

The importance of physiological data variability in wearable devices for digital health applications

Gloria Cosoli¹, Angelica Poli², Susanna Spinsante², Lorenzo Scalise¹

¹ Department of Industrial Engineering and Mathematical Sciences, Università Politecnica delle Marche, v. Breccie Bianche, 60131 Ancona, Italy

² Department of Information Engineering, Università Politecnica delle Marche, v. Breccie Bianche, 60131 Ancona, Italy

ABSTRACT

This paper aims at characterizing the variability of physiological data collected through a wearable device (Empatica E4), given that both intra- and inter-subject variability play a pivotal role in digital health applications, where Artificial Intelligence (AI) techniques have become popular. Inter-beat intervals (IBIs), ElectroDermal Activity (EDA) and Skin Temperature (SKT) signals have been considered and variability has been evaluated in terms of general statistics (mean and standard deviation) and coefficient of variation. Results show that both intra- and inter-subject variability values are significant, especially when considering those parameters describing how the signals vary over time. Moreover, EDA seems to be the signal characterized by the highest variability, followed by IBIs, contrary to SKT that results more stable.

This variability could affect AI algorithms in classifying signals according to particular discriminants (e.g. emotions, daily activities, etc.), taking into account the dual role of variability: hindering a net distinction between classes, but also making algorithms more robust for deep learning purposes thanks to the consideration of a wide test population. Indeed, it is worthy to note that variability plays a fundamental role in the whole measurement chain, characterizing data reliability and impacting on the final results accuracy and consequently on decision-making processes.

Section: RESEARCH PAPER

Keywords: Wearable devices; physiological measurements; data variability; physiological monitoring

Citation: Gloria Cosoli, Angelica Poli, Susanna Spinsante, Lorenzo Scalise, The importance of physiological data variability in wearable devices for digital health applications, Acta IMEKO, vol. 11, no. 2, article 25, **Citation:** Gloria Cosoli, Angelica Poli, Susanna Spinsante, Lorenzo Scalise, The importance of physiological data variability in wearable devices for digital health applications, Acta IMEKO, vol. 11, no. 2, article 25, June 2022, identifier: IMEKO-ACTA-11 (2022)-02-25

Section Editor: Francesco Lamonaca, University of Calabria, Italy

Received July 13, 2021; **In final form** March 21, 2022; **Published** June 2022

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Corresponding author: Gloria Cosoli, e-mail: g.cosoli@staff.univpm.it

1. INTRODUCTION

The use of wearable devices is constantly spreading all over the world, thanks to their wide accessibility and high ease of use [1] (even if further actions and improvements are still needed to overcome barriers for a larger adoption by older adults [2]). Nowadays a continuously growing number of people wear a smartwatch monitoring a plethora of physiological parameters: heart rate (HR) [3], energy expenditure (EE) [4], blood volume pulse signal (BVP) [5], electrodermal activity (EDA) [6], acceleration signal [7], sleep quality [8], respiration rate [9], stress-related indices [10], etc. These measurements can be useful for different purposes, from cardiovascular monitoring [11] to sleep tracking [12], through activity assessment [13], fitness-oriented applications [14] and blood pressure observation [15], just to cite

some. Furthermore, in the recent months wearable devices have expanded their application also to the possible detection of early symptoms related to SARS-CoV-2 pandemic [16], since this virus has stressed the importance of remote monitoring both to limit contagion and for “testing, tracking and tracing” strategies [17]. However, there are also critical aspects that should be thoroughly considered, pertaining to health-related data privacy issues and measurement accuracy of these innovative wearable instruments [5], which undoubtedly play important roles in the era of personalized medicine and digital health [18], [19].

