IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 April 2023, accepted 9 May 2023, date of publication 12 May 2023, date of current version 22 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275733

RESEARCH ARTICLE

Multi-Sensor-Based Action Monitoring and Recognition via Hybrid Descriptors and Logistic Regression

SADAF HAFEEZ[1], SAUD S. ALOTAIBI[2], ABDULWAHAB ALAZEB[3], NAIF AL MUDAWI[3], AND WOOSEONG KIM[4]

[1]Department of Computer Science, Air University, Islamabad 44000, Pakistan
[2]Information Systems Department, Umm Al-Qura University, Makkah 24382, Saudi Arabia
[3]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[4]Department of Computer Engineering, Gachon University, Seongnam 13120, South Korea

Corresponding author: Wooseong Kim (wooseong@gachon.ac.kr)

This work was supported in part by the Deanship of Scientific Research funded by Najran University under Research Group, Funding Program under Grant NU/RG/SERC/12/40; and in part by the Technology Development Program funded by the Ministry of Small and medium enterprises (SMEs) and Startups, South Korea, under Grant G21S2861919.

ABSTRACT In the fields of body-worn sensors and computer vision, current research is being done to track and detect falls and activities of daily living using the automatic recognition of human actions. In the area of human–machine communication, different combinations of sensors and communication technologies are often used to capture human action. Many researchers have also worked with artificial intelligent systems to detect actions, understand scenes, and implement systems that are more efficient in human action recognition. Although effective approaches are needed to detect outdoor activities with the combination of human actions, feature extraction can be quite a complicated task in a human activity recognition system development. Thus, this paper proposed a solution to detect human activities via hybrid descriptors based on robust features and accurate results. In this study, complex backgrounds, including multiple humans in video frames, were detected. First, inertial signal and video frames are pre-processed using denoising techniques, after which the frames are used to remove the background by detecting human motions and extracting the silhouettes. Then, these silhouettes are further used to extract the human body key points to make the human skeleton. Then the time and frequency domain features are extracted for inertial signals, and geometric features are extracted for the skeleton body points. Finally, multiple feature sets are combined and fed into a zero order optimization model, after which logistic regression is utilized to recognize each action. The proposed system has been evaluated on three benchmark datasets, including, the UP Fall dataset, the University of Rzeszow Fall dataset, and the SisFall dataset and proved its significance by achieving accuracy of 91.51%, 92.98%, and 90.23%, on the aforementioned datasets respectively.

INDEX TERMS Fall detection, geometric characteristics, human activity recognition, inertial sensors.

I. INTRODUCTION

In recent years, the demand for improvements in Human Action Recognition (HAR) systems, which aim to tackle human health and fitness conditions, has increased [1], [2], [3], [4], especially for ambient assisted-living. To increase the performance of their recognition systems, many researchers in this field have utilized inertial sensors, while others have employed depth or RGB sensors [5], [6], [7], [8]. The wide range of applications of HAR systems include

The associate editor coordinating the review of this manuscript and approving it for publication was Mohamed Elhoseny.

smart home environments [9], security surveillance [10], and human health monitoring [11]. Activities of daily living (ADL) can be easily recorded with inertial sensors i.e., accelerometers, gyroscopes, and magnetometers [12], [13] that are more reliable and less expensive. However, these sensors cannot tackle some complex and suspicious activities due to signal and background noise. Thus, hybridizing the system with the computer vision sensors like a Kinect camera, enhances the system's performance manyfold [14], [15], [16]. The collected data from cameras capture more information about the motions recorded. This capability is beneficial for extracting features from video frames [17], [18], [19], [20] and detecting various human activities. For this purpose, multi-sensor fusion [21], [22] is used to track daily life activities and falls by first dealing with heartbeat and pulse readings [23], thus helping to save human lives in emergency scenarios.

Thus far, significant progress has been made in this area, and numerous effective HAR systems have been created for a variety of uses [24]. Nonetheless, the identification and characterization of human interactions persist as a challenging predicament because of many reasons, such as variations in viewpoint angles [25], poor background lighting [26], and various body movements [27]. Despite this, due to their reduced susceptibility to variations in illumination as compared to RGB cameras, low-cost depth sensors [28] such as Microsoft Kinect are progressively being utilized. Additionally, numerous tasks that seem similar are often not categorized correctly, and in some situations, certain body parts are not properly detected. Related to this, the integration of an accelerometer with a Kinect camera helps in extracting body movements from a three-dimensional (3D) space with vital information. An accelerometer measures physical activities by continuously monitoring variations in acceleration, while the Kinect camera facilitates the accurate tracking of human actions in 3D view.

In this article, a novel approach is proposed to efficiently detect and recognize human actions using Machine Learning (ML) techniques. The proposed system consists of acquiring sensor data, denoising data captured from sensors, selecting features, optimizing the selected features, and performing classification for ADL. Inertial sensors are attached with the human body, and a Kinect camera is set in the room to capture data from multiple sensors. The collected data is then pre-processed through filtration methods. Furthermore, the inertial signals are first denoised using different filteration techniques, from which the best filter results are used for feature extraction. Meanwhile, human silhouettes are extracted from RGB video frame sequences using background subtraction. Then, the full-body silhouettes are used to extract different joint points and identify the six main body points. Geometric features are extracted using these key points. Two distinct kinds of features i.e., inertial and RGB features are combined and then optimized using optimization techniques. The logistic regression classifier is also applied to the recognition of actions. Three publicly available datasets that offer RGB and skeletal data on human motion are used in the experimentation phase of the system development. The main contributions of this research to the literature are as follows:

- Silhouettes are segmented from RGB photos using the background subtraction algorithm.
- Important key points are detected via morphological operation.
- Geometric and inertial features are extracted using important body points.
- Experimentations with three complex and sizable datasets for fall detection are conducted.

The remainder of the paper is organized into sections. Similar research works are described in Section II, while Section III discusses the proposed system's architecture. The implementation details of the proposed method and the corresponding results are presented in Section IV. The developed system is discussed in detail in Section V. Finally, the paper's conclusion and the future directions are presented in Section VI.

II. RELATED WORKS

The creation of effective HAR systems has received significant research attention in recent years. Inertial sensor- and video-based systems are the two groups into which the existing systems are separated. Below, specific details on each category's research have been covered:

A. INERTIAL SENSOR-BASED HAR SYSTEMS

Human actions can be recognized through wearable inertial sensors attached to different parts of a human body. Lei et al. [29] proposed a methodology for recognizing human activities using inertial sensors. The proposed method combines CNN with DFSMN to model long-term dependencies in temporal series. They collected data from two separate inertial sensors placed on the waist and thigh of the human body and found that the proposed method performed well in classifying routine activities. In [30], authors presented a multi-layer hybrid architecture using CNN and LSTM for non-intrusive sensor-based human activity recognition. The proposed model integrated local features and scale-invariant with activity dependencies, achieving a balance between accuracy and efficiency. The study demonstrated that the proposed model outperformed traditional machine learning and other deep learning methods. However, the study only evaluated the proposed model on one dataset, and further research was needed to evaluate its effectiveness in real-world scenarios and compare it with other state-of-the-art methods. The study in [31] presented a multi-modal 3D human pose estimation dataset named mRI, which consisted of over 160k synchronized frames from 20 subjects performing reha-

