

Sleep behavior assessment via smartwatch and stigmergic receptive fields

Antonio L. Alfeo¹ · Paolo Barsocchi² · Mario G. C. A. Cimino¹ · Davide La Rosa² · Filippo Palumbo² · Gigliola Vaglini¹

Abstract Sleep behavior is a key factor in maintaining good physiological and psychological health. A well-known approach to monitor sleep is polysomnography. However, it is costly and intrusive, which may disturb sleep. Consequently, polysomnography is not suitable for sleep behavior analysis. Other approaches are based on actigraphy and sleep diary. Although being a good source of information for sleep quality assessment, sleep diaries can be affected by cognitive bias related to subject's sleep perception, while actigraphy overestimates sleep periods and night-time disturbance compared to sleep diaries. Machine learning techniques can improve the objectivity and reliability of the observations. However, since signal morphology vary widely between people, conventional machine learning is complex to set up. In this regard, we present an adaptive,

reliable, and innovative computational approach to provide per-night assessment of sleep behavior to the end-user. We exploit heartbeat rate and wrist acceleration data, gathered via smartwatch, in order to identify subject's sleep behavioral pattern. More specifically, heartbeat rate and wrist motion samples are processed via computational stigmergy, a bio-inspired scalar and temporal aggregation of samples. Stigmergy associates each sample to a digital pheromone deposit (mark) defined in a mono-dimensional space and characterized by evaporation over time. As a consequence, samples close in terms of time and intensity are aggregated into functional structures called trails. The stigmergic trails allow to compute the similarity between time series on different temporal scales, to support classification or clustering processes. The overall computing schema includes a parametric optimization for adapting the structural parameters to individual sleep dynamics. The outcome is a similarity between sleep nights of the same subject, to generate clusters of nights with different quality levels. Experimental results are shown for three real-world subjects. The resulting similarity is also compared with the dynamic time warping, a popular similarity measure for time series.

✉ Mario G. C. A. Cimino
mario.cimino@unipi.it

Antonio L. Alfeo
luca.alfeo@ing.unipi.it

Paolo Barsocchi
paolo.barsocchi@isti.cnr.it

Davide La Rosa
davide.larosa@isti.cnr.it

Filippo Palumbo
filippo.palumbo@isti.cnr.it

Gigliola Vaglini
gigliola.vaglini@unipi.it

¹ Department of Information Engineering, University of Pisa, Largo L. Lazzarino 1, 56122 Pisa, Italy

² National Research Council, Institute of Information Science and Technologies, via G. Moruzzi 1, 56124 Pisa, Italy

Keywords Sleep monitoring · Smartwatch · Stigmergy · Neural receptive field

1 Introduction

One of the most important markers of a healthy lifestyle is represented by the quality and quantity of sleep. These factors directly affect the waking life, including productivity, emotional balance, creativity, physical vitality, and the general personal health. Indeed, poor long-term sleep patterns can lead to a wide range of health-related

problems, such as high-blood pressure, high stress, anxiety, diabetes, and depression [15]. In this context, the monitoring of sleep patterns becomes of major importance for various reasons, such as the detection and treatment of sleep disorders, the assessment of the effect of different medical conditions or medications on the sleep quality, and the assessment of mortality risks associated with sleeping patterns in adults and children [46].

Traditionally, the polysomnographic (PSG) recordings have been widely used in order to infer the sleep quality [1]. In particular, the PSG recordings include the electroencephalography (EEG), the electrooculogram (EOG), and the electromyogram (EMG) data of the patient [16]. In this regard, the quality measure is usually captured with self-reports via paper-based surveys and diaries that, although being difficult and tedious to be collected, represent a reliable source of information [12]. Nevertheless, sleep diaries can be affected by cognitive bias related to the subject's sleep perception.

In recent years, because of the development of ubiquitous technology in health care, the research effort involving non-invasive sensors to assess and report sleep patterns is actively progressing. In [46], contact-based pressure mattress and a non-contact 3D image acquisition device have been used for sleep monitoring. In [37], motion detection, sound, and vibration sensor give information about sleeping. A relevant source of information on sleeping is represented by motion data coming from worn inertial sensors (i.e., accelerometers) embedded in smartphones or wristbands [50]. Indeed, data coming from worn devices has been thoroughly exploited in different scenarios aiming at monitoring human activities [51], falls [30], or gait analysis [20, 21]. The use of this kind of information is the basis of the so-called actigraphy.

Actigraphy is a non-invasive method of monitoring human rest/activity cycles. In 1995, the Standards of Practice Committee of the American Sleep Disorders Association (ASDA) commissioned a task force to evaluate the role of actigraphy in sleep medicine. The term actigraphy refers to methods using wristband-like devices to monitor and collect data generated by movements. This involves a person wearing an accelerometer-based device on their wrist. ASDA's effort on actigraphy led to a review paper on the topic [56, 57] and a set of guidelines [4]. The acknowledgment for actigraphy as a valid tool by ASDA was an important landmark in its acceptance by sleep-related researchers and clinicians. The use of actigraphy is continuously rising in sleep research and medicine, as demonstrated by the increasing number of publications over the years [11, 56].

While commercially available activity trackers based on wearable devices can be considered valid for measuring sleep phases and heart rate (HR) during sleep [63, 64], there exist many sleep analysis algorithms that, exploiting

smartphone sensors only, have not been validated by scientific literature or studies [50]. This is even more evident when considering long-term analysis. Many people track their sleep through mobile and wearable technology, together with contextual information that may influence sleep quality, like exercise, diet, and stress. However, there is limited support to help people make sense of this wealth of data, i.e., to explore the relationship between sleep data and contextual data. In [41], authors try to bridge the gap between sleep-tracking and sense-making through the design of a web-based tool that helps individuals understand sleep quality. However, an automatic tool able to monitor the sleep over the long period and give a user-tailored quality measure is still missing.

Indeed, most sleep scoring algorithms provide a threshold-based analysis of subject's activeness during the whole night. Unfortunately, due to peculiarities of each subject's sleep, some thresholds cannot be effective for any user nor exhaustive for a sleep behavior analysis. As an example, same REM sleep ratio values can be obtained with nights characterized by different number and duration of REM-NonREM cycles, which is an important behavioral difference. In contrast with more traditional scoring algorithms, novel machine learning approaches can provide greater accuracy due to their ability to generate nonlinear classification borders [56]; moreover, they can improve the objectivity and reliability of the observations [39]. On the other hand, the use of machine learning techniques often requires a careful tuning of their structural parameters, which can be provided by employing an expert in the field or even via brute-force search [60].

In this paper, we present an automatic tool for monitoring sleep behavior that uses a commercially available smartwatch, in order to sample heart rate and wrist inertial data, and a novel detection technique based on stigmergic receptive fields (SRFs). The SRF transforms an input time series into a stigmergic trail and provides a (dis-)similarity measure against another signal trail. The dissimilarity measure can be parametrically optimized according to sleep quality annotations on a set of reference nights. The proposed approach is able to handle the continuous availability of data samples from many subjects, possibly using different models of devices among them. This can be effectively achieved by using a middleware communication platform which enables the possibility to connect different kinds of smart wristbands, thus supporting many technologies. The presence of the middleware also enables the long-term scenario and the use of environmental/domestic sensors for future developments in data fusion algorithms for the identification of correlations between sleep quality, daily activities, and the characteristics of the surrounding environment. The overall system has been experimented on data collected in 20 nights by seven subjects, covering different

population groups: a woman aged 88, affected by arterial hypertension, a man aged 72, three students (aged from 22 to 28), and two mid-adults. As a result, the system was able to detect behavioral shift caused by awakenings and an overall sleep quality.

The paper is organized as follows. Section 2 covers the related work on actigraphy-based approaches and techniques. Section 3 describes the methodological approach used in modeling the sleeping behavior, together with the details of the system from the algorithmic point of view. In Section 4, the experimental studies are presented with the obtained results from the analyzed case studies in Section 5, while Section 6 draws the final conclusions.

2 Related work

Related research on actigraphy has focused on detecting various parameters of sleep such as quality, body posture, and sleep stages. In [15], the authors propose an algorithm able to infer the sleep duration by using the smartphones. The algorithm predicts sleep duration by exploiting a collection of soft hints such as the time and length of smartphone usage or the recharge events and environmental measures such as prolonged silence and darkness. In [31], the authors proposed a smartwatch algorithm able to define dynamic thresholds that enable for a distinction between three general states: high-amplitude (i.e., daily activities), low-amplitude motion (i.e., sleeping), and doffed motion. In [17], the authors propose a wearable device platform that utilizes motion tracking to determine the user's status and measures oxygen saturation. Continuous measurement of oxygen saturation during the user's sleep allows to detect obstructive sleep apnea. In [43], an ergonomic sensor is worn by a user, and the heart rate variations and variability of sleep are collected to provide behavioral information of stressed people. In [47], a detection of roll-over movements during sleep using a wearable armband-shaped sensor is proposed. Basing on the roll-over frequency, the authors classified sleep depth into two stages: light and deep sleep. In [23], a system based on a textile sensor for ECG and respiratory assessment and a triaxial micro-electro-mechanical system (MEMS) accelerometer is proposed to monitor astronauts during sleep in micro-gravity. In [13], the authors present a benchmark dataset from an open-source wrist-worn data logger that contains high-frequency 3D inertial data from 42 sleep lab patients, along with their data from clinical polysomnography. The authors analyze the dataset with two traditional approaches for detecting sleep and wake states and propose an algorithm by using the accelerometer data, which operates on a principle of Estimation of Stationary Sleep-segments (ESS).

From a technical point of view, a number of commercial wrist-mounted actigraphy devices are available on the market such as FitBit,¹ Jawbone,² and WakeMate.³ Each device has its own characteristics and implements its sleep-wake scoring algorithms. Assessing the sleep quality estimation accuracy of these devices is a daunting task, as their hardware design are different and they are closed-source systems that have not been clinically tested. In order to evaluate the sleep quality independently of the used technology, we found in literature two main approaches.