Physiological signals can be collected through wearable devices 24 hours a day, 7 days a week, producing big amounts of data, which are analysed through Artificial Intelligence (AI) algorithms more and more frequently, in order to provide useful information for the so-called decision-making processes [20],

[21], thus supporting human choices in different fields, from Industry 4.0 [22], [23] to eHealth [24]. The purposes can be different: emotion classification [25], activity recognition [26], hypertension management [27], fall detection [28], smart living environments and well-being assessment [29], and so on. In order to be able to develop robust models, capable to provide reliable information, data quality is fundamental [30]; in this perspective, not only hardware and acquisition options (e.g. sampling frequency, signal-to-noise ratio (SNR), resolution, etc.) have a big impact, but also data variability, linked both to different sources to collect data [31], but also to the physiological variability itself. Indeed, the classification performance of AI algorithms surely depends on the variability observed in the data collected on the test population: if it is true that (physiological) variability somehow hinders a perfect discrimination among classes, on the other hand it is necessary to test a wide population in order to include its variability and avoid overfitting issues. These aspects should be thoroughly considered when developing AI algorithms for digital health applications, which cannot neglect physiological variability characterising the involved population, and consequently the measured data.

The study reported in this manuscript aims at evaluating the intra- and inter-subject variability of different physiological signals collected through a wearable wrist-worn device (Empatica E4). In particular, the authors have analysed cardiac-related parameters (i.e. heart rate variability – HRV – parameters computed on the BVP signal measured through a photoplethysmographic – PPG – sensor), features computed on EDA signal and skin temperature (SKT) values. Mean, standard deviation and coefficient of variation have been computed for each extracted parameter, considering the repeated tests on a same subject to evaluate intra-subject variability, and the whole acquired data for inter-subject variability.

The rest of the paper is organized as follows: Section 2 describes the materials and methods employed for data acquisitions and for the evaluation of data variability, Section 3 reports the intra- and inter-subject variability results, and finally in Section 4 the authors provide their considerations and conclusions.

2. MATERIALS AND METHODS

2.1. Participants

The study was conducted on 10 healthy volunteers: 3 males, 7 females; age of (33 ± 16) years with a range of (15 - 59) years; height of (169.78 ± 8.83) cm; weight of (66.55 ± 12.00) kg; BMI of (22.92 ± 2.14) kg/m^2 – data are reported as mean \pm standard deviation. They declared they did not take any medication in the 24 hours preceding the tests, nor had particular clinical histories possibly influencing the results. Before starting the tests, each participant was informed on the test purpose and procedure and signed an informed consent according to the European Regulation 2016/679, i.e., the General Data Protection Regulation (GDPR) to obtain the permission for processing personal data.

2.2. Data collection

In order to assess the inter-subject and intra-subject variability of physiological parameters, each subject repeated the acquisitions six times, for a total of 60 recordings, each lasting 5 minutes. Ambient temperature and relative humidity were equal to (20 ± 2) $^{\circ}\text{C}$ and (50 ± 5) %, respectively, to be perceived as comfortable by most of the involved individuals. The



Figure 1. Measurement setup.

participants (with a skin colour classification of Type II – Fitzpatrick scale), laying comfortably in a supine position (i.e., in rest condition) in a quiet room, were instructed to relax as much as possible, breathe normally, and not talk during recordings, in order to minimize movement artifacts. As shown in Figure 1, the physiological signals were simultaneously collected through a multisensory wearable device, namely Empatica E4 [32], placed on the dominant wrist. This acquisition device was chosen as it provides the raw data, thus resulting particularly suitable for research purposes. Firstly, the participants were allowed to adjust the device positioning to increase the comfort feeling. Then, the device placement was verified to ensure the optimal skin contact (not worn too tightly or too loosely), and consequently to guarantee the optimal conditions for reliable PPG sensor acquisition [33] and, therefore, as high as possible data quality.