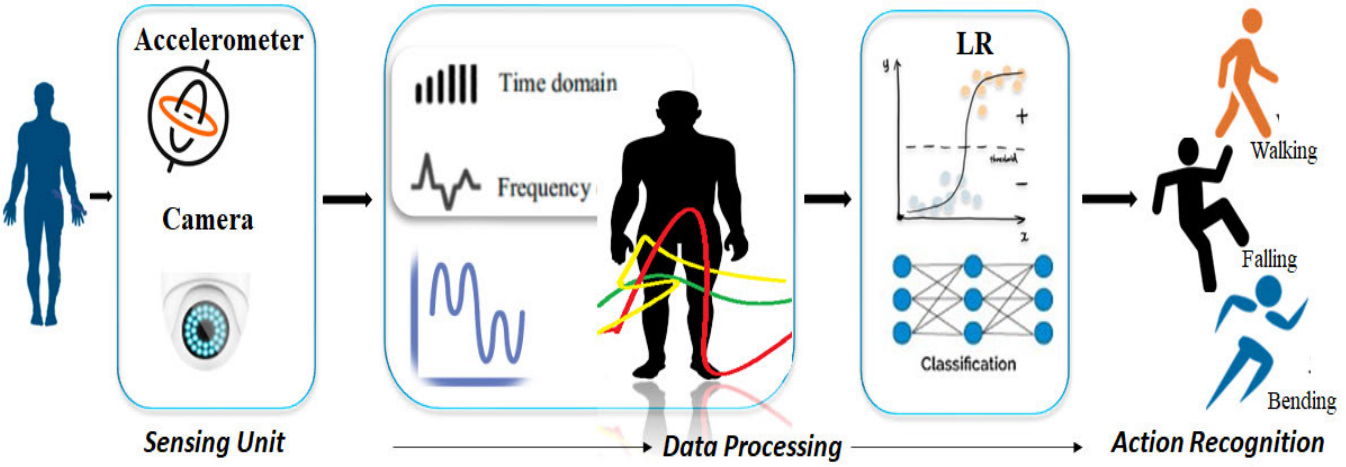


FIGURE 1. The architecture of the proposed HAR system.

bilitation exercises. The dataset included mmWave, RGB-D, and inertial sensor modalities, and was designed to fill the gap in previous datasets that did not exploit multiple modalities and focused less on home-based health monitoring. The authors conducted extensive experiments using the mRI dataset and identified the strengths of each modality. They hoped that the release of the dataset would facilitate the applications of home-based health monitoring and catalyze research in pose estimation, multi-modal learning, and action understanding.

Abidine et al. introduced a modified weighted SVM for the purpose of identifying human ADL carried out in a smart home environment [32]. They also addressed several implementation-related problems with the HAR algorithms, such as the training data's unnecessary sequence features and class imbalance. They then presented a new system to identify daily living behaviors in the surroundings of smart homes r to address these difficulties. Their technique is based on the integration of linear discriminant analysis (LDA), weighted SVM, and principal component analysis (PCA).

First, the features extracted from PCA and LDA were used to minimize the training set and obtain the most important characteristics. Then, a weighted SVM was used for every activity class to increase recognition rates and ultimately to address an unbalanced dataset of ADL. To enhance recognition, their suggested strategy was compared with ML techniques. However, their proposed system can be significantly improved with the use of deep learning algorithms.

B. VIDEO-BASED HAR SYSTEMS

In recent years, researchers have examined human actions by extracting skeleton information based on video sensors. Chen et al. [33] proposed a two-level hierarchical framework for action recognition based on 3D skeleton data. The framework was designed to address significant challenges such as high intra-class variance, variability in movement speed, and computational complexity. The proposed solution consisted of two primary modules, namely, a part-based clustering module to cluster relevant joints automatically and a motion feature extraction and action graph module to construct action graphs for recognition purposes. The accuracy of the skeletal points was a crucial factor for the system's performance, as the precision of joint point positioning had a direct influence on the accuracy of action classification. The work in [34] introduces VideoMoCo, an unsupervised video representation learning method that enhances the temporal feature representations of MoCo, a successful image representation learning method. To improve temporal robustness, VideoMoCo incorporates a generator to remove frames from the input sample and a temporal decay mechanism to model key attenuation in the memory queue when calculating the contrastive loss. This method empowers the learning of video representations without the need for pretext tasks. However, potential limitations may include the accuracy of the temporal decay model and difficulty in selecting an optimal dropout rate.

A method was proposed in [35] to improve the speed of the deep two-stream architecture for video-based action recognition by replacing the computationally expensive optical flow calculation with motion vector obtained directly from compressed videos. However, the motion vector had limitations, such as lacking fine structures and containing inaccurate motion patterns that degraded the recognition performance. To address this issue, the authors proposed transferring knowledge learned with optical flow CNN to motion vector CNN through initialization transfer, supervision transfer, and their combination. The proposed method aimed to improve the efficiency of the two-stream architecture while maintaining the recognition performance. In [36], the authors introduced a new approach called ''dynamic image'' which is a compact representation of videos that is useful for video analysis with convolutional neural networks (CNNs). The dynamic image is obtained through a ranking machine that encodes the temporal evolution of video frames and is produced by directly applying rank pooling on the raw image

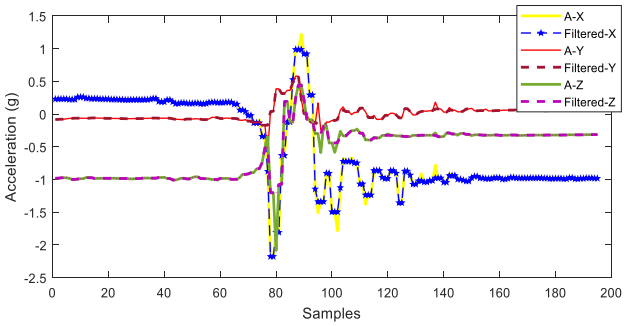


FIGURE 2. Inertial sensor signals of the x, y, and z axes with a median filter.

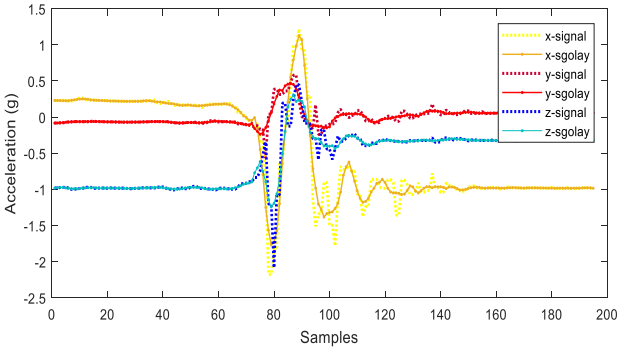


FIGURE 3. Inertial sensor signals of the x, y, and z axes with the Savitzky–Golay filter.

pixels of a video, resulting in a single RGB image per video. The authors developed an efficient and effective approximate rank pooling operator to speed up the process and allow for the generalization of dynamic images to dynamic feature maps. Their system demonstrated a limited ability to capture fine-grained details in videos due to the compact representation and the need for fine-tuning existing CNN models for video data.

III. THE PROPOSED APPROACH

The proposed system consists of four primary sections: pre-processing, feature extraction, feature optimization, and action recognition through classification. Below is a detailed discussion of each section's techniques and findings. A flowchart of the suggested system architecture is shown in Fig. 1.

A. PRE-PROCESSING

In this module, the signals obtained from inertial sensors are enhanced to reduce the inconsistency, noise, and redundancy. Different filters are applied to detect the signal of interest and remove noise without signal data loss. In RGB images, frame sequences are used for pre-processing, and the resulting image frames contain different objects in the background. Then, the background subtraction technique has been used to detect the human bodies and their motions. Multiple frames were generated in each action class, so the region of interest is extracted from each class.

1) MEDIAN FILTER

In the median filter, a digital nonlinear filtering method has been implemented in order to detect and remove noise. In particular, we used a non-static median filter and a linear method of denoising dealing without data loss of a specific motion [37]. The median filter is implemented as a window-by-window technique over the whole signal. The function of a digital median filter is to smoothen the signals by taking continuous sample points while calculating their median to replace the value of the current point. The calculation is carried out using the following relation:

$$y\,[a,\,b] = medianx\{[k,\,l]\,,\,(k,\,l)\,\epsilon w\} \tag{1}$$

where k and l belong to window w of the specified neighboring values around the point $[a,b]$ of the image. A comparison of filtered and raw data from a tri-axial accelerometer is demonstrated in Fig. 2.