The first approach is based on measure metrics such as the time taken to initially fall asleep (sleep onset latency, SOL), time awake overnight after sleep onset (wake after sleep onset, WASO), and total sleep time (TST). When all these metrics are evaluated, an accurate assessment of the person's sleep efficiency (SE) can be made. SE is an overall measurement of a person's sleep quality and it simply is a ratio of the time spent asleep (TST) to the amount of time spent in bed (SOL + WASO + TST). This method is based on metrics used by sleep clinicians to infer if the patients do not suffer sleep problems [29].

The other approach is based on the classification of the subjective quality ratings. The quality ratings are usually captured with self-reports via paper-based surveys and diaries. Examples include the Sleep Timing Questionnaire [48] and the Epworth Sleepiness Scale [38].

A complementary approach involves keeping a sleep diary. While tedious to collect, a diary-based approach has proven to be reliable [12]. Indeed, in [26], the authors demonstrate that daily self-report is a valid index of sleep disturbance. Sleep diaries have also been found to be reliable for bedtime and wake-time estimates via actigraphy [62] and ambulatory electroencephalographic monitoring [45]. Nevertheless, data collected via diaries usually present a high level of uncertainty, thus requiring adaptation. A novel trend of work investigates explicit integration of machine learning algorithms into the data collection process to accomplish adaptation. For example, machine learning methods are deployed in [36] to achieve on-line adaptation to users' multimodal temporal thresholds within a human computer interaction application framework. Some other work studies application of reinforcement learning to adaptive fusion systems to perform dynamic data reliability estimation [5, 35]. A recent work also proposed using kernel-based learning methods to achieve adaptive decision fusion rules [27]. Machine learning techniques can improve the objectivity and reliability of the observations.

¹FitBit, <http://www.fitbit.com/>.

²Jawbone, <https://jawbone.com/>.

³WakeMate. <http://www.wakemate.com/>.

However, since signal morphology vary widely between people, the use of machine learning algorithms frequently implies a careful tuning of their structural parameters. Unfortunately, this tuning is often a “black art” requiring expert experience, rules of thumb, or sometimes brute-force search [60]. There is therefore great appeal for approaches aimed to save the effort required in model parameterization and management. In this context, great advantages can be provided by employing the techniques belonging to the emergent paradigms, which offer a model-free computational approach, characterized by adaptation, autonomy, and self-organization of data [61]. Specifically, the emergent paradigm technique known as stigmergy consists of a biologically inspired model of spatiotemporal aggregation of samples. By using stigmergy, each data sample is associated with a digital pheromone deposit, providing the analysis with an information granulation process. More specifically, the pheromone deposits last for a pre-defined time interval, providing temporal granulation, but also, they are characterized by width over the space of the sensing variable (e.g., heart rate), providing spatial granulation. Thus, in contrast with other machine learning techniques (e.g., neural network), the stigmergy-based approach (i) exhibits robustness with respect to minor fluctuations occurring in data by adopting stigmergic information granulation, (ii) does not require time dynamics modeling, because it intrinsically embodies the time domain [9], and (iii) requires smaller training set, as shown in Section 5. For these reasons, we present an adaptive, reliable, and innovative stigmergy-based computational approach to provide per-night assessment of sleep behavior to the end-user.

3 Methods and functional design

The proposed solution employs a smartwatch to gather subjects’ physiological and inertial data, i.e., heartbeat rate and wrist acceleration. The smartwatch embeds a heart rate monitor based on an optical sensor to detect peaks in blood flow. Thereafter, it computes the heart rate over an interval of time established by the constructor, i.e., 1 s. The accelerometer embedded in the smartwatch measures both static (due to gravity) and dynamic acceleration on the three axes, sampling them at 10 Hz. These signals are pre-processed (see Section 4.1) and normalized between 0 and 1. Moreover, the acceleration signals are summarized by the standard deviation of the acceleration magnitude, by now on referred only as acceleration. Both acceleration and heart rate signals are processed in order to derive an assessment of subjects’ sleep quality, which to be compared with respect to a sleep quality ground truth.

The ground truth is obtained from early sleep quality evaluation, provided by the subject as a Perceived Sleep Quality (PSQ) score annotated on a sleep diary. PSQ is scored as Normal (N) or Abnormal (A). Due to possible human misperception of sleep quality, we support the ground truth extraction from the sleep diary using a Sleep Stage Estimator (SSE): a software aimed to analyze user sleeping behavior based on physiological parameters [10]. The SSE provides a Computed Sleep Quality (CSQ) score based on an estimation of wake, REM, and NREM stages and their occurrence during the night.

In contrast with the approach proposed in this paper, the SSE lacks usability and reliability. Indeed, it is a threshold-based sleep quality classifier; thus, it requires the intervention of a physiological signal analyst in order to provide an effective threshold setup. Moreover, thresholds are user specific, so each new case study requires a new threshold setup phase. In addition, it can produce a rough approximation of sleep periods and night-time disturbance because it does not take into account subjective sleep evaluation [42]. Finally, it is not suitable to assess sleep behavior, as detailed in Section 1.

Finally, the resulting CSQ and PSQ scores are compared: any night log whose computed score matches the perceived sleep quality becomes an entry of the ground truth night set. System assessment performance will be evaluated using signals collected during these nights, in order to assess sleep quality. More in details, the assessment error will be computed as the mean square distance between sleep quality assessment and PSQ for each night.

In the following, we provide a bottom-up explanation of our system architecture, designed employing the principles of stigmergy. Stigmergy is an indirect coordination strategy adopted in social insect colonies [18]. Our approach to computational stigmergy replicates this process in the field of time series processing, exploiting both spatial and temporal dynamics [6, 8].

3.1 Stigmergic receptive field

The fundamental computational unit of the proposed approach is the stigmergic receptive fields (SRFs). Each SRF periodically takes samples of a reference signal and the current signal $\bar{d}(k)$ and $d(k)$, respectively, in Fig. 1, providing a measure of their similarity. This process can be decomposed in the functional modules depicted in Fig. 1.

First, the clumping process takes normalized data as input. It acts as a sort of soft discretization with respect to a set of levels of analysis, creating clumps of samples. The clumping can be implemented as a sigmoidal function,

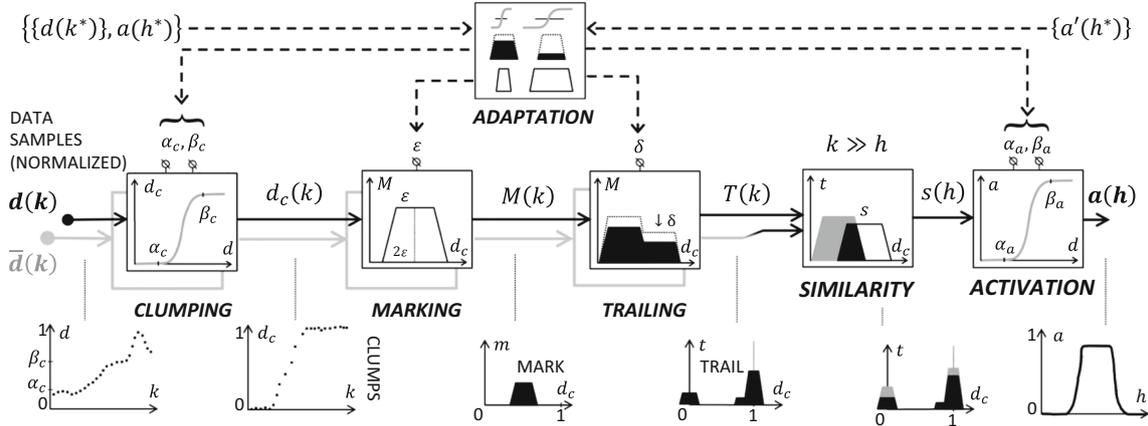


Fig. 1 Stigmergic receptive field

characterized by structural parameters α and β , and formally described in (1).

$$f(x, \alpha, \beta) = \begin{cases} 0 & \text{if } x \leq \alpha \\ 2 \left(\frac{x - \alpha}{\beta - \alpha} \right)^2 & \text{if } \alpha \leq x \leq \frac{\alpha + \beta}{2} \\ 1 - 2 \left(\frac{x - \alpha}{\beta - \alpha} \right)^2 & \text{if } \frac{\alpha + \beta}{2} \leq x \leq \alpha \\ 1 & \text{if } x \geq \beta \end{cases} \quad (1)$$

Second, the sample $d_c(k)$ enables the release of a corresponding mark $M(k)$. A mark can be implemented by a trapezoid with three attributes: intensity (height), width, and position.

Third, the trailing process accumulates marks creating the trail structure. The trail $T(k)$ is characterized by temporal decay, whose intensity decreases over time. Thus, individual mark disappears after some step of time unless new marks are released in the same position or in its proximity. Consecutive samples close to a specific value (clump) will superimpose, counteracting the decay. The trailing can then be considered as spatiotemporal sample aggregation mechanism. Fourth, current and reference trails are compared by similarity process, using Jaccard measure in (2). It provides a measure of similarity $s(h)$, between 1 (identical trails) and 0 (non-overlapping trails).

$$S = J(T_{ref}, T_{cur}) = \frac{|T_{ref} \cap T_{cur}|}{|T_{ref} \cup T_{cur}|} \quad (2)$$

In Fig. 1, $\bar{d}(k)$ and $d(k)$ are the samples of the reference and the current signals, respectively. Both signals periodically feed the SRF and are processed in parallel up to the similarity, where they are compared.

