

Human activity recognition based on multi-instance learning

Duygu Bagci Das^{1,2}  | Derya Birant³

¹Department of Computer Programming, Ege University, Izmir, Turkey

²The Graduate School of Applied and Natural Sciences, Dokuz Eylul University, Izmir, Turkey

³Department of Computer Engineering, Dokuz Eylul University, Izmir, Turkey

Correspondence

Duygu Bagci Das, Department of Computer Programming, Ege University, Izmir, Turkey.
Email: duygu.das@ceng.deu.edu.tr and duygu.bagci.das@ege.edu.tr

Abstract

Human activity recognition (HAR) is the process of classifying a person's actions, and it is an essential task for many human-centered applications. *Multi-instance learning (MIL)* is a special case of machine learning where the training examples are bags containing many instances, and a single class label is assigned for an entire bag of instances. In this study, we integrated these two concepts by introducing a novel approach: “human activity recognition based on multi-instance learning”, called HAR-MIL. Unlike previous studies, the proposed HAR-MIL method represents human activities differently: as a bag of various wearable sensors (gyroscope, magnetometer, accelerometer, and linear acceleration). HAR-MIL presents an applicable and flexible model by providing multi-instance representation and eliminating the restrictions of traditional single-instance representation. Therefore, the adverse effects of missing data, defective sensors, and biased measurement on activity classification performance were minimized. This study is the first to investigate the performance of two MIL algorithms (SimpleMI and MIWrapper) on HAR. In this study, we explored the effect of four main factors (sensor positions, sensor types, base learners, and single or multiple participants) on the multi-instance representation of relevant human daily activities. The effectiveness of the proposed HAR-MIL method was demonstrated on 50 participant-based and sensor-position-based activity recognition datasets. The experimental results showed that HAR-MIL is effective for wearable sensor-based HAR with high classification accuracy (99.32%). Furthermore, the results showed that the proposed method outperformed the state-of-the-art methods by 10% on average on the same dataset.

KEY WORDS

classification, human activity recognition, machine learning, multiple instance learning, wearable sensors

1 | INTRODUCTION

Machine learning algorithms have been widely used in different fields such as smart city applications (Tiwari et al., 2021), score prediction (Birant, 2021), medical diagnosis (Sharma et al., 2021), and fraud detection (Forough & Momtazi, 2022). Besides these topics, human activity recognition (HAR) is an emerging field, in which machine learning (ML) techniques are used to generate a classifier that learns activities from a set of observations of individuals' actions. Researchers are currently excited about investigating this field and adapting ML algorithms to this problem.

HAR is the process of automatically identifying physical activities performed by humans, such as, walking, eating, sitting, or falling. In recent years, HAR has enabled various applications, such as health care (Yassine et al., 2017), sports analysis (Zokas & Lukosevicius, 2018), transportation

(Carpineti et al., 2018), rehabilitation (Kantoch, 2017), military (Mukherjee et al., 2017), fall detection (Kerdjidj et al., 2020; Ozcan & Velipasalar, 2016), human-computer interaction (Yan et al., 2020), entertainment, gait analysis (Panahandeh et al., 2013; Weiss et al., 2019) and security (Supriyatna et al., 2019). Particularly, HAR is essential in many fields of medicine and sports to evaluate an individual's activity performance. In HAR, ML algorithms are usually applied to observations collected by cameras (Supriyatna et al., 2019), radars (Li et al., 2019), or wearable sensors (Daş & Birant, 2021; Kerdjidj et al., 2020; Tian et al., 2019; Wang et al., 2019). Camera-based HAR approaches have some drawbacks, such as requiring enough lighting, limiting the area, being an expensive solution, and potentially breaching personal privacy. Radar-based HAR approaches have a limited operating range and are generally used in military and security operations. Wearable sensor-based HAR approaches focus on action data gathered by a single sensor or a set of sensors. Wearable sensors are becoming more popular as sensor technology advances. These sensors are directly or indirectly installed on the human body. Typical examples are accelerometers, global positioning system sensors, gyroscopes, microphones, magnetometers, and biosensors. These sensors can be placed not only on watches, goggles, belts, hats, shoes, or smartphones but also on electromyography (EMG) devices composed of surface electrodes. Body temperature, pulse, speech, posture, and motion data can be collected using sensors. Using such sensors is not only useful for HAR but also efficient for other activity recognition problems, such as animal activity recognition (Birant & Yalniz, 2022). In this study, we used data collected from five smartphones placed on five body positions (left pocket, right pocket, belt, upper-arm, and wrist) with four types of sensors (accelerometer, linear acceleration, gyroscope, and magnetometer).

Generally, using raw sensor data directly in an intelligent model is not functional because it lacks significant information to specify and identify human activities. Thus, in this study, several data preparation methods were used to obtain relevant information from raw data, including signal filtering, signal segmentation based on the sliding window, and feature extraction.

Multi-instance learning (MIL) is defined as a type of supervised ML, in which a set of related training instances are grouped, called a bag, and a single class label is assigned to each bag. The purpose of MIL is to construct a classifier from a set of training bags with labels; and then; use the classifier to predict the labels of new (unseen) bags. In binary multi-instance classification, the label is set as positive if a bag has at least one positive instance; nevertheless, if all data instances in a bag is negative, the bag is labelled negative.

Several techniques for solving MIL problems have been discussed in the literature. Converting the bags to a standard representation method can be given as an example of those approaches. The foremost algorithms in MIL are SimpleMI and MIWrapper because many others are only suitable for binary multi-instance (MI) classification. Therefore, in this study, SimpleMI and MIWrapper methods were used individually with different classification algorithms as base learners individually, including decision tree (DT), k-nearest neighbours (KNN), neural networks (NN), partial decision tree algorithm (PART), support vector machines (SVM), AdaBoost (AB), and random forest (RF).

This study makes six significant contributions. (i) This study combines two concepts (HAR and MIL) differently and it proposes a novel approach, called HAR-MIL, such that human activity is represented as a bag of different wearable sensors, including an accelerometer (A), gyroscope (G), magnetometer (M), and linear acceleration (LA). (ii) This study investigates the optimal combination between MIL algorithms (SimpleMI and MIWrapper) and base classifiers (DT, SVM, NN, PART, KNN, RF, AB) for wearable sensor-based HAR problems. (iii) To the best of our knowledge, there is no MIL-based HAR study that answers the question of which sensor location (left pocket, right pocket, belt, upper arm, and wrist) provides more meaningful and key data to represent and recognize relevant daily activities. (iv) This is the first study to investigate different sensor configurations (A + G + M + LA, A + G + M, A + G, A + M, and G + M) in MIL-based HAR to determine which sensor configuration in a MI problem captures the relevant activity more accurately. (v) We also explore and compare the performances of MIL-based HAR for each participant individually (personalized classifiers) and all participants jointly (generalized classifier) in this study. (vi) The proposed method achieved higher performance (10% improvement on average) than the-state-of-the-art methods on the same dataset.

This study specifically answers the following important research questions:

1. How can participant datasets be used to improve MIL-based HAR, either individually or jointly?
2. What is the key combination of MIL algorithms and base classifiers for the best classification performance?
3. Which sensor type and position provide more useful data for multi-instance-based daily activity recognition?
4. Which combination of wearable sensors performs the key role for MIL-based HAR?

In the experimental studies, four experiments were conducted to investigate the impacts of participants, classification algorithms, sensor locations, and sensor types on MIL-based HAR for the first time. The proposed approach was assessed on 50 real-world activity recognition datasets, and the experimental results showed that the proposed HAR-MIL method effectively recognizes seven different physical human activities, including walking, standing, jogging, sitting, biking, upstairs, and downstairs.

This study is structured as follows. Section 2 presents the related literature and previous studies on HAR and MIL. Section 3 provides the background information about the HAR, MIL, and base learners used in this study. This section also explains the proposed approach with details of each step and the algorithm. Section 4 presents a short explanation of the dataset and provides the experimental results. Finally, concluding remarks and future work are presented in Section 5.

2 | RELATED WORK

2.1 | Literature review of human activity recognition

HAR has been widely used in many fields, such as (i) transportation (Carpineti et al., 2018) (i.e., vehicle type, navigation), (ii) intelligent environments (Bianchi et al., 2019; Yassine et al., 2017) (i.e., smart homes, smart traffic, smart hospital), (iii) military (Mukherjee et al., 2017) and security (Supriyatna et al., 2019) (i.e., bank, airport, school), and (iv) sports (Zokas & Lukosevicius, 2018) and health (Wang et al., 2019; Yassine et al., 2017) (i.e., gait analysis) (Panahandeh et al., 2013; Weiss et al., 2019), rehabilitation (Kantoch, 2017), fall detection (Kerdjidj et al., 2020; Ozcan & Velipasalar, 2016). Significant contributions in the last area can be explained as follows. Ozcan and Velipasalar (2016) constructed a fall detection model using wearable devices equipped with accelerometers and cameras. Panahandeh et al. (2013) proposed an approach for gait analysis in which the motion is modelled using a hidden Markov model.

Most HAR studies in the literature (Gao et al., 2019; Tian et al., 2019) have attempted the identification of basic human activities such as sitting, downstairs, walking, and upstairs. Nonetheless, some recent HAR studies focused on identifying various types of activities, such as military-based activities (Mukherjee et al., 2017) (i.e., run, jump, jump-rope), transportation-based activities (Carpineti et al., 2018) (i.e., riding a car, train, or bus), hand-oriented/non-hand-oriented activities (Das & Birant, 2020; Weiss et al., 2019) (i.e., typing, clapping, eating, drinking or kicking, jogging), and sports activities (Zokas & Lukosevicius, 2018) (i.e., push-ups, sit-ups, and squats). In addition to such activities, some studies mainly considered transitional human activities (sit-to-stand or stand-to-sit, etc.) (Wang et al., 2019). In this study, we built a multi-instance-based classifier to recognize seven different physical activities, including biking, jogging, sitting, walking, standing, upstairs, and downstairs.

A HAR system framework mainly includes the following steps: data acquisition, data pre-processing, signal segmentation, feature extraction and selection, training, model evaluation, and classification. Some earlier studies primarily focused on this issue because the feature extraction phase plays an essential role in the performance of the final system. For example, Attal et al. (2015) compared three different unsupervised learning techniques with four different supervised learning techniques in terms of precision, recall, specificity, and F-measure. Their results showed that the extracted features better characterized the activities. Similarly, Shoaib et al. (2014) considered four different feature sets to achieve more successful activity recognition results. In the literature, most HAR systems (Carpineti et al., 2018; Zokas & Lukosevicius, 2018) use the classification method as a supervised learning technique, even though other different types of ML techniques, such as semi-supervised and unsupervised (Machado et al., 2015; Yassine et al., 2017) have also been successfully used to provide information about human activities from raw sensor data. Supervised ML techniques are applied to labelled examples to construct classifiers, whereas unsupervised ML techniques are applied to unlabelled data. Some HAR applications have incorporated more than one ML technique, with clustering before classification providing sufficient evidence of both popularity and applicability. Some widely used classification algorithms for wearable sensor-based HAR include the C4.5 DT (Carpineti et al., 2018; Catal et al., 2015), SVM (Tian et al., 2019), NN (Mohamed et al., 2018), PART (Shoaib et al., 2014), KNN (Kerdjidj et al., 2020), RF (Zainudin et al., 2017), AB (Tian et al., 2019), hidden Markov model (HMM) (Panahandeh et al., 2013; Yan et al., 2020), linear/logistic regression (LR) (Catal et al., 2015), convolutional neural network (CNN) (Bianchi et al., 2019; Li et al., 2019), and Naive Bayes (NB) (Shoaib et al., 2014). Catal et al. (2015) investigated the strength of classifier ensembles in the task of human activity prediction by using DT, NN, and LR techniques.

Currently, different sensor types have been used in the design of HAR systems, such as wearable, vision-based, and environmental sensors. In this study, we used wearable sensors because they provide many advantages, such as wide coverage area, privacy protection, and high robustness when modelling activity recognizers. Wearable sensors can be effortlessly attached to various body parts (waist, arm, leg, wrist, etc.) depending on the human activities being studied. Mukherjee et al. (2017) explored the optimal sensor location for a group of human activities and the most relevant sensor types for classifying different activities. Similarly, Shoaib et al. (2014) explored different phone locations to examine whether changing smartphone positions enhances activity recognition performance or not.