2) SAVITZKY-GOLAY FILTER

Three fundamental phases are typically used in a Savitzky–Golay Filter (SGF); the moving average, convolution, and the least squares [38]. The result of the SGF is presented in Fig. 3 as the replacement of a rounded sequence, which is similar to the basic moving average approach. The least squares method minimizes the squared error on all points, thus establishing the distance between the original and the smoothed points. Furthermore, a convolution method is typically used to obtain a moving average by the direction of weighing, which is defined as a polynomial of the defined quantity of the signal.

When connected to the inertial signal, the weighted coordinates produce polynomial least squares fit that is incorporated into the filter window. It is expected that the polynomial can better maintain factual moments and lessen the filter-induced bias. The basic least-square convolution for time series smoothing is expressed as:

$$y = mA \tag{2}$$

where A is the vector, and m is the signal.

3) BACKGROUND SUBTRACTION

Using the color of data and region unification as basis, the background subtraction technique is applied over RGB data. The change in sequential images is detected using an adaptive threshold-based technique and spatiotemporal [21] differencing for human detection in the video frames. Mathematically, it is defined by the following relation:

$$M^{f}_{FH} = \left(M^{f}_{He} \leftarrow M^{f-1}_{He} + \Delta M^{q-1}_{He}\right) + M^{f}_{End} \tag{3}$$

where M^{f}_{FH} represents a human location in a video frame and M^{f}_{End} shows the bounding box of the size of the detected human. Once human is successfully detected, the next stage is to identify the different joint locations for human body

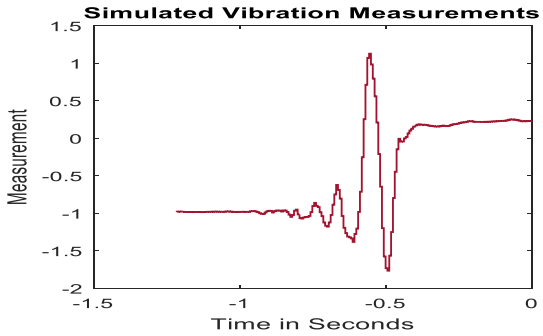


FIGURE 4. The temporal moment of the IMU signal.

movement detection. We also applied a skin detection algorithm [39], along with boundary dilation and other morphological operations to extract human silhouettes. After silhouette monitoring and detection, the proposed system is able to extract the skeleton for further processing.

4) MORPHOLOGICAL OPERATION

In the morphological process [40], the human silhouettes from the RGB images are obtained. First, threshold-based segmentation is performed to convert the depth images into binary images. The segmented images then undergo morphological operations by using the binary dilation technique. Binary dilation enlarges human edges by adding pixels, while binary erosion subtracts boundary pixels [41] from the resulting images. The mathematical description of binary dilation and binary erosion is given by Eq. 4 and Eq. 5 respectively.

$$X \oplus Y = \{q | \left(\check{Z}\right), \cap X \neq \theta\} \tag{4}$$

$$X - Y = \{q | \left(Z\right), \subseteq X\} \tag{5}$$

where q is the pixel location for element Z. Through this technique foreground objects have been extracted from the background objects in a frame.

B. FEATURE EXTRACTION

We extracted various time-domain and frequency-domain features from windows depending upon the type of the sensor. The time and frequency domain features [42] of the raw signals are explained in Algorithm 1 for each window frame of IMU sensors. Feature extraction remains a huge challenge in various human activity studies, such as walking, running, sitting, and lying. We extracted various geometric features that are explained individually in the following sections.

1) SKEWNESS

Skewness refers to the degree of asymmetry or the lack of symmetry in the distribution of different data [43]. This statistical characteristic calculates the discrepancy between the distribution of a signal and a normal distribution to represent an EMG signal using the equation:

$$SK = \frac{1}{N} \sum_{l-1}^{N} \left(\frac{a_i - a}{\sigma}\right) \tag{6}$$

Algorithm 1 Inertial Measurement Unit (IMU) Features

Input: Accelerometer (ax, ay, az)
Output: MultiFused_Features (f_1,f_2, f_3, …., f_n)
Features_vector ⟵ []
Signal_sample ⟵ GetSampleInertialSignal ()
Process: HAR(Acc)
MultiFusionVector ⟵ []
DenoiseSignal ⟵ Median_Filter(Acc)
DenoiseSignal ⟵ SavitzkyGolay_Filter(Acc)
SampleSignal (DenoiseSignal)
% While not satisfied exit condition
StatisticalSignalFeature ⟵ ExtractStatisticalFeatures(DenoiseSignal)
FrequencySignalFeature ⟵ ExtractFrequencyFeatures(DenoiseSignal)
MultiFusionVector ⟵ [StatisticalSignalFeature, FrequencySignalFeature]
end while
return MultiFusionVector

where σ is the standard deviation , N is a sample window, a is the mean, and a_i is a random value.

2) FUZZY ENTROPY

The conditional probability [44] is fuzzy entropy, whose negative natural logarithm corresponds to this probability, and it states that for comparable m + 1 points, two vectors with identical m points remain similar. The consistency of time series is measured by fuzzy entropy by using the following mathematical relation:

$$FE\left(y, n, p, m, l\right) = -\ln\left(\psi^{n+1,l}\left(m, p\right) | \psi^{n,l}\left(m, p\right)\right) \tag{7}$$

where m and n are the fuzzy coefficients. The standard deviation is reduced when alternative values of n and r, are used. Here, r and n are selected for windows.

3) TEMPORAL MOMENT

Temporal moment'' is suggested to detect abnormal human actions and fall pattern calculation by capturing the prediction changes in signals [45]. In particular, it captures any change related to the speed information occurring in the signal. It also provides an efficient way to distinguish the change between signals in accordance with the change in frequency over the time shown in Fig. 4.

This helps in avoiding the computational overhead costs and extracting the important features of the inertial signal. Temporal moments can be calculated using the following equation.

$$tm_j\left(s\right) = \int_{-\infty}^{\infty} \left(t - s\right)^j x^2(t) dt \tag{8}$$

where t and s represent the speed at two distinct temporal samples of the input signal $x(t)$.

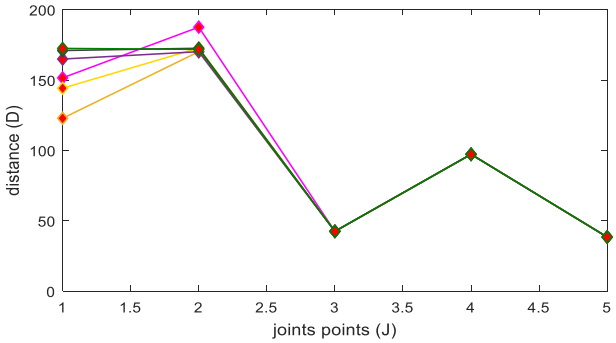


FIGURE 5. Euclidean distance from a single joint to other skeleton joints.

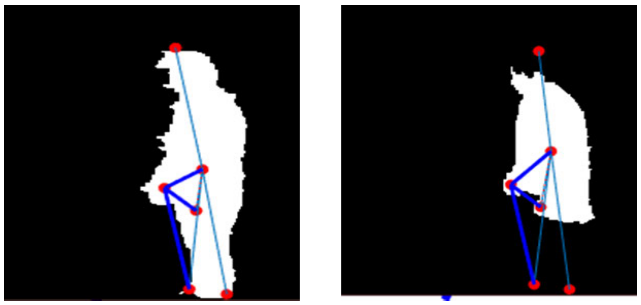


FIGURE 6. Joint-line distance representing distance of joint from line plane.

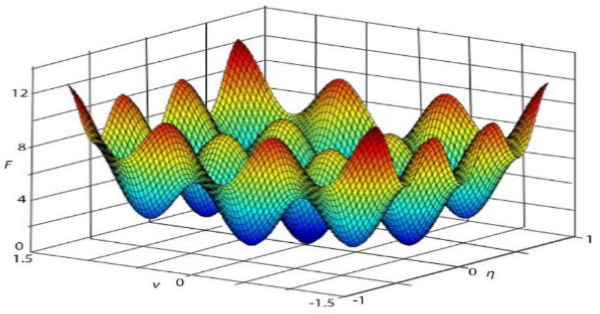


FIGURE 7. Zero order optimization of first order with high-speed convergence.

