IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 16 August 2023, accepted 13 September 2023, date of publication 22 September 2023, date of current version 2 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3317893

RESEARCH ARTICLE

Deep Ontology-Based Human Locomotor Activity Recognition System via Multisensory Devices

MADIHA JAVEED[1], (Graduate Student Member, IEEE), NAIF AL MUDAWI[2], ABDULWAHAB ALAZEB[2], SAUD S. ALOTAIBI[3], NOUF ABDULLAH ALMUJALLY[4], AND AHMAD JALAL[1]

[1]Department of Computer Science, Air University, Islamabad 44000, Pakistan
[2]Department of Computer Science, College of Computer Science and Information System, Najran University, Najran 55461, Saudi Arabia
[3]Information Systems Department, Umm Al-Qura University, Makkah 24382, Saudi Arabia
[4]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Corresponding author: Nouf Abdullah Almujally (naalmujally@pnu.edu.sa)

The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Group Funding program grant code (NU/RG/SERC/12/6). This research is supported and funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Recognition of human locomotor activities is crucial for monitoring the motion patterns. Current studies for human locomotor activities recognition focused on detecting basic motion patterns. In this study, we proposed a four-modules-based human locomotor recognition model via deep learning, which will support in identifying two signal patterns including static and kinematic motions and classifying the daily activities across different subjects. These motion patterns have been monitored through visual devices along with physical and ambient sensors to extract the complex and basic motion from distinct data forms. The four modules include processing, extraction, optimization, and recognition. Each module focuses on certain processing elements for human locomotion recognition. The processing module represents the pre-processing and segmentation stages for motion and ambient-based data along with extraction of human skeleton points from the visual data. Next, the extraction phase focuses on motion patterns identification and features extraction from the multisensors-based data. Then, the optimization module helps in prominent features selection via genetic algorithm. Furthermore, the recognition module utilized a deep learning technique called hidden Markov model to detect human locomotor activities. The average accuracy rates of 73.05% and 71.14% have been achieved for high-level and atomic-level activities over both datasets. The experimental results have shown that the proposed model outperforms the conventional multisensory systems based on deep classifiers via confidence levels for each skeleton point extracted.

INDEX TERMS Activity recognition, classification, deep learning, locomotion prediction, multisensory devices, patient monitoring, pattern recognition.

I. INTRODUCTION

Human locomotor activities [1] are important for several applications, such as patient monitoring [2], indoor-outdoor activity management [3], robotic learning [4], and muscle fatigue [5]. Therefore, locomotion recognition has been a research hotspot for decades. Wearable multisensor- [6] and vision-based [7] devices have accelerated remote monitoring via human locomotion recognition (HLR). HLR systems can be built upon data from wearable sensors [8], vision sensors [9], and the fusion of multisensory devices [10]. These systems can assess human physical and physiological status using good learning modules from machine learning (ML) and deep learning (DL) techniques.

Several HLR methods with multiple sensing technologies and learning modules have been proposed in the last five years [11], [12], [13], [14], [15]. A few systems are based on wearable sensors including accelerometer, gyroscope, and magnetometer [16], while others focused on

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

utilizing the physiological sensors, such as, electromyography (EMG), electrocardiography (ECG), and mechanomyography (MMG) [17]. Some methods preferred to apply the vision-based sensors, while others used the ambient sensors attached to different indoor devices.

In this study, we propose a multi-sensory four-modules-based HLR model to recognize multiple activity types. Enhanced locomotor activities recognition has been attained by providing missing data from a sensor in the form of data from other sensors taking advantage of multi-sensory devices. First, the proposed system recognized human locomotor activities using state-of-the-art filtration technique and motion patterns identification method in processing module. Next, the features mined in this paper have demonstrated good potential for the identification of indoor human movements. The proposed combination of features extraction in extraction module for visual data gives higher accuracy for HLR, where five skeleton body-points have been identified to further extract body-points-based features along with full-body-based features for vision-based data.

Then, a genetic algorithm has been utilized in optimization module to select relevant features. Furthermore, a hidden Markov model-based classifier has been used to evaluate the proposed model in the last module called recognition module. The analysis of results via precision, recall, F1-score, and specificity has also indicated that the proposed technique has more compliance toward complexities present in human activity motion patterns. Overall, the proposed multisensors-based HLR system outperformed the conventional single-sensor-based systems via confidence levels achieved for each skeleton body-point. This paper also presents the domain-specific ontology of the system to be deployed in a real-time environment. This study adds to the research community through the following key contributions:

- The novel signal filtration technique for wearable sensors-based data has been proposed in processing module that helps to avoid errors from the setting and orientation modifications. Hence, this study provides a novel and improved filtration of inertial measurement unit (IMU) signals.
- A state-of-the-art motion patterns identification via extraction module has been proposed for physical data resulting in enhanced recognition for complex and basic motion patterned activities.
- An innovative fitness function has been proposed in this paper to achieve improved results in terms of features optimization via genetic algorithm.

The remainder of this article is organized as follows. Section II describes the current literature related to single-sensor-based and multi-sensors-based HLR systems. Section III describes the methodology used for HLR and explains the four modules utilized in the proposed model. Section IV presents experiments performed using the proposed model along with their analysis. Section V gives details of domain-specific ontology for the proposed system and section VI discusses the future research directions and concludes the study.

II. RELATED WORK

Recently, researchers have explored different sensing technologies and classifying methods to achieve improved HLR ability [18], [19], [20], [21], [22], [23].

A. HLR VIA VISUAL SENSORS-BASED SYSTEM

First, visual sensor-based systems are considered. Batchuluun et al. [18] proposed a novel method to extract the joints and skeleton information for activity recognition using a generative adversarial network (GAN) to enhance joint and skeleton extraction. This method comprised four steps, preprocessing, joint-GAN, postprocessing, and convolutional neural network (CNN)–long short-term memory (LSTM) learning. This method focuses on converting 1-channel thermal images to 3-channel thermal images. Nevertheless, the 3-channel thermal images could not thoroughly differentiate between humans and other objects. Hence, humans were not detected correctly in images with low spatial texture information causing low performance for the propose method. Yin et al. [19] proposed a real-time model for human action detection in the healthcare field. Multichannel LSTM was used as a functional and versatile three-dimensional skeleton-based action detection system. A special loss function was used to improve accuracy. Experiments were performed on the NTU RGB+D and TST Fall detection v2 datasets. However, the frame-level error detection technique did not accomplish error-free skeleton data and was unable to assess the cause of the dimensionality issues causing the system to achieve lower accuracy rates. Similarly, another study presented in [20] focused on residual convolutional neural network (CNN) and spatial attention module for action recognition from video frames by extracting spatiotemporal features. The model helped in focusing or suppressing certain portions of the frame in order to get more specific information through features extraction from both spatial and temporal domains. Though, the proposed technique did not use optical flow maps and consequently the performance was compromised.

B. HLR VIA WEARABLE SENSORS-BASED SYSTEMS

Second, we consider wearable sensors-based systems. Mutegeki et al. [21] proposed a simple method for HLR. They employed both CNN and LSTM, instead of features engineering module, to improve the performance for a small number of activities. Different combinations of CNN and LSTM have been proposed. Nonetheless, the results obtained in [21] demonstrate that the performance decreased for the complex actions like atomic-level activities. Besides, the softmax loss increased as the model got more complex; therefore, using both CNN and LSTM layers did not improve the results. Hajjej et al. [22] used single-sensor-based data and a Quaternion based filtration technique. Then, multiple segmentation techniques have been applied to make windows

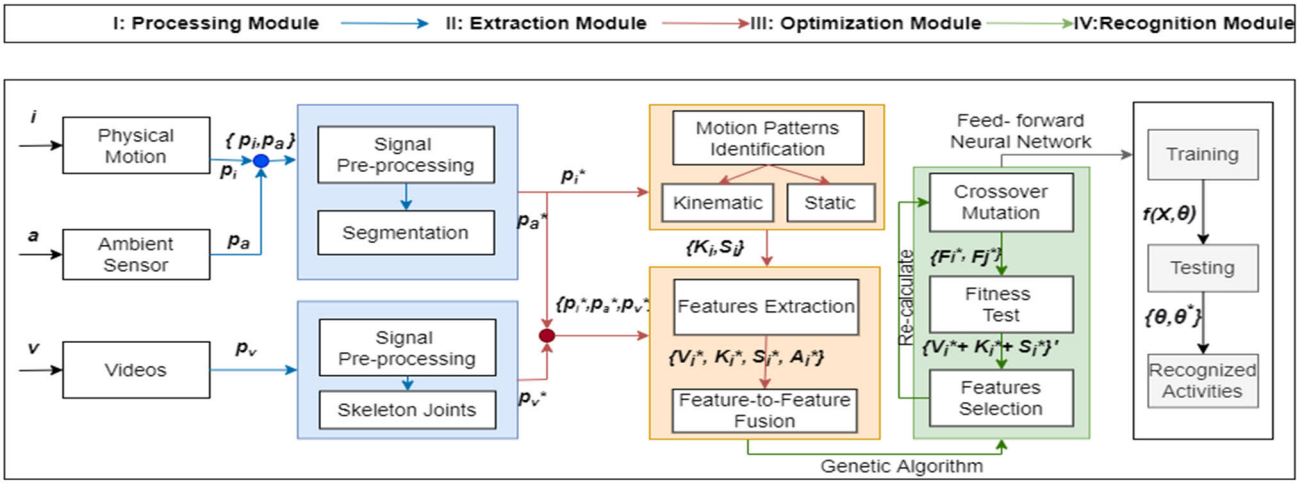

FIGURE 1. Flow architecture of the proposed MS-DLD system.

of the filtered data. Next, the patterns type has been identified and features are extracted and selected. Further, to classify the activities, LSTM has been utilized. Though, this study has few limitations including response time delay and more computational costs. In [23], authors represent a wearable human activity recognition methodology that used IMU sensors to collect data from the actions performed. Next, the data is pre-processed through multiple steps, including drop-out data, global normalization, moving average, sliding overlap windows, and data segmentation methods. Five different deep learning models, such as CNN, recurrent neural network, LSTM, BiLSTM, and gate recurrent unit, have been utilized to recognize different human activities performed. However, the system's performance in terms of time is compromised due to a total of 892,839 learning epochs.