Table 1 shows that the most commonly used sensors for activity recognition are gyroscope (G) and accelerometer (A), while the rarely used ones are the magnetometer (M) and linear acceleration (LA). Accelerometer and gyroscope sensors are prevalent selections for measuring position or movement, because of their low cost, lightweight (few grams), small size (few mm), ease of the program, and they can be easily attached to the body using either an adhesive or a strap. Tian et al. (2019) presented a two-layer multiclassifier method for HAR using a single wearable accelerometer. In our study, we used four different sensors (A, G, M, and LA) to improve human recognition performance. However, the main here is to use these sensors jointly in a bag to define a human action.

Currently, most HAR applications have been developed for smartphones because phones include several sensors (i.e., accelerometer, gyroscope), and they provide useful, wearable, easy-to-use computing platforms with computing power, and storage. For instance, Zokas and Lukosevicius (2018) constructed a recognition system to recognize certain types of human activities using a gyroscope and acceleration data which were collected using a smartphone. Nowadays, a few HAR studies focus on activity recognition on different devices such as biological devices or smartwatches (Ehatisham-Ul-Haq et al., 2020; Weiss et al., 2019). Some HAR applications use inertial sensors, while some of them use external sensors such as health sensors (Kantoch, 2017) or environmental sensors (Wang et al., 2019).

Gait analysis is an important and main branch of HAR. Calculation of joint angles from sensor data provides a successful solution to recognize gait patterns because human gait is the manifestation of change in the joint angles of the hip, knee, and ankle (Bijalwan et al., 2021). Joint angle

TABLE 1 Comparison of this study with the previous studies.

Ref	Year	#activities	#participants	Sub or base algorithms							
				DT	SVM	NN	PART	KNN	RF	AB	DL
Nair et al. (2022)	2022	6	36								+
Thakur et al. (2022)	2022	5	13	+	+			+	+	+	
Nafea et al. (2021)	2021	6	36-30								+
Alemayoh et al. (2021)	2021	8	8								+
Jain et al. (2021)	2021	7	10-30-744								+
Kerdjidj et al. (2020)	2020	2-3-7	17	+	+				+		
Ehatisham-Ul-Haq et al. (2020)	2020	6	60	+		+				+	
Tian et al. (2019)	2019	8	10		+						+
Gao et al. (2019)	2019	4	2			+				+	
Weiss et al. (2019)	2019	18	51	+					+	+	
Mohamed et al. (2018)	2018	6	4	+	+	+	+				
Mukherjee et al. (2017)	2018	6	17								
Zokas and Lukosevicius (2018)	2018	3	9			+					+
Garcia-Ceja et al. (2018)	2018	7	3								+
Carpineti et al. (2018)	2018	5	13	+	+	+					+
Kantoch (2017)	2017	4	7				+				
Zainudin et al. (2017)	2017	6	4								+
Attal et al. (2015)	2015	12	6		+				+	+	
Catal et al. (2015)	2015	6	36	+		+					
Shoaib et al. (2014)	2014	7	10	+	+			+	+	+	+
Our study		7	10	+	+	+	+	+	+	+	+
Ref		Sensor types				Sensor position					Multi-instance
		A	G	M	AL	Pocket	Belt	Arm	Wrist	Others	
Nair et al. (2022)	2022	+	+			+					+
Thakur et al. (2022)	2022	+	+					+			+
Nafea et al. (2021)	2021	+	+			+	+				+
Alemayoh et al. (2021)	2021	+	+				+				+
Jain et al. (2021)	2021	+	+				+	+	+	Chest Ankle	+
Kerdjidj et al. (2020)	2020	+	+						Chest		+
Ehatisham-Ul-Haq et al. (2020)	2020	+	+			+			Hand Bag Table		+
Tian et al. (2019)	2019	+				+					+
Gao et al. (2019)	2019	+	+	+	+			Hand			+
Weiss et al. (2019)	2019	+	+			+		+			
Mohamed et al. (2018)	2018	+				+	+	+	+		
Mukherjee et al. (2017)	2018	+	+	+		+	+	+	Thigh		+

TABLE 1 (Continued)

Ref	Year	Sensor types				Sensor position				Multi-class	Multi-label	Multi-view	Multi-task	Multi-instance
		A	G	M	AL	Pocket	Belt	Arm	Wrist					
Zokas and Lukosevicius (2018)	2018	+	+	+				+			+			
Garcia-Ceja et al. (2018)	2018	+							+	Table		+		
Carpineti et al. (2018)	2018	+	+		+	+					+			
Kantoch (2017)	2017	+	+				+				+			
Zainudin et al. (2017)	2017	+				+	+	+	+			+		
Attal et al. (2015)	2015	+								Chest Shank Ankle	+			
Catal et al. (2015)	2015	+				+					+			
Shoaib et al. (2014)	2014	+	+	+	+	+	+	+	+		+			
Our study		+	+	+	+	+	+	+	+					+

calculations can increase the adoption of wearable sensors in the field of HAR and improve the performance of the predictive model. Recent advancements in the calculation of the joint angles of the hip, knee, and ankles have motivated us to develop a reliable and robust model for gait activity recognition using wearable sensors (Bijalwan et al., 2022).

HAR problems have been addressed by researchers as different types of classification solutions, such as multi-class classification (Catal et al., 2015; Ehatisham-Ul-Haq et al., 2020), multi-label classification (Ehatisham-Ul-Haq et al., 2020; Mohamed et al., 2018; Zainudin et al., 2017), multi-view classification (Garcia-Ceja et al., 2018; Zhang et al., 2019) and multi-task classification (Gao et al., 2019). In contrast to these present types, a different classification type, called multi-instance classification, was proposed for HAR in this study. Here, an entire human action is represented as a bag of instances, where each instance is a feature vector extracted from different sensor data obtained at a certain time.

In this study, we do not consider the shallow (single-instance) classification algorithms, such as most HAR systems (Wang et al., 2019). Instead, we used a more sophisticated but practical human activity recognizer based on multi-instance learning (HAR-MIL).

2.2 | Literature review for multiple instance learning

Recently, MIL has been used in various areas, including health (Yousefi et al., 2018), bioinformatics (Bandyopadhyay et al., 2015) education (Ma et al., 2019), and security (Stiborek et al., 2018). Quellec et al. (2017) reviewed existing studies on medical image and video analysis as MIL problems in terms of modelling strategies. Campanella et al. (2018) used a deep MIL to make classifications and localizations in pathology.

The algorithms proposed for MIL can be grouped into three categories. (i) Some new algorithms were proposed particularly for MI problems, such as multi-instance diverse density (MIDD). (ii) Some standard learning algorithms were adapted to deal with MI data, such as MISVM (multi-instance SVM) and MILR (multi-instance logistic regression). (iii) Instead of adapting the algorithm, the data were used to transform it into a standard representation, including MIWrapper (Frank & Xu, 2003) and SimpleMI (Dong, 2006). In this study, we used the last algorithms because they are suitable for both binary-class and multi-class classification.

To the best of our knowledge, MIL-based HAR has never been considered in detail; only a few studies (Saha et al., 2021, Mosabbab et al., 2015, Stikic & Schiele, 2009) considered both concepts together, but in different ways. Unlike previous studies, the proposed HAR-MIL method represents the human activity phenomenon differently: as a bag of various wearable sensors (A, G, M, LA). Because MIL-based HAR has received little attention in the literature, we performed a comprehensive study, comprising the effects of four main aspects (sensor types, sensor positions, base learners, and single-multiple participants) on HAR-MIL for the first time.

Table 2 compares our study with related MIL-based HAR studies. The uniqueness of this study comes from the concepts considered in the field of MIL-based HAR (Table 2). Specifically, our study not only proposed an approach for HAR based on MIL but also investigated its

TABLE 2 Comparison of our study with the related MIL-based HAR studies.

Study	Data types		Concepts for MIL-based HAR					Method
	Visual	Signal	ST/HAR	SC/HAR	SP/HAR	SMP/HAR		
Prabono et al. (2022)	-	+	-	-	-	-	-	CNN
Saha et al. (2021)	-	+	-	-	-	-	-	MIML-KNN
Guan et al. (2016)	-	+	-	-	-	-	-	ARHMM
Yi and Lin (2016)	+	-	-	-	-	-	-	STG
Mosabbeb et al. (2015)	+	-	-	-	-	-	-	WSL
Our study	-	+	+	+	+	+	+	HAR-MIL (SimpleMI, MIWrapper)

effectiveness on different sensor types (ST) (A, G, M, LA) in a bag, sensor combinations (SC) (A + G + M + LA, A + G + M, A + G, A + M, and G + M), sensor positions (SP) (left pocket, right pocket, belt, upper arm, and wrist), and single-multiple participants (SMP) for the first time. Existing studies have contributed to the literature by proposing an approach and only measuring its effectiveness using some datasets without considering all the concepts described here. By considering all these concepts, this study extensively contributes to the related community. Our study also differs from previous studies in that we used different methods, which are SimpleMI and MIWrapper, whereas previous studies used the methods of weakly supervised learning (WSL), auto-regressive hidden Markov model (ARHMM), Spatio-temporal graph (STG), CNN, and multiple instances multiple label learning-based k-nearest neighbour (MIML-KNN).

3 | MATERIALS AND METHODS

This section includes background information about the methods used in this study and explains the proposed approach, “human activity recognition based on multi-instance learning” called HAR-MIL.

3.1 | Human activity recognition

HAR is denoted as the task of correctly detecting human action based on the observations gathered by cameras or sensors. Initially, researchers focused on human activity classification from images and videos, but as sensor technology advances, they investigated recognizing activities using sensors. Sensor-based HAR can be categorized into two groups: wearable sensor-based HAR (WSHAR) and ambient sensor-based HAR (ASHAR). ASHAR applications sense activities from fixed sensors that are embedded in an environment (i.e., home, factory) or attached to some specific fixed object in a predetermined point of interest, such as a door, fridge, wall, and so on (Yassine et al., 2017). Typical examples are light, reed switch, temperature, flow, pressure and vibrations sensors, RFID tags, infrared sensors, video cameras, and microphones. ASHAR systems are suitable, specifically for controlling security (i.e., intrusion detection) and the care demands of older people living at home. However, it only works with predetermined infrastructures in a limited area where the fixed sensors are placed. WSHAR applications, an alternative to ASHAR systems, identify activities by analysing raw data collected from sensors within wearable smart gadgets, such as smartphones, smart glasses, and smartwatches. These enable enormous applications because they cover a relatively large space (Wang et al., 2019). In this study, we focus on wearable sensor-based HAR because of its popularity.

3.2 | Multiple instance learning

In MIL, each learning sample is defined as a bag that contains multiple instances. Besides its feature values, the sole information about an instance is its membership relationship to a bag (Frank & Xu, 2003). MIL differs from traditional (single) instance learning by learning multiple instance dataset that includes bags of instances. MIL can be mathematically expressed as follows. Given χ as the instance space. The dataset D includes a set of bags and it is denoted by $D = \{(B_1A_1), (B_2A_2), \dots, (B_nA_n)\}$, where $B_j = \{x_{1j}, x_{2j}, \dots, x_{tj}\} \subset \chi$ is defined as a bag, t is the number of instances in the bag, and n is the number of training bags. The instance x_{kj} indicates the k^{th} instance in the bag B_j . The classification task occurs at the bag level. However, the multiple instance classifier takes an unseen bag as the input and predicts its class label. Even though several multiple instance learners, SimpleMI and MIWrapper combined with seven different classification algorithms were individually used on HAR datasets in this study. **SimpleMI:** SimpleMI is defined as the implementation of traditional learning algorithms on MIL applications (Dong, 2006).