Fifth, activation can increase or decrease the similarity value, depending on two activation thresholds. Here, the term ‘‘activation’’ is taken from neural sciences and it is

related to the requirement that a signal must reach a certain level before a processing layer can fire to the next layer. It can be implemented by a sigmoidal function. In order to enable an effective similarity, based on relevant dynamics, an SRF must be properly tuned. For this purpose, the adaptation mechanism uses the differential evolution (DE) algorithm [18] to adapt the structural parameters of the SRF: (i) the clumping inflection points α_C, β_C ; (ii) the mark width ϵ ; (iii) the trail evaporation δ ; (iv) the activation inflection points α_A, β_A . DE aims to minimize the mean square error, computed as difference between desired and actual output value for a given tuning set. In Fig. 1, the tuning set is denoted by asterisks: it is a sequence of (input, desired output) pairs, on the left side, together with a corresponding sequence of actual output values, on the right side.

3.2 Stigmergic perceptron

Let us consider a single time series. What is actually interesting is not the continuous variation of the raw samples over time, but the transition from one type of behavior to another, for example from bradycardia to tachycardia or from still wrist to moving wrist. Each type of behavior should be general and reusable for a broad class of subjects. More formally, each type is called archetype and is a pure form time series fragment representing a behavioral type of activity.

A collection of SRFs can be used to measure the similarity of a given signal with many archetypes, in order to classify an input time series as a series of archetypes.

More specifically, Figs. 2 and 3 show ten and six archetypes describing heart rate and arm motion. Here, ordered by increasing intensity, we present each archetype together with the corresponding stigmergic trail. For the heart rate signal, we have the following: (Fig. 2a) bradycardia that represents the lower heart rate activity level, which

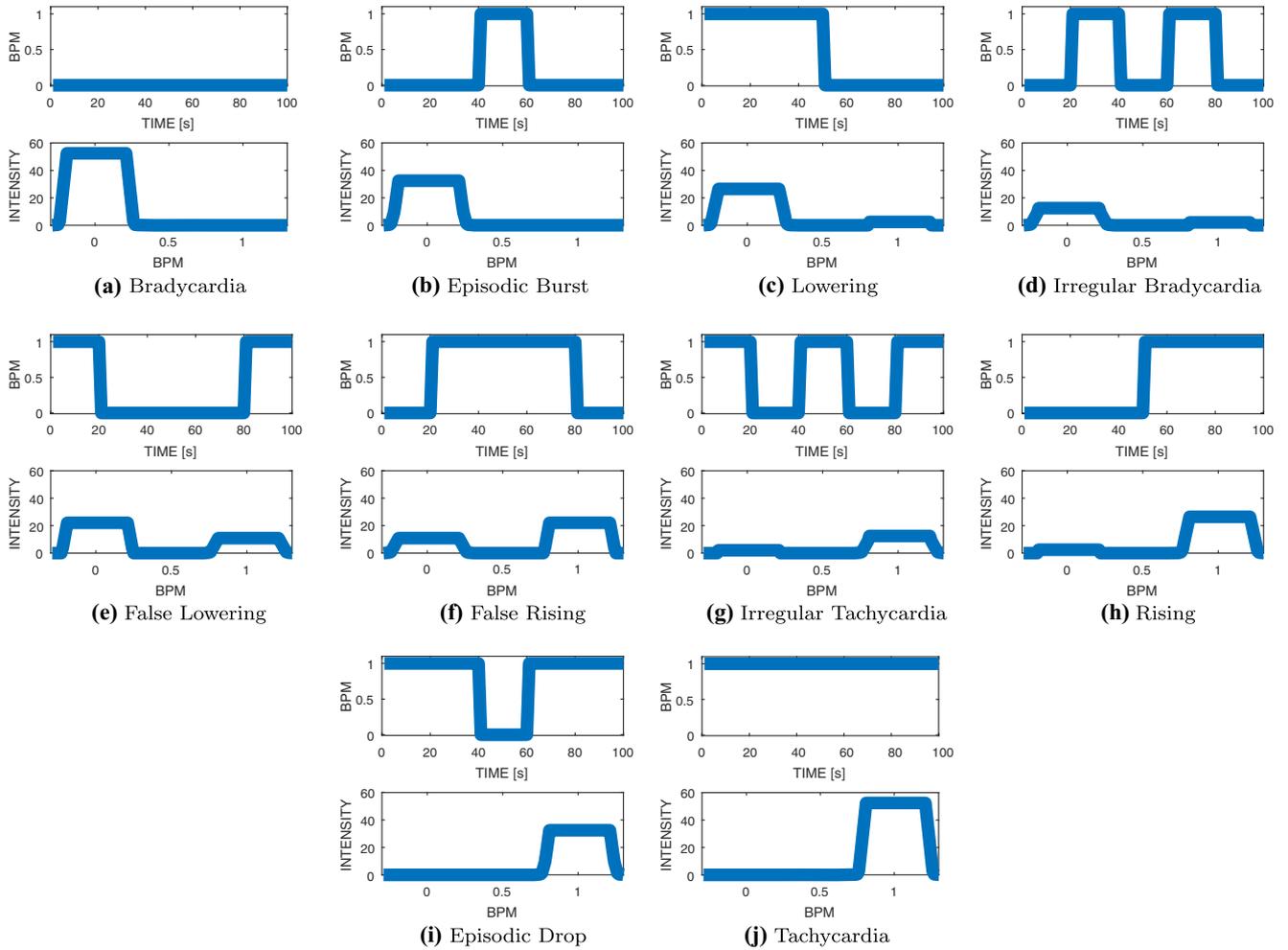


Fig. 2 Heart rate archetypes

is exhibited during early deeper NonRem sleep phase; (Fig. 2b) episodic burst, which depicts a single spike of the heart rate activity, a brief and sudden increase of the heart rate; (Fig. 2c) lowering, which usually characterize the deepen of the sleep in which the progressive drop of heart rate activity occurs; (Fig. 2d) irregular bradycardia, instead, represents mostly high but irregular heart rate activity characterizing full REM sleep phase; (Fig. 2e) false lowering and (Fig. 2f) false rising depict long-lasting irregularities in heart rate behavior; (Fig. 2g) irregular tachycardia that represents mostly low but irregular heart rate activity characterizing early REM sleep phase; (Fig. 2h) rising, which represents the progressive increase of heart rate activity characterizing the transition to REM sleep phase; (Fig. 2i) episodic drop that represents a single brief drop of the heart rate activity; (Fig. 2j) tachycardia that represents the higher heart rate activity. On the other hand, for wrist motion, we have the following: (Fig. 3a) still, which represents the complete paralysis characterizing REM sleep phase; (Fig. 3b) short move, which depicts a brief and isolated movement; (Fig. 3c) intermittent motion, which represents the

occurrence of some movements characterizing early sleep phase; (Fig. 3d) long pause, which depicts long-lasting stop of wrist motion; (Fig. 3e) awakening, representing a sudden and steady increase of wrist motion, usually occurring during transitions from sleep to wake phase; (Fig. 3f) awake, in which the high occurrence of wrist motion probes the awakens of the subject.

Since any real signal is usually similar to more than one archetype, the collection of receptive fields should be arranged into a connectionist topology, making a stigmergic perceptron (SP) [2]. If the archetypes are ordered for increasing activity, the stigmergic perceptron can combine linearly the similarity values provided by all receptive fields, providing a value between 0 and $N-1$, where N is the number of archetypes, which is further normalized between 0 and 1 and called activity level.

As a result, a stigmergic perceptron provides a new time series of activity level for a given time series. In order to calculate the overall similarity between two sleep nights, we use second-level receptive fields. It compares two activity levels as two time series.

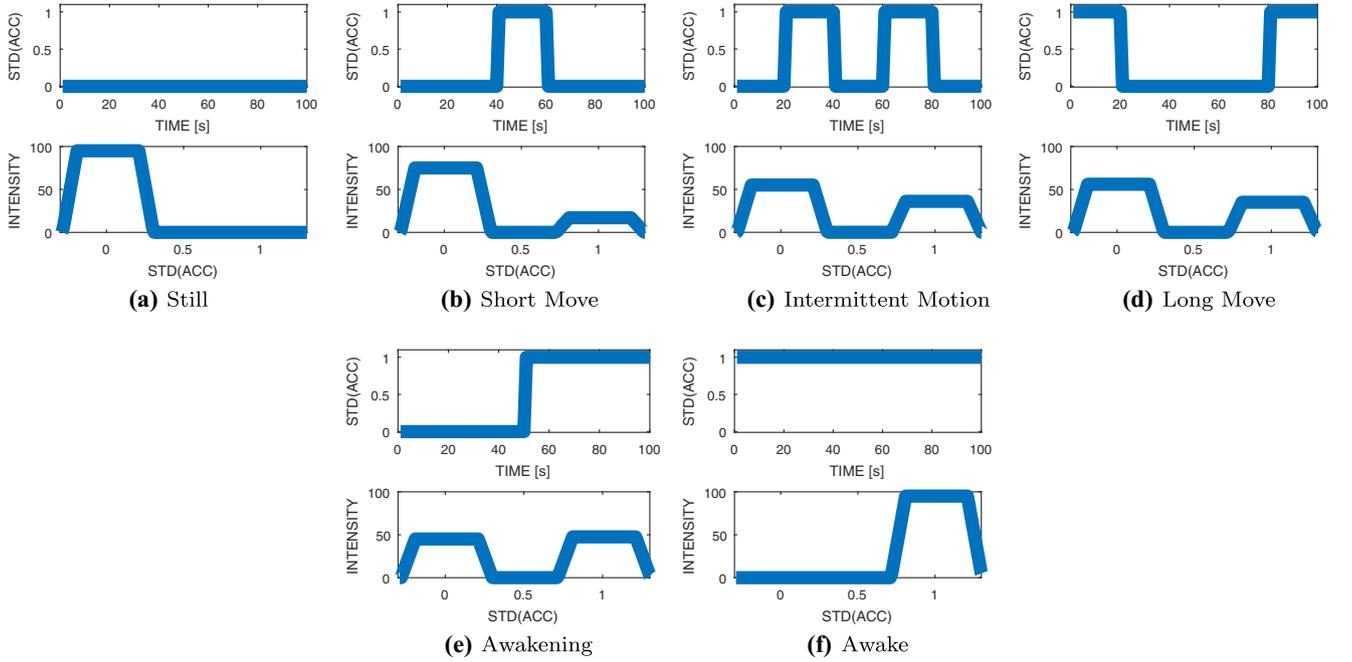


Fig. 3 Arm motion archetypes

The SRF outcomes are combined by computing their weighted mean, using the provided similarities as weights.

