Information Fusion 104 (2024) 102197

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Situation identification in smart wearable computing systems based on machine learning and Context Space Theory

Giuseppe D'Aniello [a,*], Matteo Gaeta [a], Raffaele Gravina [b], Qimeng Li [b], Zia Ur Rehman [a], Giancarlo Fortino [b]

[a] DIEM, University of Salerno, Via Giovanni Paolo II, Fisciano, 84084, Italy
[b] DIMES, University of Calabria, Via P. Bucci, Rende, 87036, Italy

ARTICLE INFO

Keywords:
Situation awareness
Activity recognition
Context Space Theory
Information fusion

ABSTRACT

Wearable devices and smart sensors are increasingly adopted to monitor the behaviors of human and artificial agents. Many applications rely on the capability of such devices to recognize daily life activities performed by the monitored users in order to tailor their behaviors with respect to the occurring situations. Despite the constant evolution of smart sensing technologies and the numerous research in this field, an accurate recognition of in-the-wild situations still represents an open research challenge. This work proposes a novel approach for situation identification capable of recognizing the activities and the situations in which they occur in different environments and behavioral contexts, processing data acquired by wearable and environmental sensors.

An architecture of a situation-aware wearable computing system is proposed, inspired by Endsley's situation-awareness model, consisting of a two-step approach for situation identification. The approach first identifies the daily life activities via a learning-based technique. Simultaneously, the context in which the activities are performed is recognized using Context Space Theory. Finally, the fusion between the context state and the activities allows identifying the complex situations in which the user is acting. The knowledge regarding the situations forms the basis on which novel and smarter applications can be realized.

The approach has been evaluated on the ExtraSensory public dataset and compared with state-of-the-art techniques, achieving an accuracy of 96% for the recognition of situations and with significantly low computational time, demonstrating the efficacy of the two-step situation identification approach.

1. Introduction and motivations

The increasing popularity of wearable devices can be attributed to their capability to unobtrusively and pervasively monitor users. This capability is crucial for various applications, such as telemedicine, safety, and security [2]. These devices are integral components of a constantly ever-growing 'smart environment', which encompasses other environmental sensors, smart assistants, and Internet of Things (IoT) devices. Collectively, these sensing devices generate vast amounts of data about users and their surroundings, which can be harnessed by applications to offer smarter functionalities.

To enhance the capabilities of wearable computing systems (WCS), they should possess the ability to monitor users based on their current situations. Therefore, these systems should be capable of perceiving the ongoing situations. Situation awareness (SA) refers to the capacity of both human and artificial agents to comprehend the events occurring in their surroundings and respond appropriately [3]. It is a crucial feature and a prerequisite for cognitive systems. SA has been formally defined by Endsley as the "perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [4]. Endsley's model of SA encompasses three distinct levels:

- *Perception*: Acquiring data regarding the status of the relevant elements of the environment.
- *Comprehension*: Understanding the meaning of the perceived data with respect to the operations and goals.
- *Projection*: Predicting and tracking the future evolution of the recognized elements' states.

A Situation-Aware Wearable Computing System (SA-WCS) is a wearable system that can perceive, understand, and adapt its behavior based on the situation it is monitoring [5]. This includes recognizing individual activities and considering factors like the environment and

* Corresponding author.
E-mail address: gidaniello@unisa.it (G. D'Aniello).

https://doi.org/10.1016/j.inffus.2023.102197
Received 17 May 2023; Received in revised form 12 December 2023; Accepted 13 December 2023
Available online 16 December 2023

user attributes, such as age, identity, location, time, and temperature. The significance and risks associated with an activity depend on the surrounding conditions. For instance, the impact of a fall can vary based on whether it involves an elderly person or a child, or if it happens in a park or a house.

The growing demand for SA-WCS systems capable of exhibiting intelligent behavior has led to increased interest in the field. In particular, these systems have the potential to enable cost-effective, pervasive, and widespread monitoring of situations involving users without the need for complex sensors or environmental setups, such as cameras.

In line with this research direction, this study introduces an approach for situation identification in smart wearable computing systems, defined within the framework of a situation-aware wearable computing system (SA-WCS) reference architecture. The architecture is defined according to Endsley's model of Situation Awareness. Particularly, within the overall architecture, the proposed situation identification approach covers the first two levels of Endsley's model, namely perception and comprehension, with the ultimate goal of identifying daily-life situations involving human beings.

The recognition of the context wherein the human activity is performed is of paramount importance for the identification of the situations. The SA-WCS exploits a context recognition approach based on Context Space Theory (CST) [6] to represent the state of the context in which the users perform their activities. Such a context is then fused with human activities recognized via a machine-learning approach in order to identify the situations.

The capability of recognizing the context using a small subset of the available features makes the approach quite robust and capable of identifying activities in-the-wild, as demonstrated by an evaluation conducted on the ExtraSensory public available dataset [7]. Such a dataset contains activities and situations gathered in real-world contexts in uncontrolled experiments.

The contributions of this paper are summarized as follows:

1. we introduce the situation-aware WCS and its importance in various applications,
2. we highlight the need for the unobtrusive, pervasive, and simple identification of activities and complex situations in in-the-wild contexts, preferably using simple sensors, without requiring cameras or other intrusive devices, even in dynamic and complex environments.
3. we propose a reference architecture for situation-aware WCS;
4. we propose an approach for situation-aware activity recognition that integrates and fuses machine learning techniques for activity recognition with an expert-based approach for context identification, enabling the identification of situations in the wild.

The rest of the article is organized as follows: Section 2 discusses the state-of-the-art situation-aware wearable computing systems and situation-aware activity recognition. Section 3 describes the proposed situation identification approach. Section 4 presents the experiments and discusses the results. Finally, Section 5 concludes the paper and draws future developments.

2. Related work

SA systems rely on a variety of sensors to perceive the environment and understand the situation. Technological advancements have led to the development of numerous sensors that have found applications in SA systems. The specific set of sensors adopted in an SA system strongly depends on the application domains. For instance, in Ambient Intelligence (AmI), SA systems prioritize high-level conceptual information over low-level sensor data. To achieve this objective, SA systems need to collect, combine, process, and abstract available sensor data from the smart environment.

Situation-aware sensor-based wearable computing systems can be considered a subset of a broader category of situation-aware sensor-based systems that mainly uses wearable devices to collect data.

In Section 2, we briefly review the literature on situation-aware wearable computing systems, with a specific focus on systems capable of identifying situations involving human users. Subsequently, in Section 2.2, we examine the major advancements in human activity recognition techniques, particularly those utilizing numerical sensor data, especially from wearable sensors.

2.1. Situation-aware wearable computing systems

Despite the growing interest in sensorized wearable devices for activity recognition, the literature on situation awareness in wearable computing systems is still not well consolidated [5].

At the time of writing, Schiele et al.'s groundbreaking work [8] anticipated many of the requirements for future situation-aware applications and introduced relevant technologies. Their work does not align with Endsely's model, but they acknowledged the necessity for a situation-aware system to encompass sensing (perception), reasoning (comprehension), and prediction (projection). The authors introduced the concept of 'scene', which is a collection of events that enables user context inference and illustrates the relationships between them. This concept corresponds to the notion of 'situation'. While the term 'situation awareness' sporadically appears in earlier works on wearable systems, it is only within the last decade that research has begun to define this domain more clearly and conduct more systematic studies. For instance, Don et al. [9] proposed a situation-aware patient monitoring system based on Endsley's model. Ardito et al. [10] introduced an additional level to Endsley's model, namely the 'Control level', positioned between 'Perception' and 'Comprehension', to enhance system reliability and safeguard the data input from tampering and spoofing. Some works, such as [11], adopt the definition of a situation proposed by Ye et al. [12], which considers the situation as an abstract representation of the state of the environment derived from the context. In [13], a situation is defined as a combination of human activities and the location where they are performed. In [14], a situation is described as a set of contexts related to individual perception modalities, and a similar definition is found in [15], where a situation concerns the identification of relevant contexts and the significant relationships between them.

Environmental data collected from the user's surroundings play a crucial role in enhancing our understanding of the situation, as they serve as the primary source of contextual information. For instance, the user's location can be roughly determined (e.g., indoors or outdoors) [16] or precisely measured outdoors using GPS [14,17] and indoors using Bluetooth beacons [13]. Other relevant data include temporal constraints such as the time of day (morning, evening, night) [16–18], social cues for context awareness and situational understanding [19,20], and additional information about the user's or patient's health status from electronic medical records (EMR) and medical domain knowledge [9,10,15,19–21]. Rule-based reasoning engines leverage the latter for comprehension in health monitoring and situation awareness applications.

The perception of data is another critical aspect in situation-aware wearable computing systems. Sensors enable perception by collecting data about users' behaviors and their surroundings. In wearable computing systems designed for situation awareness (SA-WCS), the most common sensors include accelerometers, gyroscopes, ECG sensors, and GPS. Additionally, light sensors and microphones, often integrated into the user's personal mobile device, are used to gather environmental and social contextual information. As smart wearables become more widespread, we can expect the adoption of additional sensors, such as cameras embedded in glasses, pressure sensors integrated into clothing and footwear, and ultra-wideband radars, to further enhance situation awareness.

An essential task of any situation-aware WCS is the identification and understanding of the situation. Situation identification techniques can be categorized into data-driven and knowledge-driven methods. Data-driven approaches employ machine learning techniques for recognizing situations [1,13,18]. In contrast, knowledge-driven approaches focus on formal logic, ontologies, and semantic reasoning to unveil relationships between entities and identify the situation [9,14,15,17,21,21].

Another crucial capability of situation-aware systems is predicting and tracking the evolution of a situation during the projection phase. In [11], a probabilistic model is presented for predicting the next actions based on the anomalous behavior of users with cognitive impairments. In [21], an approach for predictive analytics and preventive health services is proposed, leveraging a semantic reasoning engine for clinical decision support. However, the paper lacks detailed information on this. Contexta-CARE [17] is a multi-device system for assisted living, comprising environmental sensors and a wearable device. It supports the prediction of dangerous situations by analyzing sequences of events.

2.2. Human activity recognition

Human Activity Recognition (HAR) aims at the identification of human activity [22,23]. In the last years, wearable sensor-based HAR has gained popularity, especially in smart homes [24], healthcare [25,26], cultural heritage [27], smart industry and smart home [28], sports [29], human–robot interaction [30], and more. HAR techniques typically involve data preprocessing and action recognition with supervised, unsupervised, and semi-supervised learning methods. Supervised learning has been widely used for wearable sensor-based HAR; however, it requires labeled datasets that may be unavailable or expensive. Some supervised learning approaches used for HAR include Support Vector Machines (SVM) [31], Artificial Neural Networks (ANN) [32], K-Nearest Neighbors (KNN), functional networks and granular functional networks [33], random forests [34], and others. Deep learning (DL) classification has recently been employed for HAR using supervised and unsupervised approaches. Common DL models include Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM) networks, and autoencoders. A review of DL methods for HAR is provided in [35]. Among unsupervised learning methods for HAR [36], Gaussian Mixture Models (GMM) are popular probabilistic clustering methods [37]. K-means clustering has also been employed for HAR [38], although the K-means algorithm can sometimes converge very slowly with incorrect initialization [39]. Hidden Markov Models have been used for human activity recognition [40], particularly for short-term and simple activities [22]. To address the limitations of supervised approaches, which require large amounts of labeled data, semi-supervised learning techniques can achieve good performance. Numerous classical semi-supervised learning methods, including generative models, self-training, co-training, transfer learning [41], Gaussian process models, and deep learning methods [42,43], have been employed recently for HAR.