The basic idea behind this algorithm is to summarize each bag and build an instance representing the entire bag. In this procedure, the entire MI dataset is transformed into a single-instance dataset. Therefore, a base learner can be executed without restriction. The class label is predicted using the single-instance model directly. Three alternative summarization techniques can be employed in SimpleMI (Dong, 2006). The first method is the *arithmetic mean*, which evaluates the average of each attribute for each bag. In this method, a single instance is considered to represent the entire bag. Generally, the mean (M) of instances of each bag is calculated to compress multiple instances into an average of these instances, as shown in Equation (1). Following that, the algorithm creates a new instance where the value of each attribute is equated to the mean.

$$M(X) = \frac{1}{|X|} \sum_{x \in X} (x) \quad (1)$$

Another summarization technique is the *geometric mean*, which evaluates the geometric mean of each attribute for each bag. Attribute values are summarized by multiplication, which is then calculated as the n th root of the product, where n is the number of dimensions: Equation (2). If m_{g1} is the mean value of the first dimension, the summarization of a bag can be represented using a new instance $(m_{g1}, m_{g2}, m_{g3}, \dots, m_{gn})$.

$$M_g(X) = \left(\prod_{x \in X} (x) \right)^{\frac{1}{n}} \quad (2)$$

The final option is the *minimax* method, where the minimum and maximum values of each dimension are considered for each bag. Therefore, the summarization is represented as $(\min x_1, \min x_2, \min x_3, \dots, \min x_n, \max x_1, \max x_2, \max x_3, \dots, \max x_n)$. The arithmetic mean approach was used for SimpleMI in this study. *MIWrapper*: MIWrapper is a multiple instances classification method that considers two major concepts that provide good performance (Frank & Xu, 2003). The first concept considers an appropriate weighting system where each instance is weighted in line with the corresponding bag size. This ensures that each bag is treated equally. The weight of the k th instance (w_k) in a bag is set as follows

$$w_k = \frac{p_x}{P_B} \times \frac{1}{m_k} \quad (3)$$

where p_x is the number of instances, P_B is the number of bags, and m_k is the number of instances in the k th bag. MIWrapper uses the inheritance rule given shown in Equation (4).

$$LR(X) = \frac{1}{|X|} \sum_{x \in X} lr(x) \quad (4)$$

where $lr(x)$ is the instance-based learner, and $LR(X)$ is the bag-based learner.

The second concept considers diverse density collective assumption, instead of the standard multiple instance assumption (Frank & Xu, 2003). The collective diverse density assumption assumes that the contribution of all instances of a bag is independent and equal to the class label of a bag. Therefore, the probability of each instance of a bag can be averaged. The bag level probability (P_{LR}) for a positive bag B_j^+ is given as follows:

$$P_{LR}(x = t | B_j^+) = \frac{1}{X} \sum_k P_{lr}(x = t | B_j k^+) \quad (5)$$

it can be specified for a negative bag B_j^- ,

$$P_{LR}(x = t | B_j^-) = \frac{1}{X} \sum_k P_{lr}(x = t | B_j k^-) \quad (6)$$

where X is the number of instances in the j th bag, x is a single point existing in the feature space and t is the true concept (Dong, 2006). The candidate for the true concept is determined by measuring its closeness to different positive bags. The more positive bags mean, the more likelihood is achieved. To measure the likelihood, the probabilistic measure of diverse density (D_d) is proposed (Oded & Tomás, 1997) as follows

$$D_d = P(x = t | B_1^+, B_2^+, B_3^+, \dots, B_n^+, B_1^-, B_2^-, B_3^-, \dots, B_m^-) \quad (7)$$

Because the goal is to maximize the likelihood, a general mathematical expression of the maximum diverse density is evaluated as follows

$$D_d(\max) = \operatorname{argmax}_x \prod_j P(x=t|B_j^+) \prod_j P(x=t|B_j^-) \quad (8)$$

The candidate point must be closer to the positive bag while it should be as far as possible from the negative bag. The first product in Equation (8) is the measurement of the candidate's closeness to the positive bag, whereas the second product term is the measurement of how far the candidate is away from the negative bags.

3.3 | Proposed approach (HAR-MIL)

In traditional HAR systems, an activity label is assigned to a single feature vector, which is extracted from one or more sensor data. However, when multiple sensors are attached to a human body to recognize the activity, they are jointly sampled, and each sensor is responsible for measuring a specific type of data. Because they are strongly related to each other, their data can be unified in a MIL manner to reach a final consensus. Therefore, multi-sensors formed as multi-instances can significantly improve recognition. Based on this motivation, we proposed a novel approach called HAR-MIL in this study. Our approach integrates multiple sensors in a unified manner using by utilizing an activity-based MIL framework. Multiple sensor data of activity are considered a bag of instances, where each instance is a feature vector extracted from one sensor data at a particular time.

Figure 1 shows the flowchart of a HAR-MIL approach, including data acquisition, pre-processing, signal segmentation, feature extraction, MIL representation, training, performance evaluation, activity classification, and decision support. In the *data acquisition* step, raw data were obtained at a particular sampling rate from multiple types of sensors (A, G, M, and LA) placed on a wearable device (i.e., smartphones, smart glasses, smartwatches, and smart shoes), during human action. After that, the raw data were transferred to the data collection center through a transmission technology (i.e., Wi-Fi, Bluetooth). The *data pre-processing* phase mainly involves filtering the raw data. The *signal segmentation* phase splits a large signal data stream into smaller fixed-size chunks (windows) via a sliding window technique, and human activities are related to each window. In the *feature extraction* phase, informative features are extracted from each signal segment by additional processing to improve the generalization performance of the HAR system. These features are the number of zero crossings, RMS velocity, crest factor, root mean squared (RMS) value, kurtosis, skewness, peak-to-peak value, and signal entropy. In the *MIL representation* phase, the feature vectors, which were obtained for each sensor, are grouped to form a bag. Therefore, each instance contains a feature vector extracted from a particular sensor, and each bag includes the instances from all sensors. In the *training* phase, the bag collection is fed to two MIL algorithms (SimpleMI and MIWrapper) to build models using the individual contributions of various base learners, such as DT, NN, SVM, PART, KNN, RF, and AB. The *evaluation* phase uses the k-fold cross-validation method to assess how well the constructed models recognize human activities. The classification model that performs better with the test set is chosen for HAR. The *classification* step uses the HAR model to identify the physical human activities of individuals, such as walking,

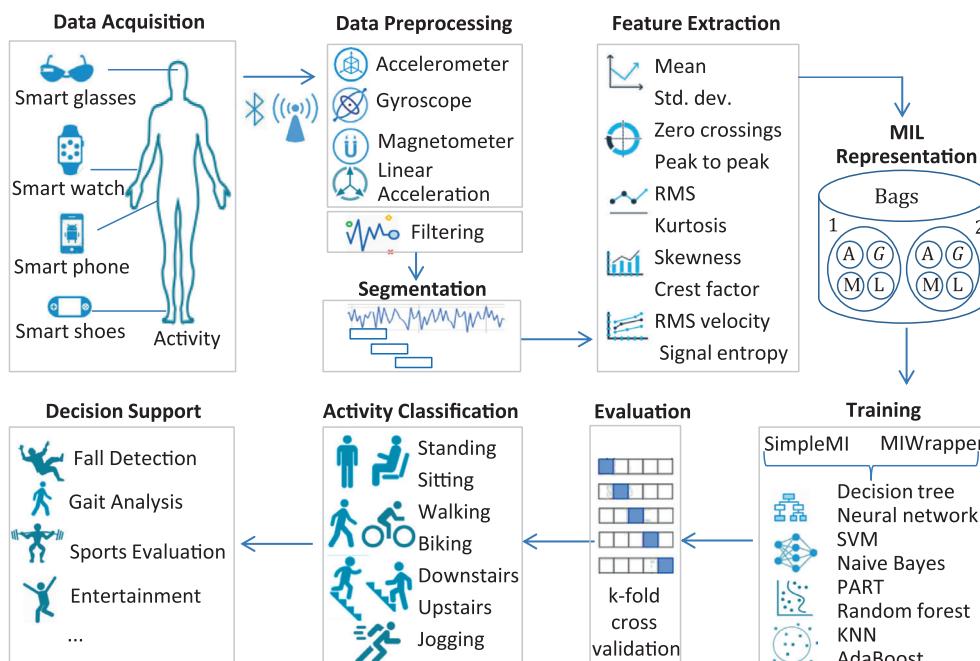


FIGURE 1 Overview of the proposed approach: Human activity recognition based on multi-instance learning (HAR-MIL).

sitting, jogging, biking, standing, downstairs, and upstairs. Each step in the workflow involves several methods and corresponds to research questions to be addressed. Furthermore, the classifier must be updated periodically by following the same sequence to achieve high accuracy.

In the proposed approach (HAR-MIL), each bag contains a set of instances, each of which belongs to an independent sensor. Let S sensor types $S = \{S_1, S_2, \dots, S_t\}$, where t is the number of sensor types. The training set D consists of a set of datasets $D = \{D_j\}_{j=1}^t$; each one contains instances belonging to a different sensor type such as A, G, M, and LA. Each dataset D_j includes a set of pairs of instances and their activity labels, which are denoted by $D_i = \{(x_{i1}, y_1), (x_{i2}, y_2), \dots, (x_{in}, y_n)\}$ for $i = 1, 2, \dots, t$, and the output attribute $Y = \{y_1, y_2, \dots, y_n\}$ has k human activities, where $y_i \in \{A_1, A_2, \dots, A_k\}$. Therefore, an instance x_{ij} is the j th element in the dataset D_i which contains samples related to S_i . An instance x_{ij} is a vector in the feature space F , such that $F = \{F_1, F_2, \dots, F_d\}$, where d is the dimension (also the number of extracted features).

We construct a MI dataset MD containing a set of bags and their activity labels such that $MD = \{\langle B_j, A_j \rangle\}_{j=1}^n$, where A_j is the activity (class label of the bag B_j). For instance, in a three-activity classification (standing, sitting, walking), the class labels of the bags are $A_1 = \text{standing}$, $A_2 = \text{sitting}$, and $A_3 = \text{walking}$. Each bag B_j consists of t instances such that $B_j = \{x_{1j}, x_{2j}, \dots, x_{tj}\}$ for $j = 1, 2, \dots, n$. A classifier Z is built to predict the class labels of unseen bags in the given test set T .

The pseudo-code of the proposed HAR-MIL method is presented in Algorithm 1. First, data preparation methods (filtering, segmentation, feature extraction) are applied to each sensor data separately in the algorithm. After that, n instances from t sensors are organized into bags. Each bag is labelled with a corresponding activity a . Here, a MI dataset MD is generated, which collects all instances from each different wearable sensor into a bag. The MI dataset is then used to train a classifier Z . In the final step, the classifier is used to predict the activity class label of each unseen bag in the test set T .

3.4 | Data preparation

To prepare data for the HAR-MIL method, the following steps were implemented: data pre-processing, segmentation, and feature extraction.

3.4.1 | Data pre-processing

The accelerometer measures the acceleration of an object along three directional axes: x-axis (vertical axis) (A_x), y-axis (horizontal axis) (A_y), and z-axis (sideway axis) (A_z). In the pre-processing stage of this study, the signal magnitude vector of acceleration, A_{mag} , was calculated using these

Algorithm 1 Human activity recognition based on multi-instance learning (HAR-MIL).

```

Inputs:  $D = \{D_1, D_2, \dots, D_t\}$  a set of datasets  $T$  is the test set
Output:  $O = \{o_1, o_2, \dots, o_p\}$  (a set of outputs that are assigned to each bag in  $T$ )
for  $j = 1$  to  $t$  do
     $D_j = \text{Filtering } (D_j)$ 
     $D_j = \text{Segmentation } (D_j)$ 
     $D_j = \text{FeatureExtraction } (D_j)$ 
end for
for  $i = 1$  to  $n$  do
    for  $j = 1$  to  $t$  do
         $x = \text{Take } x_{ij} \text{ from Dataset } D_j$ 
         $B_i = B_i \cup x$ 
    end for
     $a = \text{Take activity related with } i\text{th instance}$ 
     $MD = MD \cup \langle B_i, a \rangle$ 
end for
 $Z = \text{Training } (MD)$ 
for each bag in  $T$  do
     $o = \text{Classify } (Z, b)$ 
     $O = O \cup o$ 
end foreach
return  $O$ 
END ALGORITHM
```

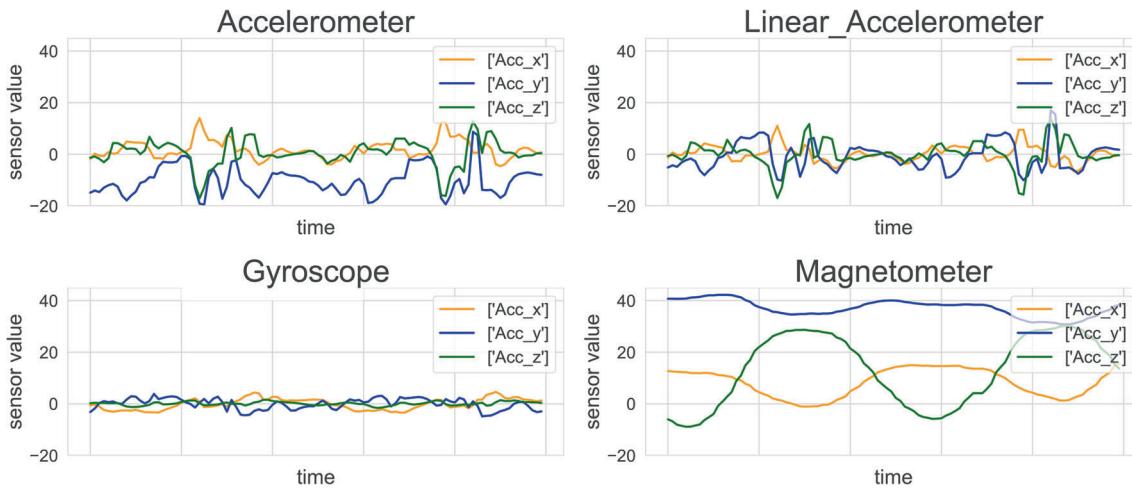


FIGURE 2 An illustration of raw HAR sensor signals located in the left pocket of a participant.

three axes' components as follows: $A_{mag} = \sqrt{A_x^2 + A_y^2 + A_z^2}$. In addition, a low-pass filter with 0.6 Hz cut-off frequency f_c was applied to the sensor raw data to isolate the gravity component. After that, the gravity value was subtracted from the original data to find the body acceleration value. An illustration of the raw sensor signals located in the left pocket of a participant is shown in Figure 2.