An important aspect is to train the SRF in a SP in order to prevent multiple activations of SRFs. For this purpose, let us consider the most sensitive SRF parameter, i.e., the evaporation δ . Indeed, high evaporation prevents mark aggregation and pattern reinforcement, while low evaporation causes the saturation of the trail. In order to handle this sensitivity, assuring a coherent processing among different SRFs inside the same SP, their adaptation is twofold:

- A global phase, which looks for an optimal range of the δ parameter among all SRF in the SP
- A local phase, which optimizes all parameters of each SRF, using range provided by global phase

Since archetypes are largely independent from the human subjects, the training set for the global and local phases is generated by using the archetypal signals as seeds and applying spatial noise and temporal shift. This way, M signals are generated for each of the N SRFs.

As an example, during the experimental setting, we generated ten signals per archetype, with a spatial noise and a temporal shift equal to 10%.

The fitness function is formally defined by (3).

$$GlobalFitness = \frac{\sum_{i=1}^N \sum_{k=1}^N \sum_{j=1}^M (|S_{jk} - \hat{S}_{jk}| * W_{ki})_i}{\sum_{i=1}^N \sum_{k=1}^N (M * W_{ki})_i} \quad (3)$$

In (3), \hat{S}_{jk} is the ideal similarity of the j th signal belonging to the k th class, which should be equal to 1 when $k = i$

or 0 when $k \neq i$; S_{jk} is the real similarity of the j th signal belonging to the k th class, computed by the i th SRF; W_{ki} is the weight of the signal belonging to the k th class (when computed by the i th SRF, it is equal to 1 when $k = i$, while it is equal to $1/(N - 1)$ when $k \neq i$). Strictly speaking, the evolution algorithm aims to find the best δ value that minimizes the weighted average of the error in similarity computation. In order to find an optimal interval for δ , we compute the following: (i) δ_{best} the value of δ corresponding to the best fitness value; (ii) $\delta_{best-10}$ the value of δ corresponding to the individual which shows a fitness equal to the 90 percentage of the best fitness value; (iii) boundaries for the range of δ is determined as $\delta_{best} \pm (\delta_{best-10} - \delta_{best})$; (iv) this procedure is repeated five times and the final boundaries of δ range are chosen as minimum and the maximum boundaries among all trials. Finally, the optimum interval $[\delta_{min}, \delta_{max}]$ is found.

The local training aims to find the optimum values for every module of each SRF. The training set is made by $2M$ signals for each SRF. Half of them belong to the archetype, while the rest belong to class of adjacent SRFs. Among the M provided for each archetype, $M/2$ are generated synthetically and the rest are selected directly from segment of subject's physiological signals.

The training set is composed of both synthetic and real signals in order to be more general, providing the system with robustness against overfitting. Indeed, by adding spatiotemporal noise to the pure archetype time series, we provide our training set with signal belonging to the same behavioral class but slightly distorted, with offset translation and temporal drift [65]. Unfortunately, it is difficult

to provide an exhaustive representation of all potential signal distortions, so we capture them directly from the real signals. The selection of the proper signals requires their visual inspection performed by a signal analyst. Moreover, depending on the computational power of the machine on which the training is performed, it could last for few hours. However, it is worth noting that this effort is required once for each class of subjects, because similar age and health conditions lead to similar behavior of physiological signal gathered during sleep [49]. Indeed, in our case studies, data gathered by subjects D, E, F, and G were analyzed by an SP trained using signals gathered with subjects D and E. Once the training set is selected and labeled, the evolution algorithm can adapt the system in order to minimize the mean square error (MSE) between computed and ideal similarities, as described in (4).

$$LocalFitness = \frac{\sum_{j=1}^M (|S_j - \hat{S}_j|^2) + \sum_{i=1}^M (|S_i - \hat{S}_i|^2)}{2M} \quad (4)$$

In (4), \hat{S}_j is the ideal similarity of the j th signal belonging to the current class, which should be equal to 1; \hat{S}_i is the ideal similarity of the i th signal, which do not belong to the current class, and should be equal to 0; S_j is the real similarity of the j th signal; and S_i is the real similarity of the i th signal.

In order to meet ideal and real similarity values, the adaptation process tunes SRF's structural parameters, evaluating the tuning effect by means of the MSE computed on the training set. In Fig. 4, the progressive reduction of the MSE provided by adaptation process, generation by generation, is shown. After the adaptation process, the MSE provided by the SP is about to be halved. Further confirmation of adaptation effectiveness is obtained by providing the trained SP

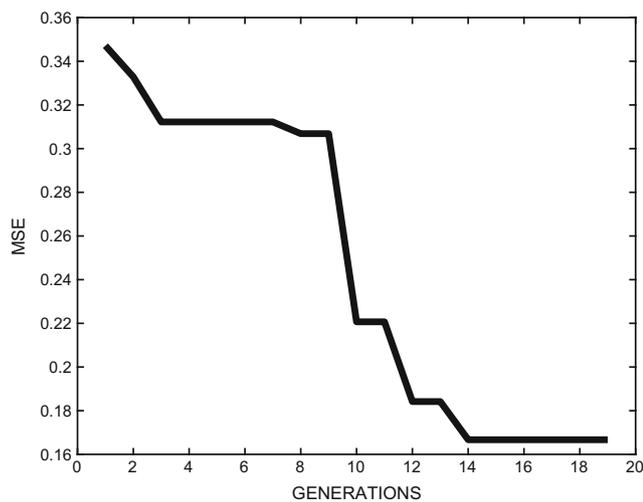


Fig. 4 MSE progress during training

with the same labeled data used as training set and collecting the obtained MSE values, shown in Tables 1 and 2.

In a properly adapted SP, the SRFs which mostly represent the input signal behavior have the main activation in the output.

As an example, in Fig. 5, we show the heart rate signals of subject B gathered during the second observation night in the minute 318 (bold line) together with the signal characterizing the “rising” archetype (gray dotted line), which appears to be the most similar with respect to the presented heart rate signal. The SP processes this signal segment producing a value of 0.7778 as activity level. Indeed, the highest similarity detected by the SP is produced by the eighth SRF (i.e., the “rising”) while a minor activation comes from the seventh one. Thus, obtained activity level measure describes the assessment of signal behavior as mostly similar to the one expressed by rising archetype and secondly by the “irregular tachycardia” archetype. In Fig. 6, the outcome of the SP with a focus on minute 318 is shown.

3.3 Multilayer architecture

SRF can be used in a multilayered architecture, thus providing higher scale of processing. In the next layer, another SRF is fed by the SP in order to provide a degree of similarity between two time series of activity level.

In order to calculate the similarity between two time series of an entire sleep night, each time series is sliced into seven partial overlapping time windows. A similarity is computed between two corresponding time windows of the two time series. Finally, the generated similarity values are aggregated.

This kind of similarity measures can be subdivided into three classes [65]: (i) similarity in time, aimed to cluster time series that vary in a similar way each step of time; (ii) similarity in shape, aimed to cluster time series according to their shape feature; and (iii) similarity in change, aimed to cluster time series according to their behavioral pattern. The similarity measure provided by our method falls in the third category which is characterized by robustness with respect to the position of irregular sleep occurrence during the night. Indeed, the similarity against different activity level time series is provided by analyzing their macro-behavior, i.e., matching the whole time series slice by slice in order to provide an overall similarity measure that considers them despite their arrangement.

In order to supply a clustering process, the second-level SRF is trained to distinguish similar and dissimilar signals, belonging to the cluster of normal sleep and abnormal sleep; thus, a training set is provided selecting M (e.g., 5) nights of known sleep quality from the ground truth. The target similarity for a couple of similar or dissimilar nights is assumed to be 1 or 0, respectively.

Table 1 MSE obtained by assessing wrist motion archetypal behavior in training set

Wrist motion archetypes	1	2	3	4	5	6
MSE	0.0041	0.0826	0.1939	0.1165	0.0800	0.0015

The fitness function adapts the mark width and the trail evaporation δ in order to minimize the MSE in the similarity, as described in (5). S_{ij} is the computed similarity between time series i and time series j , while \hat{S}_{ij} is their ideal similarity. Figure 7 describes the overall system.

$$\text{SecondLayerSRFfitness} = \frac{\sum_{i=1}^M \sum_{j=1}^M (|S_{ij} - \hat{S}_{ij}|^2)}{M^2} \quad (5)$$

3.4 Clustering of sleep nights and fusion of sensors

The similarity between pairs of signal generates a similarity matrix, which is then processed by a fuzzy relational clustering technique [19]. The fuzzy clustering generates, for each night, a membership degree to each sleep quality. In particular, let us consider the membership to the normal sleep cluster to generate a normality index for each sleep night.

Finally, the normality indexes obtained by heart rate and arm motion are combined via a weighted sum, obtaining an overall sleep quality assessment. Sleep quality assessment is defined as a real value between 0 (Abnormal) and 1 (Normal). The weights are generated by minimizing the assessment error via least square method [32]. An interesting property of the proposed approach is that the provided mapping is not explicitly modeled at design-time but achieved by the system in order to meet subject behavioral peculiarities.

4 Experimental design

In this section, we detail data gathering, signal pre-processing, and experiment description in order to provide reproducibility to our study.

