therefore, in pursuit of human activity monitoring, researchers have exploited multiple sensors integration in order to develop a more effective method of evaluating human locomotion and to enhance living comfort. In a comparative study, Shahar et al. [22] analyzed the four inertial sensors that comprise accelerometer and gyroscope signals embodied in the different body positions, including left and right wrists, chest, and waist. Additionally, statistical features are extracted for hockey playing activities, such as mean, minimum and maximum peak, and standard deviation features. Their model significantly achieves better performance. But the main limitation of this work is that the model is optimally suited for sports activities. Additionally, extracted statistical features might not be ideal as non-optimal descriptors will be acquired that do not provide optimal performance in a real-time environment. In Ref. [23], Uddin et al. introduced a feature selection technique based on the guided random forest for human daily activity recognition. Initially, a guided random forest algorithm is trained on the publicly available dataset to obtain key scores for the descriptors. The chosen scores are then incorporated into the feature selection phase. The guided random forest algorithm enables the optimal selection of descriptors that contribute to the human activity recognition model. Feng et al. [24], proposed a random forest ensemble technique for recognizing HDAR via multiple wearable inertial sensors. Additionally, the augmentation of random forest classifier delivers better monitoring capabilities for wearable sensor-based healthcare models. In Ref. [25], Jing et al. proposed a model for recognizing human daily living activities along with fall detection using multiple wearable inertial sensors. Additionally, the entire model is compared against the whole activity set comprising of static, random, and periodic actions. Furthermore, multiple features are extracted in the time and frequency domains. The limitations of their proposed method were only tested on a small number of individuals when monitoring activities and behaviors. Abidine et al. [26] introduced a weighted support vector machine (SVM) for monitoring life-log activities in a smart home environment. Additionally, they addressed various implementation issues with the HAR methods, including redundant feature descriptors and imbalance classes in the training data. To address these concerns, they presented a framework for recognizing human life-log activities in a smart home environment. Their framework is a hybrid of weighted SVM, principal component analysis (PCA), and linear discriminant analysis (LDA). Initially, the training data set is minimized via the LDA and PCA descriptors in order to attain optimal feature descriptors. Then, for each class, weighted SVM is utilized to handle unbalanced life-log activity datasets to enhance the recognition accuracy. In Ref. [27], Cillis et al. developed a pervasive solution for four human locomotion patterns, including walking, standing, and ascending and descending stairs via a wearable inertial-based triaxial accelerometer sensor. Their proposed solution incorporates minimum feature descriptors to recognize and detect four distinct locomotion patterns. The experimental results indicate a better recognition rate when dealing with static activities, but a lower rate for dynamic activities such as ascending and descending stairs.

3. Proposed solution framework

The proposed HDAR system recognizes human physical activities by exploiting three wearable inertial sensors from accelerometer, gyroscope, and a magnetometer. Initially, multi-fused signals are acquired from three inertial sensors and are filtered via a third-order median filter in order to deal with noisy signal segments emanated by abrupt motion movements. Following the denoising phase, the filtered inertial signal values are retained to provide robust and effective feature components. Thirdly, we extracted hybrid feature descriptors from different domains, including time, frequency, wavelet, and time-frequency features. Additionally, the extracted features are scaled using extreme values to avoid the possibility of any complex signal values emerging during the later phases of feature selection. Thereafter, retrieved feature descriptors are optimized by the stochastic gradient descent (SGD) algorithm for optimal feature selection. Finally, in the wearable HDAR, the optimized features acquired from SGD are catered to a random forest (RF) classifier to classify human physical activities in real-time. The proposed HDAR architecture of our complete model is depicted in Fig. 1.

3.1. Signal processing and denoising

IMU is highly sensitive to even the least amount of fluctuations, so any unintended shift is able to affect the signal shape entirely and also disrupt the feature descriptor extraction and optimization phase. Therefore, during the signal processing and denoising phase, we initiated by analyzing the noise interference associated with the inertial signals. After analyzed the abnormal signal intensity in the frame data, we opted for a third-order median filter, a denoising method that deals with the rigorous motion pattern in the framed signal without comprising any significant information.

Additionally, the third-order median filter adjusts the signal's shape to closely follow normal motion patterns. Furthermore, we also evaluated the performance of different filters by exploiting Gaussian and moving average filters to lessen the noise associated with the IMU signals. However, we adopted the third-order median filter, which delivered enhanced results when compared to the other two filters, i.e., Gaussian and the moving average filter. The raw signal and filtered signal components of the median, Gaussian, and moving average filters of the IMU sensor are illustrated in Fig. 2.

While in the final phase of signal processing and denoising, the inertial signal values are normalized by taking the signal scale into an account along with ignoring complex signal values for feature extraction.

3.2. Hybrid feature descriptor methods

Feature extraction is a key component in any machine learning-based system since it focuses on modeling data with a relevant set of attributes that encompasses the entire scenario. In this step, we proposed a hybrid feature model from four major domains to get meaningful features. These features include wavelet, time, frequency, and time-frequency domain feature descriptors.

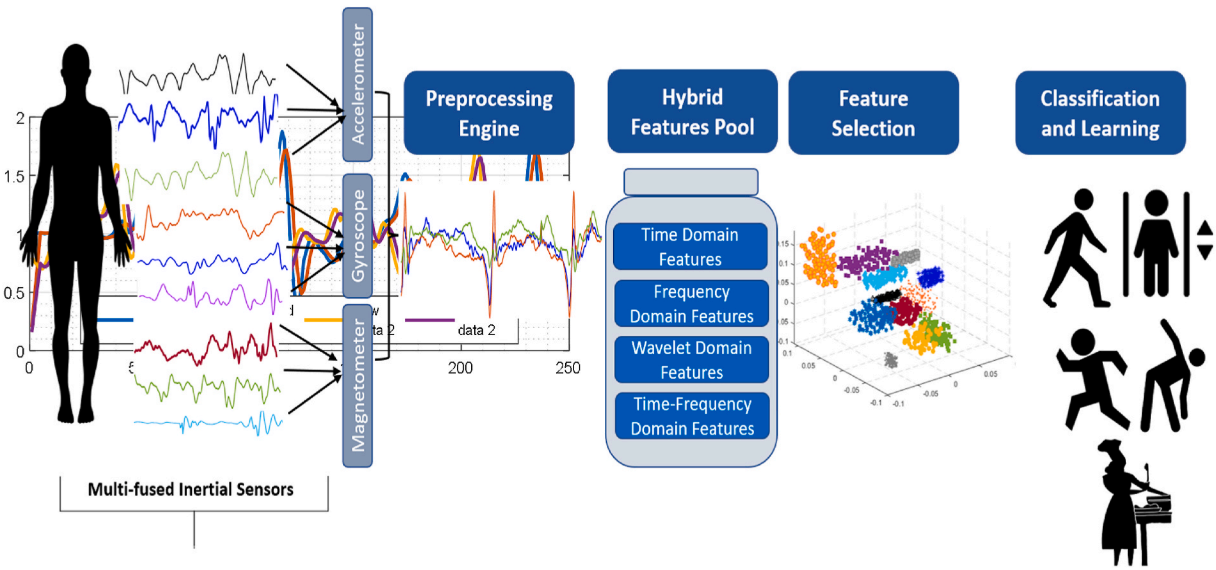

Fig. 1. System architecture of the proposed HDAR model.