3. Situation identification in smart wearable computing systems

This work presents a situation identification approach that utilizes machine learning to identify situations by combining human activities and contextual features. The approach is grounded on the reference architecture of Situation-aware Wearable Computing System (SA-WCS) depicted in Fig. 1. The primary goal of the SA-WCS architecture is to enhance applications and agents by providing them with situation awareness, enabling more intelligent behaviors. This architecture is inspired by Endsley's model [4] and is structured around three key levels of situation awareness: (L1) perception, (L2) comprehension, and (L3) projection, as depicted in Fig. 1.

The situation identification approach consists of two main phases corresponding to the perception and comprehension levels of SA-WCS, illustrated with solid lines in Fig. 1. The identified situation is then used by the projection phase (depicted with a gray dotted line in Fig. 1) to predict the future development of the situation. Applications can adjust their behavior based on the recognized situation. Specifically, the comprehension phase focuses on recognizing situations by merging contextual information (identified by the context recognition module) with human activities (identified by the high-level perception module using machine learning techniques) through the analysis of sensor data from common devices like smartphones and smartwatches.

The SA-WCS operates within a smart environment, which includes wearable devices, ambient sensors, IoT devices, and other information sources. Situation identification is achieved through the steps illustrated in Fig. 1. Specifically, the SA-WCS encompasses the following levels and functions:

- **L0 Sensing**: sensor life-cycle management, data acquisition, data cleaning, and storage.
- **L1 Perception**: data processing to identify activities, objects, events and context state. This level consists of three modules:
  - **$\mathbf{L1_{Low}}$ Low-Level Perception**: data pre-processing, segmentation and feature extraction.
  - **$\mathbf{L1_{High}}$ High-level Perception**: feature selection, object and activity recognition.
  - **$\mathbf{L1_{Context}}$ Context Recognition**: contextual data such as time, location, temperature, and other environmental features, are fused together and processed to create a representation of the context state using the Context Space Theory (CST) [6,44].
- **L2 Comprehension**: situation identification occurs through the fusion of activities recognized by the high-level perception module with the context state identified by the context recognition module.
- **L3 Projection**: tracking the possible evolution of the situations via predictive models.

The following sections offer a detailed explanation of each level of the SA-WCS, with particular emphasis on the perception and comprehension levels for the identification of situations.

3.1. L0: Sensing

The sensing phase manages the data acquisition process, which involves data cleaning and data storage activities. Acquiring real-time data from wearable devices and environmental sensors can be challenging, especially when it needs to be done in various contexts and environments, such as at home, in the office, in a car, and so on. The SA-WCS architecture considers a set $S = \{s_1, s_2, \dots, s_n\}$ of n sensors. From each sensor $s_i \in S$, a data stream Δ_i, with $i = 1 \dots n$ is acquired. Each data stream consists of values collected by the corresponding sensor at specific time intervals in accordance with the sensor's sampling frequency. Specifically:

$$\Delta_i = [\delta_i^0, \delta_i^1, \dots, \delta_i^l, \dots] \quad (1)$$

where δ_i^l is the value gathered by the sensor s_i at the time instant $t = l$.

In the sensing phase, each data stream, denoted as Δ_i, undergoes preprocessing, including data cleaning, redundancy removal, noise reduction, and more. If necessary, the data stream can also be resampled at a lower frequency to align with the subsequent processing tasks of the perception phase.

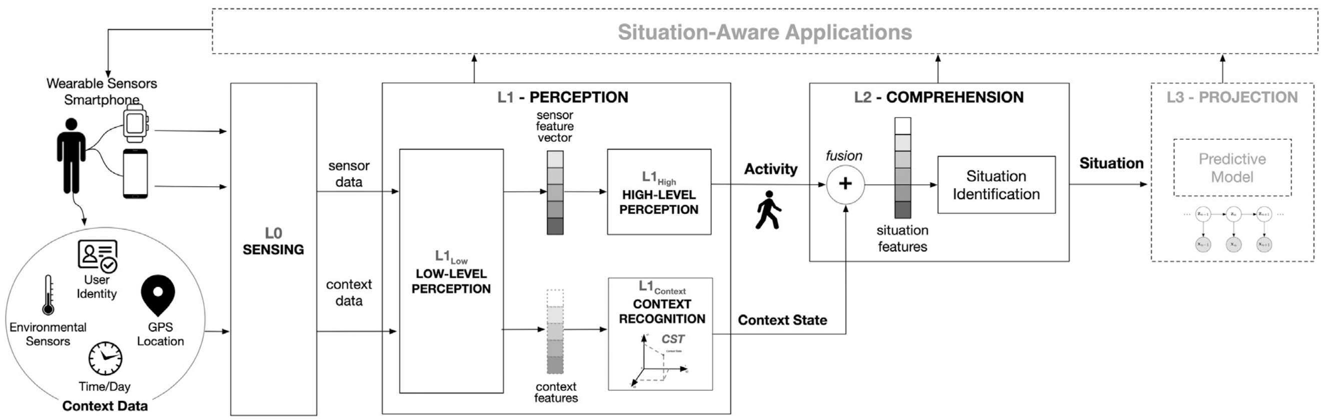

Fig. 1. Architecture of the Situation-aware Wearable Computing System.

3.2. $L1_{low}$: Low-level perception

The main task of the low-level perception phase involves extracting features from sensor data, intended for activity recognition, as well as from contextual data used to identify different context states. Features are obtained by processing the data streams within the set $\mathbf{\Delta} = \{\Delta_1, \Delta_2, \ldots, \Delta_n\}$. A set of features $\Phi = [\phi_1, \ldots, \phi_m]$ is derived from the processing and fusion of the data streams in $\mathbf{\Delta}$. To extract these features, it may be necessary to segment the data stream into time windows. Therefore, for each sensor data stream Δ_i, we can define the following data windows:

$$\begin{aligned} W^j_{\Delta_i}[\cdot] &= \Delta_i[j * size, ((j+1) * size) - 1] \\ &= [\delta_i^{j*size}, \delta_i^{(j*size)+1}, \ldots, \delta_i^{((j+1)*size)-1}] \end{aligned} \tag{2}$$

where $W^j_{\Delta_i}[\cdot]$ is the jth window containing the data points extracted from the data stream Δ_i; its dimension is equal to $size$ (which represents a time interval). Therefore, the jth window contains the data points within the time interval $[j * size, ((j+1) * size) - 1]$.

The feature value $\phi_l(j)$, $l = 1, \ldots, m$ represents the value of the lth feature for the jth time interval. It is computed by processing the data points contained in the jth window $W^j_{\Delta_i}[\cdot]$:

$$\phi_l(j) = f_{\phi_l}(W^j_{\Delta_l}) \tag{3}$$

where f_{ϕ_l} is a transformation function that computes the feature $\phi_l(j)$. Some features can be computed by aggregating and processing data from multiple windows across different sensors.

3.3. $L1_{High}$: High-level perception

In the high-level perception phase, the set of m features $\Phi = \{\phi_1, \ldots, \phi_m\}$, extracted from raw sensor data in the previous phase, is further processed to identify complex activities. This is accomplished using a classifier that takes the most relevant features as input and outputs the user's activity.

An important preliminary step before classification is feature selection. This step helps reduce the complexity of the problem, resulting in decreased computational time and improved overall performance.

The feature selection task aims to produce a subset $F^k \subseteq \Phi$ ok k features that are considered the most important for the situation identification task. Various approaches, including statistical and learning-based methods, can be used for feature selection. The SA-WCS approach uses a supervised wrapper feature selection method, which employs a supervised model to evaluate different feature subsets and choose the most suitable one. Each subset trains a model, and its performance is assessed on a separate dataset. Wrapper methods are common in activity recognition because they often yield the best feature set for a specific model [45,46].

After selecting the best feature set F^k, the next step is recognizing human activity. This involves training a classifier that takes the vector F^k as input and outputs an activity $a \in A$, where $A = \{a_1, a_2, \ldots, a_r\}$ represents the set of the r human activities to be recognized.

The choice of the classifier in this phase depends on the type of data and activities to be identified. Section 4 outlines the process for selecting the optimal classifier in high-level perception. A similar process can be applied to configure the SA-WCS approach for other recognition problems involving different data.

3.4. $L1_{Context}$: Context recognition

The identification of a situation involves recognizing the context in which a specific activity or event occurs. While the terms "situation" and "context" are often mistakenly used interchangeably, they represent two distinct concepts, with "situation" offering a greater level of expressiveness [47]. Context pertains to environmental or user-related properties, such as the user's location, the time of day, and environmental factors like temperature and humidity. On the other hand, a situation is an abstract concept defined in relation to the user's objectives. A situation not only signifies the state of a relevant phenomenon but also entails an understanding of the problem it represents and the methods for addressing it [47].

Hence, understanding and defining the state of a situation can be achieved by interpreting the user's activities in the current context. This approach provides insight into what a user is doing and why. For instance, let us consider the activity of "sitting". This activity can give rise to various situations depending on the context in which it occurs. Sitting at home in the living room with the TV on might indicate the situation "Watching TV", whereas performing the same activity in the office, in front of a laptop, during office hours, could lead to the situation "Working in the office".

The $L1_{Context}$ module, located within the architecture's perception level, is tasked with identifying and computationally representing the context. This module processes the context features extracted by the $L1_{Low}$ level from the environmental sensors present in the smart environment. The $L1_{Context}$ module is based on the Context Space Theory (CST), utilized for representing the context that will later be fused with the recognized activity in the subsequent L2 level.

CST is a specification-based technique [12] proposed by Padovitz, Boytsov et al. in [6,44]. It employs a geometrical metaphor to provide a clear and insightful representation of context. In CST, context is viewed as a multidimensional space, with each axis referred to as a context attribute, denoted as x_i, which represents the key features of the context. The value on a context space axis represents a context attribute value, and a complete set of pertinent context attribute values at a specific time constitutes the context state X. This context state identifies a point within the context space.