4.1 Data description

Nowadays, the market diffusion of wearable devices embedding physiological and inertial sensors ease the development

of health care long-term monitoring applications for both marketing and research purposes. The technology these sensors are based upon is relatively simple and low-cost but nevertheless proven to be effective and reliable [53]. The heartbeat detection is based on photoplethysmography (PPG), a volumetric measurement of an organ (plethysmogram) obtained through an optical feedback [3]. A light-emitting diode (LED) illuminates the skin while a photodetector, exploiting the blood perfusion to the subcutaneous tissues, measures the change in the absorbed light. The recognized waveform feature is the peripheral pulse, and it is synchronized to each heartbeat. The commercial smartwatches exploit the reflective method, that is, the light emitter and the detector are placed on the same side of the skin to detect the reflected radiation. Differently, other devices exploit the transmissive method: in this case, the light emitter and detector are placed each on one side of the detecting medium (e.g., fingertips, nasal septum, earlobe) to detect the amount of radiation filtering from the tissues. The green light source is the most used on the commercial devices targeting the heartbeat detection only, while red and near-infrared wavelengths are employed on medical devices when the oximetry data is detected as well [44].

Regarding the inertial sensors, the micro-electro-mechanical systems (MEMS) became the ubiquitous solution in the consumer electronic products due to their low-cost, miniaturization level, and the ability to integrate multiple sensor functions (e.g., accelerometer, gyroscope) within a single die. They are based on mass-spring mechanical systems and are composed of signal transducers and dedicated signal post-processing circuits [58]. These sensors measure the force applied to them by detecting the displacement of the sensing mass swinging within the device.

The sleeping data collection has been carried out using three smartwatch models: LG Watch Urbane (W150), LG Watch R (W110), and Samsung Gear Live. Each device is paired with a smartphone that allows the transmission of the collected data to a storage server exploiting the middleware infrastructure. The obtained data, saved in a CSV file, is composed of the heart rate and the acceleration samples together with their acquisition time stamps. Each entry of

Table 2 MSE obtained by assessing heart rate archetypal behavior in training set

Heart rate archetypes	1	2	3	4	5	6	7	8	9	10
MSE	0.1713	0.0326	0.1211	0.0637	0.0189	0.0483	0.0749	0.0445	0.0675	0.0007

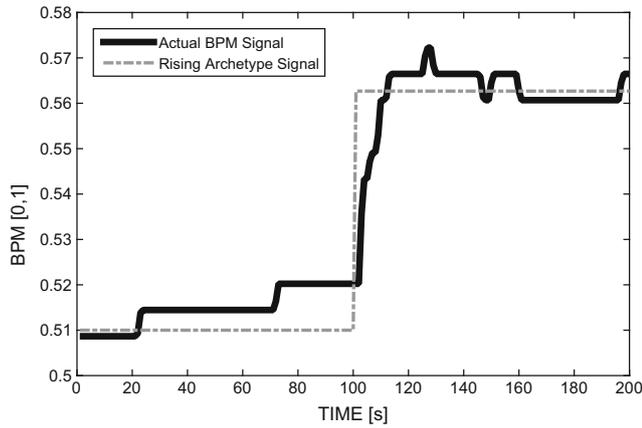


Fig. 5 Time window 635, subject B

the recorded data is formatted accordingly to this scheme:

`[ts, hr, accX, accY, accZ]`

where

- **ts** is the time stamp of the sample expressed in *ms*
- **hr** is the heart rate expressed in beats per minute [bpm]
- **accX** is the longitudinal acceleration force acting along the *x*-axis, expressed in m/s^2
- **accY** is the lateral acceleration force acting along the *y*-axis, expressed in m/s^2
- **accZ** is the acceleration force orthogonal to the screen plane and acting on the *z*-axis, expressed in m/s^2

The reference axes, valid for any android device, are shown in Fig. 8.

All data are acquired at fixed rates: the heart rate values have been sampled at 1 Hz while the acceleration sensor at 10 Hz. In order to collect reliable heart rate measurement, the smartwatch should tightly fit the user's wrist. It might

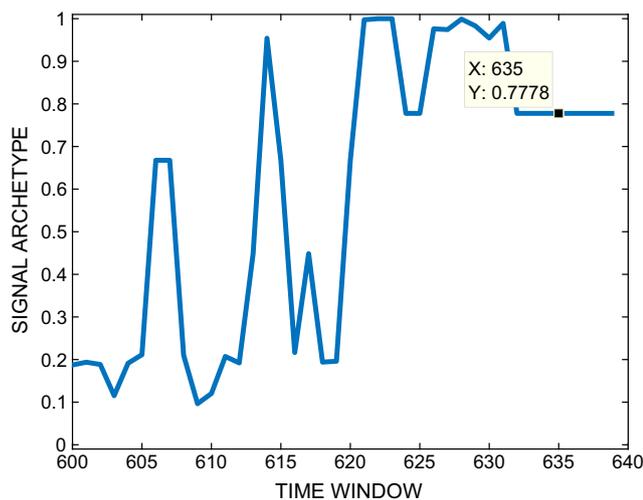


Fig. 6 Heart rate processed by SP, subject B

still happen that during the sleep, the smartwatch is kept in a position in which it misdetects the heart rate; therefore, the data post-processing task has been designed to clean the traces from any unreliable samples.

Heart rate signals are treated with (i) linear data interpolation (at 1 Hz) in order to replace null values with plausible ones, (ii) third-order median filter with a floating forward window in order to delete outliers, and (iii) resampling of the obtained series at a higher frequency (3.3 Hz).

It is worth noting that our approach does not require that the analyzed signals must have the same frequencies. Indeed, each signal can be processed independently because data fusion is applied during final computation of the overall sleep quality.

Acceleration signals a_x , a_y , and a_z are summarized computing the magnitude of the acceleration and this is treated with (i) linear data interpolation (at 3.3 Hz), in order to replace null values and match heart rate frequency, and (ii) replaced by its standard deviation computed every 3 s with a floating forward window. The use of the magnitude and its standard deviation is an established method in literature [51, 54]. Both obtained signals are further normalized in a [0,1] range by MinMax Normalization (6) procedure in order to make their values range homogeneous and therefore commensurable despite the peculiarities of the signal segment under analysis.

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

4.2 Experimental setup

The smartwatch, based on the Android Wear OS,⁴ runs a custom developed application in charge of sampling the heartbeat and the accelerometer sensors and locally storing the values. Periodically, the application connects to the smartphone via a bluetooth connection and sends all the locally stored data. Another service, running on the smartphone, receives the data from the smartwatch and sends it on a middleware infrastructure toward a main storage server.

The smartwatch solution has been designed to keep into consideration the usability and the power consumption aspects. The usability factor is important since the device is used by the elderly as well; therefore, it has to provide a clean, easy operable, and intuitive interface. Concerning the power consumption, taking into account that a continuous sampling service demands for a considerable amount of power, the carefully optimized application allows to reach an autonomy of at most 3 days with an average usage of 8 h per night.

⁴<https://developer.android.com/wear/>

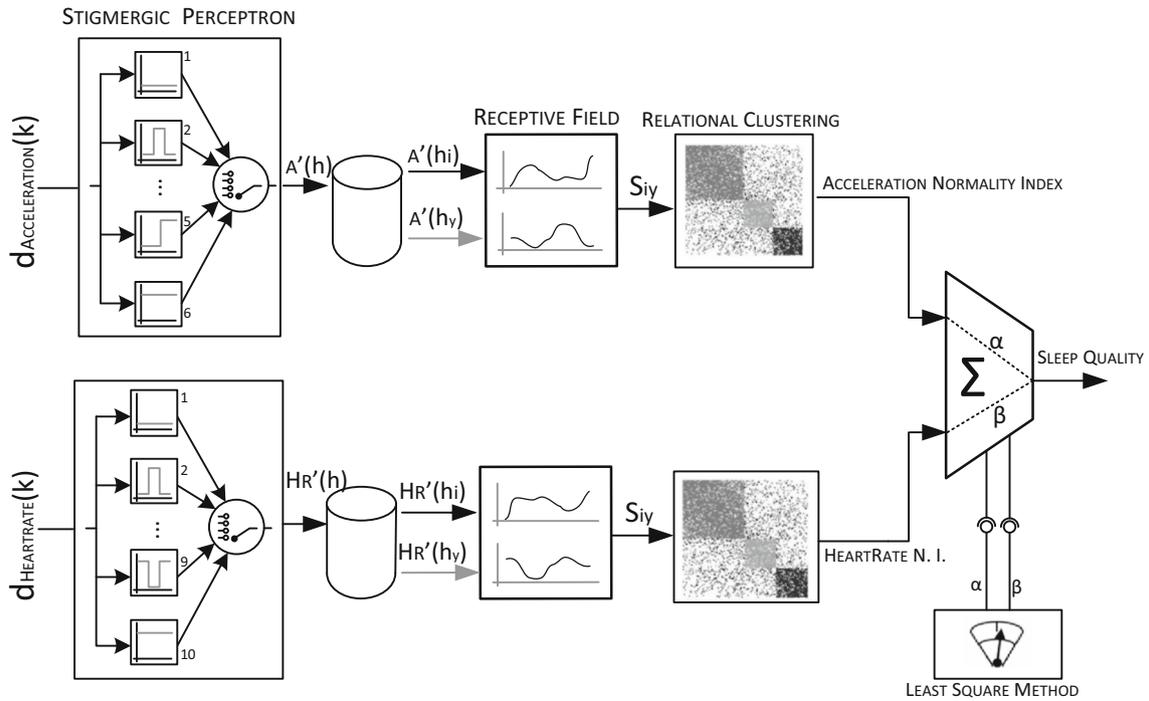


Fig. 7 Multilayer architecture

The proposed hardware and software architecture, composing the data sensing and processing system, aims at providing high flexibility and scalability. From the software point of view, a middleware layer able to collect and dispatch data among generic entities, called services, has been used. This interoperability layer allows the components, which are realized either as hardware devices and software modules, to seamlessly inter-operate with each other by using a shared representation and communication model [7, 52]. The proposed system exploits the presence of this open-source middleware communication platform, enabling the possibility to use different kinds of smart wristbands hence freeing the developers from the specific device/technology. The middleware services, running on the smartphones, act as a bridge between the wearable devices and the data storage infrastructure, providing resilience and consistency features. The presence of the middleware also enables the long-term monitoring scenario and the use of environmental/domotic sensors for future developments of data fusion algorithms for the identification of correlations between sleep quality, daily activities, and the characteristics of the surrounding environment [9].