3.2.1. *Statistical features*

The statistical attributes Sv of each frame indicate the mean, median, mode, and maximum/minimum values of the inertial signal. These feature descriptors play a significant role in analyzing the overall variations that occur as a result of each n activity as:

$$\mathrm{Sv} = \sum_{r=1}^{n} \frac{a_r}{n}, \frac{\sum_{r=1}^{n}(V-\overline{V})^2}{n-1}, mi(\mathrm{sig})(M_i), mx(\mathrm{sig})(X_i) \tag{1}$$

where n is the framed data size, a is the coefficients in the whole vector, V indicates the value of initial vector and $\overline{V}$ depicts the mean of all the framed data. Fig. 3 illustrates a triaxial plot with the augmentation of different statistical feature descriptor of walking physical activity via USC-HAD dataset (see Fig. 4).

3.2.2. *1D Haar wavelet transform (HWT)*

In the field of image and signal processing, the Haar wavelet transform (HWT) has emerged as a state-of-the-art technology. Generally, wavelets are based on mathematical methods for partitioning functions hierarchically. In our proposed approach, the Haar feature descriptor is utilized for pattern detection at a particular interval to analyze signal variability. Additionally, HWT has a wavelet-like structure, which makes it an efficient and robust signal processing tool [28,29]. Also, HWTs are represented by their indices (a, d), where 'a' indicates approximation coefficients and 'd' depicts detail coefficients. These coefficients also assists in the estimation of the strength of the IMU signal and moreover contribute into accurate restoration and segmentation. The Haar Wavelet Transform (HWT) can be defined as:

$$\psi(a) = \begin{cases} 1 & 0 \le a \le \frac{1}{2} \\ -1 & \frac{1}{2} \le a \le 1 \\ 0 & else \end{cases} \tag{2}$$

where scaling function is denoted as $\psi(a)$.

3.2.3. *Hilbert Huang transform (HHT)*

The Hilbert Huang Transform (HHT) is acknowledged as quite effective when dealing with non-linear and diverse signal data. In our case, we have a source of time series data. To deal with particular physical activities, including walking, running, which involve repeated patterns with continuously shifting frequency and amplitude. Additionally, the acquired inertial data from sensors is mostly nonlinear in nature.

The Hilbert Huang transform (HHT) splits the obtained time series of non-linear inertial data into distinct repeated components referred to as intrinsic mode functions (IMF's), and the entire process is known as intrinsic mode decomposition (IMD) [30]. Additionally, both these components produce different frequency bands and they are capable of measuring shifts in instantaneous frequencies (IF). Thus, we can evaluate the features of various activities in a meaningful way. The processed data P(t) can be determined by:

$$P(t) = \sum_{i=1}^{n} c_i + r_n \tag{3}$$

where P(t) depicted the processed signal, c_i denotes the ith IMF, and r_n is the total remainder.

3.2.4. *Wavelet packet entropy*

Wavelet packet entropy (WPE) is an efficient and robust approach for IMU signals in time-frequency domain. Firstly, WPE decompose an IMU signal into various frequency resolutions comprising detail and approximation coefficients [31]. The mathematical equation (4) of WPE can be written as:

$$d_p = \begin{cases} d_{0,0}(t) = p(t) \\ d_{i,2j-1}(r) = \sqrt{2}\sum_c h(c) d_{i-1,j}(2r-c) \\ d_{l,2j}(r) = \sqrt{2}\sum_c g(c) d_{i-1,j}(2r-c) \end{cases} \tag{4}$$

where h(c) and g(c) are two filters utilized to extract ACs and DCs, and $d_{i,j}$ depicts the reconstruction inertial signals at the ith and jth node.

3.2.5. *LEMPEL-ZIV complexity (LZC)*

The LZC method is a symbolic sequence technique based on coarse gaining estimation [32]. In our proposed approach, the inertial signal is first transformed into a finite-length sequence of symbols. Then, the transformed inertial signal is further converted into a binary sequence. Additionally, binary sequences with a median threshold T_m were employed to ensure stability of outliers. The LZ complexity along with the median threshold can be shown in equation (5) as follows:

$$\text{Inertial} \quad \mathrm{sig}(i) = \begin{cases} 1, & if\ s(i) \ge T_m \\ 0, & if\ s(i) < T_m \end{cases} \tag{5}$$