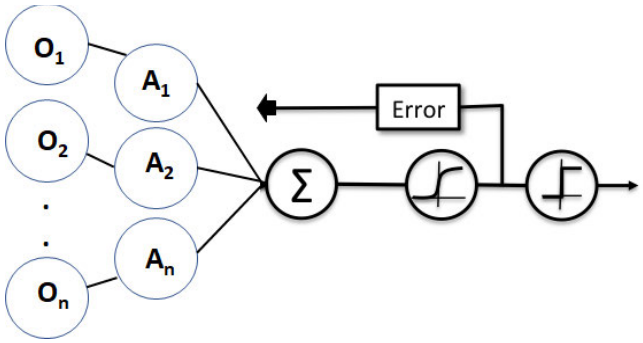


FIGURE 8. Proposed logistic regression general algorithm for action recognition.

4) GEOMETRIC FEATURES

In this study, we obtained geometric features based on different combinations of 2, 3, and 4 joint points of the skeleton in a single frame [46]. Although a large number of combinations are possible to be mined for different joints in skeleton points, it is largely a time-consuming task. Hence, we selected important planes, joints, and lines to reduce the system's computational cost. First, we extracted the Euclidean distance [47] between joint points, after which we determined the joint distance between joint j_1 and joint j_2 in the x, y coordinates. This is calculated using the equation given below:

$$Vd\,(j_1, j_2) = \sqrt{(a_1x - a_2x)^2 + (a_1y - a_2y)^2}, \tag{9}$$

where $Vd\,(j_1, j_2)$ is the distance between two joints, a_1 is the torso joint, and a_2 is the hand joint. The distances of all joints from one joint are presented in Fig. 5.

Joint line distance is another feature that extracts the distance from joint J to line (distance between two joints is represented by a line), as shown in Fig. 6.

The line forms a triangle after the intersection of the joint line with the plane line. Here, we calculated the triangle by using the Helen formula [48] given by:

$$J_{d(j_i,j_j)} = \frac{2S_{\nabla jj_ij_j}}{J_{d(j_i,j_j)}} \tag{10}$$

where joint J intersects line $Lj_{1\to}j_2$ and S is the area of the triangle.

C. FEATURE OPTIMIZATION

Zero Order (ZO) [49] is an optimization method without a gradient. It optimizes the various gradient-free emerging signals in ML problems. The ZO technique is the gradient-free [50] counterpart of the first order optimization method with ease of implementation, as shown in Fig. 7. During the last few years, researchers have increasingly worked on ZO optimization to solve big data problems and ML issues.

The relation given below, represents the difference of values along with a random direction vector in ZO optimization that leads to the random gradient estimation.

$$\nabla\mathcal{F}\,(x) = \frac{\phi(v)}{\eta}[\mathcal{F}\,(x + \eta u) - \mathcal{F}(x)]u \tag{11}$$

where $\nabla\mathcal{F}\,(x)$ is the gradient estimate error of the mean squared with respect to the true gradient. The dominant factors v and η highlight the error estimation.

D. ACTION RECOGNITION

Logistic Regression (LR) is a popular classification algorithm that models the relationship between a dependent variable and one or more independent variables using a logistic function. Its ability to predict the probability of an outcome based on input features, along with its computational efficiency and interpretability, make it a sound choice for classification tasks, where predicting the likelihood of an event is

essential [51]. It is represented by the following equation:

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)}} \tag{12}$$

where P the probability of the event occurring, $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients of the model, and $x_1, \ldots, x_n$ are the predictor variables. This classification technique is mostly used in the simplest ML algorithms for the different problems shown in Fig. 8.

During the classification of multiclass data, the LR provides the linear boundary for the decision, whereas for the nonlinear parameters, optimization is first required for the numerical data [52].

Algorithm 2 Key Body Points Extraction

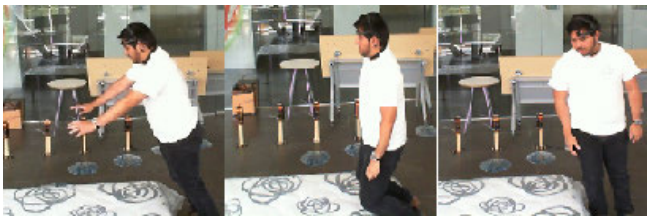
Input: human silhouettes extraction
Output: K eyBodyPoints (*jp1,jp2,jp3…jpn*):
HumanShape(Hs),
Leftfoot(Lf),
Rightfoot(Rf),
Lefthand(Lh),
Righthand(Rh),
Height(Hh),
Width(Wh)
do
For j = *cluster* to N_Clusters
OuterKeyPoint ← *GetPointsVector(UpperHeadPoint (Uh),Bottom(Bp))*
MidPoint ← *CalMidPoint(Hh,Wh) / 2*
FeetPoint ← *CalBottom(Hs) & search(Lf, Rf)*
HandsPoint ← *CalHands(Hs) & search(Lh, Rh)*
KeyPointsAppend(OuterKeyPoint, MidPoint, FeetPoint, HandsPoint)
While (search (Hs) & search (Lh, Rh) != NULL)
return *KeyPoints(jp1,jp2,jp3…jpn)*

IV. EXPERIMENTAL SETUP AND RESULTS

This section describes the specifications of the tests performed to prove the suggested system. Python and MATLAB were used in all of the processing and experiments. As for the hardware, we used a 64-bit version of Windows 10 installed in a computer running on an Intel Core i5 processor. The system contained a 5 GHz CPU and 8 GB of RAM. Three separate datasets were used to test the proposed system, and the recognition accuracy for each action class was calculated, along with accuracy, sensitivity, and F1-scores, using confusion matrices. Then, the accuracy was compared to other state of the art (SOTA) methods for additional verification. This section is divided further into two parts: dataset details and experimental results.

A. DATASETS DESCRIPTION

The UP-fall dataset [53], the University of Rzeszow (UR) fall dataset [54], and the SisFall detection dataset [55] were used in the experiments. The details of each dataset are provided in the subsections below.



(a) (b) (c)

FIGURE 9. RGB frames from the UP Fall dataset. (a) Falling forward; (b) banding knee; (c) falling sideward.

TABLE 1. A brief summary of the datasets actions.

Actions	Category	Details	Time in seconds
1	ADL	Walking	10
2		Standing	10
3		Sitting	10
4		Picking up an object	10
5		Jumping	10
6	FALL	Falling forward using hands	60
7		Falling forward using knees	60
8		Falling backwards	60
9		Falling sideward	10
10		Falling sitting in empty chair	30
11		Falling forward using hands	60

1) UP FALL DATASET

The UP fall detection dataset consists of a large amount of data for the detection of eleven fall activities. It features three trials related to each activity performed by the subjects. UP fall is a large dataset that is mainly used for fall detection. Data were obtained from 17 healthy young individuals without disabilities, who were gathered using a multimodal strategy that included wearable sensors, ambient sensors, and vision equipment. Six activities are based on ADL, including standing, sitting, walking, lying down, jumping, and picking up an object. Five are related to human fall activities, including falling forward using a hand, falling forward using knees, falling backward, falling sideward, and falling sitting in an empty chair. Each falling activity is performed within ten seconds, whereas each ADL is performed within sixty seconds, except for jumping and picking up an object, which are performed within thirty and ten seconds, respectively. Table 1 presents an overview of the time distribution between all activities over the selected dataset.