Regarding the measurement campaign, each subject has been provided with the smartwatch and a sleep diary. During each night in an observation period of 20 nights, he had to wear the smartwatch and start the data gathering software before going to sleep. In the next morning, he had to annotate his perceived sleep quality on the diary, together with light-off and wake time and any nocturnal awakening. After

the observation period, both data gathered and diary annotation are analyzed through sleep stage estimator in order to provide a ground truth for further analysis. Thereafter, the data have been pre-processed and analyzed through SRF-based approach. Further analysis is provided replacing the SRF-based approach with the dynamic time warping (DTW) to compute the distance between each couple of nights.

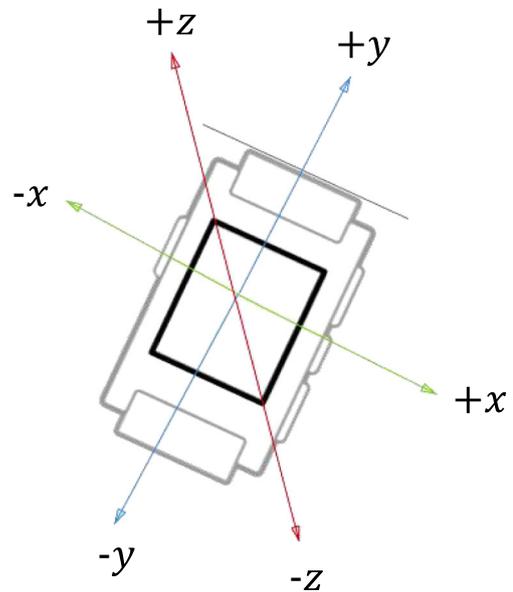


Fig. 8 Reference axes on the android smartwatches

In time series analysis, DTW is a widely used approach for measuring similarity between two temporal sequences [28]. Its main advantage is the robustness with respect to temporal scaling and temporal drift [25]. Sleep dynamics do not have a fixed structure. Nights characterized by the same perceived quality of sleep can have different number, duration, and arrangement of sleep phases. Moreover, each sleep phase corresponds to known behavior in heart rate and motion in the subject. This reason let us assume that DTW robustness will be exploited during our sleep behavior analysis. Furthermore, DTW allows to constraint the similarity computation by fixing the width of the analysis time window, increasing computation accuracy and efficiency [24]. These reasons let us consider DTW a proper competitor with respect to our approach in our analysis. Finally, DTW and SRF-based approach assessment error are compared.

5 Case study and results

The system analyzed data collected in 20 nights by seven subjects: a man aged 72 (subject A), a 22-year-old student (subject B), a woman aged 88 affected by arterial hypertension (subject C), a 36-year-old man (subject D), two 21-year-old students (subjects E and G), and a man aged 30 (subject F). Their perceived sleep quality is reported in Table 3. Using these data and the sleep stage estimator, we extract the sleep quality ground truth, reported in italics in Table 3 (i.e., the nights characterized by a reliably known sleep behavior). This knowledge has been used to select a proper training set for macro-behavior similarity: we randomly selected two nights per user (normal and abnormal, respectively), for a total of 14 nights. The signal segment used as training set for each micro-behavior (archetype) has been selected via visual inspection of each gathered signal. Data have been treated using the proposed SRF and DTW approach in order to obtain a measure of similarity (or distance) between each couple of night.

Figure 9 shows an example of similarity matrix, obtained processing heart rate signals of subject B. Each square box depicts the degree of similarity obtained by matching a couple of signals gathered in two nights. The whitest the box, the higher the degree of similarity. For example, by observing the similarities provided by matching signal gathered during night 4, it is clearly assessed as very different with respect to the most of the one gathered in remaining nights. By exploiting obtained similarity matrix, the clustering of similar signals can be provided. The similarity matrix undergoes the clustering procedure, to assess the quality of each sleep night. The assessment error is provided as the difference between sleep quality and computed quality

Table 3 Perceived sleep quality for each subject.

Night	Subject						
	A	B	C	D	E	F	G
1	A	A	A	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>
2	<i>N</i>	<i>N</i>	<i>N</i>	A	<i>N</i>	A	A
3	A	<i>N</i>	A	<i>N</i>	A	<i>N</i>	A
4	A	A	A	A	A	<i>N</i>	<i>N</i>
5	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	A	<i>N</i>	<i>N</i>
6	<i>N</i>	<i>N</i>	A	<i>N</i>	<i>N</i>	<i>N</i>	A
7	<i>N</i>	<i>N</i>	A	A	<i>N</i>	<i>N</i>	<i>N</i>
8	<i>N</i>	<i>N</i>	A	A	<i>N</i>	A	<i>N</i>
9	A	<i>N</i>	A	<i>N</i>	<i>N</i>	<i>N</i>	A
10	A	<i>N</i>	A	<i>N</i>	A	<i>N</i>	<i>N</i>
11	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	A	<i>N</i>
12	A	<i>N</i>	A	A	<i>N</i>	A	<i>N</i>
13	A	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	A	<i>N</i>
14	<i>N</i>	<i>N</i>	<i>N</i>	A	<i>N</i>	A	A
15	A	<i>N</i>	A	<i>N</i>	A	<i>N</i>	<i>N</i>
16	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	A	A	<i>N</i>
17	<i>N</i>	A	A	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>
18	<i>N</i>	<i>N</i>	A	A	A	A	<i>N</i>
19	A	<i>N</i>	A	<i>N</i>	A	A	<i>N</i>
20	<i>N</i>	<i>N</i>	<i>N</i>	<i>N</i>	A	A	A

The nights included in the ground truth are reported in italics

assessment in each night, considering as 0 the sleep quality in abnormal nights and 1 the same in normal nights.

To show the effectiveness of our approach, we present the difference between the target and the calculated normality index for both the DTW distance and the SRF stigmergic similarity, for each subject and for each sleep night (Fig. 10). In the figures, the bold line represents the normality index