It is possible to define another space linked to the context space, known as the situation space, which is designed to represent real-life situations from a user's perspective. Within this space, it is feasible to assess the context state by computing a confidence value within the range of [0;1]. This numerical value quantifies the confidence that a particular situation is occurring, based on the current context state and the user's assigned importance to each context attribute. The confidence level of a situation $S(X)$ is determined by the following equation:

$$S(X) = \sum_{i=1}^{N} w_i * contr_{Si}(x_i); contr_{Si}(x_i) = \begin{cases} a_1, & x_i \in I_{i,1} \\ a_2, & x_i \in I_{i,2} \\ \dots & \\ a_m, & x_i \in I_{i,m} \end{cases} \tag{4}$$

Where:

- $S(X)$ represents the confidence level of the situation S in the context state X;
- w_i represents the relative weight of the ith context attribute in the situation S;
- The function $contr_{Si}(x_i)$ measures the contribution of the ith context attribute value to the situation S. Usually, this function is a step function over a certain context attribute, and in this case, $I_{i,j}$ are the generic intervals related to the jth context attribute. Each interval is associated with a contribution value a_i to the situation, with $a_i \in [0, 1]$.

Hence, by establishing the context state X and the contribution functions $contr_{Si}(x_i)$, we delineate the subset of potential situations that could occur based on the present context state. The contribution of each context attribute $x_i \in X$ will serve as an additional feature in the classifier during the comprehension phase to identify the current situation.

3.5. L2: Comprehension

In the comprehension phase, the activity recognized from the $\mathrm{L1}_{\mathrm{High}}$ High-level perception module is combined with the context state identified by the $\mathrm{L1}_{\mathrm{Context}}$ module to recognize the situation.

Essentially, the primary objective of level L2 is to conduct situation identification. Situation identification involves interpreting and amalgamating various contextual elements and other pertinent information to derive situations that serve the user goal. In our architecture, situation identification integrates two primary categories of information: context states encompassing data related to the user's location, time of day, identity, environmental conditions such as temperature, etc., and human activities. Situation identification generally employs two main approaches [12]: specification-based, where expert knowledge is represented in logical rules, ontologies, or other forms of knowledge representation models; and learning-based, wherein machine learning and data mining techniques are utilized to explore relationships between sensor data, context, and situations.

The situation identification approach in the SA-WCS architecture can be defined as a hybrid approach. Specifically, a specification-based technique (Context Space Theory) is employed to identify the context, whereas machine learning techniques are utilized to recognize situations by combining context features with activities.

Specifically, a machine learning-based classifier is trained to distinguish between different situations. The input vector for the classifier is denoted as $V = [a, x_1, x_2, \dots, x_q]$ where $a \in A$ represents the activity recognized by the high-level perception phase, and $x_1, \dots, x_q \in X$ are the q context attributes identified by the context recognition module. The classifier's output is a situation $s \in S$, where $S = \{s1, s2, \dots, s_v\}$ represents the set of v possible situations that the SA-WCS system can identify.

3.6. Level 3: Projection

The final phase of SA-WCS is the projection level. The objective of the projection level is to anticipate and track the future states of situations. This step is crucial for enabling autonomous behaviors in wearable systems. An autonomous agent can make informed decisions about its next actions only if it can predict the future state of the situation arising from its actions and the current situation status. Consequently, the agent needs to assess the potential consequences of its actions on future situations and, based on this prediction, select the optimal strategy. Various techniques can be employed for this purpose, primarily relying on machine learning, Hidden Markov Models (HMM), and cognitive maps.

4. Experimental results and performance evaluation

The primary objective of the evaluation is to determine whether the proposed approach for situation identification, which follows a two-step process (first, identifying activities during the perception phase, and then, identifying situations during the comprehension phase), outperforms traditional machine learning techniques that employ a one-step approach to identify situations directly from sensor data. Specifically, the evaluation is structured in two steps:

1. The evaluation of the perception phase focuses on recognizing basic human activities.
2. The evaluation of the comprehension phase focuses on recognizing situations.

The evaluation is conducted using the Extrasensory dataset [7], which contains data related to the daily-life situations of 60 participants. This dataset, as described in Section 4.1, comprises information collected from the smartphones and smartwatches of each participant, along with contextual data. For the evaluation of the perception phase, our goal is to identify users' basic activities (sitting, standing, walking, lying down, running) using only smartphone and smartwatch data to assess the proposed approach's perception phase. To achieve this, we must determine the set F^k of the most effective sensor features and the optimal classifier for activity identification (see Section 4.2). We will test five different classifiers to identify the best-performing one, including random forest, kNN, naïve Bayes, decision tree, and artificial neural network. Furthermore, we will compare the performance of the SA-WCS perception phase with a similar approach, specifically the AAHCR method proposed by M. Ehatisham-ul-Haq and M. A. Azam in [45]. The AAHCR method represents one of the most recent state-of-the-art approaches for activity and situation identification applied to the Extrasensory dataset.

For the second phase of the evaluation, the Comprehension phase, our goal is to identify the situations in which the recognized activities occur. We consider the vector $V = [a, x_1, x_2, \dots, x_q]$ as the input for the comprehension phase, which includes the activity a and the contextual state X. The objective is to identify the situation $s \in S$. In this context, we will compare the performance of the comprehension phase with the AAHCR approach and traditional machine learning techniques. In the case of traditional machine learning techniques, the SA-WCS two-step approach (perception and comprehension) is compared with one-step classification techniques that directly classify situations from sensor features. This comparison aims to assess whether the proposed two-step architecture, inspired by Endsley's model, provides a competitive advantage in situation recognition.

Table 1 details the evaluation metrics used for the perception and comprehension tasks, which include weighted precision (PE), recall (RE), F1-score (F1), and accuracy.

The machine learning techniques were implemented in Python using the scikit-learn library.[1] We have released an implementation of

[1] https://scikit-learn.org/stable/.

Table 1
Performance evaluation metrics. $TP, TN, FP, and FN$ denote true positives, true negatives, false positives, and false negatives.

Performance metric	Definition
Precision (PR)	$PR = \frac{TP}{TP+FP}$
Recall (RE)	$RE = \frac{TP}{TP+FN}$
Accuracy (ACC)	$ACC = \frac{TP+TN}{TP+TN+FP+FN}$
F1-score (F1)	$F1 = 2 * \frac{PR*RE}{PR+RE}$

the SA-WCS situation identification approach is available on GitHub.[2] All the experiments were conducted on a workstation equipped with an Intel® Core™ i7-7600U CPU @ 2.80 GHz with 8.00 GB RAM and 64-bit operating system.

4.1. Dataset

Numerous research studies have defined methods for collecting sensor data and the associated datasets using various types of sensors, such as smartphone sensors, wearables, environmental sensors, and more, as referenced in [48–51]. Many of these datasets, however, were not gathered from real-world, in-the-wild scenarios; rather, they were obtained through controlled experiments.

An intriguing publicly available dataset that contains real-life situations with real-time data collected in the wild is the Extrasensory dataset, curated by Yonatan and Ellis [7]. This dataset includes information from 60 participants related to in-the-wild daily-life activities and utilizes diverse sensors, including accelerometers and gyroscopes from smartphones and smartwatches, GPS data, microphones, and phone status. An interesting feature of this dataset is its integration of smartphone and smartwatch data with contextual information such as location, time, phone status, and audio. This characteristic is vital for our approach as it allows us to demonstrate the advantages of context recognition in identifying situation-aware activities.

The dataset encompasses numerous situations formed by the combination of the user's activities (e.g., sitting, walking, running, lying down, standing) and contextual information. As outlined in the analysis presented in [45], some situations have a limited number of examples, and there are instances where context features and sensor readings contain a significant percentage of missing values.

Since our primary objective is to identify situations based on user activity patterns, we focus solely on activities and contexts with a substantial number of samples and temporally synchronized context labels for each selected activity, following the guidelines proposed in [45].

For this purpose, we consider the set A containing all the five basic activities (or body postures) available in the dataset, which are: standing, sitting, walking, running and lying down. Regarding situations, we selected a set S of the 14 most common situations in which the five basic activities are executed (e.g., walking indoor, shopping, etc.), reported in Table 2. The extrasensory dataset consists of 192 073 samples containing one of the 5 basic activities in A. From this dataset, in the preprocessing phase, we removed 12 599 samples (∼6.5% of the samples) because of missing sensor data such as audio not being available when the user was on call, location services being turned off by the user, some phone not having some sensors (e.g., iPhone does not have an air pressure sensor).

After removing samples with missing activity and situation labels, we obtained a refined dataset comprising 179,474 samples for activity recognition in the perception phase and 106,089 samples for situation identification, each labeled with one of the available situations in S.

[2] https://github.com/knowmis/SAWCS-Situation-Identification-in-Smart-Wearable-Computing-Systems.

Table 2
Physical activities recognized in Level 1 Perception and Situations recognized in Level 2 Comprehension extracted from the Extrasensory dataset.

Physical activity (Level 1)		Situation (Level 2)	
Code	Activity A	Code	Situation S
SIT	Sitting	SITC1	Surfing the internet
		SITC2	In a car
		SITC3	In a meeting
		SITC4	Watching TV
STN	Standing	STNC1	Indoor
		STNC2	Outdoor
LYD	Lying down	LYDC1	Sleeping
		LYDC2	Surfing the internet
		LYDC3	Watching TV
WLK	Walking	WLKC1	Indoor
		WLKC2	Outdoor
		WLKC3	Shopping
		WLKC4	Talking
RUN	Running	RUNC1	Exercise

The reduced number of samples for the situation identification task is because some samples with activity labels do not have corresponding situation labels in S.

4.2. Evaluation of the perception phase

4.2.1. Sensing and low-level perception

In the perception phase, SA-WCS utilizes sensor data to identify activities through traditional machine learning techniques. All 60 users in the dataset and the five activities listed in Table 2 are considered. For the Sensing step, we conducted preprocessing tasks, which included harmonizing activity names, data cleaning, eliminating rows with missing activity labels, merging and aligning user files, and conducting exploratory data analysis to identify feature correlations.

The Extrasensory dataset comprises smartphone sensor data acquired at 40 Hz and watch sensor data acquired at 25 Hz. The data is organized into non-overlapping 20-second windows denoted as $W^j_{\Delta_i}$. The authors of the dataset have previously published a set Φ containing 71 features extracted from these time windows. Specifically, 26 features are computed from the smartphone sensors, and 45 features are derived from the watch sensors. These features encompass various statistical measures such as energy, entropy, standard deviation, variance, third moment, arithmetic mean, median, skewness, first quartile, third quartile, maximum latency, mean of signal gradient, and minimum amplitude from all three axes of each sensor. All 71 of these features are initially considered as the output of the low-level perception phase, and a selection of the best features will be made within the high-level perception phase.

4.2.2. High-level perception

The initial step in the high-level perception phase involves determining the optimal set of features F^k from the entire feature set Φ. For the selection of the most suitable features in the extrasensory dataset, we employed a wrapper method using a recursive strategy. A wrapper method makes use of some learning scheme to appraise different subsets of features and removes the redundant features not vital for the classification task ahead [45].

We compared three distinct classifiers: RF, kNN, and SVM, all under 5-fold cross-validation settings, with different set of features of different size k, selecting the top k features (according to feature importance). Finally, the set of the top 26 features identified by the Random Forest has been selected as these achieves a good balance between accuracy and complexity, including both features from smartphone and smartwatch. Table 3 compares the three classifiers on this set of 26 features selected from the full feature set, with RF ultimately demonstrating the best performance.