We tried our best to mimic several fall scenarios, such as tripping, sitting, and falling, in different directions, and chose the most often performed ADLs. We specifically selected picking up an object because it is typical to mistake this activity for a fall. While time windows for falls were chosen based on the six seconds safe period following fall incidence,

as described in [13], time windows for daily activities were chosen to cover them in the minimum duration of time stated in similar research. A mattress was placed in the falling area for all of these activities to prevent injuries. Each task was completed by each volunteer three times (in trials) without any impairment. Three axial accelerometers and two cameras were used to collect data from the frontal and lateral views by tagging sequential actions related to the fall or ADL activities. Fig. 9 depicts the fall activities from the UP fall detection dataset.

2) UR FALL DATASET

The UR's fall detection dataset was created by gathering information from two USB-connected Kinect cameras and an IMU device worn at the waist and connected through Bluetooth. ADL events were only captured using camera 0 and an accelerometer. Devices such as the PS Move and x-IMU were used to gather sensor data. Five volunteers performed a total of 70 sequences, involving 30 falls and 40 ADLs other than falls, in an experiment set up inside an office, five people intentionally fell toward a carpet with a thickness of around 2 cm. The x-IMU was placed close to the pelvis. Each person performed the forward, backward, and lateral falls for a minimum of three times each. Each person also carried out the ADL depicted in Fig. 10, such as standing, sitting, squatting, bending over, picking up items from the ground, and resting on a settee. All deliberate falls were accurately detected. In particular, quickly sitting down was labeled as an ADL even if it was difficult to distinguish it from a purposeful fall when simply an accelerometer or even either an accelerometer and a gyroscope was used. Some of them are indoor fall-themed activities. Falling from a standing posture and falling while seated on a chair were the two different types of falls recorded. Each register held raw accelerometer data as well as depth and RGB image sequences taken by two cameras. A threshold-based falling detection technique was also used by the authors. Pictures were generated using two Microsoft Kinect cameras.

3) SisFall DATASET

This dataset was collected from 38 volunteers who were split into two groups: older persons and young adults. Eight male and seven female participants comprised the group of seniors, whereas 23 people comprised the group of young adults (11 male and 12 female). ADLs and falls were performed by young adults. Falling and the D06, D13, D18, and D19 tasks were not performed by elderly adults. We used two accelerometer models: Adxl345, which had a resolution of 13 bits and a range of 16g to +16g, and MA8451Q, which had a resolution of 14 bits and a range of 8g to +8g [20]. The subjects' waists were attached to the accelerometer gadget as demonstrated in Fig. 11. In this area, a single accelerometer device can distinguish between various activities with great accuracy. There were many activities, including walking slowly, jogging quickly, slowly sitting, standing, bending, falling forward, falling forward while walking, lateral fall while sitting, and so on. The actions were performed by 19 males and 19 females. There were 15 fall activity sessions and 19 ADL sessions.

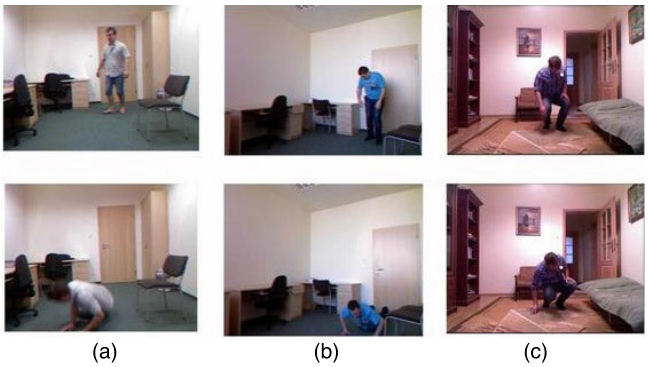


FIGURE 10. RGB frames from the UR Fall dataset. (a) fall from standing; (b) bend and pick; (c) look under furniture.

FIGURE 11. RGB frames from SisFall dataset. (a) falling forward while walking caused by a trip; (b) walking.

TABLE 2. Summary of the datasets.

Dataset	Classes	Sense Modality
UP FALL Dataset	8	RGB, Depth, Accelerometer
UR FALL Dataset	11/60	RGB, Depth, Accelerometer
SisFall Dataset	8	RGB, Depth, Accelerometer

B. EXPERIMENTS AND RESULTS

The proposed system's performance evaluation was validated. The confusion matrix of each class, along with the precision, recall, classification results, and comparison of our suggested system with other cutting-edge approaches and systems were employed as separate metrics, which were divided into two groups during the experimental phase. In Table 2, we present the summary of the action classes and sensors.

TABLE 3. The UP FALL dataset confusion matrix based on individual classes.

Action Classes	FFK	FFH	FSC	FSD	FBD	JMG	WLG	SDG	STG	PUO	LYG
FFK	**0.90**	0	0.01	0.02	0.03	0	0.04	0	0	0	0
FFH	0.01	**0.84**	0	0.05	0	0.04	0	0.05	0	0	0.01
FSC	0	0.02	**0.90**	0	0.01	0.01	0	0	0.04	0	0.02
FSD	0	0	0.01	**0.97**	0.01	0	0	0.01	0	0	0
FBD	0.02	0	0	0.02	**0.90**	0.01	0	0	0	0.03	0.02
JMG	0	0.01	0.01	0	0.02	**0.90**	0	0.02	0.04	0	0
WLG	0.01	0	0	0.02	0	0	**0.96**	0	0	0.01	0
SDG	0	0.02	0.02	0	0	0.03	0	**0.90**	0	0.01	0.02
STG	0.02	0	0	0	0.06	0	0	0.01	**0.87**	0	0.04
PUO	0	0	0.01	0	0	0	0	0	0.02	**0.96**	0.01
LYG	0	0.04	0	0.02	0	0.02	0	0.04	0	0	**0.88**

Mean recognition accuracy = 91.51%

*FFK = falling forward using knee; FFH = falling forward using hand; FSC = falling sitting in empty chair; FSD = falling sideward; FBD = falling backward; JMG = jumping; WLG = walking; SDG = Standing; STG = Sitting; PUO = Picking up an object; LYG = Laying

TABLE 4. The UR FALL dataset confusion matrix based on individual classes.

Action Classes	Standing	Sitting	Lying down	Bending	Crawling	Others
Standing	**0.84**	0.06	0.01	0.02	0.03	0.04
Sitting	0	**0.96**	0	0	0	0.04
Lying down	0	0	**0.99**	0	0.01	0
Bending	0	0.03	0.03	**0.87**	0.01	0.06
Crawling	0.02	0	0	0.08	**0.90**	0
Others	0	0.03	0.01	0	0.02	**0.94**

Mean recognition accuracy = 92.98%

TABLE 5. The SISFALL dataset confusion matrix based on individual classes.

Action Classes	FFW	FBW	LFW	FFWT	FFJT	VFW	FWH	FFU	LFU	FFD	WS
FFW	**0.85**	0	0	0.01	0.08	0	0	0	0.03	0	0.04
FBW	0	**0.98**	0.01	0	0	0	0	0	0	0.1	0
LFW	0.01	0	**0.96**	0	0	0.02	0	0	0	0	0.01
FFWT	0	0	0.04	**0.90**	0	0	0.02	0	0	0.04	0
FFJT	0.04	0.03	0	0.05	**0.80**	0.05	0	0	0.01	0	0.02
VFW	0	0.01	0.01	0	0	**0.90**	0	0.06	0	0	0.02
FWH	0	0	0	0.01	0	0	**0.98**	0	0	0	0.01
FFU	0.05	0	0.04	0	0.03	0	0	**0.82**	0.02	0.04	0
LFU	0.01	0.04	0	0.02	0	0	0.04	0	**0.86**	0	0.03
FFD	0.02	0	0.01	0	0	0.06	0	0	0	**0.89**	0.02
WS	0	0.01	0	0	0.02	0	0.01	0.4	0	0	**0.92**

Mean recognition accuracy = 90.23%

* FFW = falling forward while walking; FBW = falling backward while walking; LFW = lateral fall while walking; FFWT = fall forward while walking; FFJT = fall forward while jogging; VFW =vertical fall while walking; FWH = fall with the use of hands; FFU = fall forward when trying to get up; LFU = lateral fall when trying to get up; FFD = fall forward when trying to sit down; WS = walking slowly

The detailed experimentation results along with examples for each subgroup are provided in the preceding subsections.