C. HLR VIA AMBIENT SENSORS-BASED SYSTEMS

Third, HLR methods based on ambient sensors are also present in the literature [24], [25], [26]. Natani et al. [24] designed a framework based on multiple neural networks. They used multilayer perceptron, CNN, RNN, and CNN–RNN ensemble models. Moreover, performance is not reliable when it comes to deep learning methods and they also take large amount of time and computations. An activity recognition framework based on the detection of behavior transformation and sensor data segmentation was introduced by Chen et al. [25]. They proposed a hybrid fuzzy c-means for classifying the sensor events and change point detection-based segmentation system to give segmented sensor events. Both machine learning and deep learning were used to perform the classification. However, the events were misclassified due to using less efficient machine learning techniques. Their method was also evaluated on the CASAS dataset, which did not support complex motion patterns. Hence, this approach is not applicable to complex human locomotion events. The authors proposed a binary ambient sensors-based system in [26]. A deep CNN has been used in the model for passive infrared and door-based sensors data. Activities of daily living have been selected using adaptive boosting and Fuzzy c-means classifiers. Although the system performed well but could not recognize the complex activities like relaxing room, dining room, bedroom, and leaving home with acceptable accuracies.

D. HLR VIA MULTI-SENSORS-BASED SYSTEMS

Finally, the multi-sensors-based systems have been explored to see the literature in detail. In [27], the authors have proposed a multimodal locomotion classification system by considering Opportunity++ and HWU-USP datasets as the input. The model suggested to pre-process the multimodal data from these datasets through multiple techniques for each signal type. For visual data, human skeleton has been extracted and for inertial data, filtration along with windowing has been applied. Further, the features have been extracted via SLIF, linear prediction cepstral coefficients, and Pearson correlation. Finally, the features are fused and optimized followed by locomotion prediction by using recursive neural network. Although the system performed better than other single-sensors-based systems, however, the confidence levels achieved for each extracted skeleton body-point are not up to the mark especially for both ankle points. Another sensors-based human activity recognition system has been proposed in [28]. The system used multiple built-in sensors available in smart devices and recognized basic and complex activities performed. The raw data has been pre-processed using normalization, windowing, and filtering methods. Further, the system proposed to extract rotational, statistical, time-based, and frequency-based features along with fusion. Multiple learning algorithms have been utilized to classify both basic and complex activities including Naïve Bayes and K nearest neighbor (KNN), and neural network. Yet, these learning methodologies are prone to errors and

provide less accuracy when it comes to HLR, therefore the performance lacked acceptable results. In [29], Nafea et al. showed a human activity recognition system to monitor people remotely. Multi-sensors have been utilized to record data and deep learning models have been proposed to recognize the activities, such as CNN and gated recurrent unit. However, the proposed method could not perform well due to huge loss of training and validation sets.

III. PROPOSED HLR SYSTEM

There are a few drawbacks faced with different methods present in the literature. Misclassifications can occur due to the subjects' variations, activity conditions, sensor usage, and processing techniques used, thereby decreasing the accuracy [30], [31], [32]. Therefore, an enhanced HLR model is proposed to address such issues. Fig. 1 describes the proposed multisensor-based HLR model with processing, extraction, optimization, and recognition modules. In module I named as processing, raw signals from multisensory devices such as inertial measurement units (IMUs), ambient sensors, and videos are processed through filtration and segmentation or skeleton modeling steps. Next, the extraction module has been used to extract the motion patterns for physical motion-based data and features followed by their fusion for all three sensor types. Then, the optimization module focuses on selecting relevant features. Finally, the deep learner, RNN [33], is used to train and test human locomotor activities by dividing the extracted data into training and test sets in recognition module. Following subsections give the details of each proposed module as mentioned in Fig. 1:

A. MODULE I: PROCESSING

Raw signals consist of physically attached sensor-based signals p_i, ambient sensor-based signals p_a, and video frame sequences p_v. p_i and p_a were pre-processed using a novel filter called the gravity quaternion change normalization filter proposed by this study. Initially, the p_i (raw IMU signals) and p_a(light, switches, RF [34] etc.) represented together as p_i, p_a, have been filtered through low-pass Butterworth filter LPF_s and high-pass Butterworth filter HPF_s [35]. Then, the LPF_s and HPF_s are normalized using the Euclidean distance:

$$Norm = \sqrt{LPF_1 + LPF_2 + LPF_3} + , \sqrt{HPF_1 + HPF_2 + HPF_3}, \quad (1)$$

where LPF_1, LPF_2, and LPF_3 denote the low-pass Butterworth filtered values for the x-, y-, and z-axes, respectively, and HPF_1, HPF_2, and HPF_3 represent the high-pass Butterworth filtered values of the x-, y-, and z-axes, respectively.

The videos of multiple subjects have been converted into frame sequences p_v for pre-processing. Then, we select a delta of 50 sequences to consider for preprocessing to avoid repeated processing that can cause time delays and additional computational costs [36]. Moreover, a background sequence is extracted because the cameras were not moving, so it was easy to select it from any number of sequences. For all the selected sequences, the background sequence is removed from the frame by taking the difference between the original frame and the background frame [27]. Fig. 2 presents a frame sequence of vision-based data with the background subtracted. To reduce the noise present in the frames, the discrete wavelet transform is applied over all background-subtracted frames [37].

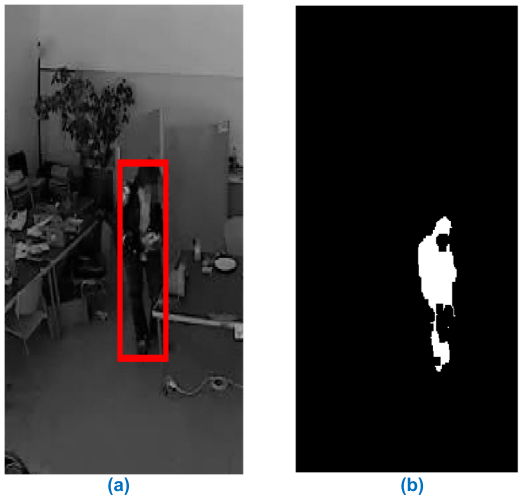

FIGURE 2. (a) Before background subtraction and (b) after background subtracted from a frame sequence of Opportunity++ dataset.

Furthermore, the physical and ambient preprocessed data are further segmented using overlapping time-based windows [38], whereas the visual data are windowed through event-based windows. The segmentation process for all three signal types $\{p_i{}^*, p_a{}^*, p_v{}^*\}$ was achieved as:

$$\{p_i{}^*, p_a^*, p_v{}^*\} = p_i{}^*.\Delta t + p_a{}^*.\partial e + p_v{}^*.\delta s, \quad (2)$$

where $p_i{}^*$ and t give the motion-sensor filtered data and time, $p_a{}^*$ and e provide the ambient-sensor filtered data and event, and $p_v{}^*$ and s denote the vision-based data and frame sequences.

B. MODULE II: EXTRACTION

The extraction module includes motion pattern identification and feature extraction, along with feature-to-feature fusion. Motion pattern identification is implemented on the physical motion-based data p_i only, and all three physical, ambient, and visual data have been processed through features extraction and fusion phases. The three types of signal-based features have been extracted separately and listed in Table 1.

For physical data p_i, this study extracted two patterns separately including kinematic and static patterned signals. For kinematic pattern-based physical signals K_i, we have extracted features via dynamic time warping, whereas for static pattern-based physical signals S_i, we have mined features by using hidden Markov random fields technique. Mel-frequency cepstral coefficients have been extracted for ambient data p_a. For visual data, we proposed to extract features through two techniques including skeleton body-points

TABLE 1. Signal types and features extracted.