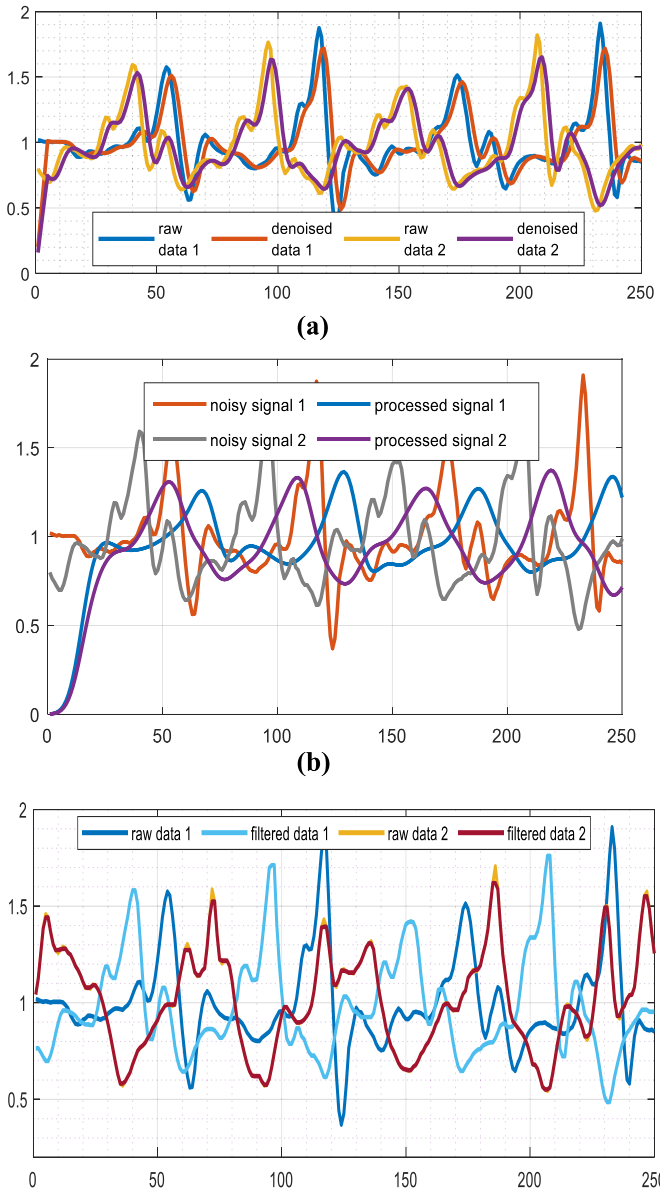

Fig. 2. Signal Processing and Denoising. IMU sensors with raw (unfiltered) and denoised signals for walking activity via (a) Median filters, (b) Gaussian Filter, and (c) Moving Average Filter on the IM-WSHA dataset.

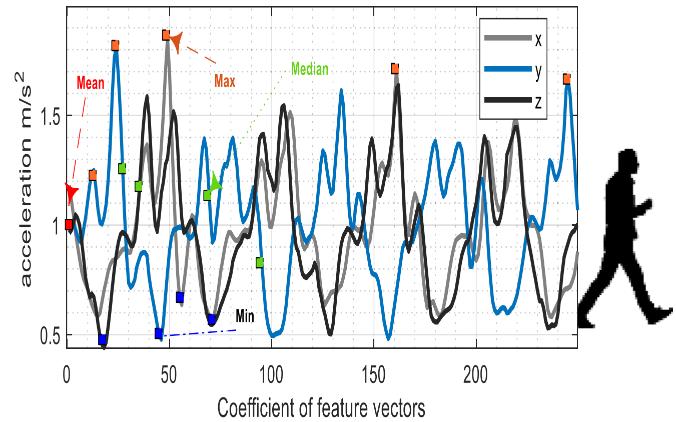

Fig. 3. Triaxial vector plot of statistical features of the walking physical activity via USC-HAD dataset.

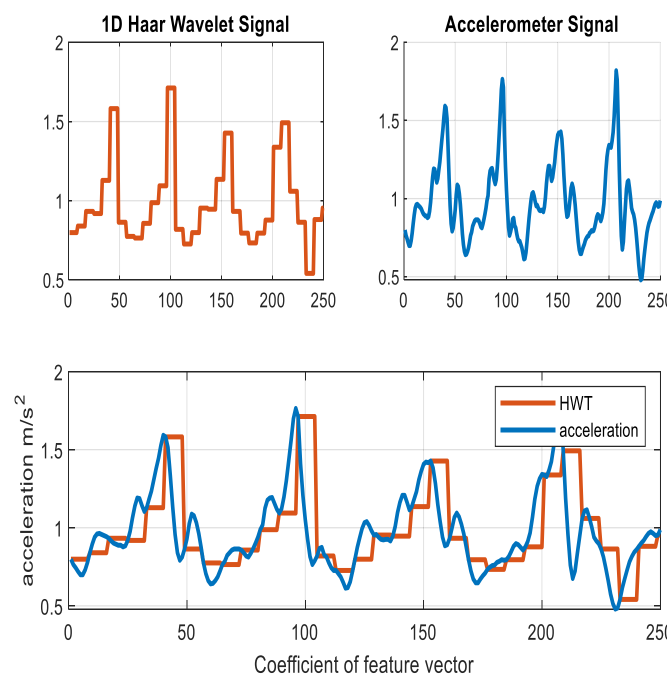

Fig. 4. 1D- Haar inertial signal feature in a vector plot over physical activity (walking) of USC-HAD dataset.

Fig. 7 depicts the relationship between signal frequency and Lempel-Ziv Complexity (see Fig. 6) (see Fig. 8) (see Fig. 9) (see Fig. 5).

3.3. Hybrid feature selection and classification

In the proposed HDAR approach, feature descriptors are optimized via a well-known optimization algorithm, stochastic gradient descent (SGD). Then, the acquired optimized vector is catered to the random forest (RF) algorithm. The results of random forest are compared with multilayer perceptron and support vector machine (SVM).

3.3.1. Optimization via stochastic gradient descent (SGD)

Gradient descent is an effective approach for finding the best solution with a minimum cost function with a linear function. Gradient descent was initially used in neural network research in order to update network gradients. However, the gradient descent method may run slowly if all of the training data is processed at each epoch or if the training data set is extensive. To address this issue, we introduced the Stochastic Gradient Descent method (SGD) with minibatch as an optimizer that does not consume all of the training data [33]. However, minibatch SGD paired with random data selection reduces the cost and variability associated with conventional stochastic gradient descent. As a result, minibatch needs careful consideration while employing adaptive learning rates with initial parameters to achieve the lowest loss function. Thus, the learning parameters are corrected and the output is acquired dependent on the learning rate. Therefore, initially, a 0.0100 learning rate is set and the total number of iterations is preset at 1000, tuned via regularization parameters, the number of passes over the training set.