2.3. Data acquisition device

Individual physiological signals were recorded with the multimodal device Empatica E4 (Class IIA Medical Device according to the 93/42/EEC Directive) – firmware version: FW 3.1.0.7124. Such a device captures the Inter-Beat-Interval (IBI), BVP, EDA, human SKT, and 3-axis accelerometer signals. In particular, BVP and IBI signals, both sampled at 64 Hz with a resolution of 0.9 nW/Digit, are derived from the PPG sensor. On the bottom of the wristband, there are two green light emitting diodes (LEDs) enabling the measurements of blood volume changes and heartbeats, and two red LEDs for reducing the motion artifacts. Additionally, two units of photodiodes (total 14 mm^2 sensitive area) measure the reflected light. On the bracelet band of Empatica E4, two Ag/AgCl electrodes allow to pass a small amount of alternating current (frequency 8 Hz, with a maximum peak-to-peak value of 100 μA) for measuring the skin conductance in μS , sampled at 4 Hz with a resolution of 900 pS in the range of [0.01, 100] μS . At the same sampling frequency (4 Hz), an infrared thermopile, placed on the back of the case, records the SKT data in $^{\circ}\text{C}$ with an accuracy of ± 0.20 $^{\circ}\text{C}$ (within the range 36 $^{\circ}\text{C}$ - 39 $^{\circ}\text{C}$), and a resolution of 0.02 $^{\circ}\text{C}$. Calibration is valid in the range [-40, 115] $^{\circ}\text{C}$. The last sensor is a 3-axial MEMS accelerometer used to collect the acceleration along the three dimensions X, Y, Z with a 32 Hz sampling frequency and a default measurement range of ± 2 g. In this case the resolution of the output signal is 0.015 g (8 bit). A dedicated mobile application (E4 Realtime) was used to stream and view data in real-time on a mobile device connected with Empatica E4 via Bluetooth Low Energy (BLE). Following each measurement session, data were automatically transferred to a cloud repository (Empatica Connect) to view, manage, and download raw data in .csv format in the post-processing phase of the study.

2.4. Data analysis

As mentioned above, in this study the data variability analysis was conducted on HRV (or, more precisely, on Pulse Rate Variability [34], [35]), EDA and SKT signals, previously processed in MATLAB environment in order to extract relevant features. Regarding the HRV evaluation, after applying a previously developed artifact correction method [36], the analysis was performed on IBIs signals by using the Kubios toolbox [37]. Seven meaningful HRV-related parameters were extracted from the corrected IBIs signals in time domain (Table 1), namely: mean and standard deviation of IBIs; mean, standard deviation, minimum and maximum values of HR; root mean square of successive RR interval differences. Frequency domain parameters were not considered in the present work, to limit the number of parameters extracted from the same signal, and also because the parameters in frequency domain can be strongly affected by spurious components linked to movement artifacts, to which wrist-worn wearable devices are prone [38], even more during intense physical activities [39]. Concerning EDA data, the Bio-SP toolbox [40] was used to pre-process the signals and to extract all the features that the toolbox permits to compute. Indeed, EDA signal is composed by the superimposition of two components, specifically skin conductance response (SCR) and skin conductance level (SCL), related to the fast response to external stimuli events and the slow changes in baseline levels, respectively. This means that the SCL depends on the individual characteristics (e.g. skin condition), and can differ markedly between individuals. Consequently, under rest condition with no external stimuli, the SCL has a higher impact than the SCR component on both EDA signal trend and amplitude. According to the literature [41], a Gaussian low-pass filter, with a 40-point window and a sigma of 400 ms, was applied to reduce noise and motion artifacts due to potential subject's wrist movements. In order to characterize the EDA signal, the following five features were computed within Bio-SP toolbox in time domain (Table 1): SCR mean duration, SCR mean amplitude, SCR mean rise-time, EDA mean signal, number of SCRs. Finally, since an inspection of the SKT data revealed slight and slow °C changes at rest, no

filters were applied. Therefore, from the raw SKT signal the following parameters were extracted (Table 1): mean and standard deviation, minimum and maximum of skin temperature.

Once the whole set of features was computed and extracted from the considered signals, both intra- and inter-subject variability was evaluated for each metric. More specifically, data variability was estimated by computing the mean (μ), standard deviation (σ) and coefficient of variation ($c_v = \sigma/\mu$) for all the extracted features. Furthermore, the normality of the parameters distributions was verified by means of Shapiro-Wilk test [42] (null hypothesis: the test population is normally distributed; p -value ≤ 0.05 considered as statistically significant).

3. RESULTS

In this section, results are reported by grouping them according to data type: cardiac related parameters (i.e. HRV analysis parameters, Subsection 3.1), EDA-related parameters (Subsection 3.2) and skin temperature parameters (Subsection 3.3).