Signal Type		Features Extracted
Physical Data p_i	Kinematic Patterns	Dynamic Time Warping
	Static Patterns	Hidden Markov Random Field
Ambient Data p_a	-	Mel-frequency Cepstral Coefficients
Visual Data p_v	Skeleton Body-points	Gray Level Co-occurrence Matrix
	Full-body	Geodesic Distance

and full-body p_v*. Gray-level co-occurrence matrix has been utilized for skeleton body-points-based features extraction and geodesic distance has been recommended for full-body-based features mining.

Motion can be of two types—kinematic and static nature [39]. In this study, we used the polynomial probability distribution [40] method to segment motion patterns from physical data into two pattern distributions as:

$$P(M, a) = w_0 + w_1 a + w_2 a^2 + \ldots + w_M a^M, \quad (3)$$

where w_M represents the unknown weight for each $m \in [0, \ldots, M]$, and the M^{th} order polynomial P on the interval $[x, y]$ are calculated. Each colored line in the Fig. 3 displays the polynomial distribution calculated via Eq. 3 for every window of kinematic physical motion data. The separation of kinematic patterned activities from static patterned activities is based on the first value of polynomial distribution between 0 and 21. Whereas, all the other values represent static motion activities. The kinematic patterns comprised of walking, jogging, kneel etc. activities, however the static patterns include lying, and sitting, etc. activities.

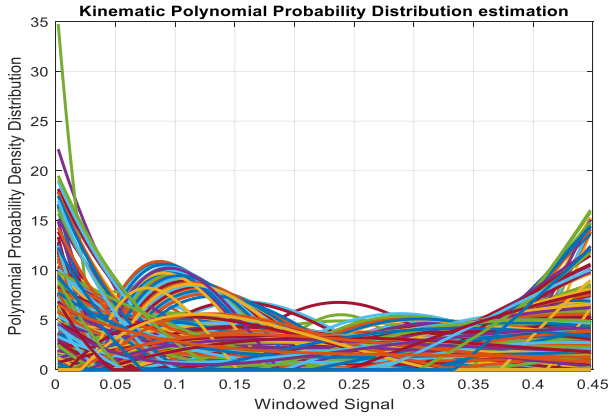

FIGURE 3. Polynomial distribution for kinematic physical activities.

After the pattern identification, different feature extraction techniques are applied to the three signal types to obtain stochastic feature matrices {Vi*, Ki*, Si*, Ai*} using the physical kinematic–static ambient visual (Pks-AV) feature extraction algorithm as given in Algorithm 1.

Algorithm 1 Pks-AV Feature Extraction Algorithm

Input: the kinematic–static physical IMU signals {Ki, Si}, ambient IMU signals {Pa*}, full-body and body-points based visual frame sequences {Pv*}, current window W;
Output: Stochastic features matrix {Vi*, Ki*, Si*, Ai*};
1: extract DTW for W in Ki to get Ki*;
2: extract HMRF for W in Si to get Si*;
3: extract MFCC for W in Pa* to get Ai*;
4: extract GLCM for Full-body Pv* to get Vi*;
5: extract G-dist for Body-points Pv* to get Vi*;
6: fuse the Ki*, Si*, Ai*, and Vi* to get {Vi*, Ki*, Si*, Ai*};

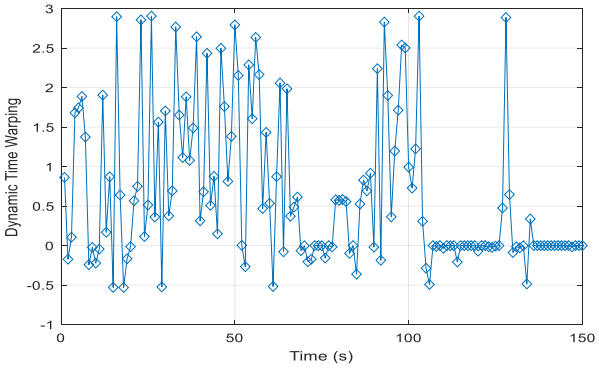

FIGURE 4. DTW calculated for different windows over Opportunity++.

where, {Vi ∗, Ki ∗, Si ∗, Ai ∗} p resents the Vi ∗ visual-data features extracted from gray level co-occurrence matrix (GLCM) and geodesic distance (G-dist) both, the Ki ∗ kinematic motion-data features extracted from dynamic time warping (DTW), the Si ∗ static motion-data features extracted from hidden Markov random fields (HMRF), and Ai ∗ ambient-data features extracted from Mel frequency cepstral coefficients (MFCC).

For kinematic physical signals, a feature extraction method used is DTW. First, DTW calculates the time-based comparison [41] via Eq. 4 and 5 between P and R windows of data using Euclidean distance formula as given in Eq. 6. Then, it searches for warping path using Eq. 7.

$$P = [p_1, p_2, \ldots, p_i, \ldots, p_m] \quad (4)$$

$$R = [r_1, r_2, \ldots, r_j, \ldots, r_n] \quad (5)$$

$$d(p, r) = \sqrt{(p - r)^2} \quad (6)$$

$$PT = [pt_1, pt_2, \ldots, pt_k] \text{ with } \max(m, n) \leq k < m + n - 1 \quad (7)$$

where each pt_k represents the grid extracted via p_m and r_n. DTW cost function has been calculated using Eq. (8). Fig. 4 shows the DTW results calculated for different windows on Opportunity++.

$$\boldsymbol{DTW(P, R) = min\sqrt{\sum_{i=1}^{k} pt_k}}. \quad (8)$$

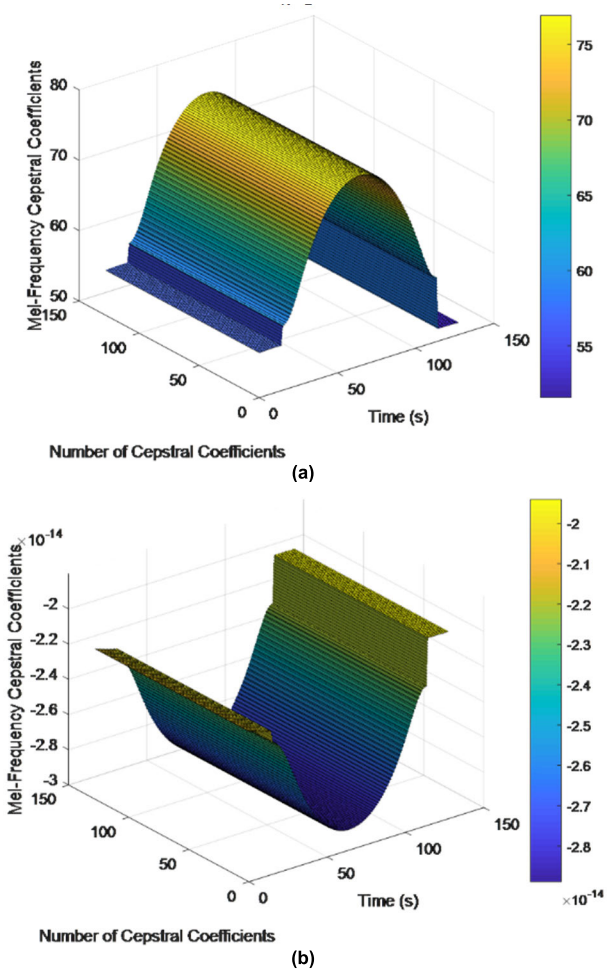

FIGURE 5. MFCC extracted for (a) open drawer and (b) open door sensors.

HMRF is proposed to work with joint likelihood probability as given in Eq. 9 as:

$$P(y|x, \Theta) = \prod_i P(y_i|x_i, \theta_{x_i}), \tag{9}$$

where Θ gives the parameters-based set and $P(y_i|x_i, \theta_{x_i})$ denotes the Gaussian distribution. Next, to estimate the labels, the MAP estimation has been used by extracting prior energy function as:

$$U(x) = \sum_{c \epsilon Cl} F_c(x), \tag{10}$$

where $F_c(x)$ gives the potential clique and Cl denotes the set of possible cliques. After extracting prior energy from cliques, the features similar to clique have been extracted.

For ambient data, MFCC are used to extract relevant features [43]. Formula given in Eq. 11 is used to calculate cepstral coefficients, where d_t is the coefficient from t frames; a typical value used for N is 2. Fig. 5 shows the MFCC features extracted over two different ambient sensor signal windows.

$$d_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2}. \tag{11}$$

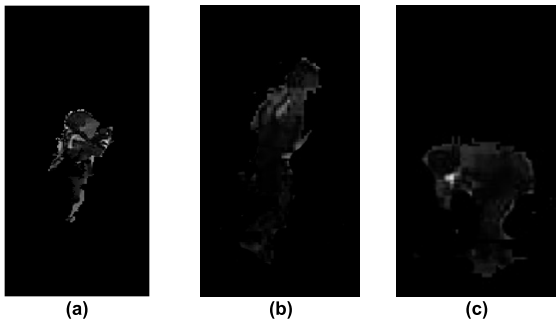

FIGURE 6. GLCM extracted for (a) drinking from cup, (b) open door, and (c) open dishwasher activities on Opportunity++.