1) INDIVIDUAL CLASS ACCURACY

The performance data for the proposed model can be effectively conveyed through the utilization of confusion matrices that distinctly present the classification results in terms of false positives, false negatives, true positives, and true negatives for each respective category. Table 3, Table 4, and Table 5 contain the confusion matrices for the UP fall, UR fall, and SisFall datasets, respectively. It is evident from these tables that the recognition rates of the action classes across all three datasets were considerably high, with mean accuracy rates of 91.51%, 92.98%, and 90.23%, respectively, thereby indicating the efficacy of the model in accurately classifying the actions.

The sitting as well as standing activities in the dataset are the examples of action classes that still have some uncertainty between them because they include similar actions.

TABLE 6. Measurements on the UP FALL dataset over precision, sensitivity, and F1 score.

Action Classes	Precision	Sensitivity	F1 score
FFK	0.88	0.90	0.88
FFH	0.84	0.84	0.84
FSC	0.91	0.90	0.90
FSD	0.95	0.97	0.96
FBD	0.91	0.90	0.90
JMG	0.89	0.90	0.89
WLG	0.97	0.96	0.96
SDG	0.91	0.90	0.90
STG	0.86	0.87	0.86
PUO	0.95	0.96	0.95
LYG	0.90	0.88	0.89
Mean	0.90	0.91	0.90

TABLE 7. Measurements on the UR FALL dataset over precision, sensitivity, and F1 score.

Action Classes	Precision	Sensitivity	F1 score
Standing	0.85	0.84	0.84
Sitting	0.95	0.96	0.95
LYD	0.95	0.99	0.97
Bending	0.85	0.87	0.86
Crawling	0.91	0.90	0.90
Others	0.93	0.94	0.93
Mean	**0.91**	**0.92**	**0.91**

The analysis of the confusion matrices depicted in Table 3 reveals that the UP Fall dataset's falling backward and falling while seated actions are confounded. Furthermore, Table 4 demonstrates that the UR Fall dataset's sitting and bending behaviors are erroneously classified. Similarly, Table 5 highlights that the SisFall dataset's actions of shaking hands and hugging are entangled, indicating the presence of ambiguity and misclassification within the model.

Table 6, Table 7, and Table 8 provide the accuracy, sensitivity, and F1 scores for each class in the UP Fall, UR Fall, and SisFall datasets, respectively. For all action and falling classes, the precision, sensitivity, and F1 scores for each dataset have been calculated as follows:

$$Precision = \frac{TPT}{TPT + FPT} \tag{13}$$

$$Sensitivity = \frac{TPT}{TPT + FNT} \tag{14}$$

$$F1score = \frac{2(Precision \times Sensitivity)}{(Precision + Sensitivity)} \tag{15}$$

where the letters TP, FP, and FN, respectively, stand for True Positives, False Positives, and False Negatives.

2) COMPARISON WITH OTHER SOTA METHODS

Within this section, a comparative analysis is performed between the proposed approach and prior methodologies in the field of human activity recognition (HAR). The accuracy of each evaluated methodology is compared to that of the suggested system, wherein the proposed method is pitted against several state-of-the-art (SOTA) systems that employ the UP fall, UR fall, and SisFall datasets, as depicted in Table 9, Table 10, and Table 11, respectively. The results indicate that the suggested hybrid descriptors exhibited superior performance in contrast to the features employed in the latest SOTA systems.

TABLE 8. Measurements on the SIS FALL Dataset over Precision, Sensitivity, and F1 Score.

Action Classes	Precision	Sensitivity	F1 score
FFW	0.84	0.85	0.84
FBW	0.96	0.98	0.96
LFW	0.95	0.96	0.95
FFWT	0.91	0.90	0.90
FFJT	0.85	0.84	0.84
VFW	0.89	0.90	0.89
FWH	0.99	0.98	0.98
FFU	0.81	0.82	0.81
LFU	0.87	0.86	0.86
FFD	0.88	0.89	0.88
WS	0.91	0.92	0.91
Mean	**0.90**	**0.90**	**0.89**

TABLE 9. Comparison of the UP FALL dataset with other state of the art methods.

System Methods	Accuracy (%)
SVM (IMU)+ EEG system [53]	90.77
MS-DLD system [56]	88.75
Naïve Bayseian Classifier. [57]	88.61
JDM [58]	88.10
Proposed Method	**91.51**

TABLE 10. Comparison of the UR FALL dataset with other state of the art methods.

System Methods	Accuracy (%)
SVM [59]	90.00
KNN [60]	89.00
LSTM [61]	85.00
LSTM + RNN [62]	92.83
Proposed Method	**92.98**

TABLE 11. Comparison of the SIS FALL dataset with other state of the art methods.

System Methods	Accuracy (%)
Fine KNN [63]	83.76
CNN [64]	90.00
RNN(C_9) [65]	91.75
RRM [66]	89.00
Proposed Method	**90.23**

V. DISCUSSION

The proposed framework represents a complete HAR system intended to work for a variety of real-world issues involving the monitoring of smart homes as well as security and

surveillance systems. Although it is designed to be used with falling and ADL datasets, the proposed logistic regression model can also be implemented with RGB and inertial datasets by omitting the model concatenation stage and using only two streams.

Each stage of the process from pre-processing through classification helps the system perform better. The described feature extraction method extracts robust features with success, which are crucial for the appropriate classification of interaction. The results were obtained by first concatenating both RGB and inertial signals and then optimizing them and training the signals with images using a logistic regression model. Additionally, because video sequences were present in all three datasets, the LR classification step yielded accurate results.

Despite producing excellent outcomes, the suggested system has certain drawbacks. Currently, the suggested skeleton model for the detection of significant body points can only be used to extract six key points. However, higher accuracy can be attained if additional important locations are discovered and their characteristics extracted. Furthermore, the proposed system is also exceedingly complex and expensive to compute.

VI. CONCLUSION

In this study, we efficiently recognized fall detection and ADL recognition in our proposed system. Our system can be deployed in various applications, including healthcare, emergency facilities, and virtual gaming. In our proposed system, we used an ML method for feature extraction, the data from which was evaluated via LR classification after the extraction of inertial and skeleton features. Owing to their dynamic nature and complex backgrounds, three challenging datasets were used to test the system's performance. Compared with other SOTA techniques, our system demonstrated the best performance. In the future, we will explore new features and modeling methods related for HAR to deep learning using complex datasets.

ACKNOWLEDGMENT

The authors would like to thank Dr. Ahmad Jalal for his support in experimentations and analytical review during this system design.

REFERENCES

[1] A. Ayman, O. Attalah, and H. Shaban, "An efficient human activity recognition framework based on wearable IMU wrist sensors," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Dec. 2019, pp. 1–5.

[2] O. S. Eyobu and D. Han, "Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network," *Sensors*, vol. 18, no. 9, p. 2892, Aug. 2018.

[3] A. Jalal, M. A. K. Quaid, and A. S. Hasan, "Wearable sensor-based human behavior understanding and recognition in daily life for smart environments," in *Proc. Int. Conf. Frontiers Inf. Technol. (FIT)*, Dec. 2018, pp. 105–110.

[4] A. Jalal, M. A. K. Quaid, and M. A. Sidduqi, "A triaxial acceleration-based human motion detection for ambient smart home system," in *Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2019, pp. 353–358.

[5] I. Alrashdi, M. H. Siddiqi, Y. Alhwaiti, M. Alruwaili, and M. Azad, "Maximum entropy Markov model for human activity recognition using depth camera," *IEEE Access*, vol. 9, pp. 160635–160645, 2021.

[6] T. F. N. Bukht, H. Rahman, and A. Jalal, "A novel framework for human action recognition based on features fusion and decision tree," in *Proc. 4th Int. Conf. Adv. Comput. Sci. (ICACS)*, Feb. 2023, pp. 1–6.

