Received December 23, 2019, accepted December 30, 2019, date of publication January 6, 2020, date of current version January 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964237

C2FHAR: Coarse-to-Fine Human Activity Recognition With Behavioral Context Modeling Using Smart Inertial Sensors

MUHAMMAD EHATISHAM-UL-HAQ [1], MUHAMMAD AWAIS AZAM [1,2], YASAR AMIN [1], AND USMAN NAEEM [3]

[1]Faculty of Telecom and Information Engineering, University of Engineering and Technology (UET), Taxila 47050, Pakistan
[2]Faculty of Information Technology, Whitecliffe Technologies, Wellington 6011, New Zealand
[3]Faculty of Science and Engineering, School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K.

Corresponding author: Usman Naeem (u.naeem@qmul.ac.uk)

This work was supported by the Queen Mary University of London, London E1 4NS, U.K., and in part by the Directorate of Advanced Studies Research and Technological Development (ASR&TD), University of Engineering and Technology, Taxila 47050, Punjab, Pakistan.

ABSTRACT Smart sensing devices are furnished with an array of sensors, including locomotion sensors, which enable continuous and passive monitoring of human activities for the ambient assisted living. As a result, sensor-based human activity recognition has earned significant popularity in the past few years. A lot of successful research studies have been conducted in this regard. However, the accurate recognition of *in-the-wild* human activities in real-time is still a fundamental challenge to be addressed as human physical activity patterns are adversely affected by their behavioral contexts. Moreover, it is essential to infer a user's behavioral context along with the physical activity to enable context-aware and knowledge-driven applications in real-time. Therefore, this research work presents ''C2FHAR'', a novel approach for *coarse-to-fine human activity recognition in-the-wild*, which explicitly models the user's behavioral contexts with activities of daily living to learn and recognize the fine-grained human activities. For addressing real-time activity recognition challenges, the proposed scheme utilizes a multi-label classification model for identifying *in-the-wild* human activities at two different levels, i.e., *coarse* or *fine-grained*, depending upon the real-time use-cases. The proposed scheme is validated with extensive experiments using heterogeneous sensors, which demonstrate its efficacy.

INDEX TERMS Activity recognition, behavioral context, context-aware, machine learning, smart sensing.

I. INTRODUCTION

The progression of the Internet of Things (IoT) and smart sensing technologies has made ubiquitous computing an indispensable platform for assisting people in their routine life. IoT offers a much-needed paradigm to ubiquitously connect different sensing modalities for enabling diverse applications relating to human activity monitoring and tracking [1]. The rapid development in smart sensing systems and technologies has encouraged the researchers to make use of these technologies for passive monitoring of people's activities. In recent years, Human Activity Recognition (HAR) has gained significant popularity and become a major research pocket in the areas of pervasive and ubiquitous computing. The goal of HAR is to provide accurate and opportune information as regards to the people's activities by processing data streams originating from different sensing modalities (such as video cameras, wearable/on-body inertial sensors, mobile sensors, and/or ambient sensors) [2]. In recent years, sensor-based HAR has become increasingly significant in wide-ranging disciplines, including human-computer interface (HCI), ambient assistive living in smart homes [3], [4], driving behavior analysis [5], [6], robotics [7], [8], and telecare for personal health monitoring [9]. Passive and continuous monitoring of human activities is necessary for human behavior cognition and understanding. Furthermore, it can provide assistance to the people in their living and working environments. Therefore, HAR is pertinent to all disciplines of life, including patient activity monitoring for health-care, children's activity tracking for their safety, and monitoring of the instructor's and students' activities in a lecture hall for improved learning.

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

Although sensor-based HAR provides accurate and reliable information about people's activities, however, identifying natural human activities *in-the-wild* is quite challenging. Complex and chaotic correlations between human beings and their behavioral contexts make it hard to recognize natural user behavior. Existing HAR systems mostly focus on recognizing the activities of daily living (ADLs) under controlled settings, where the participants are instructed to perform a series of activities in a defined pattern to train the HAR model [10]–[13]. However, natural human activities have significant variability due to rich and intricate behavioral contexts that adversely affect human activity patterns. For instance, when a person is *in a meeting* or *in a car*, the *sitting* posture is usually different. Similarly, when a person *walks alone* or *walks with a group of people*, the *walking* pattern of the person may vary. Unfortunately, the present methodologies for sensor-based HAR are not capable of fairly distinguishing *in-the-wild* human activities with vital inconsistencies in the behavioral contexts. The HAR models trained under controlled settings cannot adapt well to the changes in human activity patterns and the associated behavioral contexts. It ultimately leads to a poor understanding of human activities *in-the-wild*. Therefore, incorporating the user's context information in the HAR model is necessary to learn and recognize how people perform their ADLs in different settings. In [14], the authors presented an *in-the-wild* context recognition model that utilizes heterogeneous sensors from smartphones and smartwatches to recognize a number of human behavioral contexts. They used logistic regression (LR) classifier to infer context information based on a single model-per-label approach that only recognizes the single-label context information at one time. However, this information is not adequate to accurately model multi-label context-aware human activities *in-the-wild*. Furthermore, the explicit relationship between human activities and the associated behavioral contexts is not possible in this case, which is the core of human behavior analysis and essential for many context-aware applications and recommender systems.

In view of the existing challenges associated with HAR and its real-time applications, this research work proposes an innovative approach for *coarse-to-fine human activity recognition* (C2FHAR) *in-the-wild* using heterogeneous sensors. The novelty of the proposed scheme lies in the notion of modeling and recognizing the secondary information concerning human behavioral contexts in combination with the primary ADLs. In this aspect, the proposed method utilizes a *multi-label* classification model that assigns two types of activity labels to each data instance and enables simultaneous recognition of human activities at two different levels. The first level of classification only identifies six (06) single-label primary ADLs (e.g., *lying*, *sitting*, *standing*, *walking*, *bicycling,* and *running*) *in-the-wild*, thus providing a coarse representation of human activities. In contrast, the second level explicitly models human behavioral contexts with the ADLs to learn and recognize a fine-grained representation of human activities *in-the-wild*. For this purpose, fourteen (14) context labels are associated with six primary ADLs in different combinations to produce a set of twenty-nine (29) fine-grained human activities. These contexts signify information relating to the user's social context (*indoor/outdoor/in a car/in a meeting*), secondary activity (*sleeping/watching TV/talking/shopping/surfing the internet /exercise*), and phone position (*in hand/in pocket/in bag/on table*) *in-the-wild*. In this way, *fine-grained* HAR aims to provide the coinciding recognition of the ADLs and the associated behavioral contexts. Thus, a more precise and comprehensive illustration of *in-the-wild* human activities is obtained at the second level, which is crucial for effective decision making in real-time intelligent systems. The primary objective of employing the multi-label classification model for the proposed scheme is to enable real-time recognition of both *coarse* and *fine-grained* activities. For a given data instance to recognize in real-time, the classification is performed independently at two different levels, and the individual classification scores of both outputs are compared. Finally, the output with a higher confidence score or probability is chosen as the final system output. In this way, the proposed scheme offers either *coarse* or *fine-grained* activity representation *in-the-wild* based on real-time scenarios and use-cases. A public domain dataset, i.e., *ExtraSensory* [15], is employed for testing and validating the proposed scheme using three machine learning classifiers, including Decision Tree, Random Forest, and Neural Networks, where the best one provides efficient recognition performance. The proposed scheme can serve as a building block for wide-ranging applications and recommendation systems that utilize knowledge about human activities and their social or/and behavioral contexts to form improved decisions. Also, it can be utilized for human behavior modeling and analysis, which in turn can be used for predicting and preventing health-related risks for personal autonomy. This research work entails the following key contributions:

- A novel multi-label classification scheme is proposed for enabling real-time C2FHAR *in-the-wild*. In this aspect, diverse behavioral contexts are modeled with the ADLs to learn and recognize better the real-world use-cases.
- A systematic and reproducible approach is presented for *fine-grained* activity selection and annotation using the publicly available *ExtraSensory* dataset. This approach is generalizable to any dataset that entails multiple labels related to a single instance.
- The effect of behavioral context modeling on HAR performance is investigated and discussed in detail based on *context-independent* and *context-dependent* HAR experiments.
- A detailed experimental analysis is conducted to investigate the performance of heterogeneous sensors for the proposed C2FHAR scheme using three machine learning classifiers, i.e., Decision Tree, Random Forest, and Neural Networks.

The remaining portion of the paper is organized as follows. Section II discusses the related works for the proposed scheme. Section III explains the proposed methodology of research in detail. Section IV provides detailed experimental results and analyzes the performance of different sensors and machine learning algorithms for the proposed scheme. Finally, Section V concludes the finding of this research work and provides suggestions for future works.

II. RELATED WORK

Sensor-based HAR has undergone extensive research work in the past few years as it provides the notion of passive sensing for human-centric computing with privacy well intact [16]. The detailed surveys about sensor-based HAR approaches are presented in [11]–[13], [17], where the researchers have utilized different types of sensors for HAR, including cameras, wearable inertial sensors, smartphone sensors, and their combination. Wearable sensors are particularly advantageous because of their aptitude to be worn or placed at multiple body positions for efficient HAR. Hence, a lot of research studies focused on manipulating body motion sensors for HAR [18]–[22] and detection of abrupt activities like fall [23], [24]. Jalloul [25] discussed the applications of the wearable sensor in clinical practice and reviewed the wearable sensing systems used for monitoring movement disorders. Huang *et al.* [26] proposed a probabilistic model for risk stratification from the clinical perspectives. Liu *et al.* [27] presented an interval-based probabilistic generative model for recognizing complex hand activities (with inherited structural diversities), which is based on Bayesian network structures. The authors demonstrated the validity of their proposed scheme on the benchmark datasets as well as the self-constructed dataset. The experimental results indicated the competitiveness of their proposed method. In [28], the authors proposed a data-driven probabilistic model for recognizing low and medium level human activities. They evaluated their proposed model on the *Opportunity* dataset and obtained state-of-the-art results. Liu *et al.* [29] utilized wearable sensors for recognizing housekeeping tasks as complex activities of daily living, which achieved successful results. Noor *et al.* [30] proposed an ontology-based approach that utilizes ambient sensors in combination with the wearable sensors to recognize daily living activities in a smart home. Villalonga *et al.* [31] proposed "MIMU-Wear", an ontology-based sensor selection and reconfiguration scheme for wearable activity recognition. Yurtman and Barshan [32] proposed an orientation-invariant scheme for HAR using wearable sensors, which applies different transformations on the raw data as preprocessing to minimize the influence of sensor orientation activity recognition. Their proposed techniques can be applied to the existing wearable systems by simply transforming the time-domain sensor data at the pre-processing stage. The research work done by Sztyler *et al.* [33] proposed a position-aware system for HAR using wearable sensors, which recognizes the sensor placement based on physical activity patterns. Recently, the authors in [34] proposed HuMAn, a wearable-sensor based approach for the classification of 21 complex *at-home* activities. All these wearable sensor-based approaches tend to achieve efficient recognition performance. However, the use of on-body sensors for activity monitoring often creates inconvenience for the users as most people generally hesitate in wearing on-body sensors, which ultimately leads a person to act or behave irrationally. Thus, the aim of recognizing the user's natural behavior is not accomplished.