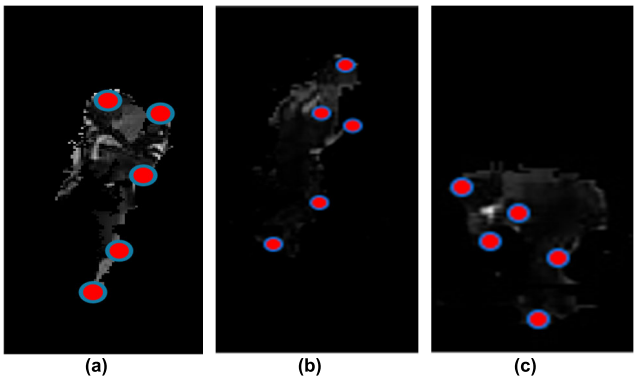

FIGURE 7. Skeleton body points for the geodesic distance calculation of (a) drinking from cup, (b) open door, and (c) open dishwasher on Opportunity++.

For visual data, the features are extracted from two categories. One aims to extract related features from the techniques focusing on detected skeleton-body points. The other emphasizes full-body feature extraction techniques. For full-body feature extraction, a technique called GLCM has been employed.

GLCM supports extracting the texture characterization [44]. If I is the matrix comprising extracted GLCM values, element I(i,j) gives the number of times a pixel of a given gray level i is found adjacent to a pixel of gray level j. GLCM features are calculated by determining the distribution of co-occurring values at a given offset. Fig. 6 shows the GLCM features extracted over a frame sequence of Opportunity++ [45].

For the skeleton-body points, geodesic distance has been calculated using Eq. 12 between the skeleton-body points [46]. It supports finding the angle between two body points and multiplying by the earth's circumference. Fig. 7 presents the skeleton-body points for calculating the geodesic distance [47], including the head, elbow, wrist, knee, and ankle.

$$\text{distance} = \text{angle} * \text{pi} * \text{radius}. \tag{12}$$

C. MODULE III: OPTIMIZATION

All extracted features {Vi ∗, Ki ∗, Si ∗, Ai ∗} have been further fused based on windows W for physical, ambient, and visual data forms. This has increased the vector size

tremendously and requires a feature optimization technique. Hence, the proposed HLR model suggests applying a genetic algorithm (GA), which is an effective strategy for high-dimensional search spaces by taking a small chunk of data to find the global minima with random biological operations, such as crossover, selection, and mutation [48]. Chromosomes are used as the basic units to perform such biological operations. A feature vector is transformed into the equivalent chromosomes by mapping features to respective genes {Fi ∗, Fj ∗} [49]. Then, an optimal search path is derived by considering features as the population of the chromosomes that later on gets filled with the fittest set of featured chromosomes described in Fig. 8. A fitness function proposed in Eq. 13 reflects the key factors for HLR, and the performance is important to keep the response time as less as possible.

$$fitness = \omega_{pd}\bar{s}_{pd} + \omega_{ad}\bar{s}_{ad} + \omega_{vd}\bar{s}_{vd} + \frac{\sigma}{f_n}, \qquad (13)$$

where $\bar{s}$ denotes the average score over all subjects for physical data pd, ambient data ad, and visual data vd. The performance was determined using the data-specific Pearson's correlation and ω is the scaling factor that determines the importance of each signal type. It is set to 0.5, 1.0, and 0.25 in the proposed HLR model for pd, ad, and vd, respectively. To counter the imbalance between signal types, such as the vd having more data points than the other two signal types, thus, we gave the smallest weight to it in order to balance the results. Similarly, ad has less data points than pd, hence assigned more weight. f_n represents the n number of features used as chromosomes and σ is a scale factor set to 0.15 to keep it less prominent in the fitness criteria. $\frac{\sigma}{f_n}$ determines the fitness from the number of features used, as it will give higher fitness with lesser features with $f_n \geq 1$. This fitness function will support the proposed HLR model to attain less number of features selected $\{\text{Vi} * + \text{Ki} * + \text{Si} * + \text{Ai} *\}'$ along with less computational time. The maximal possible fitness with these scale factors is 1.0.

D. MODULE IV: RECOGNITION

A multidimensional vector is produced from the optimization module for all three signal types on the Opportunity++ [45] and Carnegie Mellon University Multi-Modal Activity Database (CMU-MMAC) [50] datasets. Recognition results comprise two levels of locomotor activities—high- and atomic-level locomotion—for the forecast of human locomotion. The high-level locomotor activities comprise of general actions performed that can be included in more than one atomic-level actions. The atomic-level locomotor activities are granular actions that are specific to a type of action and are not being repeated in multiple activities. The sequences of data from both datasets are given as an input to the hidden Markov model (HMM)-based classifier.

IV. EXPERIMENTAL SETTINGS AND RESULTS

Multiple experiments were performed to evaluate the performance of the proposed HLR model. The first experiment

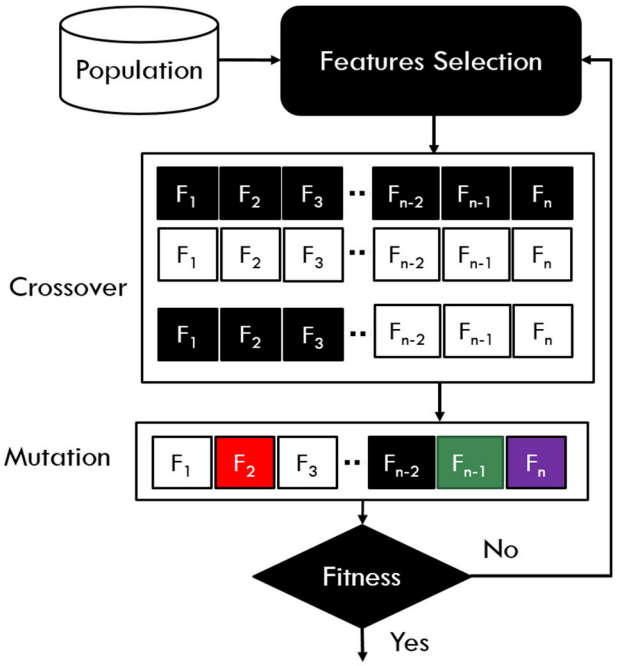

FIGURE 8. Genetic algorithm architecture for HLR system.

for the HLR model's validation is to determine the details of the selected datasets. The second experiment analyzes the proposed model's accuracy for both high- and atomic-level locomotor activities using HMM. It also compares the performance of HMM using different metrics such as precision, recall, and F1-scores over both datasets. The last experiment compares the confidence levels of five skeleton-body points extracted from video sequences with other state-of-the-art model confidence. The experiments were performed using the proposed HLR algorithm implemented on MATLAB with the hardware platform of Intel® i7 Core, 1.80GHz CPU, and 24.0-GB RAM.

Two benchmarked datasets were used to validate the claims about the HLR model's performance, namely, Opportunity++ [45] and CMU-MMAC [50]. Notably, one dataset was collected in a real environment and the other was collected in a synthetic live-in laboratory environment, indicating diversity.

A. OPPORTUNITY++ DATASET

Opportunity++ [45] is a publicly available real-life environment-based dataset containing 25 h of video data along with physical and ambient data in the form of IMU signals, where ambient data refers to 13 reed switches and 8 3D acceleration sensors attached to drawers, kitchen appliances, and doors. Data were collected from 12 subjects who performed multiple high- and atomic-level locomotor activities inside a room. The HLR model focuses on the four high-level locomotor activities—standing, walking, sitting, and lying—along with atomic-level locomotor activities—*open door, close door, open fridge, close fridge, open dishwasher, close*

TABLE 2. Confusion matrix of high-level locomotion over Opportunity++.

Locomotor Activities	Stand	Walk	Sit	Lie
Stand	**5**	4	1	0
Walk	2	**6**	1	1
Sit	1	0	**9**	0
Lie	0	0	0	**10**
Mean Accuracy = 75.0%				

dishwasher, open drawer, close drawer, clean table, drink from cup, and toggle switch. Six types of experiments called runs were performed for every subject, including five daily-life routine activities runs followed by a drill run: *grooming, relaxing, preparing food, eating food, cleaning up, and taking break*. Multiple daily-life routine activities were performed in each run. The pixel resolution for video recording was 640 × 480, and 10 frames per second speed was used. The videos were also anonymized to hide the subjects' identities. Fig. 9 gives a glance of sample frame sequences from the dataset.