The stochastic gradient descent (SGD) of all training set for $a^{(k)}$ and $b^{(k)}$ is represented as:

$$\theta = \theta - \eta \cdot \nabla_\theta J\left(\theta; a^{(k)}; b^{(k)}\right) \tag{6}$$

where η denotes minibatch size, and the minimum loss function is represented as:

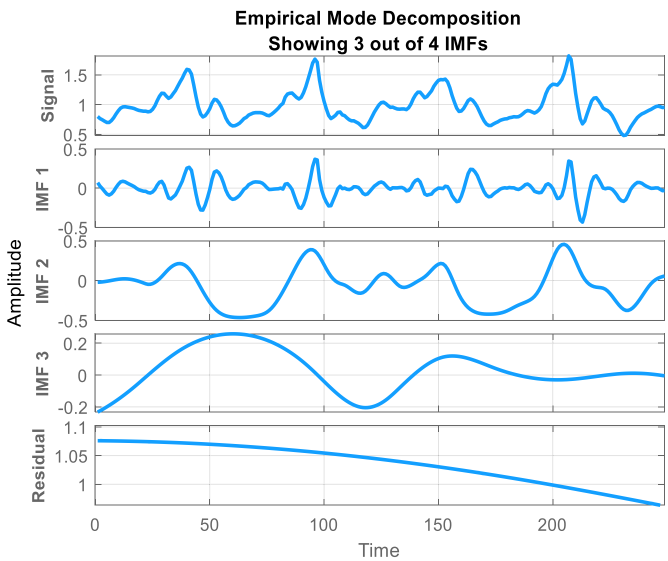

Fig. 5. Empirical model decomposition (EMD) from the inertial signal over USC-HAD dataset.

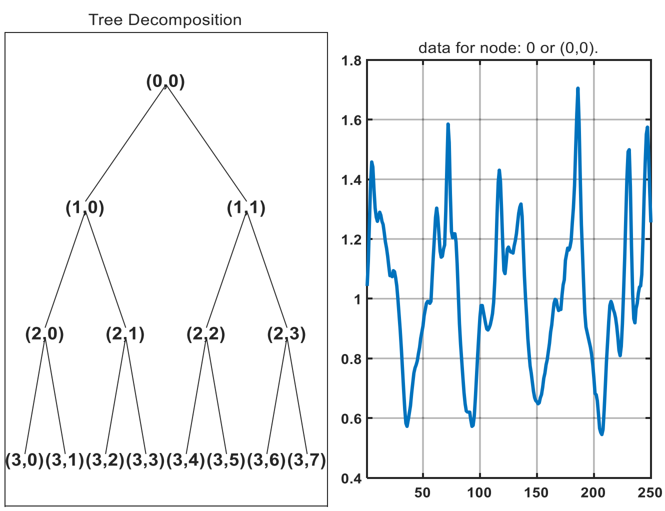

Fig. 6. Two level tree wavelet decomposition for an inertial signal data for exercise data over IM-WSHA dataset.

$$\theta = \theta - \eta \cdot \nabla_{\theta} J\left(\theta; a^{(k:k+n_{bs})}; b^{(k:k+n_{bs})}\right) \tag{7}$$

3.4. Random forest

After the hybrid feature selection step, we catered acquired optimal feature descriptors to a random forest (RF) classifier in order to classify human physical activities. Random forest is an ensemble learning method utilized for classification and other related tasks that functions by training a substantial number of decision trees and generating a class that is the mean of the particular trees [34]. Additionally, the random forest comprises a unique variant of bagged trees, a technique for generating a training set. Bagging extracts samples from all three physical activity datasets, including IM-WSHA, USC-HAD, and MOTIONSENSE datasets. For each sample, a model is developed and used to make classification-based decisions. Finally, based on maximum votes, all of the decisions are combined to get the final result and decisions. Mainly, random forest is an ensemble classifier that is both exceptionally accurate and has a rapid training period.

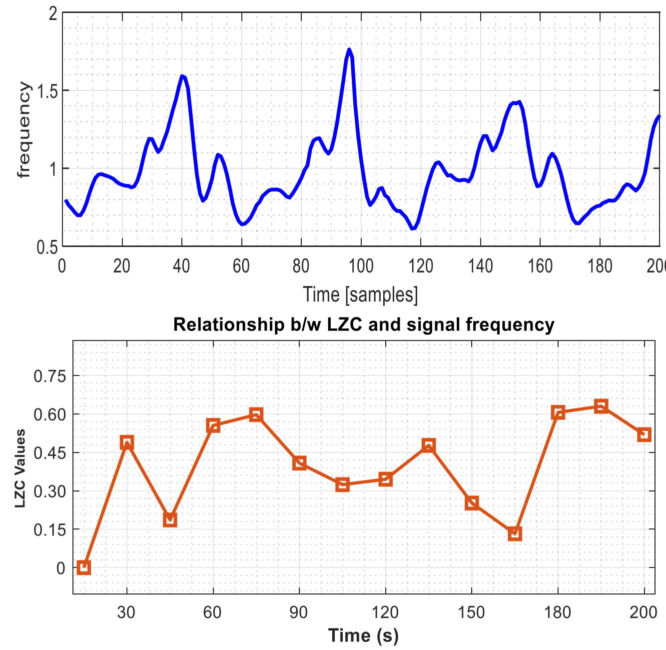

Fig. 7. The result of Lempel-Ziv complexity and signal frequency modulation.

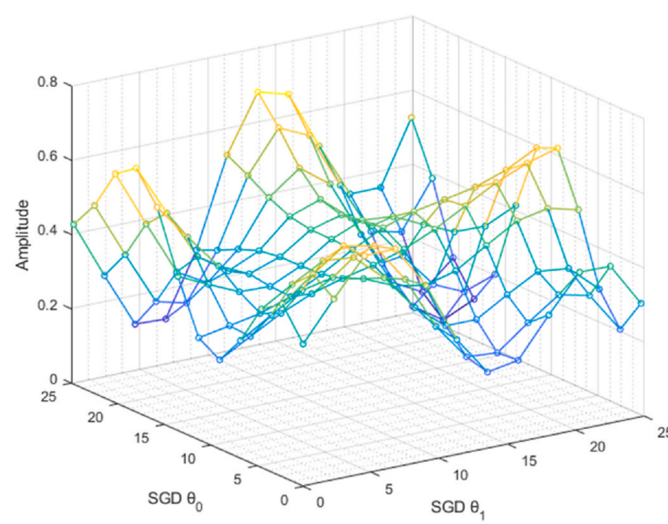

Fig. 8. Stochastic Gradient Descent Optimizer with adaptive learning for a walking forward activity over USC- HAD dataset.

$$\widehat{f} = \frac{1}{R} \sum_{r=1}^{R} f_b(z') \tag{8}$$

where, z' depicts the predictions for random samples. It is estimated by averaging the predictions made by all the various trees on z'.