In recent years, the evolution of smartphones with growing sensing capabilities has attracted a lot of researchers to utilize smartphone inertial sensors for HAR [35]–[37]. The orientation and position sensitivity of smartphone inertial sensors is challenging to tackle, which poorly affects the recognition performance. Most of the existing smartphone-based HAR methods are position-dependent, i.e., these methods require a smartphone to be placed at some fixed position on the user's body. However, in the natural living environment, people do not usually keep a smartphone at a fixed place all the time. The smartphone position keeps on changing depending upon the user's own convenience. In this case, smartphone-based position-dependent HAR schemes fail to recognize human activities with higher accuracy. A few researchers have also attempted to develop smartphone-based position-independent HAR schemes, where the HAR model is trained corresponding to the motion data of the same activity for different smartphone positions [38]–[40]. Smartphone-based position-aware HAR systems have also been developed, which recognize the phone position and human activity at two different levels usiing two or more classifiers [41]–[43]. The inadequacy of these methods is their higher computational cost owing to the use of multiple classifiers for HAR. A few research studies utilized smartphones for user context recognition as well [44]–[46]. However, smartphone-based HAR systems are not capable of recognizing hand-based movements or activities, such as *eating*, *drinking*, and *brushing*. As a result, some research studies also focused on the utilization of heterogeneous sensing for recognizing complex human activities [47]–[49], which have appeared to improve the recognition results in most of the cases. Shoaib *et al.* [50], [51] utilized the mixture of smartphone and smartwatch sensors for recognizing complex human activities and achieved effective recognition performance.

Recently, with the continuous evolution in deep learning algorithms, some research studies have also utilized deep learning schemes [52]–[55] for sensor-based HAR. In [56], [57], the authors presented the detailed surveys on recent advancements in sensor-based HAR using deep learning models and discussed their limitations and future implications. Deep learning-based schemes are computationally costly and thus cannot be implemented effectively for real-time operations on battery-constrained devices, such as smartphones and smartwatches.

III. METHODOLOGY OF RESEARCH

The proposed HAR scheme entails a supervised machine learning approach, which consists of four key steps. These steps include: 1) data acquisition, 2) data preprocessing,

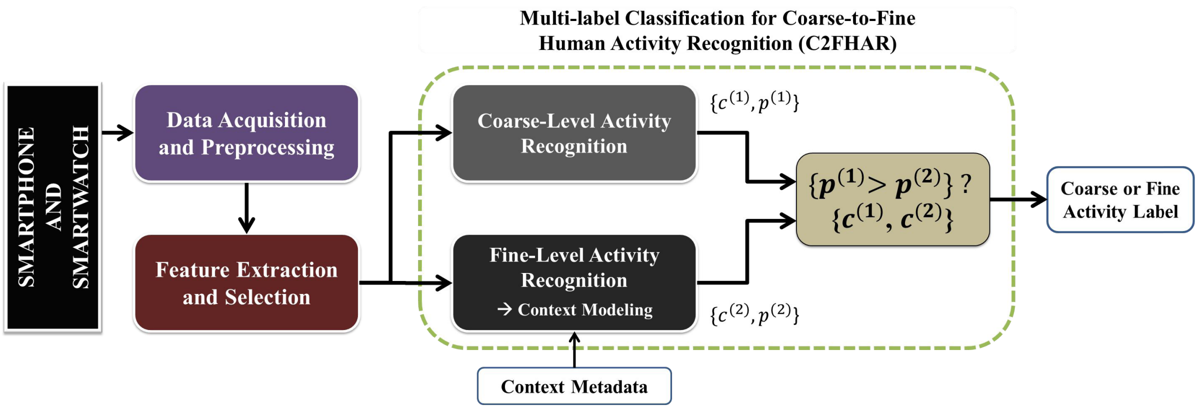

FIGURE 1. Proposed methodology for C2FHAR based on multi-label classification model. Here, $c^{(1)}$ and $c^{(2)}$ represent the predicted class labels for coarse-level and fine-level HAR, respectively, whereas $p^{(1)}$ and $p^{(2)}$ represent the predicted probabilities of $c^{(1)}$ and $c^{(2)}$ respectively.

3) feature extraction, and 4) activity recognition. Fig. 1 shows the block diagram of the proposed methodology of research. The detailed explanation relating to each step is provided in the following sections.

A. DATA ACQUISITION

For validating our proposed scheme, we opted to use the publicly available *ExtraSensory* [15] dataset that conforms to the pipeline of C2FHAR model. This specific dataset is selected on account of three key reasons: 1) the dataset is collected *in-the-wild* from 60 users, including 26 males and 34 females, without imposing any restriction on the users regarding the ADLs execution, 2) the dataset contains a large number of context labels associated with the selected ADLs, which provide supplementary information about a user's context *in-the-wild*, and 3) the dataset consists of heterogeneous data from both smartphone inertial sensors (i.e., accelerometer and gyroscope) and the watch accelerometer. These sensors are utilized individually as well as in different combinations to assess their performance for the proposed scheme.

1) FINE-GRAINED ACTIVITY ANNOTATION

The *ExtraSensory* dataset contains six primary ADLs (same as selected in our scheme) per user. Besides, it provides a large number of secondary context labels (in binary form) corresponding to each instance of these ADLs. These context labels provide auxiliary information about human behavioral contexts when executing a particular primary activity *in-the-wild*. Owing to an *in-the-wild* collection of the dataset, the contexts labels relating to each primary activity are not consistent for all the users. Consequently, the selection and annotation of fine-grained activities, i.e., *primary ADLs incorporated with behavioral context information*, is not a straightforward task. To model context information with the ADLs, relabeling of these activities is required with the joint information of the user's social or behavioral context *in-the-wild*. In this aspect, we performed a systematic analysis of the *ExtraSensory* dataset to realize the co-occurrences of the selected ADLs with different behavioral contexts (such as *user's secondary activity/ location/body state*) and phone positions. Following this, for each user's data, we calculated the frequencies of different behavioral contexts and phone positions that occur in pairs with each primary activity. Finally, we chose twenty-nine (29) most frequently occurring triplets of *primary activity*, *user's behavioral context,* and *phone position* as fine-grained activities, which have a sufficient number of total instances. In this way, a methodical approach is employed for data relabeling and annotation, which is reproducible.

Table 1 shows the list of 29 fine-grained activities (FGAs) obtained as a result of data relabeling. As the *ExtraSensory* dataset is collected *in-the-wild*, hence, based on this analysis, it can be said that these 29 combinations of the ADLs, behavioral contexts, and phone locations are the most common in daily life. This systematic analysis can be helpful for other researchers doing the same kind of work. Overall, fourteen (14) different context labels (including four (04) phone positions) are incorporated with the selected primary ADLs in different combinations, as shown in Fig. 2. These behavioral contexts are mutually exclusive, having no overlapping instances. The activities of *lying, sitting, standing,* and *walking* are modeled with three (03), ten (10), six (06), and eight (08) multi-label contexts (including phone position), respectively, as shown in Table 1. It can be observed that *bicycling* and *running* activities are only used in the context of *exercise with phone in pocket* as there was no other additional information regarding their behavioral contexts in the selected dataset. Likewise, there are a few activity instances among different users, where the data corresponding to the selected behavioral contexts is missing. These data samples are simply ignored for further processing.

B. DATA PREPROCESSING

The raw data acquired from the smartphone or smartwatch sensor may contain different types of high-frequency noise, such as including instrumentation noise or noise generated by the insentient movement of the participants. In this study,

TABLE 1. Set of 29 fine-grained activities for the proposed C2FHAR scheme.

ADLs	Code	FGAs - ADLs with Joint Information of Behavioral Contexts and Phone Positions
Lying	*a1*	*Lying when sleeping with phone on table*
	a2	*Lying when surfing the internet with phone in hand*
	a3	*Lying when watching TV with phone on table*
Sitting	*a4*	*Sitting when surfing the internet with phone in hand*
	a5	*Sitting when surfing the internet with phone on table*
	a6	*Sitting in a car with phone in bag*
	a7	*Sitting in a car with phone in hand*
	a8	*Sitting in a car with phone in pocket*
	a9	*Sitting in a meeting with phone in pocket*
	a10	*Sitting in a meeting with phone on table*
	a11	*Sitting when watching TV with phone in hand*
	a12	*Sitting when watching TV with phone in pocket*
	a13	*Sitting when watching TV with phone on table*
Standing	*a14*	*Standing indoor with phone in bag*
	a15	*Standing indoor with phone in hand*
	a16	*Standing indoor with phone in pocket*
	a17	*Standing indoor with phone on table*
	a18	*Standing outdoor with phone in hand*
	a19	*Standing outdoor with phone in pocket*
Walking	*a20*	*Walking indoor with phone in pocket*
	a21	*Walking indoor with phone on table*
	a22	*Walking outdoor with phone in bag*
	a23	*Walking outdoor with phone in hand*
	a24	*Walking outdoor with phone in pocket*
	a25	*Walking when shopping with phone in bag*
	a26	*Walking when talking with phone in pocket*
	a27	*Walking when talking with phone in hand*
Bicycling	*a28*	*Bicycling during exercise with phone in pocket*
Running	*a29*	*Running during exercise with phone in pocket*

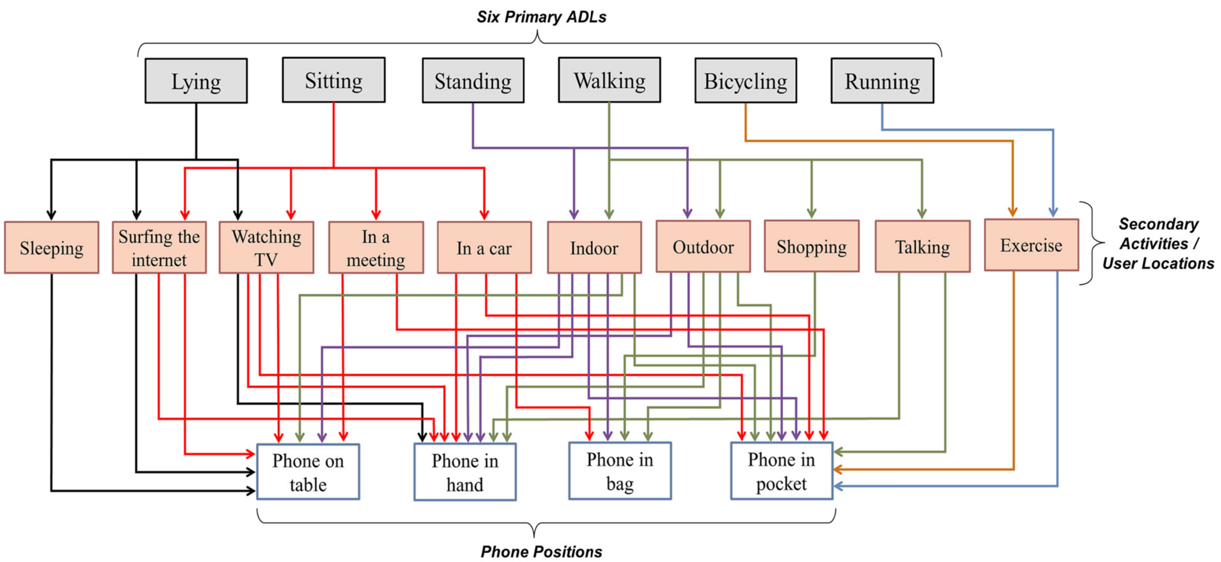

FIGURE 2. Six (06) primary activities of daily living (ADLs) and their association with overall fourteen (14) *in-the-wild* behavioral contexts (including the users' locations/secondary activities and phone positions) to produce fine-grained activities. All ADLs are associated with one or more secondary activities/user locations, which in turn are linked with one or more phone positions. Each primary activity and the associated contexts are represented with the arrow lines, starting from the top till bottom. A few examples of the fine-grained activities obtained in this study include *lying when sleeping with phone on table, sitting in a meeting with phone on table, sitting in a meeting with phone in pocket*, and *walking indoor with phone in pocket*, etc.

we used a computationally efficient time-domain smoothing filter (of order 3) for removing unwanted noise from the signal.