[7] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian J. Sci. Eng.*, vol. 41, no. 3, pp. 1043–1051, Mar. 2016.

[8] S. Hafeez, A. Jalal, and S. Kamal, "Multi-fusion sensors for action recognition based on discriminative motion cues and random forest," in *Proc. Int. Conf. Commun. Technol. (ComTech)*, Sep. 2021, pp. 91–96.

[9] S. Badar ud din Tahir, A. Jalal, and M. Batool, "Wearable sensors for activity analysis using SMO-based random forest over smart home and sports datasets," in *Proc. 3rd Int. Conf. Adv. Comput. Sci. (ICACS)*, Feb. 2020, pp. 1–6.

[10] H. Badave and M. Kuber, "Face recognition based activity detection for security application," in *Proc. Int. Conf. Artif. Intell. Smart Syst. (ICAIS)*, Mar. 2021, pp. 487–491.

[11] M. Javeed, A. Jalal, and K. Kim, "Wearable sensors based exertion recognition using statistical features and random forest for physical healthcare monitoring," in *Proc. Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2021, pp. 512–517.

[12] U. Azmat and A. Jalal, "Smartphone inertial sensors for human locomotion activity recognition based on template matching and codebook generation," in *Proc. Int. Conf. Commun. Technol. (ComTech)*, Sep. 2021, pp. 109–114.

[13] M. Javeed, M. Gochoo, A. Jalal, and K. Kim, "HF-SPHR: Hybrid features for sustainable physical healthcare pattern recognition using deep belief networks," *Sustainability*, vol. 13, no. 4, p. 1699, Feb. 2021.

[14] K. Banjarey, S. P. Sahu, and D. K. Dewangan, "A survey on human activity recognition using sensors and deep learning methods," in *Proc. 5th Int. Conf. Comput. Methodol. Commun. (ICCMC)*, Apr. 2021, pp. 1610–1617.

[15] V. D. Huszár and V. K. Adhikarla, "Live spoofing detection for automatic human activity recognition applications," *Sensors*, vol. 21, no. 21, p. 7339, Nov. 2021.

[16] A. Bagate and M. Shah, "Human activity recognition using RGB-D sensors," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS)*, May 2019, pp. 902–905.

[17] M. H. Siddiqi, N. Almashfi, A. Ali, M. Alruwaili, Y. Alhwaiti, S. Alanazi, and M. M. Kamruzzaman, "A unified approach for patient activity recognition in healthcare using depth camera," *IEEE Access*, vol. 9, pp. 92300–92317, 2021.

[18] M. Shanmughapriya, S. Gunasundari, and S. Bharathy, "Loitering detection in home surveillance system," in *Proc. 10th Int. Conf. Emerg. Trends Eng. Technol.-Signal Inf. Process. (ICETET-SIP)*, Apr. 2022, pp. 1–6.

[19] A. Jalal, M. Mahmood, and A. S. Hasan, "Multi-features descriptors for human activity tracking and recognition in indoor-outdoor environments," in *Proc. 16th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2019, pp. 371–376.

[20] F. Abdullah and A. Jalal, "Semantic segmentation based crowd tracking and anomaly detection via neuro-fuzzy classifier in smart surveillance system," *Arabian J. Sci. Eng.*, vol. 48, no. 2, pp. 2173–2190, Feb. 2023.

[21] M. Javeed and A. Jalal, "Body-worn hybrid-sensors based motion patterns detection via bag-of-features and fuzzy logic optimization," in *Proc. Int. Conf. Innov. Comput. (ICIC)*, Nov. 2021, pp. 1–7.

[22] A. A. Badawi, A. Al-Kabbany, and H. A. Shaban, "Sensor type, axis, and position-based fusion and feature selection for multimodal human daily activity recognition in wearable body sensor networks," *J. Healthcare Eng.*, vol. 2020, pp. 1–14, Jun. 2020.

[23] A. Jalal, M. Batool, and S. B. U. D. Tahir, "Markerless sensors for physical health monitoring system using ECG and GMM feature extraction," in *Proc. Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2021, pp. 340–345.

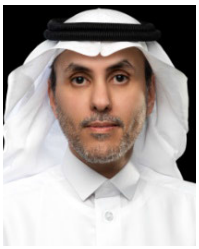
[24] X. Shu, J. Tang, G. Qi, W. Liu, and J. Yang, "Hierarchical long short-term concurrent memory for human interaction recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 1110–1118, Mar. 2021.

[25] A. Franco, A. Magnani, and D. Maio, "A multimodal approach for human activity recognition based on skeleton and RGB data," *Pattern Recognit. Lett.*, vol. 131, pp. 293–299, Mar. 2020.

[26] D. G. Ivanovich, C. Xu, and P. Julián, "Event-based hand shadow recognition with varied light intensity and background subtraction," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Oct. 2021, pp. 1121–1124.