4. Experimental setting and analysis

All experiments are conducted on a notebook (laptop) equipped with an Intel Core i5-9300H processor running at 2.40 GHz base frequency, 8 GB DDR4 RAM, and a dedicated Nvidia GeForce GTX 1650 graphics card running Windows 10 Home 64-bit and MATLAB and Google Colab (Python) tools. Additionally, a framework has been established to evaluate the performance of our proposed model HDAR on three benchmark datasets, including USC-HAD, IM-WSHA, and the MOTIONSENSE datasets. Furthermore, we also employed the leave-one-

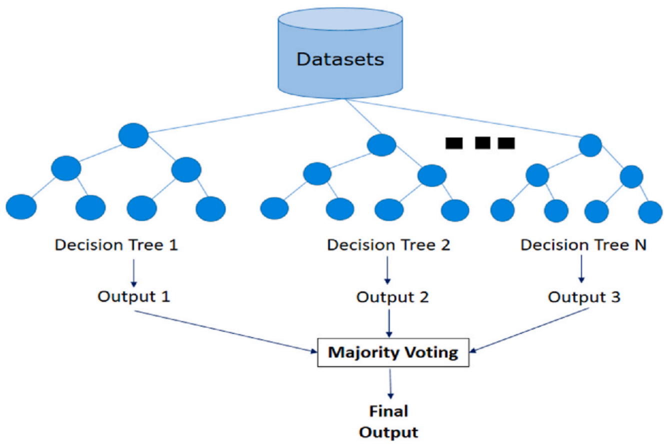

Fig. 9. Proposed HDAR model over Random Forest classifier.

subject-out cross-validation (LOSO) method to evaluate our HDAR system's validation performance in both indoor and outdoor conditions.

4.1. The university of Southern California human activity dataset (USC-HAD)

The USC-HAD dataset [35] is acquired by exploiting a motion node inertial device, which comprises a wearable network with 6° of freedom (DoF) for comprising and tracking three-dimensional motion. Additionally, it is comprised of various sensors, fused with a gyroscope and an accelerometer, that provide real-time orientation information. These inertial sensors are embodied at the front right hip in order to get relevant signal data. A group of 14 individuals engaged in 12 distinct activities (including, *jumping up, standing, sitting, sleeping, walking forward, elevator down, elevator up, running forward, walking left, walking right, downstairs, and upstairs*). The sampling rate of the sensors employed in this experiment is 100 Hz.

4.2. The IM – wearable smart home activities (IM-WSHA) dataset

The IM-WSHA database [36] involves data from three IMU sensors, such as accelerometers, gyroscopes, and magnetometers. These IMU sensors are positioned at three distinct body locations, including the thigh, wrist, and chest, to continuously capture significant features of human life-log activities. Ten respondents, including five males and five females, covered 11 distinct tasks in a smart home scenario (such as, *cooking, ironing, vacuum cleaning, drinking, exercise, phone conversation, reading book, walking, brushing hair, watching tv and using computer*).

4.3. The MOTIONSENSE dataset

The MotionSense dataset [37] is a freely available open-source public dataset that comprises 6-DoF accelerometer and gyroscope sensor data from a smartphone. The volunteer's smartphone is placed in his front pocket. Additionally, a diverse group of twenty-four volunteers, including 14 males and 10 females, covered six static and dynamic tasks in an indoor and outdoor environment, such as sitting and standing, walking and running, downstairs, and ascending.

4.4. Evaluation of parameters using recognition accuracies

In this setup, we analyze the performance of a random forest classifier by catering to optimally selected feature descriptors of wavelet, time, frequency, and time-frequency using the USC-HAD, IM-WSHA, and MOTIONSENSE datasets.

The experiment was conducted three times to assess the performance of the proposed HDAR system against three benchmark datasets. The confusion matrix for the USC-HAD dataset for twelve human physical activities is shown in Fig. 10, where an overall accuracy of 91.08% was attained. In the IM-WSHA dataset, Fig. 11 represents an average accuracy of 91.45% over eleven life-log activities. Furthermore, the confusion matrix in Fig. 12 depicts the MOTIONSENSE dataset over six static and dynamic activities, with a recognition rate of 93.16%.

The results of the comparison between the proposed HDAR approach and other sophisticated methods are shown in Table 1. In Tables 2–4, we compared the performance of the proposed HDAR system with two additional state-of-the-art approaches, including Multilayer Perceptron (MLP) and Support Vector Machine (SVM) classifiers, via accuracy, precision, recall, and F1 scores for all classes in the following datasets: USC-HAD, IM-WSHA, and MOTIONSENSE.

We also analyzed one of the dataset namely USC-HAD. As illustrated in Fig. 13, we can simply discriminate between all 12 classes contained in the USC-HAD dataset.

5. Discussions

Our implementation of the HDAR framework is being developed to attain a high F-measure score and recognition accuracy by considering all three benchmark datasets. In this work, we proposed a robust and effective model that accurately extracts features from different body locations and generates a hybrid set of features. Initially, denoising is carried out over the inertial signals to remove extra motion artifacts and noise. For denoising, we have utilized a third-order median filter to remove the noise ratio without losing any vital information. After signal processing, hybrid feature descriptors are extracted from a different domain to attain better performance. Following that, acquired features are optimized via stochastic gradient descent (SGD) with minibatch in order to select optimal feature descriptors. Finally, various classifiers are used to evaluate the performance of the proposed HDAR system. Additionally, physical activities from three benchmark datasets are classified over the random forest classifier, which has shown a significant recognition rate over other state-of-the-art methods.

The following are the limitations of the proposed HDAR model.

- While incorporating an inertial sensor is a viable approach in publicly available benchmark datasets, however, it may introduce problems associated with the stability and precision of a single sensor.
- One of the drawbacks of the HDAR framework is transfer learning, which allows knowledge acquired by a single inertial sensor positioned on a certain body position to be exploited by numerous

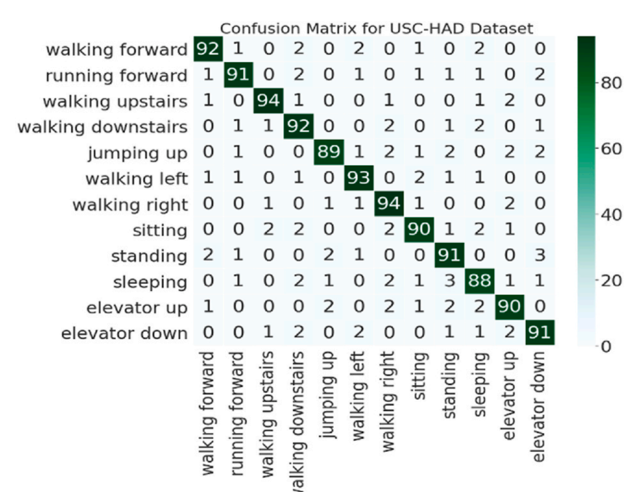

Fig. 10. Confusion Matrix of 12 physical activities on the USC-HAD dataset using Random Forest.