The window size is a crucial element in the data segmentation process, which depends on the sampling rate of the acquisition device as well as the type of activities that have to be recognized. Generally, for simple and cyclic ADLs in some controlled environments, the time duration of $2s$-$5s$ has proved to be sufficient for HAR [17], [50]. In contrast, complex and non-repetitive activities require a more significant amount of time (i.e., $15s$ to $30s$) to produce adequate results [50]. In addition to the simple ADLs, the proposed method also targets to recognize multi-label FGAs *in-the-wild*, which entails intricate and chaotic patterns. Hence, we opted to use a window size of 20-seconds in time for segmenting the raw data samples (obtained from the smartphone and smartwatch sensors at a sampling rate of 40 Hz and 25 Hz, respectively. The *ExtraSensory* dataset also contains the pre-segmented data chunks of 20-seconds in time. Accordingly, we also used the same window size for feature extraction.

C. FEATURE EXTRACTION

Feature extraction generally computes important signal attributes from the preprocessed data, which are helpful for resolving the classification problem in hand. However, the selection and choice of features to be extracted is crucial as it directly influences the classification results. As the proposed HAR model is designed to adequately distinguish between the varying patterns of the ADLs in different contexts, hence a set of robust features is required. For this purpose, we conducted an empirical analysis of different sets of hand-crafted features used in the existing HAR studies [34], [40], [58]–[60] using a supervised correlation-based feature subset selection (CfsSubetSel) approach [61]. We initially extracted a set of twenty-five (25) time-domain features from the raw sensor data. These features include *maximum and minimum amplitudes*, *mean*, *variance*, *standard deviation*, *kurtosis*, *skewness*, *3rd and 4th moments*, *peak-to-peak amplitude and time*, *peak-to-peak slope*, *minimum and maximum latencies*, *latency-amplitude ratio*, *signal percentiles* (*i.e., 25th, 50th, and 75th*), *energy and normalized energy*, *mean and normalized mean of first and second difference of the signal*, and *signal entropy*. This set of features is then subjected to the CfsSubetSel method for feature reduction. Finally, a subset of overall twelve (12) unique features is obtained from three individual sensors (smartphone accelerometer, smartphone gyroscope, and watch accelerometer) and their combination, which are finally used for classification in the next stage. The details related to the selected features are presented below, along with their mathematical representation in Eq. (1) - (13).

Arithmetic Mean ($\bar{s}$):

$$\bar{s} = \frac{1}{N}\sum\nolimits_{m=1}^{N} s_d(m), \tag{1}$$

where $s_d(m)$ represents a 1D acceleration or rotation signal, the subscript d represents the direction of the signal (either along x, y, or z-axes), m symbolizes the m^{th} sample of the signal, and N is the total number of samples in the signal $s_d(m)$.

Minimum Signal Amplitude (s_{dmin}):

$$s_{dmin} = min(s_d(m)), \tag{2}$$

where $min(s_d(m))$ represents the minimum amplitude value contained in the signal $s_d(m)$.

Maximum Signal Amplitude (s_{dmax}):

$$s_{dmax} = max(s_d(m)), \tag{3}$$

where $max(s_d(m))$ denotes the maximum amplitude value contained in the signal $s_d(m)$.

Maximum Latency (n_m): It represents the index of maximum signal amplitude value.

$$m_{s_{dmax}} = \{m \mid s(m) = s_{dmax}\} \tag{4}$$

Minimum Latency ($m_{s_{dmin}}$): It represents the index of minimum signal amplitude value.

$$m_{s_{dmin}} = \{m \mid s(m) = s_{dmin}\} \tag{5}$$

Kurtosis (K): It is computed based on the 4^{th} moment about the signal mean and represents the tailedness of the signal probability distribution.

$$K = \frac{\mathfrak{m}_4}{\sigma^2} = \frac{1}{N}\sum_{m=1}^{N} \frac{(s_d(m) - \bar{s})^4}{\sigma^4}, \tag{6}$$

where m_4 represent the fourth moment of the signal $s_d(m)$, which is given in Eq. (6).

$$\mathfrak{m}_4 = E[(s_d(m) - \bar{s})^4], \tag{7}$$

where E denotes the expected value.

First Quartile (Q_1): It represents the 25^{th} percentile of a signal.

$$Q_1 = s_d\left(\left\lfloor \frac{N+1}{4} \right\rfloor\right) \tag{8}$$

Signal Median ($\mathfrak{M}$): It represents the 50^{th} percentile or median of a signal.

$$\mathfrak{M} = s_d\left(\left\lfloor \frac{N+1}{2} \right\rfloor\right) \tag{9}$$

Third Quartile (Q_3): It represents the 75^{th} percentile of a signal.

$$Q_3 = s_d\left(\left\lfloor \frac{3(N+1)}{4} \right\rfloor\right) \tag{10}$$

Normalized Mean of Signal Gradient ($\bar{\nabla}$): It represents the normalized mean of the first difference of a signal.

$$\bar{\nabla} = \frac{1}{N}\sum_{m=1}^{N}\left(\frac{|s_d(m) - s_d(m-1)|}{s_{dmax}}\right) \tag{11}$$

Normalized Mean of Signal Laplacian ($\bar{\Delta}$): It represents the normalized mean of the second difference of a signal.

$$\bar{\Delta} = \frac{1}{N}\sum_{m=1}^{N}\left(\frac{|s_d(m+1) - 2s_d(m) + s_d(m-1)|}{s_{dmax}}\right) \tag{12}$$

Entropy ($H(s_d(m))$): It measures the rate of change in a signal.

$$H(s_d(m)) = -\sum_{m=1}^{N} p_i(s_d(m)) \log_2 p_i(s_d(m)) ((s_d(m)), \tag{13}$$

where p_i $((s_d(m))$ is the probability of $(s_d(m))$.

These features are extracted using a fixed-length non-overlapping window of 20-second in time. The size of the final feature obtained for an activity instance is $[1 \times (12 \times 3)] = [1 \times 36]$ per sensor. In the case of the sensors fusion, the final feature vector is of size $[1 \times ((36\times \text{C}))]$, where C represents the number of sensors involved in the fusion. The primary reason for using these hand-crafted features is their successful HAR performance in the existing studies on account of low computational cost. Moreover, these features do not require a large training set, and it is easier to analyze and appreciate the individual contribution of these features (or set of features) for a specific problem. Unlike high-level or deep features that are extracted automatically from the data, hand-crafted features evade any features driven by noise and artifacts.

D. ACTIVITY RECOGNITION

The final step of the proposed methodology is activity recognition that entails two types of experiments, i.e., *context-independent* or *coarse* HAR and *context-dependent* or *fine-grained* HAR. A multi-label classification model is presented for this purpose, which simultaneously classifies the given chunk of data at two different levels and provides a notion of C2FHAR. Each activity instance is assigned two types of labels. The first one represents the primary ADLs only, which is used for *coarse* HAR at the first level of the proposed model. The other signifies the primary ADLs with the joint behavioral context and phone location information, which is employed for *fine-grained* HAR at the second level. The original activity labels (as in the *ExtraSensory* dataset) are used for *coarse* HAR, whereas, for *fine-grained* HAR, the newly assigned labels are used for each instance. The fundamental purpose of implementing two different types of HAR experiments is to investigate the effect of human behavioral context modeling on HAR performance, which is discussed in the results section.

The first-level classification entails the recognition of six single-label primary ADLs (e.g., *lying*, *sitting*, *standing*, *walking*, *bicycling,* and *running*) independent of any context information, thus providing a *coarse* representation of human activities. The second-level classifier identifies 29 context-aware activities (as listed in Table 1) and provides a *fine-grained* activity representation. Decision Tree, Random Forest, and Neural Networks are used to evaluate the performance of the proposed scheme. These classifiers are trained independently at two different levels of the proposed HAR model to identify the relevant primary ADLs or FGAs. In the end, the outputs from both stages are combined based on their classification scores to generate the final output. During the testing stage, each data chunk is passed through the first and second-level classifiers individually, and the corresponding class labels are inferred along with their posterior probabilities. The output activity identified with a higher posterior probability from both levels (i.e., *coarse* HAR and *fine-grained* HAR) is selected as the final output of the system. In this way, the proposed HAR model enables C2FHAR for real-time applications as well. Algorithm 1 provides the pseudocode of C2FHAR using the proposed multi-label classification model.

Algorithm 1 Steps for C2FHAR Using the Proposed Multi-Label Classification Model

Input(s): Training data: $\emptyset = \{X_i, L_i\}_{i=1}^{N}$
Test data: x_j (for inference stage only)
N : Total number of samples in the training set X
$L_i \in \left\{G^{(1)}, G^{(2)}\right\}$: Multiple activity labels for a training instance i
$G^{(1)}$: Set of six primary (context-independent) activity labels
$G^{(2)}$: Set twenty-nine fine-grained (context-dependent) activity labels
Output: Activity Label A_{label}(Primary or Fine-grained Activity)
Procedure:
% Training Stage
1: $M^{(1)} = trainClassifier\left(\emptyset \mid L_i \in G^{(1)}\right)$
% Classifier training for recognizing *coarse-level* activities
2: $M^{(2)} = trainClassifier\left(\emptyset \mid L_i \in G^{(2)}\right)$
% Classifier training for recognizing *fine-grained* activities
% Testing Stage
3: $\left[c^{(1)}, p^{(1)}\right] = classify\left(x_j, M^{(1)}\right)$ % Inferring *coarse-level* activity label $c^{(1)}$ and the resultant maximum posterior probability $p^{(1)}$
4: $\left[c^{(2)}, p^{(2)}\right] = classify\left(x_j, M^{(2)}\right)$ % Inferring *fine-grained* activity label $c^{(2)}$ and the resultant maximum posterior probability $p^{(2)}$
5: **if** $p^{(1)} > p^{(2)}$ **then**
$A_{label} = c^{(1)}$% Primary activity label as output
else
$A_{label} = c^{(2)}$% Fine-grained activity as output
end if

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section explains the method of analysis used for the proposed C2FHAR scheme along with the detailed experimental results and performance analysis. Moreover, it investigates the effect of behavioral context modeling on HAR *in-the-wild*. The details related to each subject are given in the following sections.