3.4.2 | Signal segmentation based on sliding window

In sensor-based HAR, *signal segmentation* is a method of splitting a large signal data stream into smaller chunks for processing, where each chunk corresponds to an activity. Signal segmentation is crucial stage because the efficiency of feature extraction and recognition accuracy directly depends on it. Generally, windowing approaches are used for signal segmentation. In the windowing approach, the sensor signals are segmented into fixed-size chunks, referred to as *windows*, and a single label is assigned to all the samples within the window. A window is defined as a set of adjacent sequences such that $X = \{x_r, x_{r+1}, \dots, x_{r+w-1}\}$, where w refers to the window size and r corresponds to an arbitrary position, such as $1 \leq r \leq n - w + 1$, where n is the data size. Furthermore, each window consists of a set of events, that is, $S = \{e_1, e_2, \dots, e_p\}$, where p is the number of events. In this study, we used the *sliding window* technique to segment the sensor data because it has been useful for the recognition of periodic signals (Shoaib et al., 2014).

A critical factor in the sliding window technique is selecting a suitable window size to achieve correct recognition since the ideal window size varies according to the characteristics of the signals being evaluated. Usually, small window sizes are better for detecting faster-changing activities; however, using short windows may lead to misclassification because a signal is sliced into multiple windows and thus some vital information about an activity may not be captured by a small window, especially in the case of semi-complex and complex activities (i.e., transportation modes). Alternatively, large windows are usually useful for the classification of semi-complex and complex activities (Akbari et al., 2018). In addition, a large window may contain signals that belong to more than one activity, because a single label is assigned to all examples within the window, decreasing classification accuracy. Based on this tradeoff, users usually determine the optimal window size using empirical methods while experimenting with various fixed window sizes and evaluating classification accuracy. In this study, the length of each window was set at 1 s because the data were collected at a rate of 50 samples per second, and it is a suitable sampling value to provide a reasonable compromise in HAR performance.

Segmentation based on the sliding window technique can be performed in either non-overlapping or overlapping ways. *Overlapping* is expressed by a percentage that defines how many samples from the previous window will be repeated in the current window, that is, $X_1 \pm X_2 \neq \emptyset$. On the other hand, *non-overlapping* indicates that the values in one window do not intersect with the values of another window, that is, $X_1 \pm X_2 = \emptyset$. It is a common practice to use overlapping windows to reduce information loss, particularly at the edges of the windows. Although different overlap values could be used, in this study, we used the sliding window technique considering a 50% of overlap between consecutive windows, based on the report of the study (Shoaib et al., 2014).

3.4.3 | Feature extraction

In the raw sensor data, only one specific value at a particular time instant of observation does not carry sufficient and in-depth information to describe human activity. *Feature extraction* typically transforms the original sensor data into more elaborate and informative features for HAR

(Wang et al., 2019). Feature extraction is designed to capture a more exhaustive and useful representation of the sensor data to correctly distinguish different human activities, such as kurtosis and skewness. Therefore, the choice of these features is critical because of the selection of the classification algorithm to generate a good classifier for HAR. Features are computed from each signal segment (window) in a frequency domain or time domain. *Frequency-domain features* describe the periodical properties of a signal over the window such as entropy; whereas, *time-domain features* are statistical summaries of the data samples, such as RMS, and peak-to-peak values.

Frequency-domain and time-domain features are usually extracted from one axis of a sensor, called *single-axial features*, such as the mean value of the x-axis of the accelerometer measure. Furthermore, *multi-axial features* can be extracted as a combination of multiple axes of single or multiple sensors, such as signal magnitude. In this study, both single and multiple axial, as well as frequency-domain and time-domain features were individually extracted from four sensor data (A, G, M, and LA) by considering three axes (x, y, and z-axis). Therefore, in this study, 112 features were extracted for each sensor type individually. The types of extracted features, including their descriptions and corresponding statistical formulas, are given in Table 3.

3.5 | Base classifiers

This section describes the individual base classifiers used in the MIL task in this study for activity recognition, including DT, NN, SVM, PART, KNN, RF, and AB. We selected these algorithms because most of them appear among the top 10 data mining methods (Wu et al., 2008).

Decision tree (DT): DT classification algorithms use attribute values to divide decision space iteratively to predict unknown class labels. The structure of the DT comprises nodes (attributes), branches (attribute values), and leaves (class labels).

Neural network (NN): Inspired by a human's nervous system, ANNs are mathematical models that have been widely used in the classification. A NN contains a set of neurons that work together with weight values to ensure that classification is performed successfully. A typical NN consists of an input layer, single or multiple hidden layers, and an output layer.

Support vector machine (SVM): In this algorithm, the classification is performed by defining borders among different instance groups. These borders are drawn at the furthermost distance to the instance groups. The algorithm calculates the distances to define the borders, which can be linear or nonlinear.

Partial decision tree algorithm (PART): This classification algorithm is a combination of the separate-and-conquer (rule-based) and the divide-and-conquer (decision tree-based) strategies. Nevertheless, it is different from traditional approaches using the creation method of each rule. It uses the leaf with the largest gain to create a decision rule, and consequently, eliminates the tree for the existing set of instances.

K-nearest neighbour (KNN): In this algorithm, an unlabelled instance is assigned to a class, which has the most members when considering KNNs. The distances between the instance and its neighbours are calculated using various metrics such as Euclidean, Minkowski, and Manhattan distance measures.

Random forest (RF): It is one of the ensemble ML methods, which includes multiple decision trees. A set of DTs, which are built from the diverse subsets of the dataset, are combined to produce the final prediction.

TABLE 3 Features extracted from sensor data.

Feature	Description	Formula
Number of zero crossings	The number of times the signal crosses the zero-level within each window. It is calculated using the sign function, which gives 0 for negative arguments, and 1 for positive ones.	$ZC = \sum_{i=2}^n \text{sign}(x_i) - \text{sign}(x_{i-1}) $
Peak-to-peak value	The difference between the maximum value of the window and the minimum value of the window.	$PP = \max(x) - \min(x)$
Root mean squared (RMS)	The quadratic mean value of the sensor signal over the window. It is a measurement of signal power.	$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$
Kurtosis	The degree of peakedness (the sharpness of the peak) of the signal distribution over the window.	$KV = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^4$
Skewness	The degree of asymmetry of the signal distribution of the samples within a window.	$SV = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$
Crest factor	The ratio of the maximum sensor signal value to the RMS values over the window. It is used to describe the impulsiveness of a signal.	$CF = \frac{\max(x)}{RMS}$
Root mean squared (RMS) velocity	The quadratic mean of the speed of the signal over the window.	$RMS_{velocity} = \sqrt{\frac{1}{n} \sum_{i=1}^n \text{diffinv}(x_i)^2}$
Signal entropy	Signal entropy, which provides an estimation of the amount of information.	$E = \sum_{i=1}^n P(x_i) \log_2 P(x_i) P(x_i) = -\frac{x_i}{\sum_{j=1}^n x_j}$

AdaBoost (AB): It is an ensemble ML technique that combines multiple weak learners to form a strong learner. In each step, it improves the success of the model by reweighting instances that are difficult to classify. The outputs of weak classifiers are combined based on a weighted voting mechanism.

4 | EXPERIMENTAL STUDIES

The practicality of the proposed approach (HAR-MIL) was evaluated on real-world datasets using two MIL algorithms (SimpleMI and MIWrapper) with the combinations of seven classification algorithms (DT, SVM, NN, PART, KNN, RF, and AB) using the machine learning tool (Witten et al., 2016). Except for AdaBoost and KNN, the input parameters of the classification algorithms were considered by their default values. For AdaBoost, the C4.5 DT algorithm was chosen as the base learner, because it has good generalization capability. For KNN, The k (the number of neighbours) value was determined by considering $k \approx \log_2(n)$ based on the previous study (Dexun et al., 2013), where n indicates the number of bags in the MI dataset. The default value of k is 1. Nevertheless, it is too small for a MI dataset, where large numbers of bags are available. We compared the alternative techniques in terms of classification accuracy, which is the percentage of testing cases predicted correctly by the classifier. Classification accuracy values were evaluated using a ten-fold cross-validation method, where the data is randomly split into nonoverlapped and equal-sized divisions, and one of those divisions is used for testing, while the remaining are used to build the classifier.

We designed four experiments to investigate the effects of the following main factors (Table 4):

- *Experiment 1—Impact of participant:* We explored the activity classification performance of MIL-based HAR when both training and testing datasets come from a single participant (personalized classifier) or all participants (generalized classifier).
- *Experiment 2—Impact of algorithms:* We evaluated the performances of different MI classification algorithms with different base classifier combinations.
- *Experiment 3—Impact of sensor position:* We investigated the possible SP (left pocket, right pocket, belt, upper arm, and wrist) to determine whether changing SP improves the classification accuracy obtained by the proposed method (HAR-MIL).
- *Experiment 4—Impact of sensor type:* We measured the impact of the various combinations of four sensors (A, G, M, and LA) to determine the most effective combination that performs best for the MIL-based HAR.

4.1 | Dataset description

In this study, we used a large-scale and publicly available sensor activity recognition (SAR) dataset (Shoaib et al., 2014) to evaluate our method. The dataset was collected for seven basic activities, which are walking, standing, jogging, sitting, biking, upstairs, and downstairs. The experiments were conducted in the university buildings, except for biking. The “walking” and “jogging” activities were performed in the department corridor, while the “upstairs” and “downstairs” activities were performed in a 5-floor building. Samples were obtained from four sensors (accelerometer (A), gyroscope (G), magnetometer (M), and linear acceleration (LA)) with a 50 Hz sampling rate using Samsung Galaxy SII (i9100) smartphones. An Android application was developed to gather the data from the sensors. It was possible to work on the data fusion and sensor fusion concepts because the data were obtained from four different sensors. Those sensors were placed on the left, and right pockets, belts, upper arms, and wrists of the individuals. The first three sensor locations are the most common ones because humans mostly carry their smartphones at one of those locations. The fourth position (upper arm) is frequently used when performing activities, such as jogging. The fifth position (wrist) simulates a smartwatch or a smart wristband. For each activity, the data were gathered for all five body locations simultaneously. The dataset was collected

TABLE 4 Experiments performed for evaluation of MIL-based HAR.

Exp.	Evaluation		Evaluation scenarios						
1	Impact of participant		Single participant				All participants		
2	Impact of algorithms	MIL algorithms	SimpleMI			MIWrapper			
		Base classifiers	DT	SVM	NN	PART	KNN	RF	AB
3	ML algorithm		Extreme learning machine						
	Impact of sensor position		Left pocket	Right pocket			Belt	Upper arm	Wrist
4	Impact of sensor type		A + G + M + LA	A + G + M		A + G	A + M	G + M	

from 10 participants, who were male and between the ages of 25 and 30. Each participant performed each activity for 3–4 min. Given 5 sensor positions and 10 participants, we used 50 datasets.

Each sensor obtains three-dimensional values: x, y, and z-axis. The accelerometer sensor measures gravitational acceleration in meters per second squared (m/s^2), while the gyroscope sensor provides the rotation rate for each of the three axes in radians per second (rad/s), and the magnetometer sensor reports the magnetic field in microtesla (μT) along each axis.