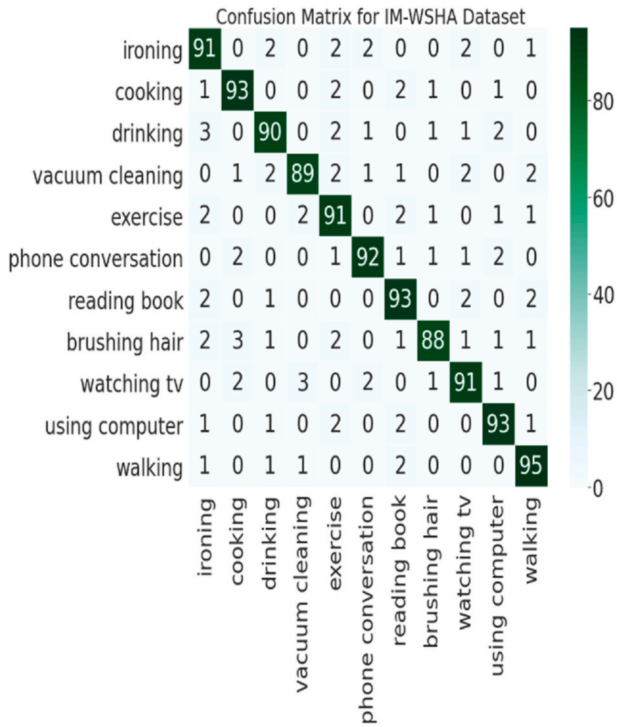

Fig. 11. Confusion Matrix of 11 human life-log activities on the IM-WSHA dataset using Random Forest.

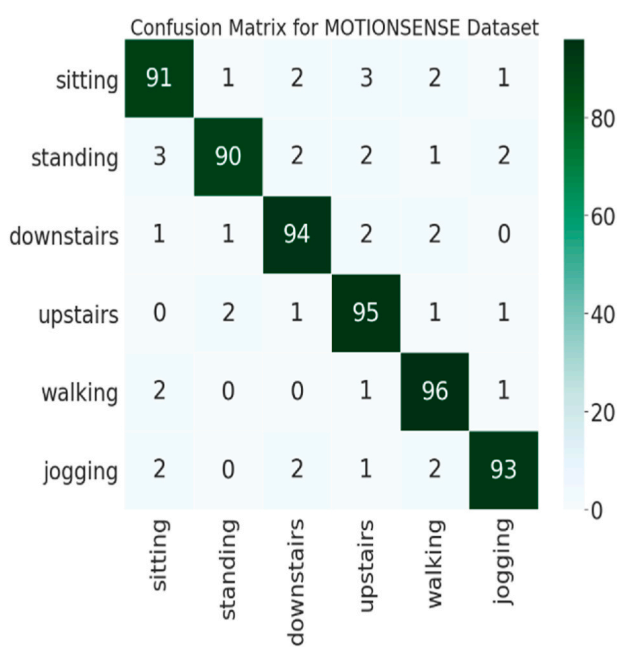

Fig. 12. Confusion Matrix of 6 static and dynamic activities on the MOTIONSENSE dataset using Random Forest.

sensors located in other positions. This prospect for transfer learning to be favored in a variety of situations has not been thoroughly analyzed.

- The proposed approach is limited to monitoring physical activities pattern that exhibit uniformity in the way motion patterns shambles. Any inconsistent information included in the proposed approach disrupts the individual's ability to recognize patterns unless the volunteer makes a smooth or consistent transition.

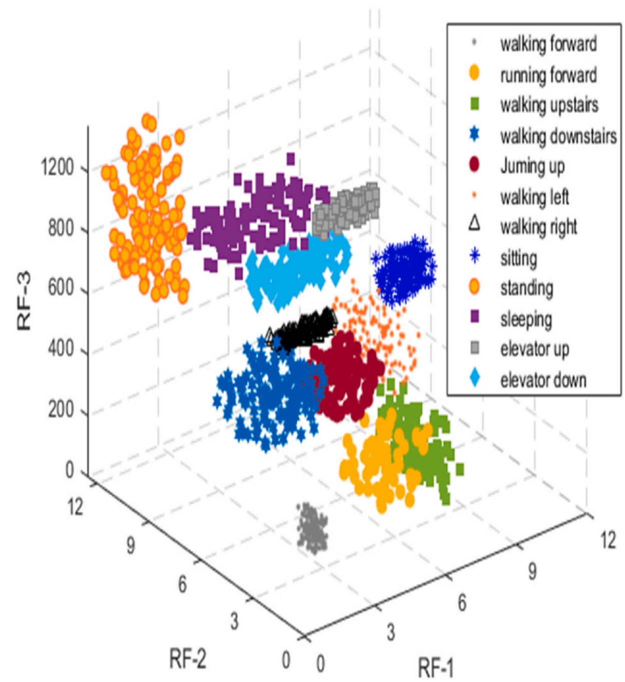

Fig. 13. Random Forest classification on the USC-HAD dataset.

Table 1
Comparison of recognition accuracy of the proposed HDAR method with other state–OF–the-art methods over USC-HAD, IM-WSHA, and MOTIONSENSE datasets.

Methods	USC-HAD (%)	IM-WSHA (%)	MOTIONSENSE (%)
Support Vector Machine	83.64	84.69	87.31
BERT MODEL [38]	–	–	79.86
Multi-fused Features [39, 40]	70	–	88.25
Symbolic Approximation [41]	84.30	–	–
Random Forest [42]	86.90	–	–
Proposed HDAR	91.08	91.45	93.16

6. Conclusion and future work

In this work, an HDAR approach based on hybrid feature descriptors, such as time-frequency, wavelet, frequency, and time features, is proposed. Extracted features are further optimized by stochastic gradient descent (SGD) and classified via ensemble learning-based random forest to improve the recognition rate of human daily living physical activities via IMU sensors. These hybrid-based descriptors assess temporal moments, relevant patterns, invariant and repeated motion patterns to optimize the performance of HDAR based systems. Additionally, this paper also compares the performance of the random forest classifier optimized by SGD with multilayer perceptron (MLP) and support vector machines (SVM). The experiments also revealed the influence of our proposed HDAR model in terms of performance measures such as accuracy, precision, recall, and F1-score. Our proposed HDAR framework assisted in the development of an ideal model of human life-log recognition.