A. METHOD OF ANALYSIS AND CLASSIFIER TUNING

The proposed C2FHAR scheme is evaluated on the *ExtraSensory* dataset using Random Forest (RF), Decision Tree (DT), and Neural Networks (NN) classifiers. These classifiers are preferred on account of their efficient performance in the existing sensor-based HAR studies [44], [62]–[65]. Furthermore, in comparison to other prevalent machine learning classifiers (such as K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), and Naïve Bayes (NB)), the selected classifiers are well-suited for the required classification tasks that entail *in-the-wild* activity data. K-NN is a *lazy-learner* that is incapable of inferring the generalized discriminative model from the training data. It memorizes the whole training data to make predictions on the test data. Hence, the use of a K-NN classifier for real-time HAR applications is not feasible. SVM performs well in the cases when the margin between different class boundaries, i.e., hyperplanes, is sufficiently large, which is not possible for the proposed HAR scheme with small inter-class variations. Moreover, the SVM classifier does not directly provide the predicted probability of each class. Instead, it provides a classification score that is transformed to predicted probability using a computationally costly cross-validation technique. Likewise, NB works based on the Bayes theorem that requires the prior and conditional probabilities of each class to be estimated. Furthermore, NB strongly assumes that the input data entails a Gaussian distribution and all the input variables (i.e., features) are entirely independent. This assumption cannot be undertaken for *in-the-wild* activity data.

A k-fold cross-validation scheme (with $k = 5$) is adopted for evaluating the performance of the selected classifiers. For this purpose, the activity instances pertaining to each user data are placed together, and the entire dataset is split into five equal parts (starting from top to bottom). Four data splits are used for training the classifiers, whereas the remaining split is used for testing purposes. This process is repeated five times until all the activity instances are used for training and testing in different iterations, which ensures fairness in the output recognition results. In each round, the classifier testing entails the whole (or largest) part of the activity instances from unseen users, which offers a notion of *leave-some-subjects-out* validation. The hyperparameters of the selected classifiers are tuned corresponding to each fold of data to minimize the training error. Consequently, a random tree (bagging using a base learner) is chosen for RF classifier with the no. of iterations set as 100. For DT classifier, a C4.5 [66] pruned tree, i.e., J48, is employed with the number of folds equal to 3. In the case of NN classifier, one fully-connected hidden layer is used with the number of neurons equal to the mean of input and output layer size. The learning rate is set equal to 0.03 with the no. of learning iterations as 109. These hyperparameter configurations provided the least training error for the proposed scheme using five-fold cross-validation.

1) PERFORMANCE EVALUATION PARAMETERS

The performance of RF, DT, and NN classifiers is validated based on commonly used performance metrics, including *accuracy* (AC), *precision* (PR), *sensitivity/recall* (SE), and *F1-score* (F1). However, using these parameters for imbalanced class distribution problem often produces biased recognition results. The *ExtraSensory* dataset used in this study for experimentation purposes is collected *in-the-wild;* hence, the number of instances is not uniform for the selected ADLs and FGAs. This inconsistency in the number of activity instances may lead to biased results in terms of AC, PR, SE, and F1. As a result, similar to our baseline study [14], the balanced accuracy (BAC) measure is used as the key performance indicator in this study. BAC represents the area under the curve (AUC) and is insensitive to the rare-labels that may cause unfairness in the output recognition results in the case of an imbalanced class dataset. In addition to BAC, Mathews Correlation Coefficient (MCC) is used to evaluate the classifier performance, which is primarily considered as a balanced measure even if the classes are of different sizes [67], [68]. MCC also signifies how well a classification model fits the data by returning a coefficient value in the range -1 to $+1$. In the case of perfection and worst prediction, a value of $+1$ and -1 is obtained, respectively, whereas, a value 0 indicates a random prediction. Eq. (14) and Eq. (15) provides the mathematical description of both these metrics, i.e., BAC and MCC, respectively. A *macro-averaging* technique is used to find out the average recognition results.

$$BAC = \frac{T_P \times (T_N + F_P) + T_n \times (T_p + F_n)}{2 \times (T_p + F_N)(T_n + F_P)} \tag{14}$$

$$MCC = \frac{(T_P \times T_N) - (F_P \times F_N)}{\sqrt{(T_P + F_P)\,(T_P + F_N)\,(T_N + F_P)\,(T_N + F_N)}} \tag{15}$$

T_P, T_N, F_P, and F_N denote true positives, true negatives, false positives, and false negatives, respectively.

B. PERFORMANCE ANALYSES OF C2FHAR

As discussed earlier, the proposed HAR model entails multi-label classification at two different levels: 1) *context-independent* ADLs recognition at the coarse level, and 2) *context-dependent* FGAs recognition at the fine level. The detailed experimental results and performance analysis for both types of HAR experiments are provided in the subsequent sections.

1) CLASSIFIER-BASED ANALYSIS OF C2FHAR

In the case of *context-independent* ADLs recognition, the selected classifiers are trained to recognize six primary activities without incorporating any context information. hence, all data samples related to a single activity are combined together independent of any behavioral context and assigned the same primary activity label. However, for *context-dependent fine-grained* har experiments, the data

TABLE 2. Average results obtained for coarse and fine-grained HAR using RF, DT, and NN classifiers.

Experiment	Classifier	Sensor(s)	AC	PR	SE	F1	BAC	MCC
Coarse-level Primary Activity Recognition (06 Context-Independent ADLs)	Random Forest (RF)	S-ACC	0.948	0.879	0.747	0.808	0.807	0.703
		S-GYRO	0.930	0.814	0.690	0.747	0.770	0.642
		W-ACC	0.944	0.845	0.701	0.766	0.773	0.687
		S-ACC + S-GYRO	0.952	0.857	0.754	0.802	0.821	0.752
		S-ACC + W-ACC	0.967	0.871	0.771	0.818	0.827	0.758
		S-GYRO + W-ACC	0.937	0.862	0.763	0.809	0.818	0.731
		S-ACC + S-GYRO + W-ACC	**0.968**	**0.909**	**0.782**	**0.841**	**0.830**	**0.743**
	Decision Tree (DT)	S-ACC	0.930	0.718	0.707	0.712	0.779	0.662
		S-GYRO	0.896	0.630	0.618	0.623	0.739	0.592
		W-ACC	0.914	0.669	0.657	0.662	0.756	0.611
		S-ACC + S-GYRO	0.944	0.771	0.761	0.766	0.801	0.668
		S-ACC + W-ACC	0.945	0.777	0.767	0.773	0.806	0.690
		S-GYRO + W-ACC	0.923	0.762	0.752	0.755	0.796	0.712
		S-ACC + S-GYRO + W-ACC	0.948	0.779	0.767	0.777	0.807	0.700
	Neural Networks (NN)	S-ACC	0.901	0.718	0.644	0.666	0.750	0.611
		S-GYRO	0.857	0.611	0.494	0.523	0.682	0.546
		W-ACC	0.913	0.698	0.615	0.641	0.735	0.577
		S-ACC + S-GYRO	0.917	0.765	0.657	0.695	0.755	0.620
		S-ACC + W-ACC	0.932	0.788	0.716	0.745	0.780	0.652
		S-GYRO + W-ACC	0.909	0.754	0.647	0.686	0.745	0.642
		S-ACC + S-GYRO + W-ACC	0.932	0.798	0.722	0.753	0.785	0.684
Fine-grained Activity Recognition (29 Context-Dependent FGAs)	Random Forest (RF)	S-ACC	0.989	0.875	0.649	0.714	0.809	0.724
		S-GYRO	0.982	0.740	0.482	0.538	0.726	0.646
		W-ACC	0.986	0.850	0.570	0.637	0.770	0.694
		S-ACC + S-GYRO	0.989	0.881	0.722	0.778	0.841	0.742
		S-ACC + W-ACC	0.991	0.794	0.733	0.760	0.850	0.758
		S-GYRO + W-ACC	0.989	0.881	0.639	0.704	0.804	0.742
		S-ACC + S-GYRO + W-ACC	**0.992**	**0.797**	**0.740**	**0.761**	**0.867**	**0.764**
	Decision Tree (DT)	S-ACC	0.983	0.628	0.612	0.620	0.792	0.680
		S-GYRO	0.975	0.457	0.433	0.444	0.704	0.640
		W-ACC	0.979	0.550	0.520	0.532	0.747	0.644
		S-ACC + S-GYRO	0.987	0.689	0.670	0.678	0.820	0.701
		S-ACC + W-ACC	0.988	0.710	0.691	0.699	0.830	0.722
		S-GYRO + W-ACC	0.984	0.624	0.608	0.615	0.790	0.737
		S-ACC + S-GYRO + W-ACC	0.988	0.721	0.704	0.712	0.837	0.740
	Neural Network (NN)	S-ACC	0.976	0.560	0.419	0.444	0.696	0.638
		S-GYRO	0967	0.372	0.206	0.229	0.591	0.583
		W-ACC	0.976	0.584	0.389	0.420	0.681	0.611
		S-ACC + S-GYRO	0.979	0.619	0.474	0.501	0.723	0.638
		S-ACC + W-ACC	0.982	0.649	0.527	0.555	0.749	0.660
		S-GYRO + W-ACC	0.981	0.657	0.497	0.532	0.734	0.678
		S-ACC + S-GYRO + W-ACC	0.984	0.726	0.580	0.622	0.775	0.712

* S-ACC, S-GYRO, and W-ACC represent the smartphone accelerometer, smartphone gyroscope, and watch accelerometer, respectively. The **highlighted** values indicate the best recognition performances for both type of HAR experiments, which are obtained using RF classifier with the fusion of all three sensor (S-ACC, S-GYRO, and W-ACC).

instances relating to six adls are split into more refined representations based on behavioral context information. All activity instances with the same primary activity and behavioral context labels are merged together and assigned a similar *fine-grained* activity label. The missing data samples from any user or sensor are simply ignored. After that, RF, DT, AND NN classifiers are trained individually corresponding to each *fine-grained* activity to recognize 29 FGAs. Three inertial sensors, including smartphone accelerometer (S-ACC), smartphone gyroscope (S-GYRO), and watch accelerometer (W-ACC), are used for HAR.

Table 2 provides the average results obtained for *coarse* and *fine-grained* HAR using RF, DT, and NN classifiers with different combinations of sensors. It can be observed from the table that the S-ACC sensor achieves an average BAC value of 80.7% for the *coarse-level* ADLs recognition using RF classifier, which is 2.8% and 5.7% more than the average BAC values of 77.9% and 75.0% attained with DT and NN classifiers, respectively. Likewise, the individual ADLs recognition performance of the S-GYRO and W-ACC sensor is also better using the RF classifier. The best performance (BAC = 83.0% and MCC = 74.3%) for ADLs recognition is achieved with RF classifier using the fusion of all three sensors (i.e., S-ACC, S-GYRO, and W-ACC). Generally, for any other combination of sensors, RF performs better than DT and NN classifiers in recognizing the selected ADLs. The worst performance is attained for ADLs recognition based on NN classifier.