The SAR dataset is in the form of raw data, including the three axial values (X, Y, Z) measured by each sensor (A, G, M, and LA). In the data preparation stage of this study, the following three steps were performed: (i) *data pre-processing*—the signal magnitude vector was calculated using these three axes components, and a low pass filter with a 0.6 Hz cut-off frequency; which was applied to isolate the gravity component and then to obtain the body component; (ii) *segmentation*—the sliding window method with 1 s window length and an overlap of 50% was used to segment the sensor data; and finally (iii) *feature extraction*—both single and multi-axial, as well as frequency-domain and time-domain features were separately extracted from four sensor data (A, G, M, and LA), including the number of zero crossings, RMS velocity, crest factor, RMS value, kurtosis, skewness, peak-to-peak value, and signal entropy. Thus, 112 features were obtained for each sensor type.

To convert single-instance data to MI data, we implemented a bag-of-sensors generator. The generator represents each action as a bag of four instances, where each instance is a feature vector extracted from different sensor data obtained at a particular period. Thus, each sensor data is represented by a feature vector and referred to as an instance, and multiple sensor data is considered a bag of instances.

4.2 | Experimental results

4.2.1 | Experiment 1: Impact of participant

In this experiment, the proposed HAR-MIL method was evaluated in terms of participant factors. We compared the recognition performance from two different view points: the sensor data comes from a single participant (personalized classifier) or all participants (generalized classifier). In the first case, each participant has his/her own classification model, implying that the dataset of each participant was used to build a classification model. However, in the second case, all participants' data are combined to construct a general classifier.

In the personalized version, the classification models are trained for a specific participant with his or her own data. Participant-based accuracy results of MIL algorithms with different base learners and sensor locations are shown in Table 5. The bold values indicate the highest values for base learners regarding each participant and each sensor position. According to the results, the accuracy values ranged between 90.42% and 99.94%. Higher accuracy values for SimpleMI were obtained using the NN algorithm, whereas higher ones for MIWrapper were achieved using the RF algorithm.

Figure 3 shows the average accuracy values achieved for each participant separately (personalized) and all participants together (generalized). These results show that it is possible to achieve high and stable recognition accuracies using the proposed HAR-MIL method for both personalized and generalized cases. The average accuracy value of each participant is different because of the differences in human behaviour. According to the results, the personal accuracies of participants were between 95.82% and 99.58% for SimpleMI, and 96.34% and 99.65% for MIWrapper. However, the overall recognition accuracy (94.53% and 93.92%) for all participants was slightly lower than that for individual ones. The results show that combining all participants' data as a single dataset decreased the classification accuracy by approximately 3% on average.

4.2.2 | Experiment 2: Impact of algorithms

In the second experiment, we evaluated the performances of two MI classification algorithms with various base classifier combinations. The classification accuracies obtained using the HAR-MIL approach depend on the classification methods being used (Figure 4). According to the results, the best accuracy was obtained for SimpleMI using the NN algorithm (99.04%), whereas it was achieved for MIWrapper using the RF algorithm (99.13%). SimpleMI + SVM combination also achieved very promising results. MIWrapper + AB combination demonstrated above 98% accuracy for all sensor positions which is because AdaBoost is an ensemble ML technique that combines multiple weak learners to form a strong learner. The results indicate that MIWrapper has slightly higher classification accuracies compared to SimpleMI.

The performance of the proposed method was compared with one of the powerful classification methods, known as extreme learning machine (ELM), which has been used for HAR in recent studies (Patil et al., 2019). The parameter tuning for ELM was performed considering alternative activation functions (*sigmoid*, *sinusoidal*, *ReLU*) and the number of hidden neurons (10, 25, 50, 100, 250, 500, 1000). The best result was obtained when using 500 hidden neurons with the sigmoid activation function. Comparing the results of ELM (94.57%) and HAR-MIL (99.32%), the proposed approach outperformed ELM for the SAR dataset. The confusion matrix obtained using the ELM is shown in Figure 5. The stable activities (standing and sitting) were perfectly classified using ELM, but it was difficult to distinguish upstairs and downstairs activities.

TABLE 5 (Continued)

		MIWrapper						
Participant/sensor position		C4.5	SVM	RF	KNN	NN	PART	AB
P9	Upper arm	99.05	99.22	99.78	99.44	99.78	99.27	99.78
	Wrist	99.55	98.83	99.78	99.66	99.61	99.61	99.78
	Left pocket	96.57	96.85	98.68	98.40	98.52	96.73	98.32
	Right pocket	97.21	92.98	98.80	98.72	97.65	97.65	98.76
	Belt	96.49	97.65	98.68	97.96	97.73	96.61	98.44
P10	Upper arm	95.09	94.29	98.32	97.49	96.81	95.81	97.85
	Wrist	96.57	93.22	98.36	97.61	97.69	96.37	98.28
	Left pocket	97.53	96.73	98.76	98.52	98.64	96.73	98.52
	Right pocket	96.73	96.73	99.12	99.04	99.16	97.13	98.52
	Belt	97.01	97.69	99.12	98.68	97.29	96.57	98.84
	Upper arm	93.97	93.34	98.28	97.05	97.73	94.41	97.29
	Wrist	93.66	92.90	97.17	95.49	96.49	93.89	97.09

Note: The bold values indicate the highest values for base learners regarding each participant and each sensor position.

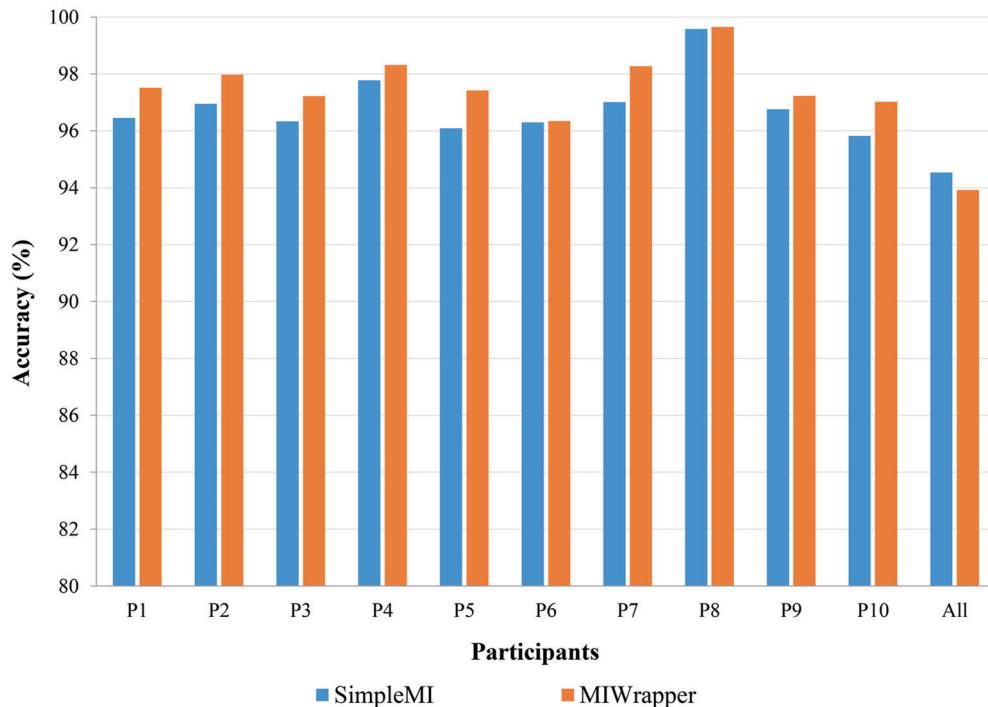


FIGURE 3 Comparison of personalized (participant-based) and generalized (all-participants-based) classifiers in terms of average classification accuracy.

4.2.3 | Experiment 3: Impact of sensor position

Wearable sensors can be placed on various body parts such as the waist, arm, or wrist. Determining the ideal body location, where the sensor will be placed, is a research-worthy problem in HAR because the sensor location considerably affects the activity recognition accuracy. The sensor location is crucial because activity dynamics may vary depending on the sensor position. For example, the sensor placement on the thigh contributes to high accuracy for jogging activity; however, the same position of the sensor is probably not useful for hand-oriented activities, such as *clapping* or *writing*. For instance, a sensor placed on a thigh produces several motions when the participant is walking, resulting in a significant fluctuation of the signal, thus the signal includes a sequence of valleys and peaks because the leg forms a series of downward and upward

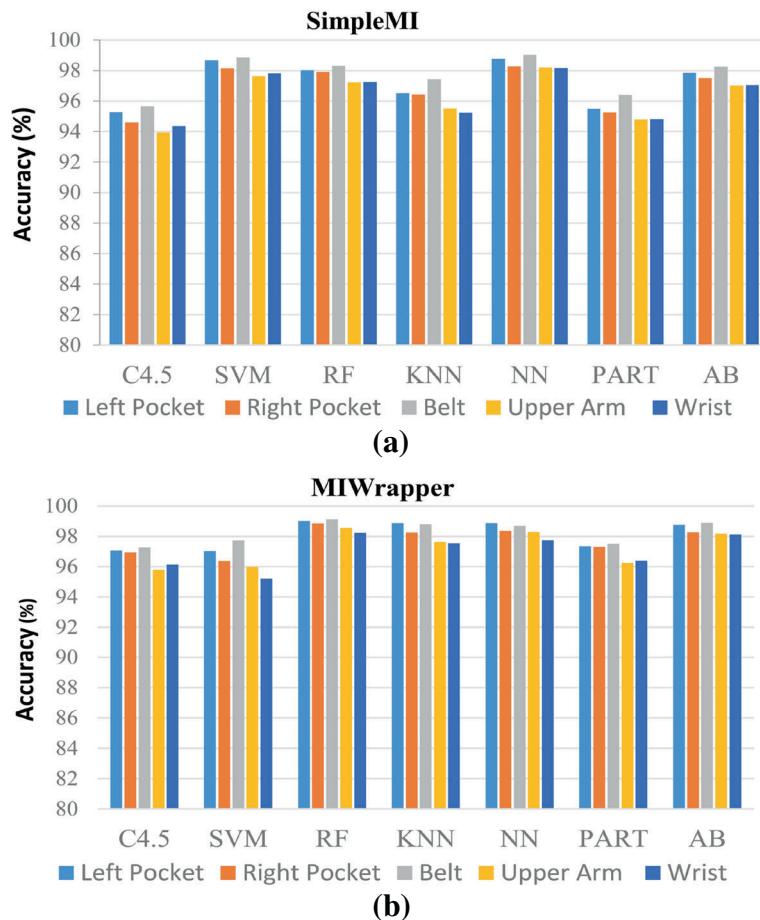


FIGURE 4 Comparison of (a) SimpleMI and (b) MIWrapper algorithms with various base learners in terms of average classification accuracy.

ELM						
a	b	c	d	e	f	g
a 98.51%	0	0	0	0	0	1.49%
b 0	100.00%	0	0	0	0	0
c 2.99%	0	95.52%	0	0	0	1.49%
d 0	0	0	100.00%	0	0	0
e 0	1.39%	0	0	97.22%	0	1.39%
f 0	0	2.47%	0	3.70%	83.95%	9.88%
g 4.41%	0	0	0	0	8.82%	86.76%
Accuracy: 94.57 %						

FIGURE 5 Confusion matrix obtained by ELM for HAR (a) walking, (b) standing, (c) jogging, (d) sitting, (e) biking, (f) upstairs, and (g) downstairs.

movements when walking. However, when the sensor is placed at the chest, the sensor signal reflects the movement of the torso. Therefore, the role of SP should be investigated in detail.

In this experiment, we evaluated the performance of the HAR-MIL method for five different SP on the body, including the left pocket, right pocket, belt, upper arm, and wrist. The first three positions are frequently used by humans when carrying smartphones. When the smartphone is in the pants pocket, the signal tends to reflect the movement of the leg. The fourth position of the sensor is commonly used when activities, such as biking are performed. The fifth position simulates a smartwatch or a smart wristband. Therefore, placing smartphones in one of these five positions does not affect the physical activities of the individual in any way. Smartphones were used to collect data for all five body locations simultaneously for each human activity. The phones were oriented portraits for the two pockets, wrist, and upper arm for the belt position. Note that the locations for these phones on users' bodies are fixed. After performing an activity, five smartphones were removed from the participant's body and the Android applications were stopped. This situation caused some abnormal signals (noise) at the start and end of each activity data. These noisy parts of the data were removed before analysis.

It is also possible to combine multiple sensors from different body positions; however, this could cause complex sensor placement, higher cost, deployment difficulty, computational complexity, and obtrusiveness for individuals. In addition, the sensors spread over the body hinder the user from performing daily living activities, which may result in the individual rejecting wearing them. As a result, the placement of sensors in a single position is more appropriate for the recognition of basic human activities, such as walking, upstairs, and downstairs. Therefore, in this study, we used the sensor data obtained from a single body position each time when training and testing classifiers.