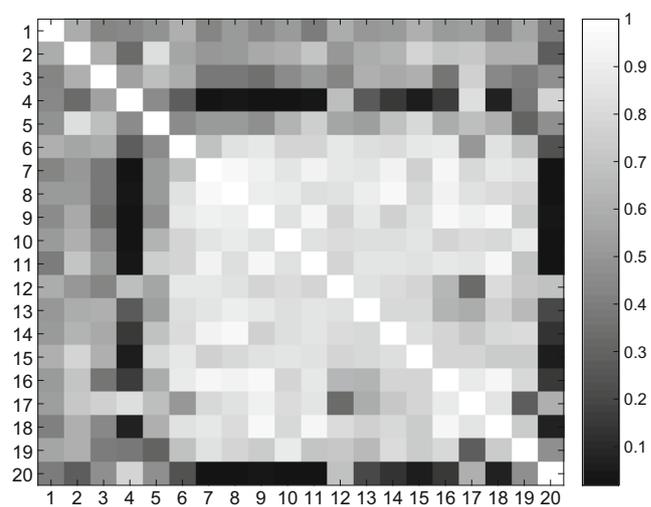


Fig. 9 Similarity matrix obtained processing heart rate of subject B per night

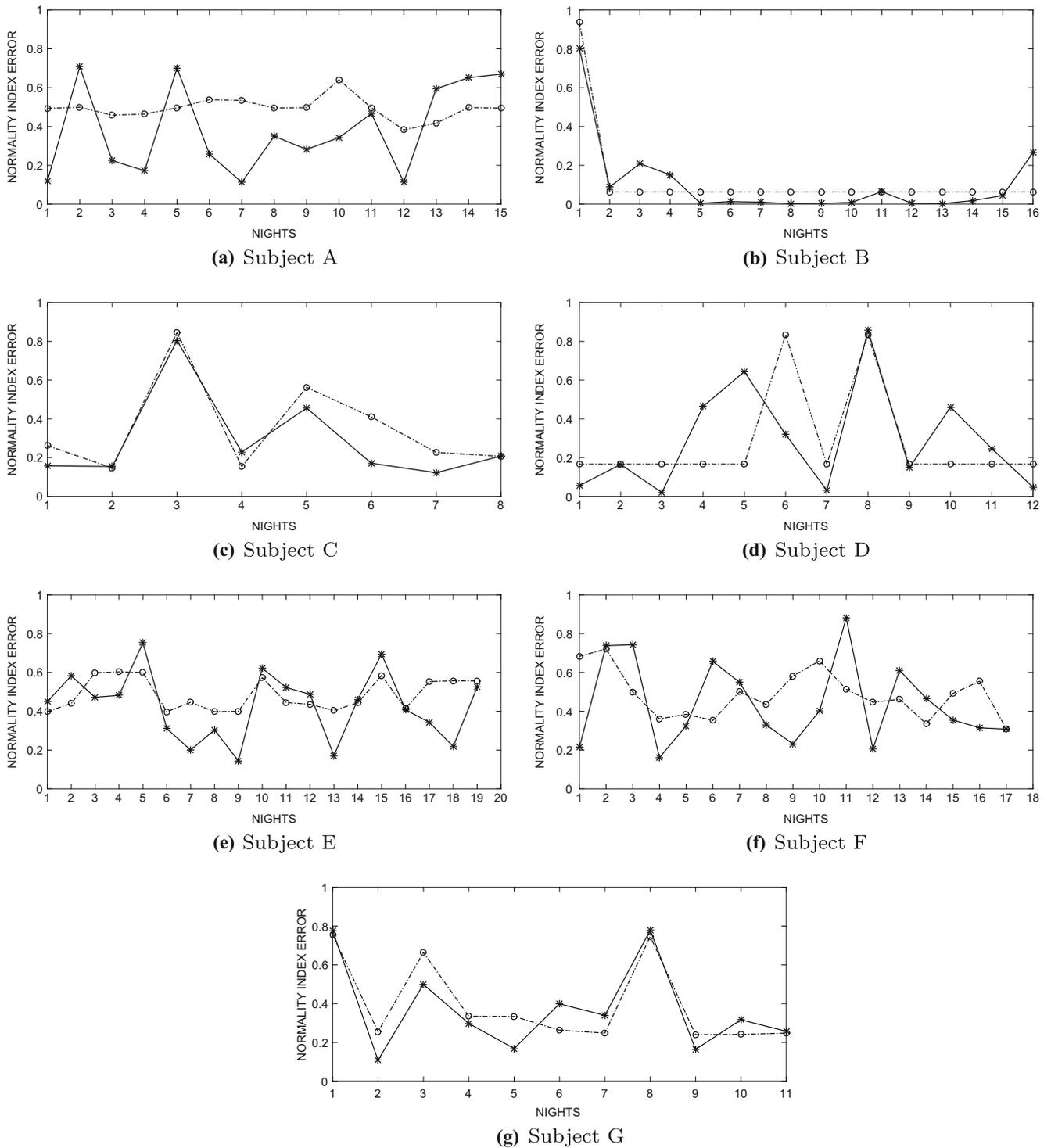


Fig. 10 DTW (dotted line) and SRF (bold line)-based sleep quality assessment error

error obtained by employing SRF, while the dotted line the DTW-based approach. As we can see, the DTW-based approach presents greater errors in most of the nights and subjects, when compared to the ones obtained with the SRF-based approach. We measure the performance obtained by

using the SRF-based approach in terms of assessment, by employing the mean square error computed on the normality index error values. Table 4 shows the mean square error for both approaches, highlighting how the SRF-based approach outperforms the DTW distance.

Table 4 Mean square error in sleep quality assessment via DTW and SRF-based approach

Approach	Subject							Mean	Std
	A	B	C	D	E	F	G		
DTW	0.247	0.059	0.176	0.138	0.243	0.251	0.196	0.187	0.070
SRF	0.195	0.052	0.120	0.112	0.212	0.238	0.185	0.159	0.066

In order to test our system in terms of percentage of correctly classified nights (i.e., accuracy), we approximate the assessment of each night to its nearest integer and compare it with target night classification. To this end, we compare our approach with state-of-the-art classifiers, besides the DTW approach. In particular, we exploited the WEKA data mining software [33] to select the best performing classifiers among the four most used families of learning schemes used in literature for similar scenarios: Bayesian frameworks, function-based (e.g., logistic regression, multilayer perceptron, SGD, SMO), rule-based, and tree-based. We trained the systems using as features the percentage of wake time, the percentage of REM time, and the time to sleep, obtained with the previously introduced sleep stage estimator [10]. The comparing systems have been selected as the best performing of each classification family using a tenfold cross-validation over the entire ground truth dataset. With this method, the original sample is randomly partitioned into ten equal-sized subsamples. Of the ten subsamples, a single subsample is retained as the validation data for testing the model, and the remaining nine subsamples are used as training data. The cross-validation process is then repeated ten times (the folds), with each of the ten subsamples used exactly once as the validation data. The resulting systems chosen for the comparison are as follows:

- Bayesian: *BayesNet*—A Bayes network learning using various search algorithms and quality measures. This algorithm considers two assumptions: nominal values and no missing values. For estimating the conditional probability tables of network, simple estimator and K2

search algorithm are used to run the BayesNet [22]. This Bayes network learning algorithm uses a hill climbing algorithm restricted by the order of the variables.

- Functions: *SGD*—A stochastic gradient descent for learning an SVM-based linear model (Hinge loss) [14, 59]. It globally replaces all missing values and transforms nominal attributes into binary ones. It also normalizes all attributes, so the coefficients in the output are based on the normalized data.
- Rules: *DecisionStump*—It builds simple binary decision “stumps” (1-level decision trees) for both numeric and nominal classification problems. It copes with missing values by extending a third branch from the stump (i.e., treating “missing” as a separate attribute value). It is highly capable of predicting the decision with single input [55].
- Trees: *DecisionTable*—Decision tables are one of the simplest machine learning techniques [40]. Basically, it consists of a hierarchical table in which each entry in the higher level table gets broken down by the values of a pair of additional features to form another table. The DecisionTable approach uses the simplest method of attribute selection: Best First. It searches the space of attributes by greedy hill climbing, augmented with a backtracking facility.

The first row of Table 5 shows the obtained performance for each selected classifier with the tenfold cross-validation method. This validation method considers as overall dataset (i.e., training, test, and validation) the nights of all users

Table 5 Accuracy obtained by each classifier for each user and their averages

Subject	SRF	DTW	BayesNet	SGD	DecisionStump	DecisionTable
10-fold cross-validation	–	–	75.73	74.76	77.67	79.61
A	66.66	80	60	53.33	60	60
B	93.75	93.75	100	100	100	100
C	87.5	75	25	25	12.5	25
D	83.33	83.33	43.75	43.75	43.75	43.75
E	68.42	57.89	52.94	70.59	64.71	64.71
F	64.71	58.82	83.33	91.67	83.33	91.67
G	81.82	72.73	78.95	73.68	78.95	84.21
Average	78.03	74.50	63.42	65.43	63.32	67.05

together. Our aim is, instead, to assess the sleep quality of each particular user independently. To this end, we validated the comparing classifiers using as test set the night of each user and as training set the night of the remaining users. Table 5 shows the accuracies obtained by each classifier for each user and their averages (last row in the table). We can see that our system, except for particular cases, performs better than supervised classifiers even if using a smaller dataset. Furthermore, the proposed solution overcomes the specificity of the selected classifiers (i.e., different classifiers performing better on different users) offering a general framework for the sleep quality assessment. Indeed, on average, it outperforms all the comparing classifiers: $\sim 4\%$ more than DTW and $\sim 11\%$ more than the best performing classifier (DecisionTable).

Summarizing the obtained results in terms of assessment and accuracy, we can observe that some subject exhibits higher variability in sleep behavior which cannot be exhaustively represented by using two behavioral classes. The number of sleep behavioral classes is peculiar for each subject and cannot be generalized, but at least two classes must be provided. Moreover, the analysis of data gathered by subject A provides a lower accuracy with respect to DTW despite having a lower MSE value. This is due to the procedure aimed to obtain a classification value by rounding the normality index [34]. For example, by applying this procedure in the analysis provided via DTW in subject A, it generates a mean rounding errors of 0.47 (defined between [0 and 0.49]), preventing a clear distinction between the two sleep behavioral classes. For this reason, the accuracy is deeply affected by micro-fluctuation in the results.

6 Conclusion

In this paper, we have presented an innovative approach to the assessment of per-night sleep quality. The input data comes from heart rate and arm motion, gathered from a smartwatch. Our approach provides a measure of sleep quality based on the similarity between normal and abnormal nights. The two categories have been generated via fuzzy clustering, using a similarity measure based on computational stigmergy, a biologically inspired model of spatiotemporal aggregation of samples. The fundamental functional unit to compute the proposed similarity is called stigmergic receptive field (SRF). Based on neurocomputing principles, SRFs are arranged into a multilayer architecture. In the first layer, a collection of SRFs is organized into a perceptron and trained to classify each time window of the input signal in terms of activity level. In the second layer, another SRF is trained to compute the similarity measure between sleep nights represented as activity level time series. The approach shows adaptivity with respect to subject's

peculiarity, based on a limited training set. Experimental studies are promising and show that in presented real-world cases, our approach outperforms the DTW distance.

Acknowledgements This work was carried out in the framework of the INTESA project, co-funded by the Tuscany Region (Italy) under the Regional Implementation Programme for Underutilized Areas Fund (PAR FAS 2007-2013) and the Research Facilitation Fund (FAR) of the Ministry of Education, University and Research (MIUR). The authors thank Giovanni Pollina, Silvio Bacci, and Silvia Volpe for their work on the subject during their thesis.

References

- Åkerstedt T, Hume K, Minors D, Waterhouse J (1994) The meaning of good sleep: a longitudinal study of polysomnography and subjective sleep quality. *J Sleep Res* 3(3):152–158
- Cimino AL, Alfeo MGCAGV (2017) Measuring physical activity of older adults via smartwatch and stigmergic receptive fields. In: INSTICC the 6th international conference on pattern recognition applications and methods (ICPRAM 2017), pp 724–730
- Allen J (2007) Photoplethysmography and its application in clinical physiological measurement. *Physiol Measur* 28(3):R1
- American Sleep Disorders Association et al (1995) Practice parameters for the use of actigraphy in the clinical assessment of sleep disorders. *Sleep* 18(4):285–287
- Ansari N, Hou ES, Zhu BO, Chen JG (1996) Adaptive fusion by reinforcement learning for distributed detection systems. *IEEE Trans Aerosp Electron Syst* 32(2):524–531
- Avvenuti M, Cesarini D, Cimino MG (2013) Mars, a multi-agent system for assessing rowers' coordination via motion-based stigmergy. *Sensors* 13(9):12,218–12,243
- Barbon G, Margolis M, Palumbo F, Raimondi F, Weldin N (2016) Taking Arduino to the internet of things: the ASIP programming model. *Comput Commun*
- Barsocchi P, Cimino MG, Ferro E, Lazzeri A, Palumbo F, Vaglini G (2015) Monitoring elderly behavior via indoor position-based stigmergy. *Pervasive Mob Comput* 23:26–42
- Barsocchi P, Ferro E, Fortunati L, Mavilia F, Palumbo F (2014) EMS@CNR: an energy monitoring sensor network infrastructure for in-building location-based services. In: 2014 International conference on high performance computing & simulation (HPCS). IEEE, pp 857–862
- Bernardeschi C, Cimino MG, Domenici A, Vaglini G (2016) Using smartwatch sensors to support the acquisition of sleep quality data for supervised machine learning. In: The 6th EAI international conference on wireless mobile communication and healthcare (MOBIHEALTH 2016). EAI, pp 1–8
- Blackwell T, Redline S, Ancoli-Israel S, Schneider JL, Surovec S, Johnson NL, Cauley JA, Stone KL (2008) Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. *Sleep-New York Then Westchester* 31(2):283
- Bootzin RR, Engle-Friedman M (1981) The assessment of insomnia. *Behav Assess* 3(2):107–126
- Borazio M, Berlin E, Kücüküydiz N, Scholl P, Van Laerhoven K (2014) Towards benchmarked sleep detection with wrist-worn sensing units. In: 2014 IEEE international conference on healthcare informatics (ICHI). IEEE, pp 125–134
- Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In: *Proceedings of COMPSTAT'2010*. Springer, pp 177–186