In the case of *context-dependent* FGAs recognition, the best performance (i.e., BAC = 86.7% and MCC = 76.4%) is also achieved using RF classifier with the feature-level fusion of S-ACC, S-GYRO, and W-ACC sensors, as indicated in Table 2. For the same set of sensors, a BAC rate of 83.7% and 77.5% is achieved with DT and NN classifier, which is 1.7% and 7.9% less than the BAC rate of RF classifier, respectively. In addition, an MCC value of 0.740 and 0.712 is obtained for DT and NN classifier, respectively, with the fusion of all sensors. Hence, it can be specified that NN underperforms as compared to RF and DT classifiers for *fine-grained* HAR as well. The individual FGAs recognition performance of each sensor is better for RF classifier as compared to both DT and NN classifiers. In the same way, all other combinations of sensors achieve better recognition performance for *fine-grained* HAR using RF classifier, which can be observed from Table 2. Therefore, based on these analyses, it can be said that the performance of RF classifier is robust than DT and NN classifier, which concludes the efficacy of choosing RF over DT and NN classifiers for *in-the-wild* recognition of ADLs and FGAs.

Generally, NN performs well with large datasets, where an ample amount of data is available for the system training. In the case of imbalanced datasets (such as *ExtraSensory*), where the number of training examples for all the classes is not equal and enough, the NN classifier provides poor performance. As the proposed scheme is based on *in-the-wild* data collected from the participants in diverse contexts; hence, there exist apparent randomness and intra-class variations in the activity patterns, which badly affect NN training. Besides, the self-collected and self-labeled data from the participants may entail labeling noise as well. As a result, the performance of NN is degraded for the proposed C2FHAR scheme. However, the RF classifier is capable of dealing with the randomness in data, thus provides better results for the proposed scheme. RF builds a collection of random decision trees (with random data samples and subsets of features), each of which individually predicts the class label, and the final decision is made on the basis of majority voting principle. In contrast to RF, the DT classifier builds a single tree on the entire dataset using the entire set of features and data samples, which may lead to overfitting. Hence, RF limits overfitting and error due to bias by aggregating the result of multiple random trees (build on different subsets of features) and thus produces more useful results. Based on these analyses, the RF classifier is determined as the best choice for *in-the-wild* HAR.

2) SENSOR-BASED ANALYSIS OF C2FHAR

This section analyzes the individual performance of different sensors for ADLs and FGAs recognition based on the results presented in Table 2. It can be investigated from these results that the S-ACC sensor provides better individual performance for *in-the-wild* HAR as compared to S-GYRO or W-ACC sensor. For RF classifier, the BAC value achieved pertaining to ADLs recognition using S-ACC is 3.7% and 3.6% better than that achieved for S-GYRO and W-ACC, respectively. Similarly, in the case of DT, the S-ACC senor achieves 4.0% and 2.3% better BAC rate as compared to the BAC values obtained with the S-GYRO and W-ACC sensors, respectively. Likewise, the average values of MCC are also better for the S-ACC sensors. When comparing the individual recognition results of the S-GYRO and W-ACC sensors, it is observed that the performance of the W-ACC sensor is better than the S-GYRO sensor. The fusion of both smartphone sensors (i.e., S-ACC and S-GYRO) achieves a BAC rate of 82.1% for ADLs recognition using RF classifier. On the other hand, combining both accelerometers (i.e., S-ACC and W-ACC) provides an average BAC value of 82.7%. The fusion of all three sensors attains the best results for ADLs recognition (using RF classifier) with an average BAC and MCC values of 0.830 and 0.743, respectively. In the case of *fine-grained* HAR, the BAC rate achieved for the S-ACC sensor is 8.3% and 3.9% more than that achieved with the S-GYRO and W-ACC sensor, respectively, using RF classifier. Also, the performance of the W-ACC sensor is 4.4% better than the S-GYRO performance when classified using RF. The fusion of both smartphone sensors (i.e., S-ACC and S-ACC) improves the recognition performance (i.e., BAC rate) of the system to 84.1%, 82.0%, and 72.3% using RF, DT, and NN classifiers, respectively. Likewise, combining both accelerometers (i.e., S-ACC and W-ACC) enhances the system performance and provides a BAC rate of 85.0%, 83.0%, and 74.9% when classified using RF, DT, and NN, respectively. The addition of the S-GYRO sensor with both accelerometers provides the best BAC and MCC values of 0.854 and 0.788, respectively, for *fine-grained* HAR using RF classifier.

Although adding the S-GYRO sensor with both accelerometers (i.e., S-ACC and W-ACC) only provides a minor performance improvement for both *coarse* and *fine-grained* HAR experiments, however, it is necessary from the practical perspective to incorporate this sensor into the proposed framework. It is owing to the reason that the S-GYRO sensor is capable of efficiently tracking the 3D rotational motion or body movements, which is very critical for *in-the-wild* HAR, where the activities can be performed in any random pattern. The accelerometer sensors do not generally provide good recognition results for rotational motion, which excites the use of S-GYRO sensor for real-time HAR. Hence, based on the obtained results and discussion, it is concluded that the fusion of S-ACC, S-GYRO, and W-ACC sensors is the best choice for *in-the-wild* recognition of the selected ADLs and FGAs.

Fig. 3 provides a comparison of the individual BAC values obtained for *context-independent* primary ADLs using the RF classifier with different sensors. The individual BAC values achieved for *bicycling*, *running*, and *walking* activities using the S-ACC sensor are 87.4%, 81.4%, and 77.8%, respectively. These values are 4.8%, 5.9%, and 4.6% better than those obtained for the same activities, respectively, using S-GYRO. The use of W-ACC provides the BAC values of 79.9%, 74.4%, and 70.6% for *bicycling*, *running*, and

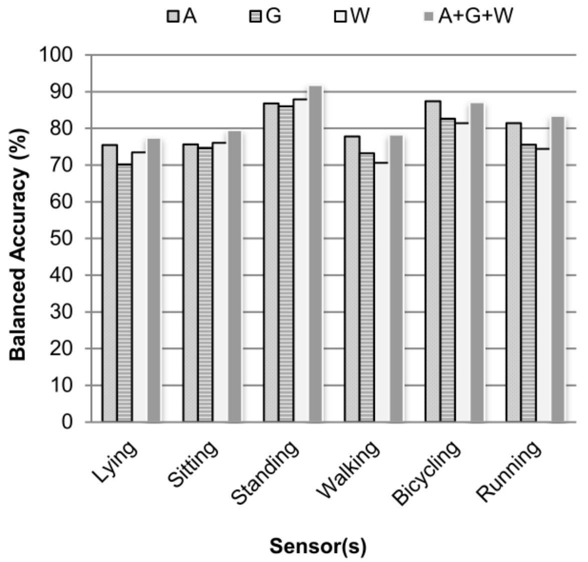

FIGURE 3. Individual recognition accuracies obtained for *coarse-level context-independent* ADLs using RF classifier. Here A, G, and W signify the smartphone accelerometer, smartphone gyroscope, and watch accelerometer, respectively.

walking activities respectively, which are lower than the corresponding BAC values obtained with the S-ACC sensor. For *standing*, *sitting*, and *lying* activities, the best BAC rates of 87.1%, 76.1%, and 75.8%, respectively, are achieved using the W-ACC. Hence, it is evident that the individual performance of S-ACC is better in recognizing *context-independent* dynamic ADLs, whereas the W-ACC sensor performs better in the case of static ADLs. Furthermore, adding W-ACC with the S-ACC and S-GYRO sensor improves the individual recognition accuracies of all primary ADLs.

Fig. 4 compares the individual recognition performance of the S-ACC, S-GYRO, and W-ACC sensor in recognizing the selected FGAs using RF classifier. This comparison is made in order to get an understanding of the role of these individual sensors in recognizing certain *context-dependent* FGAs. It can be observed from Fig. 4 that the S-ACC sensor performs better than the S-GYRO sensor in recognizing all individual FGAs. Likewise, the performance of the W-ACC senor is also better than the S-GYRO sensor in recognizing most of the individual FGAs. Hence, it is evident that the S-GYRO sensor does not provide efficient results for *fine-grained* HAR. By comparing the performance of both accelerometers (i.e., S-ACC and W-ACC), it can be observed that the S-ACC sensor performs better than the W-ACC sensor in recognizing the FGAs. In particular, for activities *a22* to *a29*(which relate to *walking*, *bicycling*, and *running* activities with different behavioral contexts), the S-ACC sensor outperforms the W-ACC sensor. In contrast, for most of the static activities, the performance of the W-ACC is better than the S-ACC. In particular, for activities *a1*, *a3*, *a5*, *a13*, *a17*, and *a21,* where the phone position is *on table,* the W-ACC sensor performs way better than the S-ACC. Also, in the case of *a2*(*lying when surfing the internet with phone in hand*),*a8*(*sitting in a car with phone in pocket*), *a12*(*sitting when watching TV with phone in pocket),* and *a19*(*standing outdoor with phone in pocket)* activities, the W-ACC sensor provides better accuracy rate than either of the smartphone sensors. It is because when these activities are performed in the associated contexts, they do not exhibit such body motion that can be tracked efficiently with the smartphone sensors in the pocket. However, these activities can be identified easily with the help of a wrist-watch based upon some definite hand, wrist, or arm movement/position. For instance, the activity *a8*(*sitting in a car with phone in pocket*) can be efficiently detected based upon the wrist and arm movements of a user when *driving a car* or his/her arm position/movement when *sitting in a car*. In the same way, the activity *a12*(*sitting when watching TV with phone on table)* can be easily identified with the hand/arm movements of a user when using the remote control or doing something else. Hence, the use of smartwatch in combination with a smartphone is essential, which increases the *in-the-wild* recognition accuracies of FGAs.