Figure 6 shows the comparative results for different sensor positions (SP). We plotted the accuracies on both dimensions to investigate the relative performance of different classification algorithms for different sensor locations. The results showed that the proposed HAR-MIL approach does not give the same accuracy rate based on the different SP. Therefore, the results indicated that the sensor position on the body is a significant issue to consider to capture correct body movements. According to the results, the highest average accuracy for both SimpleMI and MIWrapper was achieved using a belt as the smartphone position, 97.72%, and 98.29%, respectively. Thus, the most discriminant sensor location determined using the HAR-MIL method is the belt. The SimpleMI + NN combination achieved very high accuracy values, of more than 98% at all positions. Similarly, the MIWrapper + RF algorithm is the most effective method for recognizing various types of daily activities because it achieved the best accuracy on an average (98.76%) when considering all SP. The results showed that MIWrapper has slightly higher classification accuracies compared to SimpleMI. In addition, note that the recognition performance trends are similar for the left and right pocket locations. They can be used alternatively when recognizing daily human activities because they perform almost equally in many situations. All HAR-MIL algorithms achieved good performance for all sensor locations, that are above 93.94%.

4.2.4 | Experiment 4: Impact of sensor type

HAR systems typically use only one sensor type, known as an accelerometer. However, a single sensor type cannot deal with semi-complex or complex human activities in practice. This is why we investigate multiple sensory HAR applications. Different sensor types provide diverse and abundant information and varied perspectives for specific human activities. For example, an accelerometer can provide information about the

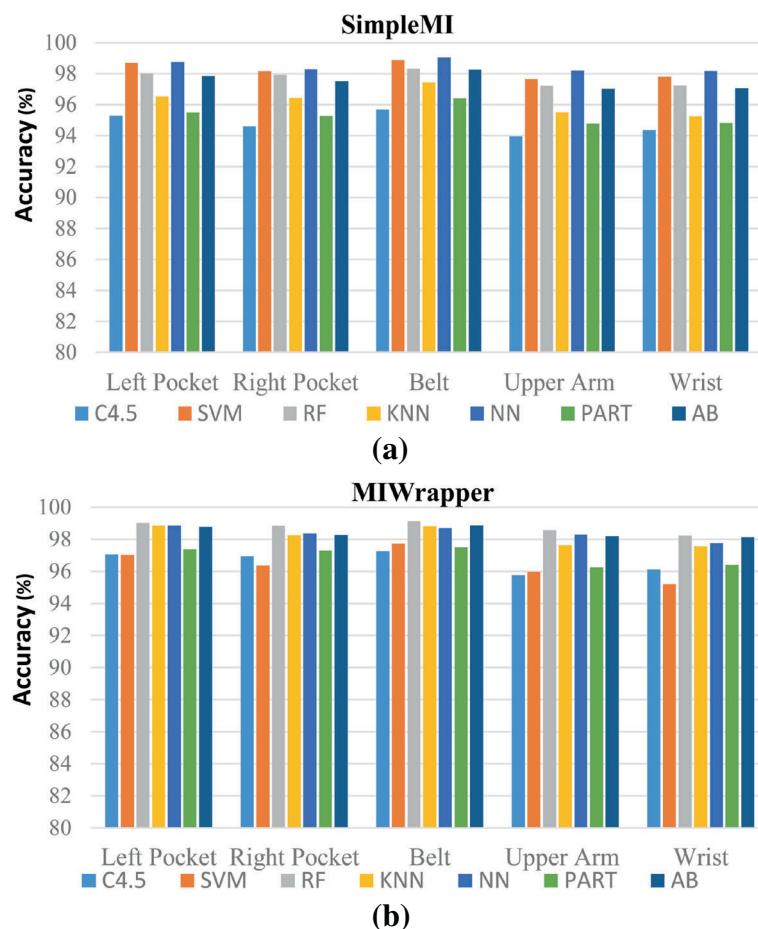


FIGURE 6 Comparison of different sensor positions in terms of classification accuracy (a) SimpleMI and (b) MIWrapper.

body translation, and a gyroscope can measure the orientation of the motion; thus, the combination of data obtained from these sensors is usually required to recognize complex human activities correctly. The combination of multiple sensor types can provide useful information about human activities, thereby improving the classification performance of a HAR system. However, using multiple sensors could increase the complexity and cost of the system compared with a single sensor type. Therefore, instead of using all available sensors, the selection of an ideal sensor combination is required to minimize data transmission overheads on a limited network.

In this experiment, we evaluated the performance of the HAR-MIL method for different sensor configurations, including A + G + M + LA, A + G + M, A + G, A + M, and G + M. This analysis was used to determine how many and which motion sensors in combination are ideal to be used for better recognition performance. We reduced the cost, complexity, and data transmission load by evaluating the performances of various sensor combinations in activity recognition to produce results comparable to the use of A + G + M + LA.

Figure 7 compares different sensor combinations when the RF algorithm is used with MIL algorithms. We used the RF algorithm due to its superiority in the previous experiments. We plotted the accuracies on both dimensions to investigate the relative performance of different sensor combinations for different sensor locations. According to the results, the best sensor combination on average is A + G for SimpleMI (97.16%), whereas it is A + G + M for MIWrapper (98.09%). The results showed that the accelerometer and gyroscope play lead roles when recognizing multi-instance-based human activities. Among the four sensor types, A + G attracts more attention because of its low cost, and flexibility in daily activities, and satisfying recognition performance. The highest accuracy (98.65%) was achieved by MIWrapper using the belt as the smartphone position. The results showed that MIWrapper has slightly higher classification accuracies than SimpleMI.

4.2.5 | Human activity evaluation

A confusion matrix is a common method for visualizing the performance of a classifier for each class. The rows and columns in the matrix represent the actual and predicted classes, respectively. An element (M_{ij}) in a confusion matrix M is the number of instances that belong to class i but are recognized as class j . The nondiagonal elements represent misclassified compounds, whereas the diagonal elements (M_{ii}) represent the number of instances correctly classified for each class.

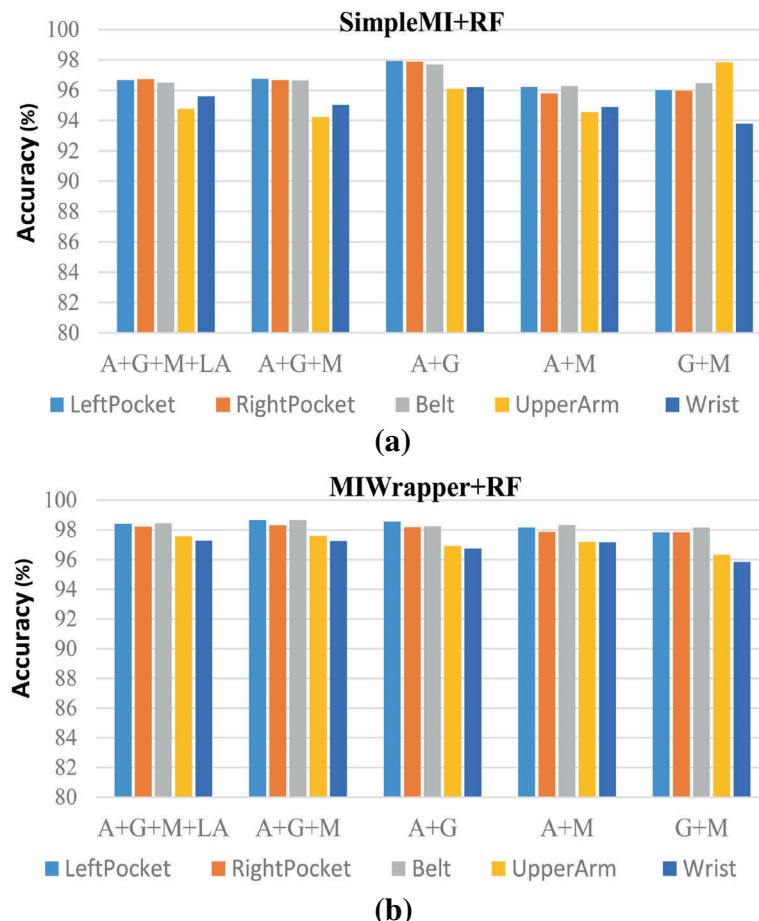


FIGURE 7 Comparison of different sensor combinations in terms of classification accuracy (a) SimpleMI + RF and (b) MIWrapper + RF.

To show the superiority of our method, we compared it with the state-of-the-art methods in the literature. The related studies that used the SAR dataset (Shoaib et al., 2014) along with the techniques and the corresponding classification accuracies are shown in Table 6. The bold value denotes the highest model accuracy. The proposed HAR-MIL approach outperformed the other methods with a 10% improvement on average Table 6. Employing HAR-MIL (SimpleMI + NN) for sensors located in the belt position achieved better accuracy (99.32%) than the state-of-the-art models on the same dataset.

5 | DISCUSSION

In this study, the performance of multi-instance learning-based human activity recognition (HAR-MIL) was investigated on daily activities (walking, standing, jogging, sitting, biking, upstairs, downstairs) using different algorithms (SimpleMI and MIWrapper) with different base learners (DT, SVM, NN, PART, KNN, RF, AB) for different sensor positions (left pocket, right pocket, belt, upper arm, and wrist) and for various sensor combinations (A + G + M + LA, A + G + M, A + G, A + M, and G + M).

TABLE 6 Comparison of the proposed method against the state-of-the-art methods on the SAR dataset.

Study	Method	Accuracy (%)
Casado et al. (2022)	Federated and continual learning (CDA-FedAvg)	81.90
Siirtola and Röning (2021)	Incremental learning	93.60
Lima et al. (2021)	NOvelty discrete data stream for human activity recognition (NOHAR)	94.85
Tao et al. (2021)	Convolutional neural network (CNN)	99.27
Braganca et al. (2020)	Human activity recognition based on symbolic representation (HAR-SR)	83.21
	Time and frequency features-based method (TF)	78.15
Bhat et al. (2020)	Neural network	95.00
Lin and Marculescu (2020)	Convolutional neural network (CNN)	96.70
Voicu et al. (2019)	Multi-layer perceptron	86.10
Chakraborty et al. (2019)	Deep neural network	82.62
Gurov et al. (2019)	Logistic regression	99.00
Shakya et al. (2018)	Convolutional neural network (CNN)	99.16
	Recurrent neural network (RNN)	95.65
	K-nearest neighbours	97.65
	Decision tree	90.57
	Random forest	95.77
Menhour et al. (2018)	K-nearest neighbours	71.00
	Support vector machine	72.50
	PCA–K-nearest neighbours–support vector machine	71.90
	LDA–K-nearest neighbours–support vector machine	73.00
Neyestani et al. (2018)	Random forest	90.72
	Naive Bayes	95.68
	Stacking learning	97.80
Zainudin et al. (2017)	MLC–binary relevance (BR)	96.00
	MLC–classifier chains (CC)	96.20
	MLC–Bayesian classifier chains (BCC)	96.20
	MLC–label combination (LC)	96.90
	MLC–RAndom k LabEL subsets (RAkEL)	96.36
	Random forest	94.83
Buenaventura and Tiglao (2017)	K-nearest neighbours	98.10
	Decision tree	74.00
Natarajasivan and Govindarajan (2016)	Support vector machine + Kalman filter	82.00
Our study	HAR based on multi-instance learning (HAR-MIL)	99.32

Note The bold value denotes the highest model accuracy.