In future work, we will incorporate more complex behaviors and activities from different settings, including healthcare clinics and professional environments, by exploiting multimodal sensors. Additionally, we will also aim to design a self-annotated dataset in complex indoor-outdoor environments.

Table 2
Measurements of evaluation metrics of the proposed HDAR method over USC-HAD dataset.

USC-HAD	Random Forest			Multilayer Perceptron			Support Vector Machine		
Activities	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
U1	0.938	0.920	0.929	0.803	0.740	0.770	0.801	0.742	0.770
U2	0.938	0.910	0.923	0.816	0.760	0.787	0.798	0.740	0.767
U3	0.949	0.940	0.944	0.826	0.750	0.786	0.812	0.760	0.785
U4	0.884	0.920	0.901	0.788	0.720	0.752	0.768	0.720	0.743
U5	0.885	0.890	0.912	0.736	0.760	0.747	0.720	0.770	0.744
U6	0.936	0.930	0.925	0.819	0.780	0.799	0.796	0.780	0.787
U7	0.920	0.940	0.917	0.803	0.770	0.786	0.800	0.775	0.787
U8	0.895	0.900	0.909	0.796	0.800	0.797	0.782	0.810	0.795
U9	0.918	0.910	0.896	0.799	0.750	0.773	0.799	0.740	0.768
U10	0.883	0.880	0.944	0.792	0.800	0.795	0.785	0.790	0.787
U11	0.880	0.900	0.969	0.786	0.880	0.830	0.772	0.750	0.760
U12	0.910	0.910	0.940	0.904	0.880	0.891	0.894	0.780	0.833

Table 3
Measurements of evaluation metrics of the proposed HDAR method over IM-WSHA dataset.

IM-WSHA	Random Forest			Multilayer Perceptron			Support Vector Machine		
Activities	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
W1	0.883	0.910	0.803	0.813	0.770	0.790	0.790	0.767	0.778
W2	0.920	0.930	0.856	0.868	0.790	0.827	0.732	0.722	0.726
W3	0.918	0.900	0.826	0.862	0.780	0.818	0.730	0.700	0.715
W4	0.936	0.890	0.833	0.898	0.810	0.851	0.740	0.690	0.714
W5	0.875	0.910	0.796	0.817	0.775	0.795	0.703	0.690	0.696
W6	0.938	0.920	0.863	0.899	0.820	0.857	0.760	0.730	0.744
W7	0.894	0.930	0.831	0.817	0.770	0.792	0.700	0.695	0.697
W8	0.946	0.880	0.832	0.909	0.830	0.867	0.730	0.740	0.734
W9	0.910	0.910	0.828	0.892	0.860	0.875	0.768	0.730	0.748
W10	0.920	0.930	0.856	0.870	0.790	0.828	0.725	0.720	0.722
W11	0.922	0.950	0.876	0.875	0.800	0.835	0.765	0.740	0.752

Table 4
Measurements of evaluation metrics of the proposed HDAR method over MOTIONSENSE dataset.

MOTIONSENSE	Random Forest			Multilayer Perceptron			Support Vector Machine		
Activities	Precision	Recall	F-measure	Precision	Recall	F-measure	Precision	Recall	F-measure
M1	0.947	0.910	0.928	0.883	0.870	0.876	0.866	0.840	0.852
M2	0.957	0.900	0.927	0.890	0.880	0.884	0.878	0.850	0.863
M3	0.930	0.940	0.935	0.879	0.865	0.871	0.872	0.835	0.853
M4	0.913	0.950	0.931	0.876	0.860	0.867	0.869	0.830	0.849
M5	0.923	0.960	0.941	0.869	0.850	0.859	0.851	0.820	0.835
M6	0.948	0.930	0.939	0.889	0.880	0.744	0.876	0.850	0.862

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ranasinghe S, Al Machot F, Mayr HC. A review on applications of activity recognition systems with regard to performance and evaluation. Int J Distr Sens Netw 2016;12(8). 1550147716665520.

[2] Chatterjee R, Moitra T, Islam SH, Hassan MM, Alamri A, Fortino G. A novel machine learning based feature selection for motor imagery EEG signal classification in Internet of medical things environment. Future Generat Comput Syst Sep. 2019;98:419–34.

[3] Gravina R, Alinia P, Ghasemzadeh H, Fortino G. Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges. Inf Fusion 2017;35:68–80.

[4] Uddin MZ, Hassan MM. Activity recognition for cognitive assistance using body sensors data and deep convolutional neural network. IEEE Sensors J 2019;19(19): 8413–9. https://doi.org/10.1109/JSEN.2018.2871203.

[5] Badar ud din Tahir S, Jalal A, Kim K. Daily life log recognition based on automatic features for health care physical exercise via IMU sensors. In: 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST); 2021. p. 494–9.

[6] Osterland S, Weber J. Analytical analysis of single-stage pressure relief valves. Int J Hydromechatron 2019;2:32–53.

[7] Qin Z, Zhang Y, Meng S, Qin Z, Choo K-K-R. Imaging and fusing time series for wearable sensor-based human activity recognition. Inf Fusion Jan. 2020;53:80–7.

[8] Li R, Lai DTH, Lee W. A survey on biofeedback and actuation in wireless body area networks (WBANs). IEEE Rev Biomed Eng 2017;10:162–73.

[9] Vogel D, Viehland D, Wickramasinghe N, Mula JM. Mobile health. Electron Market 2013;23(1):3–4.

[10] Zhu C, Miao D. Influence of kernel clustering on an RBFN. CAAI Trans Intell Technol Dec. 2019;4(4):255–60.

[11] Bonato P. Wearable sensors/systems and their impact on biomedical engineering. IEEE Eng Med Biol Mag May 2003;22(3):18–20.

[12] Casilari E, Álvarez-Marco M, García-Lagos F. A study of the use of gyroscope measurements in wearable fall detection systems. Symmetry Apr. 2020;12(4):649.

[13] Filippeschi A, et al. Survey of motion tracking methods based on inertial sensors: a focus on upper limb human motion. Sensors 2017;17(6):1257.

[14] Lara OD, Labrador MA. A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutor Third Quarter 2013;15(3):1192–209. https://doi.org/10.1109/SURV.2012.110112.00192.