Results are reported in tables as $\mu \pm \sigma (c_v)$; some examples of mean distributions are also shown by using the histogram representation.

3.1. HRV parameters

The authors analysed the variability of HRV signal at parameters level, focusing on those extracted in time domain. The Shapiro-Wilk test evidenced that RR_mean , HR_mean , HR_min and HR_max can be considered as normally distributed (p -value ≥ 0.05). An example of the distribution is reported in the histogram (Figure 2) related to RR_mean parameter. For the others (i.e. RR_std and $RMSD$), the null hypothesis cannot be rejected; the reason could be found in the limited numerosity of the test population (60 recordings on 6 subjects). Similarly, the HR_std resulted to have non-normal distribution, probably also due to the presence of one outlier subject (i.e. subject no. 6, see Table 2).

Observing the variability results in Table 2, it is possible to notice a very high variability, in particular for the parameters

Table 1. Time-domain features extracted from the physiological signals acquired in the tests.

Signal	Features	Measurement unit	Description
HRV	RR_mean	ms	Mean value of inter-beat intervals
	RR_std	ms	Standard deviation of inter-beat intervals
	HR_mean	bpm	Mean value of heart rate
	HR_std	bpm	Standard deviation of heart rate
	HR_min	bpm	Minimum value of heart rate
	HR_max	bpm	Maximum value of heart rate
	$RRMSD$	ms	Root mean square of successive inter-beat intervals
EDA	SCR_D_mean	s	Mean duration of skin conductance response signal
	SCR_A_mean	μS	Mean amplitude of skin conductance response signal
	SCR_RT_mean	s	Mean rise time of skin conductance response signal
	EDA_mean	μS	Mean value of EDA signal
	SCR_n	-	No. of skin conductance response peaks
SKT	SKT_mean	°C	Mean value of skin temperature
	SKT_std	°C	Standard deviation of skin temperature
	SKT_min	°C	Minimum value of skin temperature
	SKT_max	°C	Maximum value of skin temperature

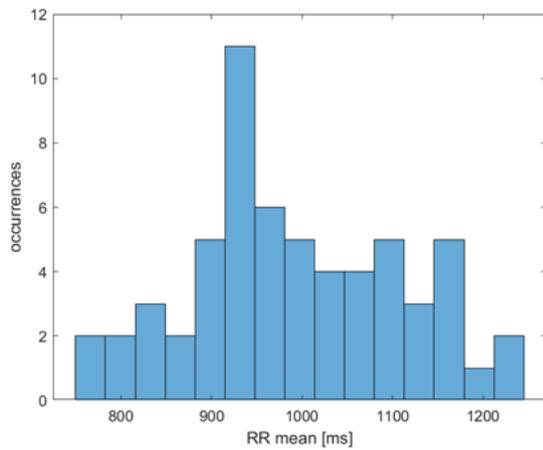


Figure 2. Histogram related to RR_mean parameter (HRV signal).

describing how the measurement oscillates around its mean value, i.e. the standard deviation values of *RR*, *HR* and *RMSSD*, reporting inter-subject variabilities of 55.8 %, 126.9 % and 65.7 %, respectively. This seems to underline the physiological variability, hence a subject's condition of interest cannot be described (and classified) without properly considering such data variability. A particular remark should be made on the extremely high inter-variability of *HR_std* parameter; indeed, this could be linked to the subject no. 6 reporting an extremely high variability (i.e., 125.1 %), as already mentioned above. Indeed, by performing a visual inspection of data collected on subject no. 6, among the tests conducted, one measurement on this subject resulted particularly noisy, hindering a reliable HRV analysis despite the use of proper artifact correction methods in the pre-processing phase. However, if this test is discarded from the variability analysis, the intra-variability of *HR_std* reduces from (12 ± 15) bpm (125.1 %) to (6 ± 3) bpm (50.0 %) – while the remaining parameters do not vary substantially; in this way, the inter-subject variability related to *HR_std* parameter would be (4 ± 3) bpm (81.1 %). The observed noise, which quite often characterises signals acquired through PPG sensors of wearable devices, could be an effect of subjects' wrist movements [38].