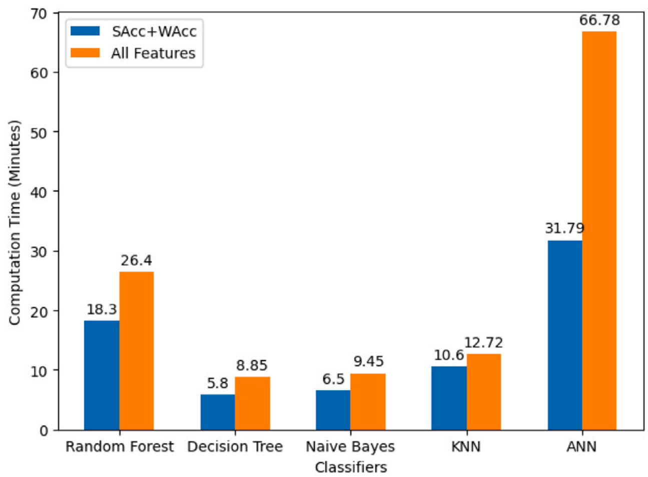

Fig. 2. Comparison of the training time for the five classifiers: (i) using the features SAcc+WAcc (blue); (ii) using all the features (orange).

Table 3
Validation metrics with 26 features selected by different algorithms.

Algorithm	Accuracy	Recall	Precision	F1-score
RF	**0.8573**	**0.6773**	**0.8356**	**0.7355**
SVM	0.8100	0.6197	0.7911	0.6777
kNN	0.8410	0.6613	0.8226	0.7199

Table 4
Feature selection. SAcc refers to the smartphone accelerometer features, WAcc to the smartwatch accelerometer features.

Set	Selected features
SAcc	Magnitude standard deviation (std), magnitude moment 4, magnitude 25-percentile, magnitude 75-percentile, magnitude entropy, magnitude log energy band 2, magnitude log energy band 4, mean-x, mean-y, mean-z, std-x, std-y, std-z
WAcc	Magnitude standard deviation (std), magnitude moment 4, magnitude 75-percentile, magnitude log energy band 2, magnitude log energy band 3, magnitude log energy band 4, mean-x, mean-y, mean-z, std-x, std-y, std-z, direction cosine similarity

The selected 26 features includes 13 features from smartphones and 13 from watches and is enumerated in Table 4. These are categorized into two groups: Smartphone accelerometer features (SAcc) and smartwatch accelerometer features (WAcc).

After feature selection, the high-level perception phase handles activity recognition using a machine learning technique. To identify the best technique for the extrasensory dataset, we assessed the performance of five different ML techniques: Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), and Naïve Bayes (NB). These classifiers have been chosen to compare both deterministic and probabilistic classifiers, as well as ensemble learning method like RF. These classifiers have been often used to recognize human activities from sensor data, and in particular accelerometer, especially in case of relatively small datasets [45,52]. To select the best classifier for the activity recognition task, each technique has been fine-tuned by selecting the suitable best parameters. A brute-force approach, namely the Grid Search Cross Validation (GridSearchCV) with five folds [53], has been used to find the best parameters for each technique. GridSearchCV uses a different combination of all the parameters with different values and calculates the performance for each combination and selects the best value for the parameters. Since the approach needs to check many combinations of the different parameters for each technique, it is time-consuming and expensive, but it allows to reach higher accuracy.

Table 5
Performance of the high-level perception phase (activity recognition) with different classifiers and sets of features.

Classifier	Sensors	Accuracy	Precision	Recall	F1-score
Decision Tree	SAcc	0.7371	0.7238	0.7371	0.7209
	WAcc	0.7452	0.7268	0.7452	0.7251
	SAcc+WAcc	0.7636	0.7515	0.7636	0.7457
	ALL	0.7636	0.7515	0.7636	0.7457
Random Forest	SAcc	0.8652	0.8567	0.6996	0.7587
	WAcc	0.8063	0.7794	0.5853	0.6449
	SAcc+WAcc	0.8304	0.8243	0.6318	0.6951
	ALL	0.8304	0.8243	0.6318	0.6951
KNN	SAcc	0.7533	0.7702	0.7733	0.7715
	WAcc	0.9362	0.9351	0.9362	0.9353
	SAcc+WAcc	0.9365	0.9354	0.9365	0.9359
	ALL	0.9369	0.9358	0.9369	0.9361
Naïve Bayes (NB)	SAcc	0.5381	0.5409	0.5382	0.5443
	WAcc	0.5119	0.5707	0.5119	0.4964
	SAcc+WAcc	0.5515	0.5985	0.5515	0.5443
	ALL	0.5846	0.6296	0.5846	0.5927
ANN	SAcc	0.5991	0.5880	0.5990	0.5782
	WAcc	0.6577	0.6355	0.6577	0.6319
	SAcc+WAcc	0.6821	0.6717	0.6821	0.6649
	ALL	0.6175	0.6245	0.6175	0.6084

These are the best configurations of the parameters of the five techniques found using the GridSearchCV approach.

- Decision tree: maximum depth of the tree = 10 (`max_depth==10`); minimum number of samples to be at a leaf node = 4 (`min_samples_leaf=4`); minimum number of samples required to split a node = 5 (`min_samples_split=5`); Gini impurity is used to measure the quality of a split (`criterion=Gini`).
- Random forest: number of trees in the forest = 100 (`n_estimators=100`); maximum depth of a tree = 300 (`max_depth=300`); minimum number of samples to split a node = 6 (`min_samples_split=6`); minimum number of samples to be at a leaf node = 4 (`min_samples_leaf=4`).
- ANN: number of neurons in the hidden layer = 100 (`hidder_layer_sizes=(100,100)`); the activation function for the hidden layer is the rectified linear unit function ReLu (`activation=ReLU`); the learning rate is adaptive (`learning_rate=adaptive`); maximum number of iterations is 1000 (`max_iter=1000`).
- KNN: the number of neighbors is 7 (`k (n_neighbors)=7`); the weight functions is distance, so points are weighted by the inverse of their distance (`weights=distance`), and the distance metric is the euclidean distance (p=2).
- Naïve Bayes (NB): a categorical naïve Bayes (NB) algorithms has been used, with no prior probabilities of classes specified (`class_prior=false`) and letting the algorithm to determine the number of categries automatically from training data (`min_cateogories=none`).

Table 5 presents the results of five classifiers in relation to four feature sets:

- SAcc (Smartphone Accelerometer) – comprising the 13 optimal smartphone features.
- WAcc (Watch Accelerometer) — encompassing the 13 top watch features.
- SAcc+WAcc — utilizing the best 26 features, 13 from each devices.
- ALL — which uses all 71 features available in the dataset.

Among all classifiers, KNN demonstrates remarkable proficiency with watch accelerometer features, achieving an accuracy of 93.69%

and an F1 score of 93.61%. Furthermore, for the SAcc+WAcc configuration, KNN surpasses other classifiers with an accuracy of 93.65% and an F1 score of 93.59%. Interestingly, this performance is almost equal to when all 71 features are used. This showcases that the 26 features have been properly identified, providing an opportunity to reduce computational effort without compromising on performance.

Focusing solely on the smartphone, the Random Forest (RF) classifier surpasses KNN in accuracy (86.52% compared to 75.33%). However, it falls short regarding the F1 score due to its lower recall. Lastly, the performance of the Artificial Neural Network (ANN) does not meet the standard in all scenarios when compared to the other four classifiers. Specifically, its accuracy rates for the smartwatch, smartphone, and combined best features are 59.91%, 65.77%, and 68.21%, respectively. The classifier that achieves the lowest performance is the Naïve Bayes.

We also compare the training time of the five classifiers in Fig. 2, again considering the SAcc+WAcc set of features. It can be analyzed from the figure that ANN and Random Forest results provide high training time, and the decision tree has a lower training time. However, in terms of accuracy, DT performance is lower than RF and KNN. Therefore, considering both accuracy and training time, we conclude that KNN is the best choice for the high-level perception phase, considering the feature set SAcc+WAcc of 26 best features.

Fig. 3 compares the confusion matrix of the proposed approach for activity recognition (configured with KNN and the best 26 features) with the AAHCR approach proposed in [45]. It can be observed that the SA-WCS outperforms the AAHCR approach for all the classes except for running. In particular, all the static activities (sitting, standing, and lying down) have many instances that have genuinely identified. Moreover, the approach performs better on the classification of the walking activity. The AAHCR approach perform slightly better on the running activity. However, when considering the context data, the SA-WCS approach will be able to identify the running situation with a greater accuracy with respect to AAHCR.

Overall, the SA-WCS approach performs better in terms of activity recognition if compared with AAHCR, as showed in Fig. 4. It can be seen that SA-WCS provides better accuracy and F1-score with respect to AAHCR, with an increase of 9.19% of accuracy and 8.71% F1-score.

4.3. Definition of the context space

The Extrasensory dataset provides the following contextual information: timestamp, user location (via GPS), audio data, phone state (including phone position, battery status, charging state, and WiFi connection), as well as additional environmental sensor data, which is rarely available and reported with lower frequency. Considering the set S of situations reported in Table 2, we conducted exploratory data analysis to pinpoint the key contextual attributes for consideration. Additionally, we assessed the significance of each feature using the same selection process applied in the high-level perception phase (Section 4.2). Notably, the timestamp feature emerged as one of the most crucial contextual factors for discerning daily-life activities. This is because certain situations occur more frequently on particular days and at specific hours, aligning with people's typical daily routines. The focus is not so much on the exact time but rather on the day of the week, as well as specific hours or moments. This allowed us to define the following context state related to the timestamp attribute. Let x_{day} be the *day of the week* context attribute, where $x_{day} = extract\_day(timestamp) \in [0, 6]$, corresponding to the days of the week (0 = Sunday to 6 = Saturday). This value is obtained by extracting the day of the week from the timestamp using the function $extract\_day(\cdot)$.