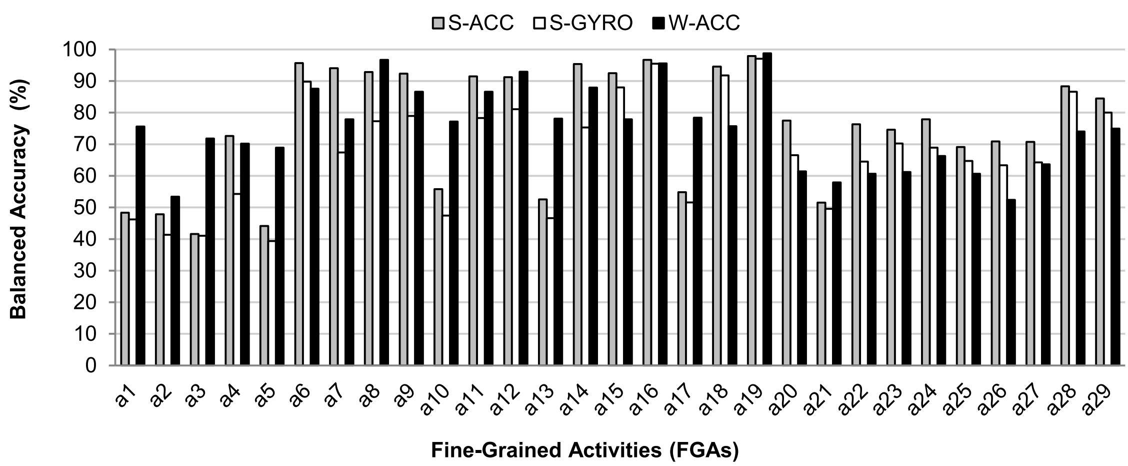

FIGURE 4. Comparison of individual sensor performance in recognizing the selected FGAs using Random Forest classifier. Here, S-ACC, S-GYRO, and W-ACC represent the smartphone accelerometer, smartphone gyroscope, and watch accelerometer, respectively.

a1	18675	12	6	2	129	1	0	1	0	0	0	0	41	0	1	3	7	0	0	17	0	10	8	30	3	3	0	49	3
a2	341	478	1	0	0	0	0	0	0	8	0	0	2	0	0	0	0	0	0	5	1	1	52	8	0	4	0	30	4
a3	36	30	200	1	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	2	0	1	0	4	0
a4	109	3	11	324	6	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	5	0	0	0	1	0
a5	1168	4	0	4	5789	1	0	0	0	0	0	0	3	0	0	5	6	0	0	1	0	0	3	33	0	3	0	14	3
a6	49	0	0	0	8	930	3	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a7	17	0	0	0	0	4	169	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a8	18	0	0	0	0	2	1	222	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a9	26	0	0	0	0	0	0	0	173	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a10	18	0	0	0	2	0	0	0	28	821	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a11	37	0	0	0	4	0	0	0	0	9	501	7	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
a12	0	0	0	0	0	0	0	0	0	0	11	157	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a13	201	0	0	1	6	0	0	0	0	0	0	28	4657	7	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
a14	26	0	0	0	0	0	0	0	0	0	0	0	7	393	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
a15	19	0	0	0	2	0	0	0	0	0	0	0	0	1	370	5	0	0	0	0	0	0	0	0	0	0	0	0	0
a16	15	0	0	0	0	0	0	0	0	0	0	0	0	0	7	1968	22	0	0	0	0	0	0	0	0	0	0	1	0
a17	150	0	0	0	6	0	0	0	0	0	0	0	0	0	0	17	3459	9	0	0	0	0	0	0	0	0	0	0	0
a18	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	10	170	7	0	0	0	0	0	0	0	0	0	0
a19	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	430	0	0	0	0	0	0	0	0	0	0
a20	7	3	2	0	6	0	1	0	0	2	0	0	0	1	0	0	0	0	0	139	1	12	0	23	12	4	16	32	40
a21	21	3	14	0	5	0	3	0	0	12	0	4	6	0	0	0	7	0	0	10	74	2	0	12	1	7	3	31	18
a22	12	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	80	13	14	3	11	36	43	21
a23	12	13	0	0	1	0	0	0	1	0	0	0	0	0	0	2	0	0	0	11	3	11	320	9	21	13	37	51	117
a24	23	9	0	0	15	0	0	0	0	0	0	0	1	0	0	0	0	0	0	37	4	12	3	497	21	21	28	113	71
a25	11	0	3	0	6	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0	11	4	12	45	4	4	19	22
a26	9	1	0	0	10	0	0	0	0	0	0	0	0	0	0	0	3	0	0	17	0	0	3	21	0	115	13	24	34
a27	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	0	3	7	4	7	0	48	31	27
a28	194	22	14	11	4	0	0	2	0	3	0	0	0	0	0	0	0	0	0	7	8	13	14	25	15	11	36	2918	191
a29	24	5	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	12	21	16	20	27	18	14	56	482
	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20	a21	a22	a23	a24	a25	a26	a27	a28	a29

FIGURE 5. Confusion matrix for the best-case *fine-grained* HAR results achieved with Random Forest classifier using the combination of S-ACC, S-GYRO, and W-ACC sensors. The rows represent the ground truths, and the columns show the predicted outputs.

Fig. 5 shows the confusion matrix for the best-case *context-dependent fine-grained* HAR results obtained with the fusion of S-ACC, S-GYRO, and W-ACC sensors using RF classifier. It can be observed from the figure that almost half of the FGAs (including *a6* to *a19*) are recognized with a true positive rate of more than 90%. These FGAs are related to the static activities of *sitting* and *standing* with the corresponding behavioral contexts. Their effective recognition results indicate that it is easier to explicitly recognize the *sitting* and *standing* activities in combination with their behavioral contexts. Generally, the *sitting* and *standing* patterns of a person vary with change in the behavioral context, e.g., *sitting* posture is usually different when *working on a computer* or *watching TV*. Similarly, the stance of *standing indoor* may quite differ from *standing outdoor.* Phone position and other parallel activities being carried out also affect the *sitting* and *standing* patterns. These differences are recorded by the S-ACC and W-ACC sensors to effectively learn and identify the FGAs related to *sitting* and *standing*(i.e., *a4* to *a19*), which can be observed from the confusion matrix reported in Fig. 5. The activities *a1* (*lying when sleeping with phone on table*) and *a2* (*lying when surfing the internet with phone on table*) are truly recognized with lower accuracy as compared to other activities (such as *a3* and *a4-a19*) due to the hostile phone position, i.e., *on table*, and the lack of any viable wrist/arm movement pattern. Hence, the false positives and false negatives are also higher for *a1* and *a2,* as indicated in Fig. 5. The activities *a4 (sitting when surfing the internet with phone in hand*) and *a5 (sitting when surfing the internet with phone on table*) are also misclassified as *a1* with a high proportion of 23.5% and 16.6%, respectively. Nevertheless, the overall recognition performance obtained for the selected FGAs is sufficient and practicable.

3) COARSE VS. FINE HAR IN-THE-WILD

Human activity patterns are not always consistent and vary with the change in behavioral context. For creating a better model for *in-the-wild* HAR, it is critical to take into account the contextual parameters related to human activities. However, it is apparent that the classification between different primary ADLs is easier in comparison to the FGAs, as the inter-class variations are usually significant even in diverse contexts. On the other hand, distinguishing between the varying patterns of the same activity in different contexts is difficult owing to the insignificant intra-class variations, which makes the recognition of FGAs more laborious and difficult. Moreover, different people have peculiar ways of conducting themselves in different contexts, which makes it harder to model *context-dependent* FGAs *in-the-wild*. To analyze the effect of behavioral context modeling on HAR performance, we compare the performance of *coarse-level* ADLs recognition with *fine-grained* HAR. The maximum average recognition rate (i.e., BAC value) achieved for ADLs and FGAs recognition is 83.0% and 86.7%, respectively, using RF classifier. As follows, the proposed scheme identifies 29 FGAs (i.e., six primary ADLs linked with diverse behavioral contexts) on account of a 3.7% increase in the BAC rate as compared to the *coarse-level* recognition of six ADLs.

As discussed earlier, recognizing the FGAs *in-the-wild* is a much harder task than identifying the simple ADLs due to the small inter-class variations. However, *fine-grained* HAR still provides a better BAC rate than simple ADLs recognition, which demonstrates the efficacy of the proposed C2FHAR approach.

Fig. 6 compares the best-case BAC values obtained corresponding to each primary activity at the *coarse* and *fine* classification level. The balanced accuracies achieved for the individual FGAs related to each primary activity are averaged to compute the *fine-level* recognition results for six ADLs. It can be analyzed from Fig. 6 that the overall recognition accuracies of the static activities (such as *lying*, *sitting*, and *standing*) are increased with the explicit modeling of the behavioral context information, which indicates that it is easier to recognize these activities in consort with their behavioral contexts. These static activities usually do not reveal any definite body motion (except upper limb movements) due to which it becomes challenging to distinguish between them *in-the-wild*, without having any prior knowledge of the user's context. As an example, the classification between the *sitting* and *lying* activities is difficult if the user's phone position or secondary activity is unknown. However, the explicit modeling of the behavioral context information (such as *user's location*, *secondary activity*, and *phone position*) with static activities helps in their identification and provides better recognition results for these activities, as indicated by the results reported in Fig. 3-5. The overall recognition accuracy of the *walking* activity is decreased as a result of behavioral context modeling. It is due to the chaotic nature of human behavior that results in the varying motion patterns even in the same settings, such as the emotional state (such as *sad* or *happy*) of a person may unconsciously change his/her gait pattern. As a result, the proficient modeling of *walking* activity in diverse behavioral contexts appears impractical, which reduces its recognition accuracy. The individual recognition accuracies of *bicycling* and *running* activities are nearly the same for both types of HAR experiments because these activities are only modeled in a single context. Therefore, the overall experimental results achieved for the proposed scheme demonstrate the effectiveness of behavioral context modeling for *in-the-wild* HAR, which is critical for human behavior modeling and understanding as well.

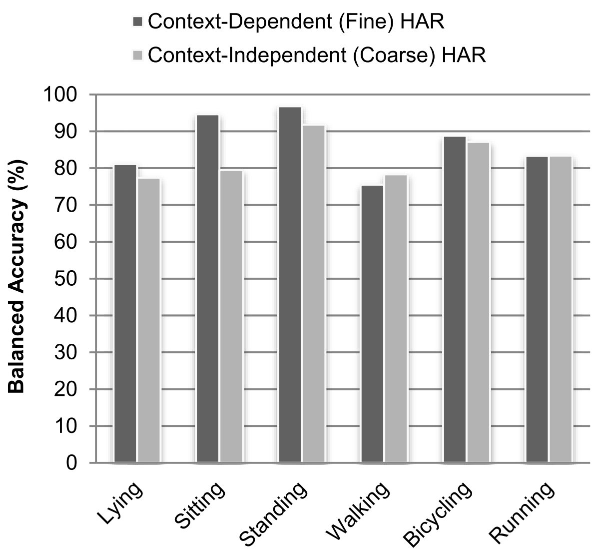

FIGURE 6. **Comparison of average HAR results obtained for six ADLs using *context-dependent* and *context-independent* recognition method. The overall performance of *fine-grained* HAR is better than *context-independent* ADLs recognition, which demonstrates the effectiveness of modeling behavioral context information for HAR *in-the-wild*.**

C. COMPARISON WITH STATE-OF-THE-ARTS

This section provides a discussion on comparative analysis of the proposed C2FHAR scheme with the existing-state-of-the-arts. In this aspect, Table 3 compares the main characteristics of a few well-known sensor-based HAR approaches with the proposed scheme. The comparison is comprehended based on the type and number of activities recognized, subjects used for experimentation, sensing modalities, machine learning classifiers used for empirical analysis, and the achieved results. It can be observed from Table 3 that the existing studies mostly target *simple/ repetitive* or *complex/non-repetitive* ADLs for recognition purposes, which are single-labeled and do not entail any user's context information. These HAR schemes are trained under controlled settings, where the occupancy of data collection experiments is generally a single location with a static context. As a result, these schemes cannot adapt to natural user behavior and perform poorly in real-world scenarios. In addition, most of the existing studies are dependent on the fusion of multiple sensing modalities (such as smartphone sensors, wearable sensors, and ambient sensors) to achieve the successful recognition results, which make them computationally expensive and infeasible. However, our proposed scheme is capable of efficiently recognizing 06 primary ADLs as well as 29 complex and multi-label FGAs *in-the-wild* based on the smartphone and smartwatch accelerometer sensors.