Intra-subject variability shows similar results, evidencing a very high variability, especially for the standard deviation parameters, describing the variations over time. On the other hand, mean value parameters show a quite low intra-subject

variability, with values often lower than 10 % (e.g. for *RR_mean* parameter, lower than 10 % with the exception of 3 subjects out of 10).

3.2. EDA parameters

As stated above, in rest conditions SCL is the predominant component of EDA signal; this can result in very low intensity signals related to SCR component (more linked to eventual stimuli), and consequently the *EDA_mean* parameter values are expected to be low.

In fact, in Table 3, *EDA_mean* parameters show very low values, up to 0.0005 μ S for the subject no. 9. Such very low mean values, together with high signal variability (i.e. high standard deviation), result in extremely high coefficients of variation (see for example subjects no. 3 and 9, where c_v is extremely high due to the fact that the mean value of signal is an order of magnitude lower than its standard deviation).

More in general, the parameters related to the EDA signals show a very high variability, with coefficient of variation values related to inter-subject variability often over 100 %. Also, intra-subject variability seems to be extremely high, evidencing that EDA signal is not stable over time, hence it should be considered in this long-term evolution, instead of limiting to use descriptive statistics. Such a high variability could be attributable to the fact that EDA measurements at total rest, with no external stimuli, seem to be quite complicated, especially when performed by means of wearable devices. In fact, there are multiple subjective causes influencing the measurement results. Furthermore, it should be considered that wrist EDA results to be quite different from standard finger EDA [43]. Regarding the type of distribution, no features extracted from EDA can be considered as normally distributed. The reason could be attributed again to the restricted test population. An example of distribution is reported in the histogram (Figure 3) for *EDA_mean* parameter.

3.3. Skin temperature parameters

Contrarily to the previously reported parameters, skin temperature (Table 4) shows measures slowly varying over time (with the exception of the standard deviation value, evidencing a higher variability – up to 87.1 % in intra-subject results), hence providing a more precise footprint of a subject in a determined condition. On the other hand, this could mean that the wrist skin temperature has a slow dynamic, thus it could be not suitable to rapidly mirror possible changes in the subject's psycho-physical

Table 2. Variability of HRV parameters in time domain. Results are reported as $\mu \pm \sigma (c_v)$.