The main results of this research are concluded as follows:

- The highest accuracy value was obtained by SimpleMI + NN for belt location as 99.32%. Therefore, the proposed HAR-MIL method outperformed the state-of-the-art methods by 10% on average on the same dataset.
- The overall average accuracy was approximately 97% when considering all participants, sensor locations, and classification algorithms.
- Participant-based (personalized) training achieves higher classification accuracy than all-participant-based (generalized) training because of the differences in human behaviour. Combining all participants' data as a single dataset decreased the classification accuracy by approximately 3% on average.
- The best classification results are usually obtained by MIWrapper; however, SimpleMI also classifies correctly as MIWrapper does. Although MIWrapper usually gives approximately 1% higher accuracy results than SimpleMI, its execution time takes approximately ten times longer than that of SimpleMI.
- The highest average accuracy results considering all sensor locations were obtained by the combinations of SimpleMI + NN (98.49%) and MIWrapper + RF (98.76%). Thus, using the NN algorithm or the RF algorithm can provide better performance among the seven base learners.
- When the sensor position is considered, the belt (97.72%), left pocket (97.23%), right pocket (96.88%), wrist (96.39%), and upper arm (96.33%) were the most effective to least effective locations for SimpleMI on average. MIWrapper has the same average ranking as SimpleMI, which is the belt (98.29%), left pocket (98.14%), right pocket (97.76%), upper arm (97.24%), and wrist (97.06%). The *belt* position yielded the highest accuracy, followed by the *left pocket* position. Therefore, the *belt* position can be successfully considered as the best sensor location when working on MI HAR data to recognize various types of human activities. However, different HAR systems could place sensors on different parts of the human body targeting specific aims, activities, and applications, instead of a belt.
- On average, the best sensor combination for SimpleMI (97.16%) is the accelerometer and gyroscope (A + G), whereas that for MIWrapper (98.09%) is the accelerometer, gyroscope, and magnetometer (A + G + M). These results show that the accelerometer and gyroscope play lead roles when recognizing multi-instance-based human activities. Therefore, instead of using all available sensors, the accelerometer and gyroscope can be used together to minimize cost, complex sensor placement, data transmission load, computational complexity, and obtrusiveness for individuals. In addition, finding the optimal HAR sensor combination may decrease the cost of the design, management, and maintenance of wearable devices.
- When considering the type of human activity is considered, the classifiers are better at recognizing the *standing*, *sitting*, and *biking* activities. Furthermore, the generated classifiers can correctly distinguish very similar activities, such as downstairs and upstairs. In terms of sensor locations, the belt position mostly takes the leading role in recognizing human activities.

These results demonstrated that the proposed HAR-MIL approach can appropriately classify human activities of daily living activities because it usually provides very high accuracy values. The results obtained in the four experiments in this study showed the feasibility of our method in terms of cost and applicability.

In this study, we demonstrated the results of MIL-based HAR for some specific scenarios, such as algorithm combinations, sensor positions, and sensor combinations. However, these results may differ if different MIL algorithms, different sensor positions, or feature sets are used. Furthermore, we left the input parameters of the classification algorithms as their default values, except for KNN and AdaBoost. Changing the default parameters may lead to different classification results. Moreover, in this study, the orientation of the smartphone is fixed, and changing the orientation may also lead to different results.

Using MIL for HAR has several advantages. Because multiple sensors are strongly related to each other, it can be useful to unify their data in a MI manner to reach a final consensus. In this way, multi-sensors formed as MI can significantly improve recognition. In addition, using the MIL may reduce the negative effects of classification accuracy caused by sensors that are out of order for some reason. The proposed method can be used in many fields and for many applications such as exercise tracking, physiotherapy and rehabilitation studies, entertainment, gait analysis, and fall detection.

6 | CONCLUSIONS AND FUTURE WORK

HAR is challenging because of the numerous ways that an activity can be represented. To accurately identify activities from the data collected using different sources, it is crucial to represent the activity properly. In addition, it is essential to minimize bias due to missing data, faulty sensors, or measurement errors. This study proposes an approach, "HAR-MIL" which combines HAR and MIL concepts to overcome such issues and eliminate the limits of a traditional single-instance representation by bringing multi-instance representation. Unlike traditional HAR approaches, the proposed method can handle multiple sensor data simultaneously, with instances obtained from several sensors appearing together in a bag to define an activity. For the first time, two multi-instance algorithms (SimpleMI and MIWrapper) with different base learners (DT, SVM, NN, PART, KNN, RF, AB) were used in the field of wearable sensor-based HAR. It is the first study to explore the effects of sensor positions (left pocket, right

pocket, belt, upper arm, and wrist), sensor combinations (A + G + M + LA, A + G + M, A + G, A + M, and G + M), and the number of participants (single or multiple) on the MI representation of the relevant activity. The performance of the proposed approach was investigated in daily human activities, such as upstairs, downstairs, walking, standing, jogging, sitting, and biking.

In the experimental studies, the effectiveness of the proposed HAR-MIL method was demonstrated on 50 participant-based and sensor-position-based activity recognition datasets. The results showed that the proposed method is effective for wearable sensor-based HAR with high classification accuracy (99.32%). Furthermore, the results showed that our method outperformed the state-of-the-art methods by 10% on average on the same dataset.

In the future, the performance of MIL-based HAR can be investigated on a dataset, which includes specific human activities, such as transportation, sports, military, and hand-oriented activities. Moreover, the impact of smartphone orientation (i.e., landscape or portrait) on the human body can be explored in a MI manner. Like many of the existing HAR approaches, the proposed method is position-dependent. However, it requires a phone to be placed at some fixed location on the body. In daily living activities, people do not always keep their phones in a fixed location. Therefore, a MIL-based HAR model can be trained on the sensor datasets of the same activity for different phone locations at the same time. In future works, the HAR-MIL approach can be investigated by considering synchronized multi-activity recognition, activity transition recognition, and age-specific (i.e., young, adult, senior) HAR. Furthermore, the concept of data fusion can be combined with the HAR-MIL approach by taking advantage of its positive aspects.

ACKNOWLEDGEMENTS

We are grateful to Ege University Planning and Monitoring Coordination of Organizational Development and Directorate of Library and Documentation for their support in editing and proofreading service of this study.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Duygu Bagci Das  <https://orcid.org/0000-0003-4519-3531>

REFERENCES

- Akbari, A., Wu, J., Grimsley, R., & Jafari, R. (2018). *Hierarchical signal segmentation and classification for accurate activity recognition*. Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers (pp. 1596–1605). <https://doi.org/10.1145/3267305.3267528>
- Alemayoh, T. T., Lee, J. H., & Okamoto, S. (2021). New sensor data structuring for deeper feature extraction in human activity recognition. *Sensors*, 21, 2814. <https://doi.org/10.3390/s21082814>
- Attal, F., Mohammed, S., Dedabirshvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y. (2015). Physical human activity recognition using wearable sensors. *Sensors*, 15, 31314–31338. <https://doi.org/10.3390/s151229858>
- Bandyopadhyay, S., Ghosh, D., Mitra, R., & Zhao, Z. (2015). MBSTAR: Multiple instance learning for predicting specific functional binding sites in microRNA targets. *Scientific Reports*, 5, 8004. <https://doi.org/10.1038/srep08004>
- Bhat, G., Tran, N., Shill, H., & Ogras, U. Y. (2020). w-HAR: An activity recognition dataset and framework using low-power wearable devices. *Sensors*, 20, 1–26. <https://doi.org/10.3390/s20185356>
- Bianchi, V., Bassoli, M., Lombardo, G., Fornacciari, P., Mordonini, M., & Munari, I. D. (2019). IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment. *IEEE Internet of Things Journal*, 6, 8553–8562. <https://doi.org/10.1109/JIOT.2019.2920283>
- Bijalwan, V., Semwal, V. B., & Gupta, V. (2022). Wearable sensor-based pattern mining for human activity recognition: Deep learning approach. *Industrial Robot: The International Journal of Robotics Research and Application*, 49, 21–33. <https://doi.org/10.1108/IR-09-2020-0187>
- Bijalwan, V., Semwal, V. B., & Mandal, T. K. (2021). Fusion of multi-sensor-based biomechanical gait analysis using vision and wearable sensor. *IEEE Sensors Journal*, 21, 14213–14220. <https://doi.org/10.1109/JSEN.2021.3066473>
- Birant, D., & Yalniz, K. (2022). Animal activity recognition from sensor data using ensemble learning. In *Emerging trends in IoT and integration with data science, cloud computing, and big data analytics* (pp. 165–180). IGI Global. <https://doi.org/10.4018/978-1-7998-4186-9.ch009>
- Birant, K. U. (2021). Multi-view rank-based random forest: A new algorithm for prediction in esports. *Expert Systems*, 39, e12857. <https://doi.org/10.1111/exsy.12857>
- Braganca, H., Colonna, J. G., Lima, W. S., & Souto, E. (2020). A smartphone lightweight method for human activity recognition based on information theory. *Sensors*, 20, 1–28. <https://doi.org/10.3390/s20071856>
- Buenaventura, C. V. S., & Tiglao, N. M. C. (2017). *Basic human activity recognition based on sensor fusion in smartphones*. Proceedings of the 1st workshop on protocols, applications and platforms for enhanced living environments (pp. 1182–1185). <https://doi.org/10.23919/INM.2017.7987459>
- Campanella, G., Silva, V. W. K., & Fuchs, T. J. (2018). Terabyte-scale deep multiple instance learning for classification and localization in pathology. *arXiv*: 1805.06983.
- Carpineti, C., Lomonaco, V., Bedogni, L., Felice, M. D., & Bononi, L. (2018). *Custom dual transportation mode detection by smartphone devices exploiting sensor diversity*. Proceedings of the international conference on pervasive computing and communications workshops (pp. 1–6).
- Casado, F. E., Lema, D., Criado, M. F., Iglesias, R., Regueiro, C. V., & Barro, S. (2022). Concept drift detection and adaptation for federated and continual learning. *Multimedia Tools and Applications*, 81, 3397–3419. <https://doi.org/10.1007/s11042-021-11219-x>