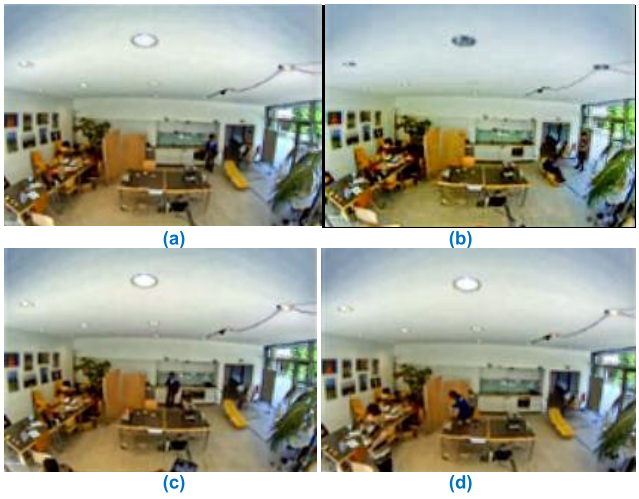

FIGURE 9. Few sample frame sequences from the Opportunity++ dataset showing activities, such as, (a) open drawer, (b) lying, (c) open dishwasher, and (d) clean table.

B. CMU-MMAC DATASET

The second dataset, CMU-MMAC [50], contains multiple sensor modalities, including an accelerometer, gyroscope, color video cameras, and 12 infrared MX-40 cameras. These sensors are located all over the human body, including forearms, upper arms, left calf, right calf, left thigh, right thigh, abdomen, and both wrists. A total of 43 subjects performed daily life activities related to cooking different recipes, such as a brownie, eggs, pizzas, salads, and sandwiches. The high-level locomotor activities include *stand still, walk, sit, turn, bend, kneel, stand up, sit up, and sit down*. The atomic-level locomotor activities include *close, open, clean, fill, out, stir, and shake*. Fig. 10 shows different frame sequences from CMU-MMAC dataset.

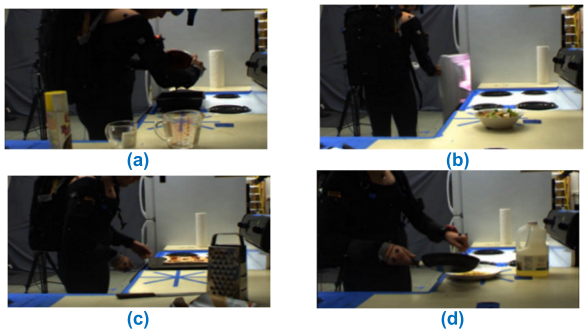

FIGURE 10. Sample frame sequences from the CMU-MMAC dataset showing activities, such as, (a) put, (b) walk, (c) stir, and (d) fill.

FIGURE 11. Confidence levels for models, the model of [45] and the proposed HLR model, on Opportunity++.

C. EXPERIMENT I: VIA OPPORTUNITY++ AND CMU-MMAC DATASETS

The proposed HLR model is validated through confusion matrices reported in Tables 2-5. We fixed the data for training and testing to 80% and 20% split, respectively. The high-level locomotor activities are presented in Tables 2 and 4 for Opportunity++ and CMU-MMAC, respectively. The low-level locomotion has been given in Tables 3 and 5 over Opportunity++ and CMU-MMAC, respectively. We have calculated the accuracy rates through confusion matrices as given in Eq. 14.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{14}$$

where, TP gives true positives, TN shows true negatives, FP provides false positives, and FN represents the false negatives from confusion matrices [51]. The proposed multimodal system has achieved the accuracy rates of 75.0% and 71.11% over high-level locomotion recognition for Opportunity++ and CMU-MMAC datasets, respectively. Similarly, accuracy rates of 73.53% and 68.75% have been achieved for atomic-level locomotion over Opportunity++ and CMU-MMAC datasets, respectively.

D. EXPERIMENT II: VIA EVALUATION METRICS

A summary of the proposed HLR model evaluation metrics is presented in Tables 6 and 7 for Opportunity++ and CMU-MMAC, respectively. The evaluation has been performed in terms of precision [52], recall [53], F1-score [54], and

TABLE 3. Confusion matrix of atomic-level locomotion in Opportunity++ dataset.

Locomotor Activities	OD1	OD2	CD1	CD2	OF	CF	ODW	CDW	ODR 1	CDR 1	ODR 2	CDR 2	ODR 3	CDR 3	CT	DC	TS
OD1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
OD2	1	7	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
CD1	0	0	6	0	0	0	2	0	0	0	0	0	0	0	2	0	0
CD2	0	0	0	9	0	0	0	0	1	0	0	0	0	0	0	0	0
OF	0	2	0	0	6	0	0	0	0	0	0	0	2	0	0	0	0
CF	0	0	0	0	0	8	0	0	0	0	2	0	0	0	0	0	0
ODW	0	0	1	0	0	0	7	0	0	0	0	0	1	0	0	0	1
CDW	0	1	0	0	0	0	0	5	0	0	0	0	0	3	0	0	1
ODR1	0	0	0	1	0	0	0	0	8	0	0	0	0	0	0	1	0
CDR1	0	1	0	0	0	0	0	0	0	7	0	0	0	2	0	0	0
ODR2	0	0	0	4	0	0	0	0	0	0	6	0	0	0	0	0	0
CDR2	0	0	0	0	0	0	1	0	0	0	0	9	0	0	0	0	0
ODR3	0	2	0	0	0	0	0	0	0	0	0	0	7	0	0	1	0
CDR3	1	0	0	0	0	0	1	0	0	0	0	0	0	8	0	0	0
CT	0	0	0	0	0	0	0	0	1	0	0	0	0	0	9	0	0
DC	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	8	0
TS	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0	0	5
Mean Accuracy = 73.53%																	

OD1=Open Door1, OD2=Open Door2, CD1=Close Door1, CD2=Close Door2, OF=Open Fridge, CF=Close Fridge, ODW=Open Dishwasher, CDW=Close Dishwasher, ODR1=Open Drawer1, CDR1=Close Drawer1, ODR2=Open Drawer2, CDR2=Close Drawer2, ODR3=Open Drawer3, CDR3=Close Drawer3, CT=Clean Table, DC=Drink from Cup, TS = Toggle Switch

FIGURE 12. Confidence levels for models, the model of [50] and the proposed HLR model, on CMU-MMAC.

TABLE 4. Confusion matrix of high-level locomotion over CMU-MMAC dataset.

Locomotor Activities	Stand still	Walk	Sit	Turn	Bend	Kneel	Stand up	Sit up	Sit down
Stand still	5	2	1	0	0	0	1	1	0
Walk	0	6	0	1	0	2	0	1	0
Sit	0	1	8	0	0	0	0	0	1
Turn	0	0	0	7	0	0	3	0	0
Bend	0	1	0	0	6	0	1	0	2
Kneel	1	0	0	0	0	9	0	0	0
Stand up	0	0	2	0	0	0	7	0	1
Sit up	2	0	0	0	1	0	0	6	1
Sit down	0	0	0	0	0	0	0	0	10
Mean Accuracy = 71.11%									

specificity [55], which showed that the proposed HLR model gave satisfactory results. Eq. 15-18 show the formulas for precision, recall, F1-score, and specificity, respectively.

$$precision = \frac{TP}{TP + FP}, \tag{15}$$

TABLE 5. Confusion matrix of atomic-level locomotion over CMU-MMAC dataset.

Locomotor Activities	Close	Clean	Open	Fill	Put	Stir	Shake	Other
Close	6	0	2	0	1	0	0	1
Clean	1	5	0	4	0	0	0	0
Open	0	0	8	0	0	1	1	0
Fill	0	0	0	10	0	0	0	0
Put	0	0	2	0	6	0	0	2
Stir	0	1	0	2	0	7	0	0
Shake	0	0	1	0	1	0	8	0
Other	0	1	0	2	0	2	0	5
Mean Accuracy = 68.75%								

$$recall = \frac{TP}{TP + FN}, \tag{16}$$

$$F1 - score = \frac{2 * (recall * precision)}{recall + precision}, \tag{17}$$

$$specificity = \frac{TN}{TN + FP}, \tag{18}$$

where, TP gives true positives, TN shows true negatives, FP provides false positives, and FN represents the false negatives from confusion matrices. From the specificity of the system, we can conclude that the proposed four-modules-based locomotion recognition model is able to detect with 93.54% and 94.0% specificities over Opportunity++ and CMU-MMAC datasets, respectively. Similarly, the system can correctly identify the locomotor activities for 73.5% and 69.0% over Opportunity++ and CMU-MMAC datasets.