As the proposed C2FHAR scheme is implemented and validated based on different sets of data and annotations, hence a fair comparison with the existing state-of-the-art is not possible. However, the efficacy of the proposed scheme can be demonstrated by comparing the output recognition results with the existing state-of-the-art. Table 3 shows that the accuracy rates achieved for the proposed C2FHAR scheme are comparable or better than the reported results of the existing sensor-based HAR studies. Additionally, the proposed scheme offers apparent advantages for effective decision making over the existing schemes, which is very crucial for real-time applications.

D. IMPLICATIONS OF C2FHAR IN REAL-TIME SETUPS

The main goal of the proposed C2FHAR scheme is the provision of real-time HAR based on the available context metadata. The accessibility and reliability of human context information in real-time scenarios cannot be guaranteed. It is

TABLE 3. Comparison of the proposed C2FHAR scheme with some well-known existing HAR studies.

Study	Activity Type	# of Activities & # of Subjects	Sensing Modality	Classifiers	Results
[65]	Simple ADLs (Single-label)	06 & 02	Wearable (Acc.)	RF	Accuracy = 90%
[14]	Behavioral Contexts (Single-label)	25 & 60	Smartphone (Acc., Gyro., Mag., GPS); Wearable (Acc.)	LR	Accuracy = 87% BAC = 80%
[34]	Complex at-home ADLs (Single-label)	21 &10	Wearable (Acc., Gyro.); Ambient (B, T, H); GPS; Bluetooth	NB, RF	Accuracy = 92%
[37]	Simple ADLs with Transitional Activities (Single-label)	12 & 30	Smartphone (Acc., Gyro.)	SVM, DBN, ANN	Accuracy = 89.61% (DBN)
[69]	Elderly People Activities (Single-label)	17 & 21	Wearable (T, Acc., Gyro., Mag., B); Ambient (PIR)	SVM	Accuracy = 98.32%
[22]	Simple ADLs (Single-label)	12 & 06	Wearable	KNN, RF, SVM, HMM, GMM	Accuracy = 99.25% (KNN)
[70]	Firefighting Activities (Single-label)	17 & 11	Wearable (Acc., Gyro., Mag.) Ambient (B, H, T)	GBT, KNN, SVM	Accuracy = 97.68% (GBT)
Proposed	**Single-label ADLs (*in-the-wild*)**	**06 & 60**	**Smartphone (Acc., Gyro.) Wearable (Acc.)**	**RF, DT, NN**	**Accuracy = 96.80% BAC = 83% (RF)**
	Multi-label Fine-grained ADLs (*in-the-wild*)	**29 & 60**	**Smartphone (Acc., Gyro.) Wearable (Acc.)**	**RF, DT, NN**	**Accuracy = 99.20% BAC = 86.7% (RF)**

* Acc.: Accelerometer; ANN: Artificial Neural Network; B: Barometer; BAC: Balanced Accuracy; DBN: Deep Belief Network; GBT: Gradient Boosting Tree; GMM: Gaussian Mixture Model; Gyro.: Gyroscope; H: Humidity Sensor; HMM: Hidden Markov Model; KNN: K-Nearest Neighbor; Mag.: Magnetometer; NB: Naïve Bayes; T: Temperature Sensor; PIR: Passive Infrared Sensor; SVM: Support Vector Machine;

quite possible that no context information is available to the system at all, which may lead to the complete failure of a *context-dependent* HAR scheme. However, the proposed scheme addresses this challenge using a multi-label classification model that offers HAR at two different levels, i.e., *coarse* and *fine*, depending upon the user's context information. For ideal use-cases in real-time, where the required context information is available to the system, the proposed scheme can effectively provide the *fine-grained* representation of human activities based on behavioral context modeling. However, in the absence of any context information, the proposed HAR model can still efficiently recognize the *context-independent* primary ADLs at the coarse level. The final output of the system can be chosen by comparing the activity classification scores obtained at both levels and selecting the output activity predicted with a higher probability. For instance, with the availability of essential context information, the proposed system can provide higher prediction scores for FGAs as compared to ADLs, which is indicated by the obtained results. However, in the absence of any context information, there is a great likelihood that the proposed system will recognize the *context-independent* ADLs with higher confidence. In this way, the proposed scheme assists real-time HAR and provides either a *coarse* or *fine-grained* representation of *in-the-wild* human activities based on the presence/absence of the required context information.

V. CONCLUSION

In this paper, a novel multi-label classification scheme is presented for sensor-based human activity recognition, which models human contexts (such as *user's location*, *secondary activity*, and *phone position*) with the daily living activities to obtain a *coarse-to-fine* representation of human activities *in-the-wild*. For this purpose, a set of 29 *fine-grained* human activities is obtained by adding context information with six primary activities (including *lying*, *sitting*, *standing*, *running*, *bicycling*, and *running*). Decision Tree (DT), Random Forest (RF), and Neural Networks (NN) are utilized to recognize six *context-independent* activities as well as 29 *context-dependent* activities based on the inertial sensor data extracted from smartphone and smartwatch. A public domain *ExtraSensory* dataset is used for validating the proposed scheme, which shows that RF performs better than DT and NN classifiers in recognizing both types of activities. The performance of *context-dependent* activity recognition is better, which state the efficacy of behavioral context modeling with the primary activities of daily living. The best recognition results are achieved with the fusion of all inertial sensors.

Although, the proposed solution offers satisfactory performance for *fine-grained* activity recognition using the smartphone and smartwatch inertial sensors, however exploring new substitutes of the proposed method can be advantageous in improving the accuracy. The proposed idea of multi-label classification can be extended to incorporate multi-level hierarchical activity classification based on the available context information. In this aspect, a large number of activities and the behavioral contexts can be modeled at different levels to enable real-time HAR. More sensors, such as GPS and microphone, can be added with the inertial sensors for better recognition of human activities *in-the-wild*. The effect of different behavioral contexts on human activity patterns can be investigated to find out the factor contributing to abnormal human behavior. The relationship between human physical and physiological activities can be investigated, which may lead to efficient human behavior cognition. Furthermore, monitoring and understanding human activities and their behavior in different contexts can help in improving the productivity of a human being.

REFERENCES

[1] A. Stisen, "Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition with realworld HAR dataset," in *Proc. 13th ACM Conf. Embedded Netw. Sensor Syst. (SenSys)*, vol. 15, 2015, pp. 127–140.

[2] Z. S. Abdallah, M. M. Gaber, B. Srinivasan, and S. Krishnaswamy, "Activity recognition with evolving data streams," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 1–36, Jul. 2018.

[3] J. Rafferty, C. D. Nugent, J. Liu, and L. Chen, "From activity recognition to intention recognition for assisted living within smart homes," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 3, pp. 368–379, Jun. 2017.

[4] W. Chi, J. Wang, and M. Q.-H. Meng, "A gait recognition method for human following in service robots," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 9, pp. 1429–1440, Sep. 2018.

[5] E. I. Vlahogianni and E. N. Barmpounakis, "Driving analytics using smartphones: Algorithms, comparisons and challenges," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 196–206, Jun. 2017.

[6] Y. Guo, B. Guo, Y. Liu, Z. Wang, Y. Ouyang, and Z. Yu, "CrowdSafe: Detecting extreme driving behaviors based on mobile crowdsensing," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Aug. 2017, pp. 1–8.

[7] R. Mojarad, F. Attal, A. Chibani, S. R. Fiorini, and Y. Amirat, "Hybrid approach for human activity recognition by ubiquitous robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 5660–5665.

[8] I. E. Olatunji, "Human activity recognition for mobile robot," *J. Phys., Conf. Ser.*, vol. 1069, Aug. 2018, Art. no. 012148.

[9] D. M. Hilty and S. Chan, "Human behavior with mobile health: Smartphone/devices, apps and cognition," *Psychol. Cogn. Sci. Open J.*, vol. 4, no. 2, pp. 36–47, Feb. 2019.

[10] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, "GCHAR: An efficient group-based context—Aware human activity recognition on smartphone," *J. Parallel Distrib. Comput.*, vol. 118, pp. 67–80, Aug. 2018.

[11] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Inf. Fusion*, vol. 46, pp. 147–170, Mar. 2019.

[12] S. R. Ramamurthy and N. Roy, "Recent trends in machine learning for human activity recognition—A survey," *WIREs Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1254, Jul. 2018.

[13] S. K. Polu, "Human activity recognition on smartphones using machine learning algorithms," *Int. J. Innov. Res. Sci. Technol.*, vol. 5, no. 6, pp. 31–37, 2018.

[14] L. G. V. Yonatan and E. Katherine, "Recognizing detailed human context in-the-wildfrom smart phones and smartwatches," *IEEE Pervasive Comput.*, vol. 16, no. 4, pp. 62–74, Oct./Dec. 2017.

[15] Y. Vaizman, K. Ellis, G. Lanckriet, and N. Weibel, "ExtraSensory app: Data collection in-the-wild with rich user interface to self-report behavior," in *Proc. CHI*, 2018, pp. 1–12.

[16] Y. Xu, Z. Shen, X. Zhang, Y. Gao, S. Deng, Y. Wang, Y. Fan, and E. I.-C. Chang, "Learning multi-level features for sensor-based human action recognition," *Pervasive Mobile Comput.*, vol. 40, pp. 324–338, Sep. 2017.

[17] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, "A survey of online activity recognition using mobile phones," *Sensors*, vol. 15, no. 1, pp. 2059–2085, Jan. 2015.

[18] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sensors J.*, vol. 17, no. 2, pp. 386–403, Jan. 2017.

[19] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015.

[20] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," *Appl. Soft Comput.*, vol. 62, pp. 915–922, Jan. 2018.

[21] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1192–1209, 3rd Quart., 2013.

[22] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, Dec. 2015.

[23] A. Sucerquia, J. López, and J. Vargas-Bonilla, "Real-life/real-time elderly fall detection with a triaxial accelerometer," *Sensors*, vol. 18, no. 4, p. 1101, Apr. 2018.

[24] F. Hussain, F. Hussain, M. Ehatisham-ul-Haq, and M. A. Azam, "Activity-aware fall detection and recognition based on wearable sensors," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4528–4536, Jun. 2019.

[25] N. Jalloul, "Wearable sensors for the monitoring of movement disorders," *Biomed. J.*, vol. 41, no. 4, pp. 249–253, Aug. 2018.

[26] Z. Huang, W. Dong, and H. Duan, "A probabilistic topic model for clinical risk stratification from electronic health records," *J. Biomed. Informat.*, vol. 58, pp. 28–36, Dec. 2015.

[27] L. Liu, S. Wang, B. Hu, Q. Qiong, J. Wen, and D. S. Rosenblum, "Learning structures of interval-based Bayesian networks in probabilistic generative model for human complex activity recognition," *Pattern Recognit.*, vol. 81, pp. 545–561, Sep. 2018.