Let x_{hour} be the context attribute *hours of the day*, with $x_{hour} = extract\_hour(timestamp) \in [0, 23]$. Lastly, $x_{duration}$ is the context attribute *duration of activities*, denoting the duration of an activity in seconds. A contribution function is defined for each context attribute in Eqs. (5)–(6) and depicted in Fig. 5(a)–(c).

$$Contr_S(x_{day}) = \begin{cases} 0.0, x_{day} = 1 \\ 0.2, x_{day} = 2 \\ 0.4, x_{day} = 3 \\ 0.6, x_{day} = 4 \\ 0.8, x_{day} = 5 \\ 1.0, x_{day} = 0, x_{day} = 6 \end{cases}$$

$$Contr_S(x_{hour}) = \begin{cases} 0.0, x_{hour} \in (0, 3] \\ 0.1, x_{hour} \in (3, 5] \\ 0.2, x_{hour} \in (5, 7] \\ 0.3, x_{hour} \in (7, 8] \\ 0.5, x_{hour} \in (8, 9] \\ 0.75, x_{hour} \in (9, 12] \\ 1.0, x_{hour} \in (12, 14] \\ 0.9, x_{hour} \in (14, 17] \\ 0.8, x_{hour} \in (17, 19] \\ 0.6, x_{hour} \in (19, 21] \\ 0.1, x_{hour} \in (21, 0] \end{cases} \tag{5}$$

$$Contr_S(x_{duration}) = \begin{cases} 0.0, x_{duration} \in (0, 1] \\ 0.1, x_{duration} \in (1, 5] \\ 0.2, x_{duration} \in (5, 10] \\ 0.3, x_{duration} \in (10, 20] \\ 0.4, x_{duration} \in (20, 30] \\ 0.5, x_{duration} \in (30, 60] \\ 0.6, x_{duration} \in (60, 120] \\ 0.7, x_{duration} \in (120, 180] \\ 0.8, x_{duration} \in (180, 360] \\ 0.9, x_{duration} \in (360, 540] \\ 1.0, x_{duration} > 540 \end{cases} \tag{6}$$

The location information within the Extrasensory dataset does not include absolute coordinates. Instead, the dataset authors chose to use relative location features to prevent overfitting to the location patterns found in the dataset's samples. Information regarding location variability and average values is provided for specific time windows $W^j_{\Delta_i}$ with size = 20 s (see Eq. (2)). However, this data does not directly reveal whether activities take place indoors or outdoors or specify the exact location (e.g., home, office, car, etc.). Given that the activities listed in Table 2 differ in terms of the extent of movement (ranging from static activities like watching TV to activities in expansive areas like being in a car or exercising) we consider the diameter of the movements (a context feature already available in the dataset), as an additional context attribute. It is defined as the maximum distance between any two recorded locations within the considered window $W^j_{\Delta_i}$. Let $x_{diameter}$ be the context attribute *diameter of movements*. The contribution function for $x_{diameter}$ is defined in Eq. (7) and depicted in Fig. 5(d).

$$Contr_S(x_{diameter}) = \begin{cases} 0.00, x_{diameter} \leq 0 \\ 0.05, x_{diameter} \in (0, 3.75] \\ 0.10, x_{diameter} \in (3.75, 7.5] \\ 0.15, x_{diameter} \in (7.5, 10] \\ 0.20, x_{diameter} \in (10, 15] \\ 0.25, x_{diameter} \in (15, 18] \\ 0.30, x_{diameter} \in (18, 22] \\ 0.40, x_{diameter} \in (22, 30] \\ 0.50, x_{diameter} \in (30, 37] \\ 0.60, x_{diameter} \in (37, 45] \\ 0.70, x_{diameter} \in (45, 52] \\ 0.80, x_{diameter} \in (52, 67] \\ 0.90, x_{diameter} \in (67, 75] \\ 1.00, x_{diameter} > 75 \end{cases} \tag{7}$$

The final context attribute under consideration is x_{audio}, associated with the feature *max audio magnitude*. Audio data was recorded from the phone's microphone at a rate of 22,050 Hz for the entire duration of each 20-second window, and the maximum magnitude value is stored as part of the respective feature. The contribution function is depicted in Fig. 5(e) and defined as follows.

$$Contr_S(x_{audio}) = \begin{cases} 0.00, x_{audio} \leq -10.0 \\ 0.05, x_{audio} \in (-10.0, -7.0] \\ 0.10, x_{audio} \in (-7.0, -6.5] \\ 0.15, x_{audio} \in (-6.5, -6.0] \\ 0.20, x_{audio} \in (-6.0, -5.0] \\ 0.25, x_{audio} \in (-5.0, -4.0] \\ 0.30, x_{audio} \in (-4.0, -3.0] \\ 0.35, x_{audio} \in (-3.0, -2.0] \\ 0.40, x_{audio} \in (-2.0, -1.0] \\ 0.45, x_{audio} \in (-1.0, 0.0] \\ 0.50, x_{audio} \in (0.0, 0.5] \\ 0.55, x_{audio} \in (0.5, 1.0] \\ 0.60, x_{audio} \in (1.0, 1.5] \\ 0.65, x_{audio} \in (1.5, 2.0] \\ 0.70, x_{audio} \in (2.0, 3.0] \\ 0.75, x_{audio} \in (3.0, 4.0] \\ 0.80, x_{audio} \in (4.0, 5.0] \\ 0.85, x_{audio} \in (5.0, 6.0] \\ 0.90, x_{audio} \in (6.0, 7.0] \\ 0.95, x_{audio} \in (7.0, 10.0] \\ 1.00, x_{audio} > 10 \end{cases} \tag{8}$$

The context space $\mathbf{X}$ for the Extrasensory dataset is therefore a five dimensional space, comprising the following dimensions $\mathbf{X} = \{x_{day}, x_{hour}, x_{duration}, x_{diameter}, x_{audio}\}$. The contribution functions are defined in Eqs. (5)–(8) and depicted in Fig. 5.

The context attributes in the Extrasensory dataset define a point within the dataset's context space for each sample. This enables the identification of 5-dimensional regions within this space where specific sets of situations are more likely to occur. Illustrations of such regions can be seen in Fig. 6, which represents a 3D subspace of the context space for simplicity. This subspace includes the dimensions x_{hour}, $x_{diameter}$, and x_{audio}. For instance, a region near the origin indicates situations with low values for $x_{diameter}$ (implying static scenarios), low values for x_{audio} (indicating silence or minimal noise), and low values for x_{hour} (representing nighttime hours). In this context, it is more likely that the user's activity is either *sleeping* or *surfing the internet*. The recognition of the specific activity, determined in the high-level perception phase, can further assist in identifying the current situation.

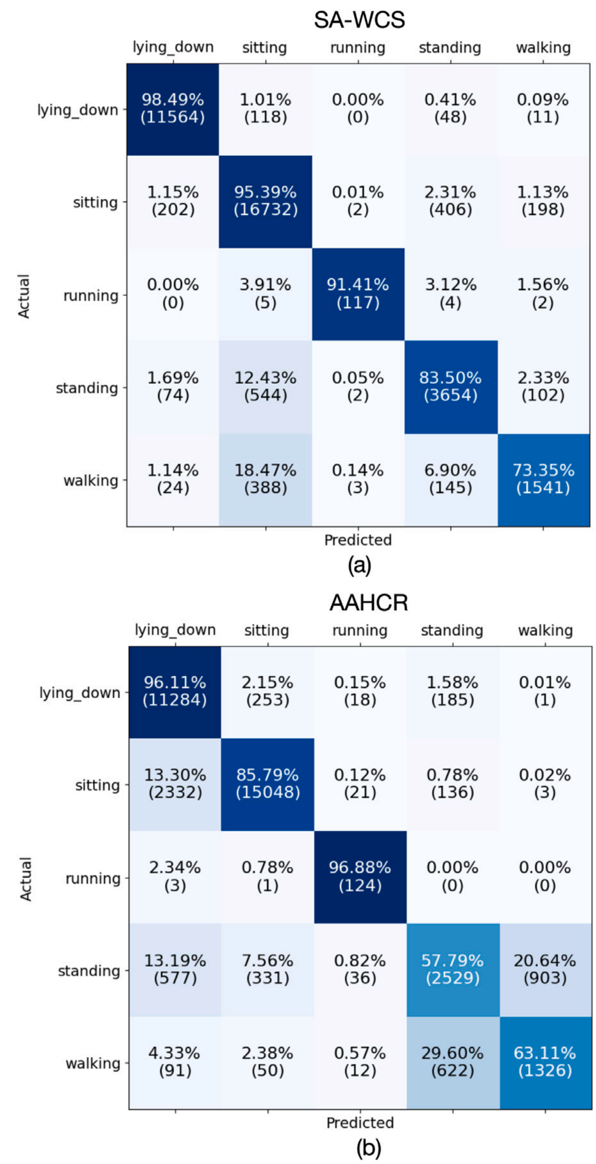

Fig. 3. Comparison of the confusion matrices for activity recognition between (a) SA-WCS approach with KNN and 26 best features, and (b) AAHCR approach [45].

Conversely, during the same hours of the day and with the same diameter but higher audio levels, it is more probable that the user is *watching TV*. On the other end of the spectrum, during daylight hours, with a wide range of movements and high audio levels, it is more likely that the user is engaged in *being in a car*, *exercising*, or *outdoor activities*.

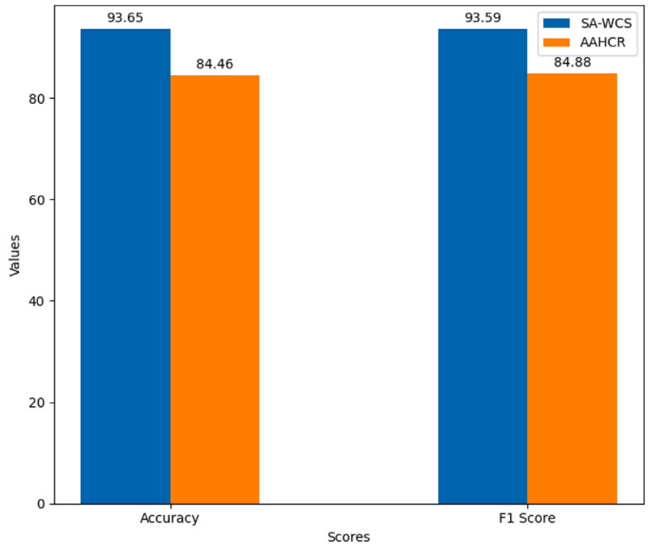

Fig. 4. Comparison of the proposed high-level perception technique with AAHCR approach [45] for the activity recognition.

4.4. Evaluation of the comprehension phase

The comprehension phase is responsible for the identification of the specific situation from the 14 situations listed in Table 2. These situations involve the performance of one of the five activities in different contexts.

The evaluation involved creating two datasets: a baseline dataset and a situation dataset, detailed in Fig. 7. Both datasets comprise the same 106,089 rows, as outlined in Section 4.1. The baseline dataset includes the top 26 features from smartphone and smartwatch data (as identified in Section 4.2.2), along with the best five context features (identified in Section 4.3), encompassing timestamp, duration, location diameter, audio magnitude, and userID. This dataset is utilized to assess the performance of the one-step baseline approach, against which the proposed SA-WCS approach is compared. The one-step baseline approach aims to classify situations by processing sensor and context features available in the Extrasensory dataset.

From the baseline dataset, the second dataset, known as the situation dataset depicted in Fig. 7, was formulated by processing data using the context recognition module. The situation dataset contains the high-level perception phase-identified activity (derived from processing the 26 smartphone and smartwatch sensor features), six context attributes (identified using the contribution functions from Section 4.3), user ID, and situation label. This situation dataset serves the comprehension phase, enabling the identification of situations using the proposed two-step approach implemented in the SA-WCS architecture.