Subject	<i>RR_mean</i> in ms	<i>RR_std</i> in ms	<i>HR_mean</i> in bpm	<i>HR_std</i> in bpm	<i>HR_min</i> in bpm	<i>HR_max</i> in bpm	<i>RMSSD</i> in ms
1	1044 \pm 66 (6.3 %)	70 \pm 43 (61.0 %)	58 \pm 4 (6.5 %)	4 \pm 2 (49.9 %)	50 \pm 4 (8.8 %)	65 \pm 6 (8.7 %)	94 \pm 64 (68.6 %)
2	1152 \pm 30 (2.6 %)	36 \pm 12 (33.7 %)	52 \pm 1 (2.5 %)	2 \pm 1 (48.5 %)	49 \pm 2 (3.3 %)	58 \pm 4 (6.1 %)	46 \pm 16 (33.7 %)
3	934 \pm 44 (4.7 %)	40 \pm 13 (33.6 %)	64 \pm 3 (4.8 %)	3 \pm 1 (39.9 %)	59 \pm 5 (7.8 %)	76 \pm 5 (6.1 %)	49 \pm 18 (36.8 %)
4	1008 \pm 80 (8.0 %)	83 \pm 51 (61.1 %)	60 \pm 5 (8.1 %)	6 \pm 5 (87.3 %)	49 \pm 7 (13.6 %)	70 \pm 5 (7.8 %)	114 \pm 69 (60.1 %)
5	1027 \pm 80 (7.8 %)	39 \pm 20 (51.9 %)	59 \pm 5 (8.0 %)	3 \pm 2 (72.0 %)	53 \pm 3 (5.0 %)	64 \pm 6 (9.7 %)	54 \pm 30 (56.1 %)
6	991 \pm 133 (13.5 %)	72 \pm 26 (35.5 %)	62 \pm 9 (14.2 %)	12 \pm 15 (125.1 %)*	52 \pm 8 (15.1 %)	74 \pm 12 (16.5 %)	97 \pm 42 (43.2 %)
7	938 \pm 31 (3.4 %)	68 \pm 28 (40.5 %)	64 \pm 2 (3.4 %)	6 \pm 5 (81.3 %)	55 \pm 5 (9.1 %)	74 \pm 4 (4.8 %)	90 \pm 44 (48.9 %)
8	873 \pm 44 (5.0 %)	48 \pm 2 (4.9 %)	69 \pm 3 (5.0 %)	4 \pm 1 (10.4 %)	61 \pm 3 (4.2 %)	79 \pm 6 (6.9 %)	45 \pm 2 (4.7 %)
9	1083 \pm 139 (12.9 %)	31 \pm 7 (23.7 %)	56 \pm 7 (12.8 %)	2 \pm 1 (24.8 %)	53 \pm 7 (12.8 %)	61 \pm 7 (11.5 %)	39 \pm 7 (18.6 %)
10	885 \pm 130 (14.7 %)	49 \pm 19 (39.6 %)	69 \pm 10 (14.7 %)	4 \pm 1 (16.9 %)	62 \pm 9 (15.3 %)	83 \pm 7 (8.7 %)	43 \pm 66 (39.0 %)
Tot.	993 \pm 117 (11.8 %)	54 \pm 30 (55.8 %)	61 \pm 7 (12.0 %)	4 \pm 6 (126.9 %)*	54 \pm 7 (12.9 %)	70 \pm 10 (14.2 %)	67 \pm 4 (65.7 %)

* Results affected by a particularly noisy test performed on subject no.6

Table 3. Variability of EDA parameters. Results are reported as $\mu \pm \sigma$ (c_v).

Subject	SCR_D_mean in s	SCR_A_mean in μ S	SCR_RT_mean in s	EDA_mean in μ S	SCR_n
1	10.2 \pm 5.0 (49.4 %)	0.0097 \pm 0.0034 (34.8 %)	5.3 \pm 2.6 (48.7 %)	0.0022 \pm 0.0032 (144.5 %)	23 \pm 7 (30.5 %)
2	26.1 \pm 26.1 (99.8 %)	0.0138 \pm 0.0056 (72.7 %)	13.1 \pm 10.6 (80.3 %)	0.0030 \pm 0.0052 (177.1 %)	14 \pm 10 (71.2 %)
3	8.3 \pm 5.4 (65.0 %)	0.0095 \pm 0.0056 (59.3 %)	4.3 \pm 2.9 (68.6 %)	0.0010 \pm 0.019 (1980.1 %)	12 \pm 9 (77.7 %)
4	12.4 \pm 17.5 (140.9 %)	0.0098 \pm 0.0079 (80.8 %)	9.4 \pm 15.8 (167.9 %)	0.0078 \pm 0.010 (133.0 %)	21 \pm 16 (76.6 %)
5	18.9 \pm 19.2 (101.7 %)	0.0115 \pm 0.0044 (38.1 %)	5.2 \pm 3.4 (65.5 %)	0.0027 \pm 0.0044 (131.4 %)	22 \pm 13 (59.5 %)
6	15.2 \pm 10.0 (65.6 %)	0.0112 \pm 0.0051 (45.6 %)	9.3 \pm 7.4 (79.5 %)	0.0043 \pm 0.0051 (50.1 %)	18 \pm 7 (41.0 %)
7	14.5 \pm 14.4 (99.2 %)	0.0105 \pm 0.0090 (85.7 %)	10.6 \pm 12.2 (115.1 %)	0.0066 \pm 0.0090 (97.4 %)	16 \pm 14 (85.5 %)
8	5.7 \pm 1.3 (22.5 %)	0.0257 \pm 0.0160 (62.7 %)	3.1 \pm 0.7 (23.1 %)	0.0029 \pm 0.0160 (126.9 %)	23 \pm 9 (38.0 %)
9	4.9 \pm 0.9 (18.1 %)	0.0116 \pm 0.0061 (52.9 %)	2.7 \pm 0.5 (18.4 %)	0.0005 \pm 0.0061 (862.7 %)	30 \pm 9 (31.3 %)
10	5.6 \pm 1.2 (20.5 %)	0.0144 \pm 0.0065 (45.6 %)	2.9 \pm 0.6 (19.3 %)	0.0029 \pm 0.0065 (137.9 %)	30 \pm 8 (27.9 %)
Tot.	12.2 \pm 13.7 (112.3 %)	0.0128 \pm 0.0089 (69.3 %)	6.6 \pm 7.9 (120.0 %)	0.0034 \pm 0.0076 (224.3 %)	21 \pm 11(54.6 %)