[28] P. Foudeh, A. Khorshidtalab, and N. Salim, "A probabilistic data-driven method for human activity recognition," *J. Ambient Intell. Smart Environ.*, vol. 10, no. 5, pp. 393–408, Sep. 2018.

[29] K.-C. Liu, C.-Y. Yen, L.-H. Chang, C.-Y. Hsieh, and C.-T. Chan, "Wearable sensor-based activity recognition for housekeeping task," in *Proc. IEEE 14th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, May 2017, pp. 67–70.

[30] M. H. M. Noor, Z. Salcic, and K. I. K. Wang, "Ontology-based sensor fusion activity recognition," *J. Ambient Intell. Hum. Comput.*, pp. 1–15, Jan. 2018. [Online]. Available: https://link.springer.com/article/10.1007/s12652-017-0668-0, doi: 10.1007/s12652-017-0668-0.

[31] C. Villalonga, H. Pomares, I. Rojas, and O. Banos, "MIMU-Wear: Ontology-based sensor selection for real-world wearable activity recognition," *Neurocomputing*, vol. 250, pp. 76–100, Aug. 2017.

[32] A. Yurtman and B. Barshan, "Activity recognition invariant to sensor orientation with wearable motion sensors," *Sensors*, vol. 17, no. 8, p. 1838, Aug. 2017.

[33] T. Sztyler, H. Stuckenschmidt, and W. Petrich, "Position-aware activity recognition with wearable devices," *Pervasive Mobile Comput.*, vol. 38, pp. 281–295, Jul. 2017.

[34] P. Bharti, D. De, S. Chellappan, and S. K. Das, "HuMAn: Complex activity recognition with multi-modal multi-positional body sensing," *IEEE Trans. Mobile Comput.*, vol. 18, no. 4, pp. 857–870, Apr. 2019.

[35] M. Gadaleta and M. Rossi, "IDNet: Smartphone-based gait recognition with convolutional neural networks," *Pattern Recognit.*, vol. 74, pp. 25–37, Feb. 2018.

[36] J. Wannenburg and R. Malekian, "Physical activity recognition from smartphone accelerometer data for user context awareness sensing," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 12, pp. 3142–3149, Dec. 2017.

[37] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, Jan. 2016.

[38] S. A. Antos, M. V. Albert, and K. P. Kording, "Hand, belt, pocket or bag: Practical activity tracking with mobile phones," *J. Neurosci. Methods*, vol. 231, pp. 22–30, Jul. 2014.

[39] A. Khan, M. Siddiqi, and S.-W. Lee, "Exploratory data analysis of acceleration signals to select light-weight and accurate features for real-time activity recognition on smartphones," *Sensors*, vol. 13, no. 10, pp. 13099–13122, Sep. 2013.
[40] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, "Fusion of smartphone motion sensors for physical activity recognition," *Sensors*, vol. 14, no. 6, pp. 10146–10176, Jun. 2014.
[41] D. Shi, R. Wang, Y. Wu, X. Mo, and J. Wei, "A novel orientation- and location-independent activity recognition method," *Pers. Ubiquitous Comput.*, vol. 21, no. 3, pp. 427–441, Jun. 2017.
[42] H. Martín, A. M. Bernardos, J. Iglesias, and J. R. Casar, "Activity logging using lightweight classification techniques in mobile devices," *Pers. Ubiquitous Comput.*, vol. 17, no. 4, pp. 675–695, Apr. 2013.
[43] D. Coskun, O. D. Incel, and A. Ozgovde, "Phone position/placement detection using accelerometer: Impact on activity recognition," in *Proc. IEEE 10th Int. Conf. Intell. Sensors, Sensor Netw. Inf. Process. (ISSNIP)*, Apr. 2015.
[44] A. M. Otebolaku and M. T. Andrade, "User context recognition using smartphone sensors and classification models," *J. Netw. Comput. Appl.*, vol. 66, pp. 33–51, May 2016.
[45] G. J. Nalepa, K. Kutt, and S. Bobek, "Mobile platform for affective context-aware systems," *Future Gener. Comput. Syst.*, vol. 92, pp. 490–503, Mar. 2019.
[46] S. A. Hoseini-Tabatabaei, A. Gluhak, and R. Tafazolli, "A survey on smartphone-based systems for opportunistic user context recognition," *CSURACM Comput. Surv.*, vol. 45, no. 3, pp. 1–51, Jun. 2013.
[47] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4566–4578, Jun. 2016.
[48] F. B. A. Ramos, A. Lorayne, A. A. M. Costa, R. R. D. Sousa, H. Almeida, and A. Perkusich, "Combining smartphone and smartwatch sensor data in activity recognition approaches: An experimental evaluation," in *Proc. 28th Int. Conf. Softw. Eng. Knowl. Eng.*, Jul. 2016, pp. 267–272.
[49] M.-S. Dao, T.-A. Nguyen-Gia, and V.-C. Mai, "Daily human activities recognition using heterogeneous sensors from smartphones," *Procedia Comput. Sci.*, vol. 111, pp. 323–328, Jan. 2017.
[50] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, "Complex human activity recognition using smartphone and wrist-worn motion sensors," *Sensors*, vol. 16, no. 4, p. 426, Mar. 2016.
[51] M. Shoaib, S. Bosch, H. Scholten, P. J. M. Havinga, and O. D. Incel, "Towards detection of bad habits by fusing smartphone and smartwatch sensors," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2015, pp. 591–596.
[52] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, Sep. 2018.
[53] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol. 59, pp. 235–244, Oct. 2016.
[54] J. Monteiro, R. Granada, R. C. Barros, and F. Meneguzzi, "Deep neural networks for kitchen activity recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2048–2055.
[55] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep residual bidir-LSTM for human activity recognition using wearable sensors," *Math. Problems Eng.*, vol. 2018, pp. 1–13, Dec. 2018.
[56] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognit. Lett.*, vol. 119, pp. 3–11, Mar. 2019.
[57] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 1533–1540.
[58] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Gener. Comput. Syst.*, vol. 81, pp. 307–313, Apr. 2018.
[59] Y. Vaizman and N. Weibel, "Context recognition in-the-wild: Unified model for multi-modal sensors and multi-label classification," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 4, 2017, Art. no. 168.
[60] M. Ehatisham-ul-Haq, M. A. Azam, J. Loo, K. Shuang, S. Islam, U. Naeem, and Y. Amin, "Authentication of smartphone users based on activity recognition and mobile sensing," *Sensors*, vol. 17, no. 9, p. 2043, Sep. 2017.
[61] M. A. Hall and L. A. Smith, "Feature subset selection: A correlation based filter approach," *Prog. Connectionist-Based Inf. Syst.*, vol. 1, no. 2, pp. 855–858, 1997.
[62] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. Int. Conf. Pervasive Comput.* Berlin, Germany: Springer, 2004, pp. 1–17.
[63] K. Lee and M.-P. Kwan, "Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results," *Comput., Environ. Urban Syst.*, vol. 67, pp. 124–131, Jan. 2018.
[64] S. Mehrang, J. Pietilä, and I. Korhonen, "An activity recognition framework deploying the random forest classifier and a single optical heart rate monitoring and triaxial accelerometer wrist-band," *Sensors*, vol. 18, no. 3, p. 613, Feb. 2018.
[65] L. Xu, W. Yang, Y. Cao, and Q. Li, "Human activity recognition based on random forests," in *Proc. 13th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD)*, Jul. 2017, pp. 548–553.
[66] R. Quinlan, *C4. 5: Programs For Machine Learning*, vol. 240. San Mateo, CA, USA: Morgan Kaufmann, 1993.
[67] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
[68] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using matthews correlation coefficient metric," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, Art. no. e0177678.
[69] Y. Wang, S. Cang, and H. Yu, "A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition," *IEEE Sensors J.*, vol. 18, no. 16, pp. 6874–6888, Aug. 2018.
[70] S. Scheurer, S. Tedesco, K. N. Brown, and B. O'Flynn, "Human activity recognition for emergency first responders via body-worn inertial sensors," in *Proc. IEEE 14th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, May 2017, pp. 5–8.

MUHAMMAD EHATISHAM-UL-HAQ received the B.Sc. and M.Sc. degrees in computer engineering from the University of Engineering and Technology (UET), Taxila, Pakistan, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree in computer engineering.

His field of specialization is pervasive and ubiquitous computing. His research interests are within the areas of signal, image and video processing, biomedical signal processing, mobile sensing, machine learning, human activity and emotion recognition, and human behavior analysis. He received the Gold Medal for his B.Sc. degree and the Chancellor's Gold Medal for his M.Sc. degree.

MUHAMMAD AWAIS AZAM received the B.Sc. degree in computer engineering from the University of Engineering and Technology (UET), Taxila, Pakistan, in 2006, the M.Sc. degree (Hons.) in wireless networks from Queen Mary University, London, U.K., in 2008, and the Ph.D. degree in pervasive and ubiquitous computing from London, in 2012. From 2006 to 2007, he was a Lecturer with UET Taxila. From May 2012 to July 2013, he was the Head of Academics with the Cromwell College of IT and Management, London. Since August 2013, he has been an Assistant Professor with the Department of Computer Engineering, UET Taxila, where he leads a Research Team of M.Sc. and Ph.D. students in the area of pervasive and ubiquitous computing. He is currently a Visiting Research Scholar with the Victoria University of Wellington and a Lead Tutor with Whitecliffe Technologies, New Zealand. His research interest includes network architecture, the IoT, network security, embedded systems, ambient intelligence, wireless communications, opportunistic networks, and recommender systems. He currently has active research collaborations in technically advanced countries, including U.K., USA, South Korea, and Germany, to find innovative and remarkable solutions for the problems in hand. He has received several national and international funding grants for projects and presenting his research at different international forums. He has also served as a reviewer for various peer-reviewed journals and technical program committee member of international conferences. He received the Gold Medal for his B.Sc. degree.

YASAR AMIN received the B.Sc. degree in electrical engineering with specialization in telecommunication, the M.Sc. degree in electrical engineering with specialization in system-on-chip design, and the Ph.D. degree in electronic and computer systems with the research focus on printable green RFID antennas for embedded sensors from the Royal Institute of Technology (KTH), Sweden, in 2001 and 2003, respectively, and the M.B.A. degree in innovation and growth from the Turku School of Economics, University of Turku, Finland. He is currently the Chairman and an Associate Professor of Telecommunication Engineering Department, University of Engineering and Technology Taxila, Pakistan. He is also the Founder of ACTSENA Research Group, UET Taxila. He has been serving as a leading Guest Editor at two International Journals and an active Reviewer of more than a dozen well reputed International journals.

USMAN NAEEM received the B.Sc. degree (Hons.) in information and communication technology and the Ph.D. degree from the Queen Mary University of London, U.K., in 2005 and 2009, respectively.

He is currently a Lecturer with the School of Electronic Engineering and Computer Science, Queen Mary University of London. His research interests are within the areas of pervasive and ubiquitous computing. Much of his research is focused on assistive technologies to support independent living for the elderly community. His other research interests also include context awareness, machine learning techniques, mobile sensing, educational technologies, gamification, and ambient intelligent environments.