Consequently, the L2 Comprehension module receive the following feature vector V:

$$V = \left[\text{activity}, x_{\text{day}}, x_{\text{hour}}, x_{\text{duration}}, x_{\text{diameter}}, x_{\text{audio}}, \text{userID}\right] \tag{9}$$

where $activity \in A$ is the activity identified by the high-level perception phase, $x_{\text{day}}, x_{\text{hour}}, x_{\text{duration}}, x_{\text{diameter}}, x_{\text{audio}}$ denote the context attribute values determined by the context recognition module. Additionally, userID serves to distinguish individual users in the Extrasensory dataset.

By narrowing our focus to just seven features, rather than the hundreds available in the dataset, we significantly reduced the computational time required for the comprehension phase. Furthermore, the utilization of the context space theory approach, which identifies the user's active region (as shown in Fig. 6), has enhanced our ability to accurately discern various situations. To determine the most suitable classifier for this dataset, we evaluated five different classifiers with a five fold cross validation approach (with the same approach used for the evaluation of the perception level described in Section 4.2.2).

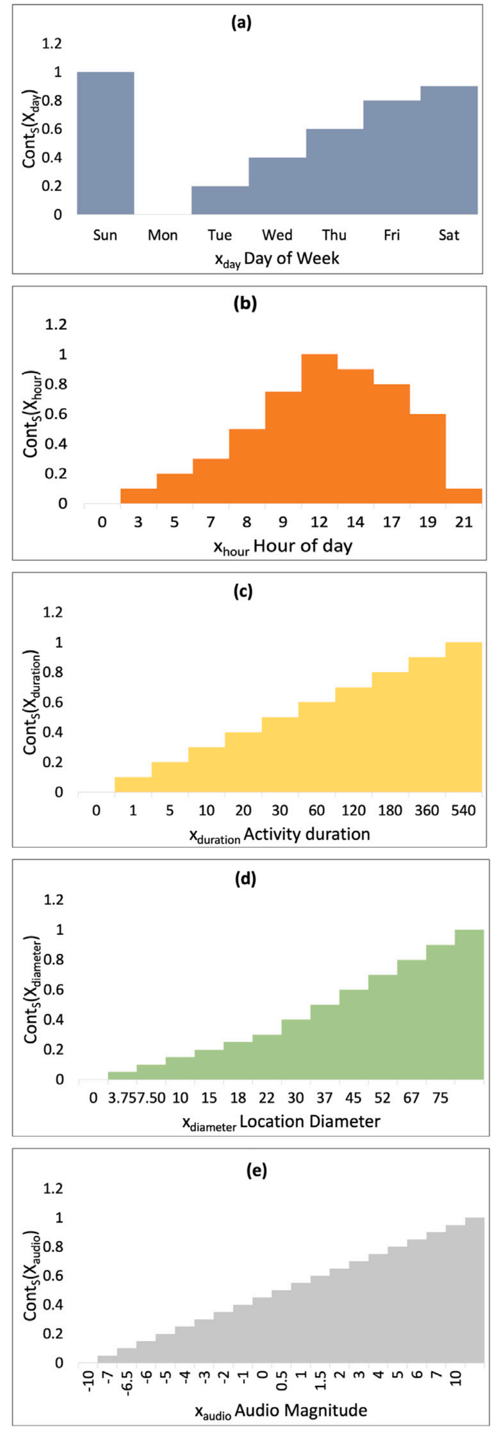

Fig. 5. Contribution functions for the context attributes defined for the extrasensory dataset. (a) x_{day} (b) x_{hour} (c) $x_{duration}$ (d) $x_{diameter}$ (e) x_{audio}.

Table 6 shows the performance of the five classifiers used in the comprehension phase of SA-WCS. Moreover, the table provides a comparison of the SA-WCS approach, which identifies situations through the two-step process (perception and comprehension) applied to the situation dataset, with the same techniques applied in a single step on the baseline dataset. This means that the classifiers employed by SA-WCS

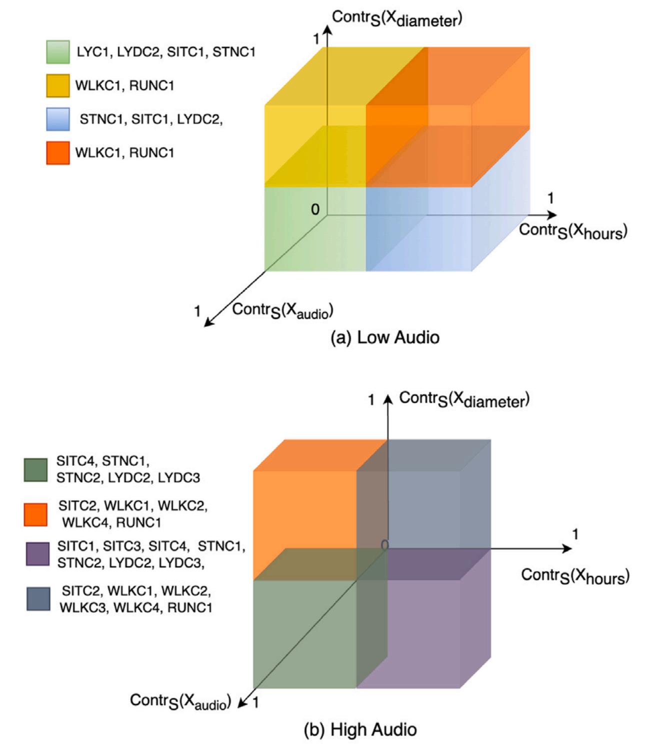

Fig. 6. Context space for the extrasensory dataset. Different regions according to the values of the contribution functions for three dimensions: x_{hour}, $x_{diameter}$, and x_{audio}. (a) Context space regions for low values of x_{audio}; (b) Context space regions for high values of x_{audio};.

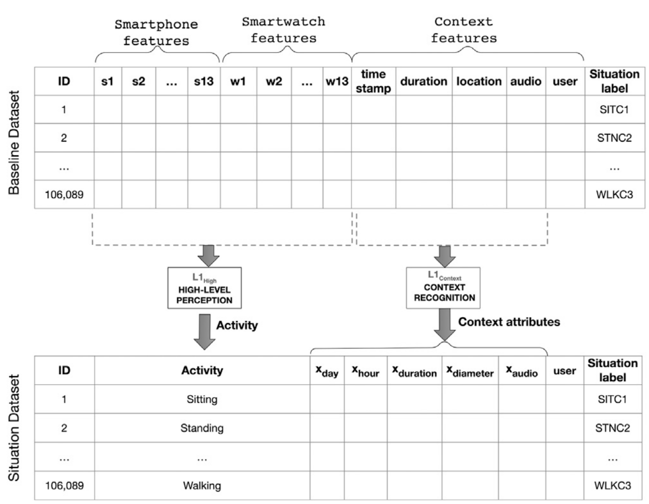

Fig. 7. Baseline and situation datasets used for the evaluation of the comprehension phase.

in the comprehension phase are also used in a traditional approach for situation classification based on sensor and context features, without the intermediate activity recognition phase as envisioned by SA-WCS.

We underline that, for the one-step techniques, we have used the context features as they are available in the baseline dataset without further processing them with the proposed CST approach. In this way we were be able to understand the advantage given by the SA-WCS architecture.

It can be observed that the SA-WCS approach outperforms the one-step classification approach with all the classifiers, according to all the considered metrics. This demonstrates that, independently by the adopted classifier, the combination of the perception phase to identify the activities and the comprehension phase to identify the situations by interpreting the activities in the light of the recognized context, results in a more effective differentiation of situations involving the users. Notably, the Random Forest (RF) classifier outperforms the other classifiers, achieving an accuracy of 95.65% and an F-Score of 95.55%, representing an increase of approximately 4.75% in accuracy and 4.45% in F1-score, compared to the one-step version of the same classifier (90.9% of accuracy and 91.10% of F1-score).

Table 6
Situation identification performance: comparison between SA-WCS approach using different classifiers and one-step machine learning techniques.

Classifier	Approach	Accuracy	Precision	Recall	F1-score
Decision Tree	SA-WCS	0.8705	0.8779	0.8706	0.8673
	One-step	0.8515	0.8512	0.8515	0.8514
Random Forest	SA-WCS	**0.9565**	**0.9561**	**0.9564**	**0.9555**
	One-step	0.9090	0.9130	0.9090	0.9110
KNN	SA-WCS	0.8454	0.8448	0.8454	0.8449
	One-step	0.6909	0.6699	0.6909	0.6802
Naïve Bayes	SA-WCS	0.8054	0.7996	0.8054	0.8025
	One-step	0.6031	0.6051	0.6031	0.6042
ANN	SA-WCS	0.87662	0.8003	0.7149	0.7436
	One-step	0.8639	0.7124	0.7049	0.7078

Table 7
Comparison between SA-WCS approach with Random Forest used in the comprehension phase and the AAHCR approach [45] for the classification of the same set of situations S.

Approach	Accuracy	Precision	Recall	F1-score
SA-WCS (RF)	0.9565	0.9561	0.9564	0.9555
AAHCR	0.8357	0.8065	0.8186	0.8126

Table 7 compares the optimal SA-WCS configuration (utilizing KNN in the perception phase and Random Forest in the comprehension phase) with the AAHCR approach. Notably, the SA-WCS approach demonstrates superior performance in classifying the same situations. It is important to note that the AAHCR approach employs a different set of features compared to SA-WCS. In Table 7, we specifically compare AAHCR without the inclusion of phone position as a feature. This comparison aligns more closely with our approach, as we also do not use phone position as a feature. However, as reported in [45], even in its best configuration, AAHCR achieves a maximum accuracy of 0.895, using more features, including phone position (which is not utilized in our approach). In this scenario, our proposed SA-WCS approach consistently outperforms AAHCR across the same set of situations with a reduced set of features.

Fig. 8.a and .b display the confusion matrices for the SA-WCS approach, employing Random Forest in the comprehension phase, and the AAHCR approach [45], respectively. This allows to evaluate the behavior of the two approaches with respect to the specific situations.

Examining the confusion matrix of SA-WCS in Fig. 8.a, it is evident that situations involving the 'Lying Down' activity, such as LYDC1 (Sleeping), are identified with higher accuracy compared to others. On the contrary, the 'Walking Indoor' (WLKC1) situation exhibits the lowest performance and often gets confused with 'Standing Indoor' (STNC1) or 'Sitting Indoor Surfing the Internet' (SITC1). Some classification errors also occur for 'Standing Outdoor' (STNC2), which is occasionally confused with 'Standing Indoor' (STNC1). This confusion arises because the contextual features lack exact GPS location information, including only the diameter of movements, which does not always distinguish between indoor and outdoor settings. However, the performance for other situations remains above 90% for all classes.