15. Chen Z, Lin M, Chen F, Lane ND, Cardone G, Wang R, Li T, Chen Y, Choudhury T, Campbell AT (2013) Unobtrusive sleep monitoring using smartphones. In: 2013 7th International conference on pervasive computing technologies for healthcare and workshops. IEEE, pp 145–152
16. Chesson A, Ferber RA, Fry JM, Grigg-Damberger M, Hartse K, Hurwitz T, Johnson S, Littner M, Kader G, Rosen G et al (1997) Practice parameters for the indications for polysomnography and related procedures. *Sleep* 20(6):406–422
17. Choi DJ, Choi MS, Kang SJ (2016) A wearable device platform for the estimation of sleep quality using simultaneously motion tracking and pulse oximetry. In: 2016 IEEE international conference on consumer electronics (ICCE). IEEE, pp 49–50
18. Cimino MG, Lazzeri A, Vaglini G (2015) Improving the analysis of context-aware information via marker-based stigmergy and differential evolution. In: International conference on artificial intelligence and soft computing. Springer, pp 341–352
19. Cimino MG, Lazzerini B, Marcelloni F (2006) A novel approach to fuzzy clustering based on a dissimilarity relation extracted from data using a TS system. *Pattern Recog* 39(11):2077–2091
20. Cola G, Avvenuti M, Musso F, Vecchio A (2016) Gait-based authentication using a wrist-worn device. In: Proceedings of the 13th international conference on mobile and ubiquitous systems: computing, networking and services. ACM, pp 208–217
21. Cola G, Avvenuti M, Vecchio A, Yang GZ, Lo B (2015) An on-node processing approach for anomaly detection in gait. *IEEE Sensors J* 15(11):6640–6649
22. Cooper G, Herskovits E (1992) A Bayesian method for the induction of probabilistic networks from data. *Mach Learn* 9(4):309–347
23. Di Rienzo M, Vaini E, Lombardi P (2015) Wearable monitoring: a project for the unobtrusive investigation of sleep physiology aboard the international space station. In: 2015 computing in cardiology conference (CinC). IEEE, pp 125–128
24. Ding H, Trajcevski G, Scheuermann P, Wang X, Keogh E (2008) Querying and mining of time series data: experimental comparison of representations and distance measures. *Proc VLDB Endow* 1(2):1542–1552
25. Esling P, Agon C (2012) Time-series data mining. *ACM Comput Surv (CSUR)* 45(1):12
26. Espie CA, Lindsay WR, Espie LC (1989) Use of the Sleep Assessment Device (Kelley and Lichstein, 1980) to validate insomniacs' self-report of sleep pattern. *J Psychopathol Behav Assess* 11(1):71–79
27. Fabek G, Mathar R (2008) Kernel-based learning of decision fusion in wireless sensor networks. In: 2008 11th international conference on information fusion. IEEE, pp 1–7
28. Figo D, Diniz PC, Ferreira DR, Cardoso JM (2010) Preprocessing techniques for context recognition from accelerometer data. *Pers Ubiquit Comput* 14(7):645–662
29. Frankel B, Coursey R, Buchbinder R, Snyder F (1976) Recorded and reported sleep in primary chronic insomnia. *Arch Gen Psychiat* 33:615–23
30. Gjoreski M, Gjoreski H, Luštrek M, Gams M (2016) How accurately can your wrist device recognize daily activities and detect falls? *Sensors* 16(6):800
31. Haescher M, Trimpop J, Bieber G, Urban B (2016) Smartmove: a smartwatch algorithm to distinguish between high-and low-amplitude motions as well as doffed-states by utilizing noise and sleep. In: Proceedings of the 3rd international workshop on sensor-based activity recognition and interaction. ACM, p 1
32. Hager GD (2012) Task-directed sensor fusion and planning: a computational approach, vol 99. Springer Science & Business Media
33. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *ACM SIGKDD Explor Newslett* 11(1):10–18
34. Heath MT (2002) Scientific computing: an introductory survey, vol 2. McGraw-Hill, New York
35. Hossain MA, Atrey PK, El Saddik A (2009) Learning multisensor confidence using a reward-and-punishment mechanism. *IEEE Trans Instrum Meas* 58(5):1525–1534
36. Huang X, Oviatt S (2005) Toward adaptive information fusion in multimodal systems. In: International workshop on machine learning for multimodal interaction. Springer, pp 15–27
37. Jeong CW, Joo SC, Jeong YS (2013) Sleeping situation monitoring system in ubiquitous environments. *Person Ubiquit Comput* 17(7):1357–1364
38. Johns MW et al (1991) A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 14(6):540–545
39. Khaleghi B, Khamis A, Karray FO, Razavi SN (2013) Multisensor data fusion: a review of the state-of-the-art. *Inform Fus* 14(1):28–44
40. Kohavi R (1995) The power of decision tables. In: European conference on machine learning. Springer, pp 174–189
41. Liang Z, Ploderer B, Liu W, Nagata Y, Bailey J, Kulik L, Li Y (2016) Sleepexplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Pers Ubiquit Comput* 20(6):985–1000
42. Lockley SW, Skene DJ, Arendt J (1999) Comparison between subjective and actigraphic measurement of sleep and sleep rhythms. *J Sleep Res* 8(3):175–183
43. Majoe D, Bonhof P, Kaegi-Trachsel T, Gutknecht J, Widmer L (2010) Stress and sleep quality estimation from a smart wearable sensor. In: 2010 5th international conference on pervasive computing and applications (ICPCA). IEEE, pp 14–19
44. Mannheim PD, Cascini J, Fein ME, Nierlich SL (1997) Wavelength selection for low-saturation pulse oximetry. *IEEE Trans Biomed Eng* 44(3):148–158
45. Maquet P (2001) The role of sleep in learning and memory. *Science* 294(5544):1048–1052
46. Metsis V, Kosmopoulos D, Athitsos V, Makedon F (2014) Non-invasive analysis of sleep patterns via multimodal sensor input. *Person Ubiquit Comput* 18(1):19–26
47. Miwa H, Sasahara SI, Matsui T (2007) Roll-over detection and sleep quality measurement using a wearable sensor. In: 2007 29th annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 1507–1510
48. Monk T, Buysse D, Kennedy K, Potts J, DeGrazia J, Miewald J (2003) Measuring sleep habits without using a diary: the sleep timing questionnaire (STQ). *Sleep* 26(2):208–12
49. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV (2004) Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep-New York Then Westchester-* 27:1255–1274
50. Ong AA, Gillespie MB (2016) Overview of smartphone applications for sleep analysis. *World J Otorhinolaryngol-Head Neck Surg* 2(1):45–49
51. Palumbo F, Gallicchio C, Pucci R, Micheli A (2016) Human activity recognition using multisensor data fusion based on reservoir computing. *J Ambient Intell Smart Environ* 8(2):87–107
52. Palumbo F, La Rosa D, Chessa S (2014) Gp-m: mobile middleware infrastructure for ambient assisted living. In: 2014 IEEE symposium on computers and communications (ISCC). IEEE, pp 1–6

53. Phan D, Siong LY, Pathirana PN, Seneviratne A (2015) Smart-watch: performance evaluation for long-term heart rate monitoring. In: 2015 International symposium on bioelectronics and bioinformatics (ISBB). IEEE, pp 144–147
54. Preece SJ, Goulermas JY, Kenney LP, Howard D (2009) A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Trans Biomed Eng* 56(3):871–879
55. Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1:81–106
56. Sadeh A (2011) The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev* 15(4):259–267
57. Sadeh A, Hauri PJ, Kripke DF, Lavie P (1995) The role of actigraphy in the evaluation of sleep disorders. *Sleep* 18(4):288–302
58. Shaeffer DK (2013) MEMS inertial sensors: a tutorial overview. *IEEE Commun Mag* 51(4):100–109
59. Shalev-Shwartz S, Singer Y, Srebro N (2007) Pegasos: primal estimated sub-gradient solver for SVM. In: Proceedings of the 24th international conference on machine learning. ACM, pp 807–814
60. Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. In: *Advances in neural information processing systems*, pp 2951–2959
61. Vernon D, Metta G, Sandini G (2007) A survey of artificial cognitive systems: implications for the autonomous development of mental capabilities in computational agents. *IEEE Trans Evol Comput* 11(2):151–180
62. Wilson KG, Watson ST, Currie SR (1998) Daily diary and ambulatory activity monitoring of sleep in patients with insomnia associated with chronic musculoskeletal pain. *Pain* 75(1):75–84
63. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, Colrain IM (2016) Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiol Behav* 158:143–149
64. de Zambotti M, Claudatos S, Inkelis S, Colrain IM, Baker FC (2015) Evaluation of a consumer fitness-tracking device to assess sleep in adults. *Chronobiol Int* 32(7):1024–1028
65. Zhang X, Liu J, Du Y, Lv T (2011) A novel clustering method on time series data. *Expert Syst Appl* 38(9):11,891–11,900