E. EXPERIMENT III: EVALUATION USING STATE-OF-THE-ART SYSTEMS

For the proposed HLR model, we chose two datasets, which have been used by different researchers for HLR. To prove the proposed model's superiority for human locomotion-related applications, we compared it with state-of-the-art

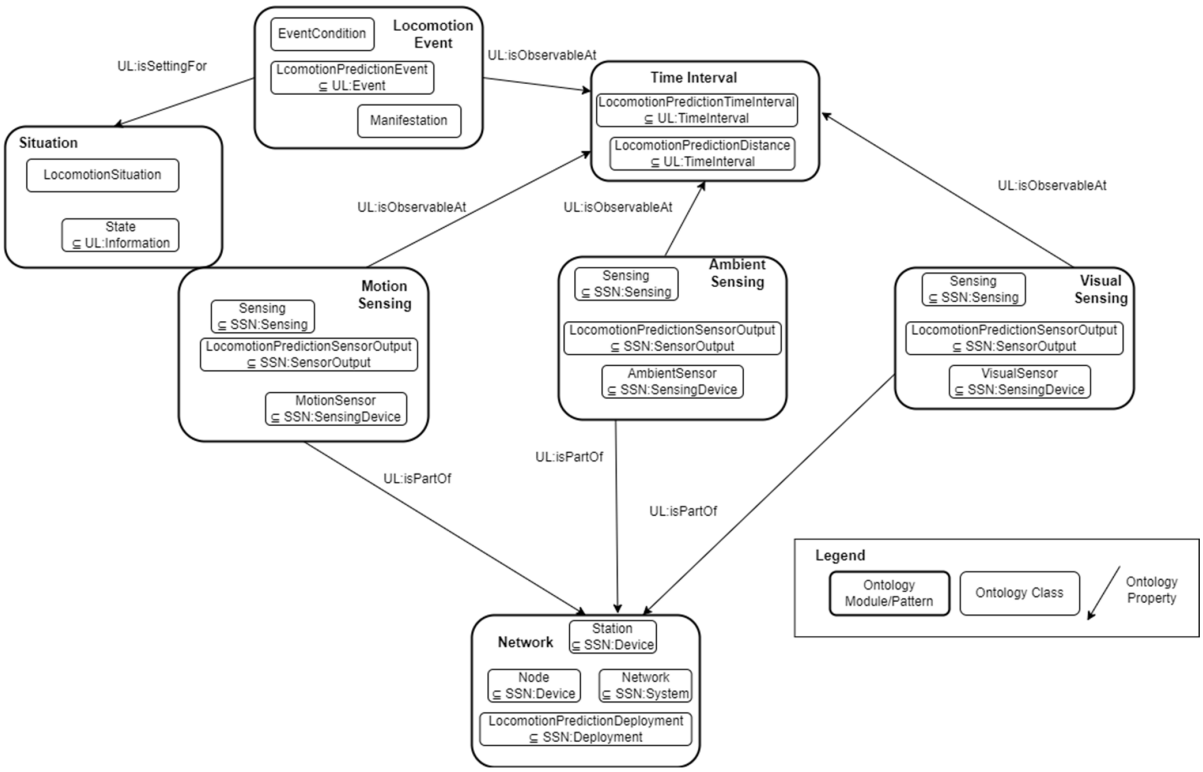

FIGURE 13. Overview of the LocomotionPrediction ontology consisting of 7 ontology modules linked together. SSN, Semantic Sensor Network; UL, Ultra-Light.

TABLE 6. Performance metrics of HMM classifier for high- and atomic-level locomotion in Opportunity++ datasets.

Locomotion	HMM			
HIGH-LEVEL LOCOMOTION				
Activities	**Precision**	**Recall**	**F1-score**	**Specificity**
Stand	0.50	0.62	0.55	0.83
Walk	0.60	0.60	0.60	0.85
Sit	0.90	0.82	0.86	0.95
Lie	1.00	0.91	0.95	1.00
Mean	**0.75**	**0.74**	**0.74**	**0.90**
ATOMIC-LEVEL LOCOMOTION				
Activities	**Precision**	**Recall**	**F1-score**	**Specificity**
OD1	1.00	0.83	0.91	1.00
OD2	0.70	0.54	0.61	0.97
CD1	0.60	0.86	0.71	0.96
CD2	0.90	0.60	0.72	0.99
OF	0.60	0.86	0.71	0.96
CF	0.80	1.00	0.89	0.98
ODW	0.70	0.64	0.67	0.97
CDW	0.50	0.71	0.58	0.96
ODR1	0.80	0.80	0.80	0.98
CDR1	0.70	0.87	0.77	0.97
ODR2	0.60	0.75	0.67	0.96
CDR2	0.90	0.82	0.86	0.99
ODR3	0.70	0.78	0.74	0.97
CDR3	0.80	0.57	0.66	0.98
CT	0.90	0.82	0.86	0.98
DC	0.80	0.80	0.80	0.98
TS	0.50	0.71	0.59	0.96
Mean	**0.73**	**0.76**	**0.73**	**0.97**

TABLE 7. Performance metrics of HMM classifier for high- and atomic-level locomotion in CMU-MMAC datasets.

Locomotion	HMM			
HIGH-LEVEL LOCOMOTION				
Activities	**Precision**	**Recall**	**F1-score**	**Specificity**
Stand still	0.50	0.62	0.55	0.92
Walk	0.60	0.60	0.60	0.93
Sit	0.80	0.73	0.76	0.96
Turn	0.70	0.87	0.77	0.95
Bend	0.60	0.86	0.71	0.93
Kneel	0.90	0.82	0.86	0.98
Stand up	0.70	0.58	0.63	0.95
Sit up	0.60	0.75	0.67	0.93
Sit down	1.00	0.67	0.80	1.00
Mean	**0.71**	**0.72**	**0.70**	**0.95**
ATOMIC-LEVEL LOCOMOTION				
Activities	**Precision**	**Recall**	**F1-score**	**Specificity**
Close	0.60	0.86	0.71	0.92
Clean	0.50	0.71	0.59	0.90
Open	0.80	0.61	0.69	0.95
Fill	1.00	0.56	0.72	1.00
Put	0.60	0.75	0.67	0.92
Stir	0.70	0.70	0.70	0.94
Shake	0.80	0.89	0.84	0.95
Other	0.50	0.62	0.55	0.90
Mean	**0.69**	**0.71**	**0.68**	**0.93**

models having skeleton modeling confidence values between 0 and 1 [27]. Fig. 11 depicts the confidence levels achieved in [45] and the proposed HLR model over Opportunity++. Fig. 12 displays the comparison by extracting confidence levels from [49] and the proposed HLR model on CMU-MMAC. The yellow dots in both Fig. 11 and 12 represent the higher confidence levels achieved by the proposed HLR model, and the blue cylinders indicate the confidence levels achieved in conventional methods [45], [50].

V. DOMAIN-SPECIFIC ONTOLOGY OF HLR SYSTEM

Ontology is a representation of a context-aware system using is-a relationships between different modules of the system along with their set of relations [56], [57], [58], [59]. There are types of ontologies, such as top-level ontology [60],

general ontology [61], and domain-specific ontology [62], [63], [64]. For the proposed HLR system, Fig. 13 shows an overview of domain-specific ontology for Locomotion-Prediction system. This ontology presents the overview of proposed HLR system in terms of physical and actual implementation for real-life projects.

VI. CONCLUSION

In summary, an HLR model based on human locomotor activity is proposed in this study. A combination of physical, ambient, and vision-based sensors is used to recognize applications. The acquired data are preprocessed through unique techniques and then segmented. Further, humans are human locomotion, which is important in many real-time detected via skeleton modeling and patterns are detected in the form of dynamic and static categories. The features are extracted from multiple stochastic extraction methods and fused. Moreover, a GA is used as the feature optimization technique. The optimized features are further fed to an RNN for classification into high- and atomic-level locomotor activities.

The proposed model can be implemented for real-time applications such as gait analysis, animal motion detection, exergaming, robot learning, disease recognition, and educational purposes. However, it still needs improvement in pattern identification and vision-based data preprocessing. The classification results also showed different divergences in values that caused the misclassification of several locomotion activities.

ACKNOWLEDGEMENT

This research is supported and funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R410), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

[1] F. Gu, K. Khoshelham, S. Valaee, J. Shang, and R. Zhang, "Locomotion activity recognition using stacked denoising autoencoders," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2085–2093, Jun. 2018, doi: 10.1109/JIOT.2018.2823084.

[2] M. Batool, S. S. Alotaibi, M. H. Alatiyyah, K. Alnowaiser, H. Aljuaid, A. Jalal, and J. Park, "Depth sensors-based action recognition using a modified K-ary entropy classifier," *IEEE Access*, vol. 11, pp. 58578–58595, 2023, doi: 10.1109/ACCESS.2023.3260403.

[3] N. Khalid, M. Gochoo, A. Jalal, and K. Kim, "Modeling two-person segmentation and locomotion for stereoscopic action identification: A sustainable video surveillance system," *Sustainability*, vol. 13, no. 2, p. 970, Jan. 2021.

[4] X. Hu, Q. Kuang, Q. Cai, Y. Xue, W. Zhou, and Y. Li, "A coherent pattern mining algorithm based on all contiguous column bicluster," *J. Artif. Intell. Technol.*, vol. 2, pp. 80–92, May 2022.

[5] M. Zheng, K. Zhi, J. Zeng, C. Tian, and L. You, "A hybrid CNN for image denoising," *J. Artif. Intell. Technol.*, vol. 2, pp. 93–99, Apr. 2022.

[6] M. Mahmood, A. Jalal, and K. Kim, "WHITE STAG model: Wise human interaction tracking and estimation (WHITE) using spatio-temporal and angular-geometric (STAG) descriptors," *Multimedia Tools Appl.*, vol. 79, nos. 11–12, pp. 6919–6950, Mar. 2020, doi: 10.1007/s11042-019-08527-8.