Comparing this confusion matrix with the one from AAHCR in Fig. 8.b, it becomes evident that SA-WCS makes significantly fewer errors across all classes. Notably, it enhances the classification accuracy

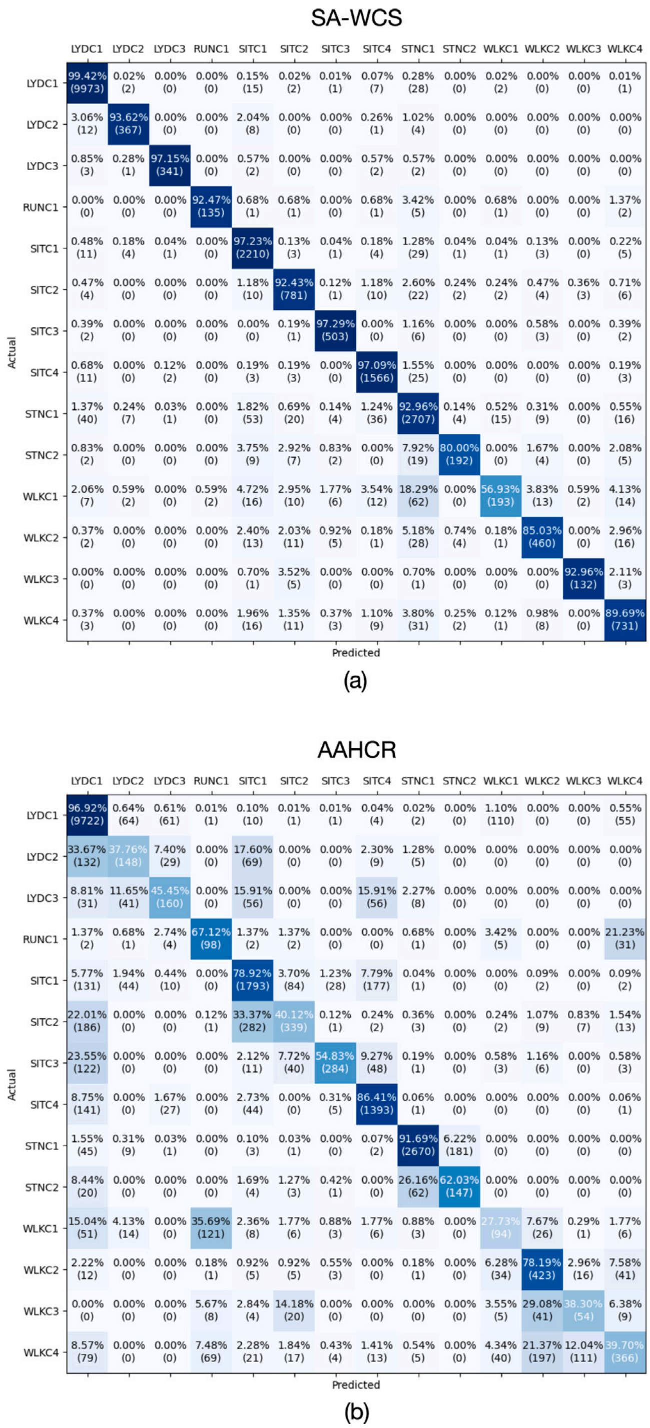

Fig. 8. Confusion matrix related to the situation identification task. Values are in percentage and in number of samples (in parentheses). (a) SA-WCS Approach. (b) AAHCR Approach.

for sitting situations (SITC1, SITC2, and SITC3) and standing activities. Additionally, it improves the recognition of the running situation, surpassing AAHCR's performance, which was less accurate in classifying activities when context was not considered in the perception phase.

Finally, we analyze the training time of SA-WCS in comparison to the one-step traditional machine learning techniques (utilizing the same hardware configuration described in Section 4.2.2). In Fig. 9, we compare the training times of all the classifiers used in the comprehension phase of SA-WCS with those of one-step traditional machine

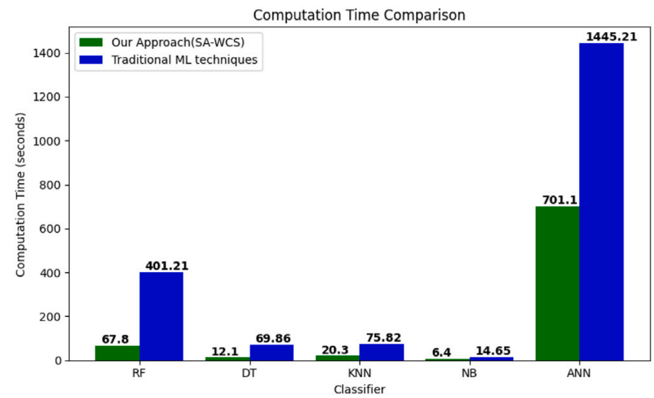

Fig. 9. Comparison of the training time between our approach and one-step traditional ML approaches.

learning techniques. SA-WCS achieves a shorter training time due to the reduced number of features and the two-step approach adopted by all the classifiers. Notably, the best classifier, Random Forest, has a training time of 67.8 s, representing an 83% reduction compared to traditional RF (401.2 s).

These experiments indicate that our approach (SA-WCS), particularly when utilizing Random Forest, is efficient in terms of computation time.

5. Conclusions

In-the-wild situation recognition is still a daunting open research challenge. This work presents an architecture of a situation-aware wearable computing system based on a two-steps approach for in-the-wild situation-aware activity recognition inspired by the model of Endsley. It involves initially identifying human activities using a learning-based technique and then determining the situation by combining the activity and the context state. The context state is established through an approach grounded in Context Space Theory.

Compared to traditional machine learning techniques that operate with a one-step approach, this proposed approach demonstrates superior performance. Additionally, by computing the context state using the geometrical metaphor, the need for numerous features to identify situations is significantly reduced, resulting in a notable reduction in computational time.

A practical implication of this outcome is that the SA-WCS approach can serve as a foundational framework for developing situation-aware applications capable of adapting to environmental conditions. These applications can be implemented as mobile applications, with smartphones as the primary sensor for both activity recognition and context identification. Furthermore, the utilization of data from smartwatches can substantially enhance performance. The potential application domains are diverse, encompassing areas such as healthcare, sports, security, safety, and military applications.

While the results so far are promising, there are still challenges and limitations that need addressing in future research. One key limitation is the manual definition of contribution functions for the context state. This process demands a deep understanding of the domain, environmental characteristics, and available contextual features, typically involving experts. As the quality and accurate modeling of the context state significantly affect situation identification performance, the context definition phase is a critical and time-consuming aspect of the SA-WCS approach.

To overcome this limitation in future work, we will explore the use of machine learning techniques to establish an initial version of the contribution functions using a data-driven approach. For instance, a decision tree technique could be employed to learn an appropriate division of contextual values for classifying existing context states, which can then be employed to determine the components of the contribution function. With a prototype of each contribution function in place, a human operator could subsequently refine and fine-tune the functions, thereby enhancing classification performance with significantly less effort and time.

In this work, the perception and comprehension phases have been defined, implemented, and evaluated within the SA-WCS approach. A future aspect of this approach includes the introduction of a projection phase, which aims to predict the future development of situations. This phase will enable the system to anticipate forthcoming situation changes, facilitating adaptive and coherent decision-making based on the predicted situation. Furthermore, the foresight of future situations can enhance both the perception and comprehension phases. It allows the perception phase to focus on a subset of available information in line with the system's expectations, leveraging attentive mechanisms. Meanwhile, the comprehension phase can improve situation identification by narrowing down potential scenarios based on the likely projection, or it can use the projection as an additional feature to differentiate between multiple possible situations, especially in cases of uncertainty.

Lastly, in future works, we intend to assess the SA-WCS approach in diverse contexts, with a specific emphasis on its application in the healthcare sector for intelligent patient monitoring, including telemedicine applications.

CRediT authorship contribution statement

Giuseppe D'Aniello: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. **Matteo Gaeta:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. **Raffaele Gravina:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. **Qimeng Li:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Zia Ur Rehman:** Data curation, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing, Formal analysis. **Giancarlo Fortino:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

The dataset used for evaluating the results of this research study is publicly available and known as "ExtraSensory" [1].

Acknowledgment

This work was supported in part by the Italian MUR through the PRIN 2020 Project "COMMON-WEARS" and in part by the project AGE-IT (PE00000015) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

References

[2] M. Zheng, P.X. Liu, R. Gravina, G. Fortino, An emerging wearable world: New gadgetry produces a rising tide of changes and challenges, IEEE Syst. Man Cybern. Mag. 4 (4) (2018) 6–14, http://dx.doi.org/10.1109/MSMC.2018.2806565.

[3] G. D'Aniello, Fuzzy logic for situation awareness: a systematic review, J. Ambient Intell. Humaniz. Comput. 14 (4) (2023) 4419–4438, http://dx.doi.org/10.1007/s12652-023-04560-6.

[4] M.R. Endsley, Toward a theory of situation awareness in dynamic systems, Hum. Factors 37 (1) (1995) 32–64, http://dx.doi.org/10.1518/001872095779049543.

[5] G. D'Aniello, R. Gravina, M. Gaeta, G. Fortino, Situation-aware sensor-based wearable computing systems: A reference architecture-driven review, IEEE Sens. J. (2022) 1–11, http://dx.doi.org/10.1109/JSEN.2022.3180902.

[6] A. Padovitz, S.W. Loke, A. Zaslavsky, Towards a theory of context spaces, in: Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops, PERCOMW '04, IEEE Computer Society, USA, 2004, p. 38.

[7] Y. Vaizman, K. Ellis, G. Lanckriet, N. Weibel, ExtraSensory app: Data collection in-the-wild with rich user interface to self-report behavior, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 1–12, http://dx.doi.org/10.1145/3173574.3174128.

[8] B. Schiele, T. Starner, B. Rhodes, B. Clarkson, A. Pentland, Situation aware computing with wearable computers, in: Fundamentals of Wearable Computers and Augmented Reality, CRC Press, 2001, pp. 511–537.

[9] S. Don, E. Choi, D. Min, A situation aware framework for activity based risk analysis of patient monitoring system, in: 2011 3rd International Conference on Awareness Science and Technology (ICAST), 2011, pp. 15–19, http://dx.doi.org/10.1109/ICAwST.2011.6163087.

[10] C. Ardito, T. Di Noia, C. Fasciano, D. Lofu, N. Macchiarulo, G. Mallardi, A. Pazienza, F. Vitulano, Towards a situation awareness for ehealth in ageing society, in: Italian Workshop on Artificial Intelligence for an Ageing Society (AIxAS), 2020. Co-Located with AI*IA 2020, the International Conference of the Italian Association for Artificial Intelligence, 2020.

[11] A. Coronato, G. Paragliola, An anomalous situation detection system for cognitive impaired people, Comput. Syst. Sci. Eng. 30 (2015) 33–41.

[12] J. Ye, S. Dobson, S. McKeever, Situation identification techniques in pervasive computing: A review, Pervasive Mob. Comput. 8 (1) (2012) 36–66, http://dx.doi.org/10.1016/j.pmcj.2011.01.004.