- Catal, C., Tufekci, S., Pirmit, E., & Kocabag, G. (2015). On the use of ensemble of classifiers for accelerometer-based activity recognition. *Applied Soft Computing*, 37, 1018–1022. <https://doi.org/10.1016/j.asoc.2015.01.025>
- Chakraborty, B., Nakano, K., Tokoi, Y., & Hashimoto, T. (2019). An approach for designing low cost deep neural network based biometric authentication model for smartphone user. Proceedings of the international technology conference of IEEE region (TENCON) (pp. 772–777). <https://doi.org/10.1109/TENCON.2019.8929241>
- Das, D. B., & Birant, D. (2020). Classification of hand-based and non-hand-based physical activities using wearable sensors. Proceedings of the 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT) (pp. 1–5). <https://doi.org/10.1109/ISMSIT50672.2020.9255171>
- Daş, D. B., & Birant, D. (2021). Ordered physical human activity recognition based on ordinal classification. *Turkish Journal of Electrical Engineering & Computer Science*, 29, 2416–2436. <https://doi.org/10.3906/elk-2010-75>
- Dexun, J., Peijun, M., Xiaohong, S., & Tiantian, W. (2013). Detection and refactoring of bad smell caused by large scale. *International Journal of Software Engineering & Applications*, 4, 1–13. <https://doi.org/10.5121/ijsea.2013.4501>
- Dong, L. (2006). A comparison of multi-instance learning algorithms. University of Waikato, Hamilton, New Zealand.
- Ehatisham-Ul-Haq, M., Azam, M. A., Amin, Y., & Naeem, U. (2020). C2FHAR: Coarse-to-fine human activity recognition with behavioral context modeling using smart inertial sensors. *IEEE Access*, 8, 7731–7747. <https://doi.org/10.1109/ACCESS.2020.2964237>
- Forough, J., & Momtazi, S. (2022). Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach. *Expert Systems*, 39, e12795. <https://doi.org/10.1111/exsy.12795>
- Frank, E., & Xu, X. (2003). Applying propositional learning algorithms to multi-instance data. University of Waikato, Hamilton, New Zealand.
- Gao, Z., Liu, D., Huang, K., & Huang, Y. (2019). Context-aware human activity and smartphone position-mining with motion sensors. *Remote Sensing*, 11, 2531. <https://doi.org/10.3390/rs11212531>
- Garcia-Ceja, E., Galván-Tejada, C. E., & Brena, R. (2018). Multi-view stacking for activity recognition with sound and accelerometer data. *Information Fusion*, 40, 45–56. <https://doi.org/10.1016/j.inffus.2017.06.004>
- Guan, X., Raich, R., & Wong, W.-K. (2016). Efficient multi-instance learning for activity recognition from time series data using an auto-regressive hidden Markov model. Proceedings of the 33rd international conference on machine learning (Vol. 48, pp. 2330–2339).
- Gurov, N., Khan, A., Hussain, R., & Khattak, A. (2019). Human activity recognition using deep models and its analysis from domain adaptation perspective. In M. Mazzara, J. M. Bruel, B. Meyer, & A. Petrenko (Eds.), *Software technology: Methods and tools. Tools 2019. Lecture notes in computer science* (pp. 189–202). Springer. https://doi.org/10.1007/978-3-030-29852-4_15
- Jain, R., Semwal, V. B., & Kaushik, P. (2021). Deep ensemble learning approach for lower extremity activities recognition using wearable sensors. *Expert Systems*, 39, e12743. <https://doi.org/10.1111/exsy.12743>
- Kantoch, E. (2017). Human activity recognition for physical rehabilitation using wearable sensors fusion and artificial neural networks. *Computing in Cardiology*, 44, 1–4. <https://doi.org/10.22489/CinC.2017.296-332>
- Kerdjidj, O., Ramzan, N., Ghanem, K., Amira, A., & Chouireb, F. (2020). Fall detection and human activity classification using wearable sensors and compressed sensing. *Journal of Ambient Intelligence and Humanized Computing*, 11, 349–361. <https://doi.org/10.1007/s12652-019-01214-4>
- Li, X., He, Y., & Jing, X. (2019). A survey of deep learning-based human activity recognition in radar. *Remote Sensing*, 11, 1068. <https://doi.org/10.3390/rs11091068>
- Lima, W. S., Bragança, H. L., & Souto, E. J. (2021). NOHAR—Novelty discrete data stream for human activity recognition based on smartphones with inertial sensors. *Expert Systems with Applications*, 166, 114093. <https://doi.org/10.1016/j.eswa.2020.114093>
- Lin, C.-Y., & Marculescu, R. (2020). Model personalization for human activity recognition. Proceedings of the international conference on pervasive computing and communications workshops (PerCom workshops) (pp. 1–7). <https://doi.org/10.1109/PerComWorkshops48775.2020.9156229>
- Ma, Y., Cui, C., Nie, X., Yang, G., Shaheed, K., & Yin, Y. (2019). Pre-course student performance prediction with multi-instance multi-label learning. *Science China Information Sciences*, 62, 29101. <https://doi.org/10.1007/s11432-017-9371-y>
- Machado, I. P., Gomes, A. L., Gamboa, H., Paixão, V., & Costa, R. M. (2015). Human activity data discovery from triaxial accelerometer sensor: Non-supervised learning sensitivity to feature extraction parametrization. *Information Processing & Management*, 51, 204–214. <https://doi.org/10.1016/j.ipm.2014.07.008>
- Menhour, I., Abidine, M. B., & Fergani, B. (2018). A new framework using PCA, LDA and KNN-SVM to activity recognition based smartphone's sensors. Proceedings of the 2018 6th international conference on multimedia computing and systems (ICMCS). <https://doi.org/10.1109/ICMCS.2018.8525987>
- Mohamed, R., Zainudin, M. N. S., Sulaiman, M. N., Perumal, T., & Mustapha, N. (2018). Multi-label classification for physical activity recognition from various accelerometer sensor positions. *Journal of Information and Communication Technology*, 17, 209–231. <https://doi.org/10.32890/jict2018.17.2.3>
- Mosabbeb, E. A., Cabral, R., la Torre, F. D., & Fathy, M. (2015). Multi-label discriminative weakly-supervised human activity recognition and localization. In D. Cremers, I. Reid, H. Saito, & M. H. Yang (Eds.), *Computer Vision—ACCV 2014* (pp. 241–258). Springer. https://doi.org/10.1007/978-3-319-16814-2_16
- Mukherjee, A., Misra, S., Mangulkar, P., Rajarajan, M., & Rahulamathan, Y. (2017). SmartARM: A smartphone-based group activity recognition and monitoring scheme for military applications. Proceedings of the IEEE international conference on advanced networks and telecommunications systems (ANTS). <https://doi.org/10.1109/ANTS.2017.8384149>
- Nafea, O., Abdul, W., Muhammad, G., & Alsulaiman, M. (2021). Sensor-based human activity recognition with spatio-temporal deep learning. *Sensors*, 21, 2141. <https://doi.org/10.3390/s21062141>
- Nair, R., Ragab, M., Mujallid, O. A., Mohammad, K. A., Mansour, R. F., & Viju, G. K. (2022). Impact of wireless sensor data mining with hybrid deep learning for human activity recognition. *Wireless Communications and Mobile Computing*, 2022, 9457536. <https://doi.org/10.1155/2022/9457536>
- Natarajasan, D., & Govindarajan, M. (2016). Human physical activity recognition using smartphone sensors. *International Journal of Computer Science and Telecommunications*, 7, 26–31.
- Neyestani, M. S., Sajedi, H., & Tawosi, V. (2018). Identification of human activities using mobile phone sensors based on stacking learning algorithm. Proceedings of the third international conference on intelligent decision science.
- Oded, M., & Tomás, L.-P. (1997). A framework for multiple-instance learning. In M. Jordan, M. Kearns, & S. Solla (Eds.), *Advances in neural information processing systems* (Vol. 10). MIT Press. <https://proceedings.neurips.cc/paper/1997/file/82965d4ed8150294d4330ace00821d77-Paper.pdf>
- Ozcan, K., & Velipasalar, S. (2016). Wearable camera- and accelerometer-based fall detection on portable devices. *IEEE Embedded Systems Letters*, 8, 6–9. <https://doi.org/10.1109/LES.2015.2487241>

- Panahandeh, G., Mohammadiha, N., Leijon, A., & Handel, P. (2013). Continuous hidden markov model for pedestrian activity classification and gait analysis. *IEEE Transactions on Instrumentation and Measurement*, 62, 1073–1083. <https://doi.org/10.1109/TIM.2012.2236792>
- Patil, P., Kumar, K. S., Gaud, N., & Semwal, V. B. (2019). Clinical human gait classification: Extreme learning machine approach. Proceedings of the 1st international conference on advances in science, engineering and robotics technology (pp. 1–6). <https://doi.org/10.1109/ICASERT.2019.8934463>
- Prabono, A. G., Yahya, B. N., & Lee, S.-L. (2022). Multiple-instance domain adaptation for cost-effective sensor-based human activity recognition. *Future Generation Computer Systems*, 133, 114–123. <https://doi.org/10.1016/j.future.2022.03.006>
- Quellec, G., Cazuguel, G., Cochener, B., & Lamard, M. (2017). Multiple-instance learning for medical image and video analysis. *IEEE Reviews in Biomedical Engineering*, 10, 213–234. <https://doi.org/10.1109/RBME.2017.2651164>
- Saha, J., Ghosh, D., Chowdhury, C., & Bandyopadhyay, S. (2021). Smart handheld based human activity recognition using multiple instance multiple label learning. *Wireless Personal Communications*, 117, 923–943. <https://doi.org/10.1007/s11277-020-07903-0>
- Shakya, S. R., Zhang, C., & Zhou, Z. (2018). Comparative study of machine learning and deep learning architecture for human activity recognition using accelerometer data. *International Journal of Machine Learning and Computing*, 8, 577–582. <https://doi.org/10.18178/ijmlc.2018.8.6.748>
- Sharma, M., Patel, S., & Acharya, U. R. (2021). Expert system for detection of congestive heart failure using optimal wavelet and heart rate variability signals for wireless cloud-based environment. *Expert Systems*, e12903. <https://doi.org/10.1111/exsy.12903>
- Shoaib, M., Bosch, S., Incel, O., Scholten, H., & Havinga, P. (2014). Fusion of smartphone motion sensors for physical activity recognition. *Sensors*, 14, 10146–10176. <https://doi.org/10.3390/s140610146>
- Siirtola, P., & Röning, J. (2021). Context-aware incremental learning-based method for personalized human activity recognition. *Journal of Ambient Intelligence and Humanized Computing*, 12, 10499–10513. <https://doi.org/10.1007/s12652-020-02808-z>
- Stiborek, J., Pevný, T., & Rehák, M. (2018). Multiple instance learning for malware classification. *Expert Systems with Applications*, 93, 346–357. <https://doi.org/10.1016/j.eswa.2017.10.036>
- Stikic, M., & Schiele, B. (2009). Activity recognition from sparsely labeled data using multi-instance learning. In T. Choudhury, A. Quigley, T. Strang, & K. Sugiyama (Eds.), *LoCA 2009: Location and context awareness* (pp. 156–173). Springer. https://doi.org/10.1007/978-3-642-01721-6_10
- Supriyatna, T. B., Nasution, S. M., & Nugraheni, R. A. (2019). Human activity recognition using support vector machine for automatic security system. *Journal of Physics: Conference Series*, 1192, 012017. <https://doi.org/10.1088/1742-6596/1192/1/012017>
- Tao, W., Chen, H., Moniruzzaman, M., Leu, M. C., Yi, Z., & Qin, R. (2021). Attention-based sensor fusion for human activity recognition using imu signals. [arXiv:abs/2112.11224](https://arxiv.org/abs/2112.11224).
- Thakur, S. S., Poddar, P., & Roy, R. B. (2022). Real-time prediction of smoking activity using machine learning based multi-class classification model. *Multimedia Tools and Applications*, 81, 14529–14551. <https://doi.org/10.1007/s11042-022-12349>
- Tian, Y., Wang, X., Chen, L., & Liu, Z. (2019). Wearable sensor-based human activity recognition via two-layer diversity-enhanced multiclassifier recognition method. *Sensors*, 19, 2039. <https://doi.org/10.3390/s19092039>
- Tiwari, S., Jain, A., Alkwai, C. L. M., Dafhalla, A. K. Y., & Hamad, S. A. S. (2021). Machine learning-based model for prediction of power consumption in smart grid- smart way towards smart city. *Expert Systems*, 39, e12832. <https://doi.org/10.1111/exsy.12832>
- Voicu, R.-A., Dobre, C., Bajenaru, L., & Ciobanu, R.-I. (2019). Human physical activity recognition using smartphone sensors. *Sensors*, 19, 1–18. <https://doi.org/10.3390/s19030458>
- Wang, Y., Cang, S., & Yu, H. (2019). A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137, 167–190. <https://doi.org/10.1016/j.eswa.2019.04.057>
- Weiss, G. M., Yoneda, K., & Hayajneh, T. (2019). Smartphone and smartwatch-based biometrics using activities of daily living. *IEEE Access*, 7, 133190–133202. <https://doi.org/10.1109/ACCESS.2019.2940729>
- Witten, I., Frank, E., & Hall, M. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.
- Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, 1–37. <https://doi.org/10.1007/s10115-007-0114-2>
- Yan, H., Zhang, Y., Wang, Y., & Xu, K. (2020). WiAct: A passive WiFi-based human activity recognition system. *IEEE Sensors Journal*, 20, 296–305. <https://doi.org/10.1109/JSEN.2019.2938245>
- Yassine, A., Singh, S., & Alamri, A. (2017). Mining human activity patterns from smart home big data for health care applications. *IEEE Access*, 5, 13131–13141. <https://doi.org/10.1109/ACCESS.2017.2719921>
- Yi, Y., & Lin, M. (2016). Human action recognition with graph-based multiple-instance learning. *Pattern Recognition*, 53, 148–162. <https://doi.org/10.1016/j.patcog.2015.11.022>
- Yousefi, M., Krzyżak, A., & Suen, C. Y. (2018). Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. *Computers in Biology and Medicine*, 96, 283–293. <https://doi.org/10.1016/j.combiomed.2018.04.004>
- Zainudin, M., Mohamed, R., Sulaiman, N., Perumal, T., Mustapha, N., & Naziri, A. S. A. (2017). *Multi-label classification using label combination to recognize human activity based on various sensor positions*. Proceedings of the 6th international conference on computing and informatics (pp. 669–674).
- Zhang, X., Wong, Y., Kankanhalli, M. S., & Geng, W. (2019). Hierarchical multi-view aggregation network for sensor-based human activity recognition. *PLoS One*, 14, e0221390. <https://doi.org/10.1371/journal.pone.0221390>
- Zokas, B., & Lukosevicius, M. (2018). *Human sport activities recognition and registration from portable device*. Proceedings of the symposium for young scientists in technology, engineering and mathematics (pp. 61–65).

AUTHOR BIOGRAPHIES

Duygu Bagci Das is a lecturer in the Computer Programming Department at Ege University, Turkey. She is about to complete her Ph.D. in the Computer Engineering Department at Dokuz Eylül University, Turkey. Her research interests are in sensor-based intelligent systems in the field of engineering, computer science, and healthcare. She has several publications indexed in SCI-E, ESCI, and SCOPUS.

Derya Birant received her B.Sc., M.Sc., and Ph.D. in Computer Engineering from the Dokuz Eylul University, Turkey, in 2000, 2002, and 2006, respectively. Since 2017, she has been an associate professor at the same place. She was a visiting lecturer at the South East European University in 2006 and Ege University between 2010 and 2012. Dr. Birant has supervised more than 30 Ph.D. and M.Sc. students. She has been involved in more than 20 long-term interdisciplinary R&D projects on data science. She has published over 120 research papers in journals, books, and conferences. Her research interests include machine learning and data mining.

How to cite this article: Bagci Das, D., & Birant, D. (2023). Human activity recognition based on multi-instance learning. *Expert Systems*, e13256. <https://doi.org/10.1111/exsy.13256>