[7] A. Ahmed, A. Jalal, and K. Kim, "RGB-D images for object segmentation, localization and recognition in indoor scenes using feature descriptor and Hough voting," in *Proc. 17th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Islamabad, Pakistan, Jan. 2020, pp. 290–295, doi: 10.1109/IBCAST47879.2020.9044545.

[8] J. Li, L. Han, C. Zhang, Q. Li, and Z. Liu, "Spherical convolution empowered viewport prediction in 360 video multicast with limited FoV feedback," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 1, pp. 1–23, Jan. 2023.

[9] Y. Zheng, Y. Zhang, L. Qian, X. Zhang, S. Diao, X. Liu, J. Cao, and H. Huang, "A lightweight ship target detection model based on improved YOLOv5s algorithm," *PLOS ONE*, vol. 18, no. 4, 2023, Art. no. e283932, doi: 10.1371/journal.pone.0283932.

[10] A. Nadeem, A. Jalal, and K. Kim, "Automatic human posture estimation for sport activity recognition with robust body parts detection and entropy Markov model," *Multimedia Tools Appl.*, vol. 80, no. 14, pp. 21465–21498, Jun. 2021, doi: 10.1007/s11042-021-10687-5.

[11] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *IEEE Access*, vol. 8, pp. 83791–83820, 2020, doi: 10.1109/ACCESS.2020.2991891.

[12] J. Meng, Y. Li, H. Liang, and Y. Ma, "Single image dehazing based on two-stream convolutional neural network," *J. Artif. Intell. Technol.*, vol. 2, pp. 100–110, Jun. 2022.

[13] J. Zhang, G. Ye, Z. Tu, Y. Qin, Q. Qin, J. Zhang, and J. Liu, "A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 46–55, Mar. 2022.

[14] F. S. Hassan and A. Gutub, "Improving data hiding within colour images using hue component of HSV colour space," *CAAI Trans. Intell. Technol.*, vol. 7, no. 1, pp. 56–68, Mar. 2022.

[15] F. Ahmad, "Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 200–218, Jun. 2022.

[16] L. Yan, Y. Shi, M. Wei, and Y. Wu, "Multi-feature fusing local directional ternary pattern for facial expressions signal recognition based on video communication system," *Alexandria Eng. J.*, vol. 63, pp. 307–320, Jan. 2023, doi: 10.1016/j.aej.2022.08.003.

[17] S. Yang, Q. Li, W. Li, X. Li, and A.-A. Liu, "Dual-level representation enhancement on characteristic and context for image-text retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 8037–8050, Nov. 2022, doi: 10.1109/TCSVT.2022.3182426.

[18] G. Batchuluun, J. K. Kang, D. T. Nguyen, T. D. Pham, M. Arsalan, and K. R. Park, "Action recognition from thermal videos using joint and skeleton information," *IEEE Access*, vol. 9, pp. 11716–11733, 2021, doi: 10.1109/ACCESS.2021.3051375.

[19] J. Yin, J. Han, R. Xie, C. Wang, X. Duan, Y. Rong, X. Zeng, and J. Tao, "MC-LSTM: Real-time 3D human action detection system for intelligent healthcare applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, no. 2, pp. 259–269, Apr. 2021, doi: 10.1109/TBCAS.2021.3064841.

[20] B. Chen, F. Meng, H. Tang, and G. Tong, "Two-level attention module based on spurious-3D residual networks for human action recognition," *Sensors*, vol. 23, no. 3, p. 1707, Feb. 2023, doi: 10.3390/s23031707.

[21] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Feb. 2020, pp. 362–366, doi: 10.1109/ICAIIC48513.2020.9065078.

[22] A.-A. Liu, Y. Zhai, N. Xu, W. Nie, W. Li, and Y. Zhang, "Region-aware image captioning via interaction learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3685–3696, Jun. 2022, doi: 10.1109/TCSVT.2021.3107035.

[23] I. E. Jaramillo, J. G. Jeong, P. R. Lopez, C.-H. Lee, D.-Y. Kang, T.-J. Ha, J.-H. Oh, H. Jung, J. H. Lee, W. H. Lee, and T.-S. Kim, "Real-time human activity recognition with IMU and encoder sensors in wearable exoskeleton robot via deep learning networks," *Sensors*, vol. 22, no. 24, p. 9690, Dec. 2022, doi: 10.3390/s22249690.

[24] A. Natani, A. Sharma, and T. Perumal, "Sequential neural networks for multi-resident activity recognition in ambient sensing smart homes," *Appl. Intell.*, vol. 51, no. 8, pp. 6014–6028, 2021.

[25] D. Chen, S. Yongchareon, E. M.-K. Lai, J. Yu, and Q. Z. Sheng, "Hybrid fuzzy C-means CPD-based segmentation for improving sensor-based multiresident activity recognition," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11193–11207, Jul. 2021, doi: 10.1109/JIOT.2021.3051574.

[26] G. Mohmed, A. Lotfi, and A. Pourabdollah, "Employing a deep convolutional neural network for human activity recognition based on binary ambient sensor data," in *Proc. 13th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, Jun. 2020, pp. 1–7, doi: 10.1145/3389189.3397991.

[27] Z. Hu, L. Ren, G. Wei, Z. Qian, W. Liang, W. Chen, X. Lu, L. Ren, and K. Wang, "Energy flow and functional behavior of individual muscles at different speeds during human walking," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 294–303, 2023, doi: 10.1109/TNSRE.2022.3221986.

[28] M. A. Hanif, T. Akram, A. Shahzad, M. A. Khan, U. Tariq, J.-I. Choi, Y. Nam, and Z. Zulfiqar, "Smart devices based multisensory approach for complex human activity recognition," *Comput., Mater. Continua*, vol. 70, no. 2, pp. 3221–3234, 2022, doi: 10.32604/cmc.2022.019815.

[29] O. Nafea, W. Abdul, and G. Muhammad, "Multi-sensor human activity recognition using CNN and GRU," *Int. J. Multimedia Inf. Retr.*, vol. 11, no. 2, pp. 135–147, Jun. 2022, doi: 10.1007/s13735-022-00234-9.

[30] F. Wang, H. Wang, X. Zhou, and R. Fu, "A driving fatigue feature detection method based on multifractal theory," *IEEE Sensors J.*, vol. 22, no. 19, pp. 19046–19059, Oct. 2022, doi: 10.1109/JSEN.2022.3201015.

[31] 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)*, Islamabad, Pakistan, Jan. 2021, pp. 512–517, doi: 10.1109/IBCAST51254.2021.9393014.

[32] J. Xu, S. Pan, P. Z. H. Sun, S. Hyeong Park, and K. Guo, "Human-factors-in-driving-loop: Driver identification and verification via a deep learning approach using psychological behavioral data," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 3383–3394, Mar. 2023, doi: 10.1109/TITS.2022.3225782.

[33] J. Zhang, C. Zhu, L. Zheng, and K. Xu, "ROSEFusion: Random optimization for online dense reconstruction under fast camera motion," *ACM Trans. Graph.*, vol. 40, no. 4, pp. 1–17, Aug. 2021, doi: 10.1145/3450626.3459676.

[34] S. Jiang, C. Zhao, Y. Zhu, C. Wang, and Y. Du, "A practical and economical ultra-wideband base station placement approach for indoor autonomous driving systems," *J. Adv. Transp.*, vol. 2022, pp. 1–12, Mar. 2022, doi: 10.1155/2022/3815306.

[35] A. Nadeem, A. Jalal, and K. Kim, "Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness," *Symmetry*, vol. 12, no. 11, p. 1766, Oct. 2020.

[36] U. Azmat, S. S. Alotaibi, N. Al Mudawi, B. I. Alabduallah, M. Alonazi, A. Jalal, and J. Park, "An elliptical modeling supported system for human action deep recognition over aerial surveillance," *IEEE Access*, vol. 11, pp. 75671–75685, 2023.

[37] A. Osadchiy, A. Kamenev, V. Saharov, and S. Chernyi, "Signal processing algorithm based on discrete wavelet transform," *Designs*, vol. 5, no. 3, p. 41, Jul. 2021, doi: 10.3390/designs5030041.

[38] C. Zhang, P. Xiao, Z.-T. Zhao, Z. Liu, J. Yu, X.-Y. Hu, H.-B. Chu, J.-J. Xu, M.-Y. Liu, Q. Zou, L. Zhang, Q. Liu, and G.-S. Li, "A wearable localized surface plasmons antenna sensor for communication and sweat sensing," *IEEE Sensors J.*, vol. 23, pp. 11591–11599, 2023, doi: 10.1109/JSEN.2023.3266262.

[39] A. Alazeb, U. Azmat, N. Al Mudawi, A. Alshahrani, S. S. Alotaibi, N. A. Almujally, and A. Jalal, "Intelligent localization and deep human activity recognition through IoT devices," *Sensors*, vol. 23, no. 17, p. 7363, Aug. 2023.