[13] S.-Y. Lee, F.J. Lin, Situation awareness in a smart home environment, in: 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), 2016, pp. 678–683, http://dx.doi.org/10.1109/WF-IoT.2016.7845412.

[14] R.M. Hegde, J. Kurniawan, B.D. Rao, On the design and prototype implementation of a multimodal situation aware system, IEEE Trans. Multimed. 11 (4) (2009) 645–657, http://dx.doi.org/10.1109/TMM.2009.2017631.

[15] J.R. Celestrini, A.M. Baldi, R.V. Andreão, J.G.P. Filho, C.A.S. Santos, Flow-based situation-aware approach for eHealth data processing, in: 2020 IEEE International Conference on E-Health Networking, Application Services (HEALTHCOM), 2021, pp. 1–7, http://dx.doi.org/10.1109/HEALTHCOM49281.2021.9398992.

[16] A. Anzanpour, I. Azimi, M. Götzinger, A.M. Rahmani, N. TaheriNejad, P. Liljeberg, A. Jantsch, N. Dutt, Self-awareness in remote health monitoring systems using wearable electronics, in: Design, Automation Test in Europe Conference Exhibition (DATE), 2017, 2017, pp. 1056–1061, http://dx.doi.org/10.23919/DATE.2017.7927146.

[17] D. Merico, R. Bisiani, H. Ali, Demonstrating contexta-CARE: A situation-aware system for supporting independent living, in: 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, 2013, pp. 309–310, http://dx.doi.org/10.4108/icst.pervasivehealth.2013.252048.

[18] A. Coronato, G. De Pietro, G. Paragliola, A situation-aware system for the detection of motion disorders of patients with Autism Spectrum Disorders, Expert Syst. Appl. 41 (17) (2014) 7868–7877, http://dx.doi.org/10.1016/j.eswa.2014.05.011.

[19] A. Saad, H. Fouad, A. Mohamed, Situation-aware recommendation system for personalized healthcare applications, J. Ambient Intell. Humaniz. Comput. (2021) 1–15, http://dx.doi.org/10.1007/s12652-021-02927-1.

[20] S. Katayama, A. Mathur, M. Van den Broeck, T. Okoshi, J. Nakazawa, F. Kawsar, Situation-aware emotion regulation of conversational agents with kinetic earables, in: 2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019, 2019, pp. 725–731, http://dx.doi.org/10.1109/ACII.2019.8925449.

[21] A. Henaien, H. Elhadj, L. Fourati, Combined machine learning and semantic modelling for situation awareness and healthcare decision support, in: The Impact of Digital Technologies on Public Health in Developed and Developing Countries: 18th International Conference, ICOST 2020, Hammamet, Tunisia, June 24–26, 2020, Proceedings, 2020, pp. 197–209, http://dx.doi.org/10.1007/978-3-030-51517-1_16.

[1] P. Delir Haghighi, M. Indrawan-Santiago, A. Perera, T. Huynh, Situation-aware mobile health monitoring, in: MobiQuitous 2014 - 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, 2014, pp. 248–256, http://dx.doi.org/10.4108/icst.mobiquitous.2014.257974.

[22] F. Serpush, M.B. Menhaj, B. Masoumi, B. Karasfi, T.R. G, Wearable sensor-based human activity recognition in the smart healthcare system, Intell. Neurosci. 2022 (2022).

[23] Q. Li, R. Gravina, Y. Li, S.H. Alsamhi, F. Sun, G. Fortino, Multi-user activity recognition: Challenges and opportunities, Inf. Fusion 63 (2020) 121–135, http://dx.doi.org/10.1016/j.inffus.2020.06.004.

[24] G. Fortino, A. Guzzo, M. Ianni, F. Leotta, M. Mecella, Predicting activities of daily living via temporal point processes: Approaches and experimental results, Comput. Electr. Eng. 96 (2021) 107567, http://dx.doi.org/10.1016/j.compeleceng.2021.107567.

[25] A. Greco, G. Percannella, P. Ritrovato, A. Saggese, M. Vento, A deep learning based system for handwashing procedure evaluation, Neural Comput. Appl. 35 (22) (2023) 15981–15996, http://dx.doi.org/10.1007/s00521-022-07194-5.

[26] S.A. Khowaja, P. Khuwaja, K. Dev, G. D'Aniello, VIRFIM: an AI and Internet of Medical Things-driven framework for healthcare using smart sensors, Neural Comput. Appl. (2021).

[27] A. Ferrato, C. Limongelli, M. Mezzini, G. Sansonetti, Using deep learning for collecting data about museum visitor behavior, Appl. Sci. (Switz.) 12 (2) (2022) http://dx.doi.org/10.3390/app12020533.

[28] M. Casillo, F. Colace, B.B. Gupta, A. Lorusso, F. Marongiu, D. Santaniello, C. Valentino, A situation awareness approach for smart home management, in: 2021 International Seminar on Machine Learning, Optimization, and Data Science, ISMODE 2021, 2022, pp. 260–265, http://dx.doi.org/10.1109/ISMODE53584.2022.9742901.

[29] Z. Zhuang, Y. Xue, Sport-related human activity detection and recognition using a smartwatch, Sensors 19 (22) (2019) http://dx.doi.org/10.3390/s19225001.

[30] J. Berg, T. Reckordt, C. Richter, G. Reinhart, Action recognition in assembly for human-robot-cooperation using hidden Markov models, Proc. CIRP 76 (2018) 205–210, 7th CIRP Conference on Assembly Technologies and Systems (CATS 2018).

[31] K.G. Manosha Chathuramali, R. Rodrigo, Faster human activity recognition with SVM, in: International Conference on Advances in ICT for Emerging Regions (ICTer2012), 2012, pp. 197–203.

[32] S. Oniga, J. Sütő, Human activity recognition using neural networks, in: Proceedings of the 2014 15th International Carpathian Control Conference (ICCC), 2014, pp. 403–406.

[33] F. Colace, V. Loia, W. Pedrycz, S. Tomasiello, On a granular functional link network for classification, Neurocomputing 398 (2020) 108–116.

[34] N.R. Nurwulan, G. Selamaj, Random forest for human daily activity recognition, in: Journal of Physics: Conference Series, Vol. 1655, IOP Publishing, 2020, 012087.

[35] S. Gupta, Deep learning based human activity recognition (HAR) using wearable sensor data, Int. J. Inf. Manage. Data Insights 1 (2) (2021) 100046.

[36] P. Ariza Colpas, E. Vicario, E. De-La-Hoz-Franco, M. Pineres-Melo, A. Oviedo-Carrascal, F. Patara, Unsupervised human activity recognition using the clustering approach: A review, Sensors 20 (9) (2020).

[37] F. Najar, S. Bourouis, N. Bouguila, S. Belghith, Unsupervised learning of finite full covariance multivariate generalized Gaussian mixture models for human activity recognition, Multimedia Tools Appl. 78 (13) (2019) 18669–18691.

[38] Y. Kwon, K. Kang, C. Bae, Unsupervised learning for human activity recognition using smartphone sensors, Expert Syst. Appl. 41 (14) (2014) 6067–6074.

[39] S. Zhao, W. Li, J. Cao, A user-adaptive algorithm for activity recognition based on K-means clustering, local outlier factor, and multivariate Gaussian distribution, Sensors 18 (6) (2018).

[40] M.V. Albert, A. Sugianto, K. Nickele, P. Zavos, P. Sindu, M. Ali, S. Kwon, Hidden Markov model-based activity recognition for toddlers, Physiol. Meas. 41 (2) (2020) 025003.

[41] S. Oh, A. Ashiquzzaman, D. Lee, Y. Kim, J. Kim, Study on human activity recognition using semi-supervised active transfer learning, Sensors 21 (8) (2021) 2760.

[42] H. Bi, M. Perello-Nieto, R. Santos-Rodriguez, P. Flach, I. Craddock, An active semi-supervised deep learning model for human activity recognition, J. Ambient Intell. Humaniz. Comput. (2022) 1–17.

[43] S. Rahimi Taghanaki, M.J. Rainbow, A. Etemad, Self-supervised human activity recognition by learning to predict cross-dimensional motion, in: 2021 International Symposium on Wearable Computers, Association for Computing Machinery, New York, NY, USA, 2021, pp. 23–27.

[44] A. Boytsov, A. Zaslavsky, From sensory data to situation awareness: Enhanced context spaces theory approach, in: 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, 2011, pp. 207–214, http://dx.doi.org/10.1109/DASC.2011.55.

[45] M. Ehatisham-ul Haq, M. Azam, Opportunistic sensing for inferring in-the-wild human contexts based on activity pattern recognition using smart computing, Future Gener. Comput. Syst. 106 (2020) 374–392, http://dx.doi.org/10.1016/j.future.2020.01.003.

[46] J. Suto, S. Oniga, P.P. Sitar, Comparison of wrapper and filter feature selection algorithms on human activity recognition, in: 2016 6th International Conference on Computers Communications and Control (ICCCC), 2016, pp. 124–129, http://dx.doi.org/10.1109/ICCCC.2016.7496749.

[47] M. Kim, M. Kim, A formal definition of situation towards situation-aware computing, in: M.D. Lytras, E. Damiani, J.M. Carroll, R.D. Tennyson, D. Avison, A. Naeve, A. Dale, P. Lefrere, F. Tan, J. Sipior, G. Vossen (Eds.), Visioning and Engineering the Knowledge Society. a Web Science Perspective, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 553–563.

[48] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, A public domain dataset for human activity recognition using smartphones, in: ESANN, 2013.

[49] C. Chen, R. Jafari, N. Kehtarnavaz, UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor, in: 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 168–172, http://dx.doi.org/10.1109/ICIP.2015.7350781.

[50] P. Esfahani, H.T. Malazi, PAMS: A new position-aware multi-sensor dataset for human activity recognition using smartphones, in: 2017 19th International Symposium on Computer Architecture and Digital Systems (CADS), 2017, pp. 1–7, http://dx.doi.org/10.1109/CADS.2017.8310680.

[51] T. Sztyler, J. Carmona, J. Völker, H. Stuckenschmidt, Self-tracking reloaded: Applying process mining to personalized health care from labeled sensor data, in: M. Koutny, J. Desel, J. Kleijn (Eds.), Transactions on Petri Nets and Other Models of Concurrency XI, Springer Berlin Heidelberg, Berlin, Heidelberg, 2016, pp. 160–180, http://dx.doi.org/10.1007/978-3-662-53401-4_8.

[52] F. Kulsoom, S. Narejo, Z. Mehmood, H.N. Chaudhry, A. Butt, A.K. Bashir, A review of machine learning-based human activity recognition for diverse applications, Neural Comput. Appl. 34 (21) (2022) 18289–18324, http://dx.doi.org/10.1007/s00521-022-07665-9.

[53] D. Wang, E. Candinegara, J. Hou, A.-H. Tan, C. Miao, Robust human activity recognition using lesser number of wearable sensors, in: 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), IEEE, 2017, pp. 290–295.