[27] B. Xu, J. Li, Y. Wong, Q. Zhao, and M. S. Kankanhalli, "Interact as you intend: Intention-driven human-object interaction detection," *IEEE Trans. Multimedia*, vol. 22, no. 6, pp. 1423–1432, Jun. 2020.
[28] A. Jalal and S. Kamal, "Real-time life logging via a depth silhouette-based human activity recognition system for smart home services," in *Proc. 11th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2014, pp. 74–80.
[29] Z. Lei, J. Xie, and L. Xiao, "Inertial sensor-based human activity recognition using hybrid deep neural networks," in *Proc. 14th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2021, pp. 1–7.
[30] S. Perez-Gamboa, Q. Sun, and Y. Zhang, "Improved sensor based human activity recognition via hybrid convolutional and recurrent neural networks," in *Proc. IEEE Int. Symp. Inertial Sensors Syst. (INERTIAL)*, Mar. 2021, pp. 1–4.
[31] S. An, Y. Li, and U. Ogras, "MRI: Multi-modal 3D human pose estimation dataset using mmWave, RGB-D, and inertial sensors," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 1–20.
[32] B. M. Abidine, L. Fergani, B. Fergani, and M. Oussalah, "The joint use of sequence features combination and modified weighted SVM for improving daily activity recognition," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 119–138, Feb. 2018.
[33] H. Chen, G. Wang, J.-H. Xue, and L. He, "A novel hierarchical framework for human action recognition," *Pattern Recognit.*, vol. 55, pp. 148–159, Jul. 2016.
[34] T. Pan, Y. Song, T. Yang, W. Jiang, and W. Liu, "VideoMoCo: Contrastive video representation learning with temporally adversarial examples," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11205–11214.
[35] B. Zhang, L. Wang, Z. Wang, Y. Qiao, and H. Wang, "Real-time action recognition with enhanced motion vector CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2718–2726.
[36] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, "Dynamic image networks for action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3034–3042.
[37] F. J. Zampella, A. R. Jiménez, F. Seco, J. C. Prieto, and J. I. Guevara, "Simulation of foot-mounted IMU signals for the evaluation of PDR algorithms," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Sep. 2011, pp. 1–7.
[38] S. R. Krishnan and C. S. Seelamantula, "On the selection of optimum Savitzky–Golay filters," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 380–391, Jan. 2013.
[39] D. Baldissera, L. Nanni, S. Brahnam, and A. Lumini, "Postprocessing for skin detection," *J. Imag.*, vol. 7, no. 6, p. 95, Jun. 2021.
[40] S. Salazar-Colores, E. Cabal-Yepez, J. M. Ramos-Arreguin, G. Botella, L. M. Ledesma-Carrillo, and S. Ledesma, "A fast image dehazing algorithm using morphological reconstruction," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2357–2366, May 2019.
[41] B. Xiong, S. D. Jain, and K. Grauman, "Pixel objectness: Learning to segment generic objects automatically in images and videos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2677–2692, Nov. 2019.
[42] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," in *IEEE Access*, vol. 7, pp. 44317–44328, 2019.
[43] Y. Zhou, S. Hong, J. Shang, M. Wu, Q. Wang, H. Li, and J. Xie, "Addressing noise and skewness in interpretable health-condition assessment by learning model confidence," *Sensors*, vol. 20, no. 24, p. 7307, 2020.
[44] W. Deng, S. Zhang, H. Zhao, and X. Yang, "A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing," *IEEE Access*, vol. 6, pp. 35042–35056, 2018, doi: 10.1109/ACCESS.2018.2834540.
[45] F. Sekak, F. Elbahhar, and M. Haddad, "Study and evaluation of the vital signs detection based on the third order cyclic temporal moment and cumulant," *IEEE Access*, vol. 10, pp. 59603–59611, 2022.
[46] S. Ali, R. Khan, A. Mahmood, M. Hassan, and A. Jeon, "Using temporal covariance of motion and geometric features via boosting for human fall detection," *Sensors*, vol. 18, no. 6, p. 1918, Jun. 2018.
[47] P. Tabaghi, I. Dokmanic, and M. Vetterli, "Kinetic Euclidean distance matrices," *IEEE Trans. Signal Process.*, vol. 68, pp. 452–465, 2020.
[48] Z. Xu, Z. Liang, Z. Zhou, Z. Li, G.-M. Du, X. Wang, and Y.-K. Song, "A precise 3D positioning approach based on UWB with reduced base stations," in *Proc. IEEE 15th Int. Conf. Anti-Counterfeiting, Secur., Identificat. (ASID)*, Oct. 2021, pp. 145–149.
[49] J. I. Poveda and N. Li, "Robust hybrid zero-order optimization algorithms with acceleration via averaging in time," *Automatica*, vol. 123, Jan. 2021, Art. no. 109361.
[50] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multiagent optimization," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 269–281, Mar. 2021.
[51] Y. Wu, Z. Wu, and C. Fu, "Continuous arm gesture recognition based on natural features and logistic regression," *IEEE Sensors J.*, vol. 18, no. 19, pp. 8143–8153, Oct. 2018.
[52] I. J. Tsang, F. Corradi, M. Sifalakis, W. Van Leekwijck, and S. Latré, "Radar-based hand gesture recognition using spiking neural networks," *Electronics*, vol. 10, no. 12, p. 1405, Jun. 2021.
[53] L. Martínez-Villaseñor, H. Ponce, J. Brieva, E. Moya-Albor, J. Núñez-Martínez, and C. Peñafort-Asturiano, "UP-fall detection dataset: A multimodal approach," *Sensors*, vol. 19, no. 9, p. 1988, Apr. 2019.
[54] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Comput. Methods Programs Biomed.*, vol. 117, no. 3, pp. 489–501, Dec. 2014.
[55] A. Sucerquia, J. López, and J. Vargas-Bonilla, "SisFall: A fall and movement dataset," *Sensors*, vol. 17, no. 12, p. 198, Jan. 2017.
[56] Y. Y. Ghadi, M. Javeed, M. Alarfaj, T. A. Shloul, S. A. Alsuhibany, A. Jalal, S. Kamal, and D.-S. Kim, "MS-DLD: Multi-sensors based daily locomotion detection via kinematic-static energy and body-specific HMMs," *IEEE Access*, vol. 10, pp. 23964–23979, 2022.
[57] T. M. Le, L. V. Tran, and S. V. T. Dao, "A feature selection approach for fall detection using various machine learning classifiers," *IEEE Access*, vol. 9, pp. 115895–115908, 2021.
[58] C. Li, Y. Hou, P. Wang, and W. Li, "Joint distance maps based action recognition with convolutional neural networks," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 624–628, May 2017.
[59] B. Kwolek and M. Kepski, "Improving fall detection by the use of depth sensor and accelerometer," *Neurocomputing*, vol. 168, pp. 637–645, Nov. 2015.
[60] F. Merrouche and N. Baha, "Fall detection based on shape deformation," *Multimedia Tools Appl.*, vol. 79, nos. 41–42, pp. 30489–30508, Nov. 2020.
[61] D. R. Beddiar, M. Oussalah, and B. Nini, "Fall detection using body geometry and human pose estimation in video sequences," *J. Vis. Commun. Image Represent.*, vol. 82, Jan. 2022, Art. no. 103407.
[62] M. Batool and M. Javeed, "Fundamental recognition of ADL assessments using machine learning engineering," in *Proc. 19th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Aug. 2022, pp. 195–200.
[63] F. A. Rashid, K. Sandrasegaran, and X. Kong, "Simulation of SisFall dataset for fall detection using MATLAB classifier algorithms," in *Proc. 12th Int. Symp. Parallel Archit., Algorithms Program. (PAAP)*, Dec. 2021, pp. 62–68.
[64] X. Yu, H. Qiu, and S. Xiong, "A novel hybrid deep neural network to predict pre-impact fall for older people based on wearable inertial sensors," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 63, Feb. 2020.
[65] M. Musci, D. De Martini, N. Blago, T. Facchinetti, and M. Piastra, "Online fall detection using recurrent neural networks on smart wearable devices," *IEEE Trans. Emerg. Topics Comput.*, vol. 9, no. 3, pp. 1276–1289, Jul. 2021.
[66] F. Heidarivincheh, M. Mirmehdi, and D. Damen, "Action completion: A temporal model for moment detection," 2018, *arXiv:1805.06749*.

SADAF HAFEEZ received the B.S. degree in computer science from the University of Central Punjab, in 2018. She is currently pursuing the M.S. degree with the Computer Science Program, Air University, Islamabad. Her research interests include digital signal/image processing, machine learning, and artificial intelligence.



SAUD S. ALOTAIBI received the bachelor's degree in computer science from King Abdul Aziz University, in 2000, the master's degree in computer science from King Fahd University, Dhahran, in May 2008, and the Ph.D. degree in computer science from Colorado State University, Fort Collins, USA, in August 2015, under the supervision of Dr. Charles Anderson. After his master's degree, he was a Deputy of the IT-Center for E-Government and Application Services, Umm Al-Qura University, Makkah, Saudi Arabia, in January 2009, where he started his career as an Assistant Lecturer, in July 2001. From 2015 to 2018, he was with the Deanship of Information Technology to improve the IT services that are provided to Umm Al-Qura University, where he is currently an Assistant Professor in computer science. He is also with the Computer and Information College as the Vice Dean for Academic Affairs. His current research interests include AI, machine learning, natural language processing, the neural computing IoT, knowledge representation, smart cities, wireless, and sensors.

ABDULWAHAB ALAZEB received the B.S. degree in computer science from King Khalid University, Abha, Saudi Arabia, in 2007, the M.S. degree in computer science from the Department of Computer Science, University of Colorado Denver, USA, in 2014, and the Ph.D. degree in cybersecurity from the University of Arkansas, USA, in 2021. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. He received the Graduate Certificate in cybersecurity from the University of Arkansas. His research interests include cybersecurity, cloud and edge computing security, machine learning, and the Internet of Things.

NAIF AL MUDAWI received the master's degree in computer science from Australian La Trobe University, in 2011, and the Ph.D. degree from the Collage of Engineering and Informatics, University of Sussex, Brighton, U.K., in 2018. During his academic journey to obtain the master's degree, he was a member of the Australian Computer Science Committee. He is currently an Assistant Professor with the Department of Computer Science and Information System, Najran University. He has published many research and scientific articles in many prestigious journals in various disciplines of computer science.

WOOSEONG KIM received the Ph.D. degree in computer science from UCLA. He was a Researcher with Samsung Electronics, Hyundai Motor, LG Electronics, and SK Hynix Semiconductor. He is currently a Professor with the Computer Engineering Department, Gachon University, South Korea. He had standardization activity in several SDOs, such as 3GPP, ITU, TTA, and ETSI. His research interests include multi-hop ad hoc networks, LTE and 5G wireless telecommunication systems, wireless LAN, SDN/NFV, and the IoT protocols.