[40] B. Cheng, M. Wang, S. Zhao, Z. Zhai, D. Zhu, and J. Chen, "Situation-aware dynamic service coordination in an IoT environment," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2082–2095, Aug. 2017, doi: 10.1109/TNET.2017.2705239.

[41] M. Alonazi, H. Ansar, N. Al Mudawi, S. S. Alotaibi, N. A. Almujally, A. Alazeb, A. Jalal, J. Kim, and M. Min, "Smart healthcare hand gesture recognition using CNN-based detector and deep belief network," *IEEE Access*, vol. 11, pp. 84922–84933, 2023.

[42] Y. Yang, Z. Lin, B. Li, X. Li, L. Cui, and K. Wang, "Hidden Markov random field for multi-agent optimal decision in top-coal caving," *IEEE Access*, vol. 8, pp. 76596–76609, 2020, doi: 10.1109/ACCESS.2020.2984786.

[43] Y. Zhuang, S. Chen, N. Jiang, and H. Hu, "An effective WSSENet-based similarity retrieval method of large lung CT image databases," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 7, pp. 2359–2376, 2022, doi: 10.3837/tiis.2022.07.013.

[44] M. Batool, A. Jalal, and K. Kim, "Telemonitoring of daily activity using accelerometer and gyroscope in smart home environments," *J. Electr. Eng. Technol.*, vol. 15, no. 6, pp. 2801–2809, Nov. 2020.

[45] M. Ciliberto, V. F. Rey, A. Calatroni, P. Lukowicz, and D. Roggen, "Opportunity++: A multimodal dataset for video- and wearable, object and ambient sensors-based human activity recognition," *Frontiers Comput. Sci.*, vol. 3, p. 2624, Dec. 2021.

[46] T. Oda, Y. Itoh, W. Nakai, K. Nomura, Y. Kitamura, and F. Kishino, "Interactive skeleton extraction using geodesic distance," in *Proc. 16th Int. Conf. Artif. Reality Telexistence-Workshops (ICAT)*, Dec. 2006, pp. 275–281, doi: 10.1109/ICAT.2006.76.

[47] D. Li, S. S. Ge, and T. H. Lee, "Fixed-time-synchronized consensus control of multiagent systems," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 1, pp. 89–98, Mar. 2021, doi: 10.1109/TCNS.2020.3034523.

[48] M. A. K. Quaid and A. Jalal, "Wearable sensors based human behavioral pattern recognition using statistical features and reweighted genetic algorithm," *Multimedia Tools Appl.*, vol. 79, nos. 9–10, pp. 6061–6083, Mar. 2020, doi: 10.1007/s11042-019-08463-7.

[49] M. Batool, A. Jalal, and K. Kim, "Sensors technologies for human activity analysis based on SVM optimized by PSO algorithm," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Aug. 2019.

[50] F. D. Torre, J. K. Hodgins, A. W. Bargteil, X. Martin, J. Macey, A. T. Collado, and P. Beltran, "Guide to the Carnegie Mellon University multimodal activity (CMU-MMAC) database," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-08-22, 2008.

[51] H. Liu, H. Yuan, J. Hou, R. Hamzaoui, and W. Gao, "PUFA-GAN: A frequency-aware generative adversarial network for 3D point cloud upsampling," *IEEE Trans. Image Process.*, vol. 31, pp. 7389–7402, 2022, doi: 10.1109/TIP.2022.3222918.

[52] J. Ahmad and M. Mahmood, "Students' behavior mining in E-learning environment using cognitive processes with information technologies," *Educ. Inf. Technol.*, vol. 24, pp. 2797–2821, Mar. 2019.

[53] 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.

[54] 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. IEEE Conf. Int. Conf. Frontiers Inf. Technol.*, Dec. 2018.

[55] X. Hou, L. Zhang, Y. Su, G. Gao, Y. Liu, Z. Na, Q. Xu, T. Ding, L. Xiao, L. Li, and T. Chen, "A space crawling robotic bio-paw (SCRBP) enabled by triboelectric sensors for surface identification," *Nano Energy*, vol. 105, Jan. 2023, Art. no. 108013, doi: 10.1016/j.nanoen.2022.108013.

[56] A. Olivares-Alarcos, D. Beßler, A. Khamis, P. Goncalves, M. K. Habib, J. Bermejo-Alonso, M. Barreto, M. Diab, J. Rosell, J. Quintas, J. Olszewska, H. Nakawala, E. Pignaton, A. Gyrard, S. Borgo, G. Alenyà, M. Beetz, and H. Li, "A review and comparison of ontology-based approaches to robot autonomy," *Knowl. Eng. Rev.*, vol. 34, p. E29, Dec. 2019, doi: 10.1017/S0269888919000237.

[57] N. Bao, T. Zhang, R. Huang, S. Biswal, J. Su, and Y. Wang, "A deep transfer learning network for structural condition identification with limited real-world training data," *Struct. Control Health Monit.*, vol. 2023, Jul. 2023, Art. no. 8899806, doi: 10.1155/2023/8899806.

[58] Q. She, R. Hu, J. Xu, M. Liu, K. Xu, and H. Huang, "Learning high-DOF reaching-and-grasping via dynamic representation of gripper-object interaction," *ACM Trans. Graph.*, vol. 41, no. 4, pp. 1–14, Jul. 2022.

[59] P. Pham and P. Do, "The approach of using ontology as a pre-knowledge source for semi-supervised labelled topic model by applying text dependency graph," *Int. J. Bus. Intell. Data Mining*, vol. 18, no. 4, pp. 488–523, 2021, doi: 10.1504/IJBIDM.2021.115477.

[60] C. Choi and J. Choi, "Ontology-based security context reasoning for power IoT-cloud security service," *IEEE Access*, vol. 7, pp. 110510–110517, 2019, doi: 10.1109/ACCESS.2019.2933859.

[61] T. Derave, T. P. Sales, F. Gailly, and G. Poels, "Sharing platform ontology development: Proof-of-concept," *Sustainability*, vol. 14, no. 4, p. 2076, Feb. 2022, doi: 10.3390/su14042076.

[62] W. H. Bangyal, N. U. Rehman, A. Nawaz, K. Nisar, A. A. A. Ibrahim, R. Shakir, and D. B. Rawat, "Constructing domain ontology for Alzheimer disease using deep learning based approach," *Electronics*, vol. 11, no. 12, p. 1890, Jun. 2022, doi: 10.3390/electronics11121890.

[63] S. M. Khabour, Q. A. Al-Radaideh, and D. Mustafa, "A new ontology-based method for Arabic sentiment analysis," *Big Data Cognit. Comput.*, vol. 6, no. 2, p. 48, Apr. 2022, doi: 10.3390/bdcc6020048.

[64] A. Ławrynowicz, A. Wróblewska, W. T. Adrian, B. Kulczyński, and A. Gramza-Michałowska, "Food recipe ingredient substitution ontology design pattern," *Sensors*, vol. 22, no. 3, p. 1095, Jan. 2022, doi: 10.3390/s22031095.

MADIHA JAVEED (Graduate Student Member, IEEE) received the M.S. degree in computer sciences from the Lahore University of Management Sciences, Lahore, Pakistan. She is currently pursuing the Ph.D. degree in computer science with Air University, Islamabad, Pakistan. From 2010 to 2013, she was a Software Engineer. Since 2013, she has been an Assistant Professor with the University of Management and Technology, Lahore (currently on study leave). Her research interests include artificial intelligence, machine learning algorithms, deep learning classification, human locomotion analysis, inertial signals filtration, multi-sensors technologies, and daily-life patterns recognition.

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

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. His research interests include cybersecurity, cloud and edge computing security, machine learning, and the Internet of Things. He received the Graduate Certificate in cybersecurity from the University of Arkansas.

SAUD S. ALOTAIBI received the Bachelor of Computer Science degree from King Abdulaziz 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, CO, USA, in August 2015, under the supervision of Dr. Charles Anderson. He started his career as an Assistant Lecturer with Umm Al-Qura University, Makkah, Saudi Arabia, in July 2001. After that, he was the Deputy of the IT-Center for E-Government and Application Services, Umm Al-Qura University, in January 2009. 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. He is currently an Assistant Professor of computer science with Umm Al-Qura University. He is working right now with the Computer and Information College as the Vice Dean for Academic Affairs. His current research interests include AI, machine learning, natural language processing, neural computing, the IoT, knowledge representation, smart cities, wireless, and sensors.

NOUF ABDULLAH ALMUJALLY received the Ph.D. degree in computer science from the University of Warwick, U.K. She is currently an Assistant Professor of computer science with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University (PNU), Riyadh, Saudi Arabia. Her research interests include human–computer interaction (HCI), artificial intelligence (AI), machine learning, deep learning, and computer-based applications.

AHMAD JALAL received the Ph.D. degree from the Department of Biomedical Engineering, Kyung Hee University, Republic of Korea. He was a Postdoctoral Research Fellow with POSTECH. He is currently an Associate Professor with the Department of Computer Science and Engineering, Air University, Pakistan. His research interests include multimedia contents, artificial intelligence, and machine learning.