[15] Espinosa R, Ponce H, Gutiérrez S, Martínez-Villaseñor L, Brieva J, Moya-Albor E. A vision-based approach for fall detection using multiple cameras and convolutional neural networks: a case study using the UP-fall detection dataset. Comput Biol Med Dec. 2019;115.

[16] Anitha U, Narmadha R, Sumanth DR, Kumar DN Kumar. Robust human action recognition system via image processing. Procedia Comput Sci 2020;167:870–7.

[17] Sharif Muhammad, Khan Muhammad Attique, Akram Tallha, Younus Javed Muhammad, Saba Tanzila, Rehman Amjad. A framework of human detection and action recognition based on uniform segmentation and combination of Euclidean distance and joint entropy-based features selection. EURASIP J Image Video Process 2017 2017;1:89.

[18] Hu T, Zhu X, Wang S, Duan L. Human interaction recognition using spatial-temporal salient feature. Multimed Tool Appl Oct. 2019;78(20):28715–35.

[19] Kong Y, Kit D, Fu Y. A discriminative model with multiple temporal scales for action prediction. Proc 13th Eur Conf Comput Vis 2014:596–611.

[20] Ince ÖF, Ince IF, Yıldırım ME, Park JS, Song JK, Yoon BW. Human activity recognition with analysis of angles between skeletal joints using a RGB-depth sensor. ETRI J Feb. 2020;42(1):78–89.

[21] Wang Z, Wang J, Xiao J, Lin K, Huang T. Substructure and boundary modeling for continuous action recognition. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012. p. 1330–7.

[22] Shahar N, Ghazali NF, As'ari MA, M A, Swee TT. Wearable inertial sensor for human activity recognition in field hockey: influence of sensor combination and sensor location. 2. In: Journal of Physics: Conference Series., vol. 1529; 2020. 022015.

[23] Uddin MT, Uddiny MA. A guided random forest based feature selection approach for activity recognition. In: 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT); 2015. p. 1–6.

[24] Feng Z, Mo L, Li M. A Random Forest-based ensemble method for activity recognition. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2015. p. 5074–7.

[25] Jing L, Cheng Z. Recognition of daily routines and accidental event with multipoint wearable inertial sensing for seniors home care. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 2017. p. 2324–89.

[26] Abidine BM, Fergani L, Fergani B, Oussalah M. The joint use of sequence features combination and modified weighted SVM for improving daily activity recognition. Pattern Anal Appl Feb. 2018;21(1):119–38.

[27] De Cillis F, De Simio F, Setola R. Long-term gait pattern assessment using a tri-axial accelerometer. J Med Eng Technol Jul. 2017;41(5):346–61.

[28] Zhou S, Sun B, Shi J. An SPC monitoring system for cycle-based waveform signals using haar transform. In: Proc. of IEEE Transactions on Automation Science and Engineering; 2006. p. 60–72.

[29] Schimmack M, Nguyen S, Mercorelli P. Anatomy of haar wavelet filter and its implementation for signal processing. In: Proceedings of IFAC-PapersOnLine; 2016. p. 99–104.

[30] Sinuraya EW, Rizal A, Soetrisno YAA, Denis. Performance improvement of human activity recognition based on ensemble empirical mode decomposition (EEMD). In: Proc. 5th Int. Conf. Inf. Technol., Comput., Elect. Eng. (ICITACEE); Sep. 2018. p. 359–64.

[31] Pathan NS, Talukdar MTF, Quamruzzaman M, Fattah SA. A machine learning based human activity recognition during physical exercise using wavelet packet transform of PPG and inertial sensors data. In: 2019 4th International Conference on Electrical Information and Communication Technology (EICT); 2019. p. 1–5.

[32] Yin J, Xu M, Zheng H. Fault diagnosis of bearing based on symbolic aggregate approximation and Lempel-Ziv. Measurement 2019;138:206–16.

[33] Bottou L. 'Stochastic gradient descent tricks. In: Neural Networks: Tricks of the Trade (Lecture Notes in Computer Science). Berlin, Germany: Springer; 2012.

[34] Pal M. Random forest classifier for remote sensing classification. Int J Rem Sens 2005;26:217–22.

[35] M Z, Sawchuk A. USC-HAD a daily activity dataset for ubiquitous activity recognition using wearable sensors. In: Proc. of the ACM Conference on Ubiquitous Computing; 2012. p. 1036–43.

[36] S B. Ud Din Tahir, Intelligent Media-Wearable Smart Home Activities. IM-WSHA; Nov. 2020 [online] Available: https://portals.au.edu.pk/imc/Pages/Datasets.aspx.

[37] Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H. Mobile sensor data anonymization. In: proc. of IEEE conf. on Internet-of-Things Design and Implementation (IoTDI'19); April 2019.

[38] Haresamudram H, Beedu A, Agrawal V, Grady PL, Essa I, Hoffman J, et al. Masked reconstruction based self-supervision for human activity recognition. In: Proceedings - International Symposium on Wearable Computers ISWC; 2020. p. 45–9.

[39] Jordao A, Nazare AC, Sena J, Schwartz WR. Human activity recognition based on wearable sensor data: a standardization of the state-of-the-art. in arXiv: 1806.05226, 2018, [online] Available: https://arxiv.org/abs/1806.05226.

[40] Jalal A, Batool M, Kim K. Stochastic recognition of physical activity and healthcare using tri-axial inertial wearable sensors. Applied Sciences; 2020.

[41] Sivakumar A. Geometry Aware Compressive Analysis of Human Activities: Application in a Smart Phone Platform''. Ph.D dissertation. Tempe, AZ, USA: Arizona State University; May 2014.

[42] Nguyen HD, Tran KP, Zeng X, Koehl L, Tartare G. Wearable sensor data based human activity recognition using machine learning: A new approach. arXiv: 1905.03809. 2019 [online] Available: http://arxiv.org/abs/1905.03809.

NURKHOLISH HALIM received his B.S. degree in information system from College of Informatics and Computer Management Professional, Makassar, Indonesia, in 2011. Since 2019 until now, he has worked as a system and technology architect at Bank OCBC-NISP, Jakarta, Indonesia. With previous work experience in various enterprise organization, startup and telecommunication. Focus on delivering solutions with machine learning, IoT and big data to address security risks and performance issues. Also, have a working knowledge of low-level network protocols (e.g. HTTP, IPv4, TCP, UDP, ICMP, Ethernet, and 802.11). His research interests are in the areas of cybersecurity, blockchain, machine learning, IoT, cyberlaw and computer vision with emphasis on mathematical modeling and performance analysis.