conditions. None of the parameters extracted from SKT signal can be considered normally distributed, according to the Shapiro-Wilk test. The reason could be the same indicated for the other signal (i.e. narrow test population). An example is reported in the histogram (Figure 4) for *SKT_mean* (where the distribution skewness is markedly < 0).

4. DISCUSSION AND CONCLUSIONS

The use of wearable devices in a growing number of application fields emphasizes the need of considering the metrological aspects determining the reliability of measurement results. In recent years, AI algorithms have known unprecedented developments, providing extremely powerful tools to support decision-making processes and thus prevent serious health-issues in a variety of digital health applications, among which affective states classification, eHealth, smart living environments and ambient assisted living. In order to obtain good performances from AI algorithms, data accuracy and data quality are of uttermost importance, along with data variability that undoubtedly represents a key factor in this scenario. Furthermore, only a part of variability can be minimised (e.g. by correcting the sensor positioning in the data acquisition phase), but another part is inevitable and uncontrollable, given that there is a physiological variability, whose values cannot be disregarded. It is a matter of fact that all the steps of the measurement chain influence the final results of AI algorithms: from the sensors

uncertainties to the data variability and accuracy, all influencing the reliability of the output information. In a real-life context, this contributes to reveal a corrupted output with a poor information quality, which could be used for different final purposes (e.g. support to decision-making processes in digital health scenarios) [29].

The results obtained in this study have highlighted the physiological data variability among different subjects and intra-subject, considering data acquired by means of a wearable device. In particular, HRV and EDA signals have been firstly analysed, observing that HRV parameters in time domain exhibit higher inter-subject variability when considering measures describing their variations over time (i.e. standard deviation values), with respect to average values, which seem more stable. Furthermore, EDA signals appear to be extremely changeable even in a same subject, evidencing the intrinsic variable nature of this type of data. Indeed, this type of signal is referred to wrist skin conductance, instead of finger one, which is the site generally used for standard measurements. Previous studies underlined that the measurement is not reliable if compared to finger/palm skin conductivity [44]; in fact, thermoregulatory processes would affect the results more than psychophysiological phenomena, which on the contrary are more influencing in the standard measurement sites [45].

On the other hand, other types of physiological data, such as SKT, can show a quite limited variability, resulting more stable than HRV and EDA. However, the slow changes could be

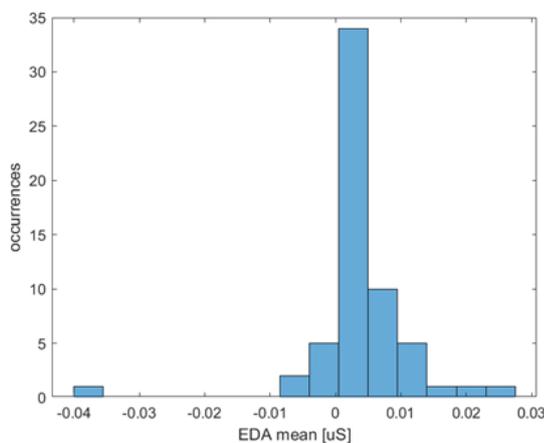


Figure 3. Histogram related to EDA_mean parameter (EDA signal).

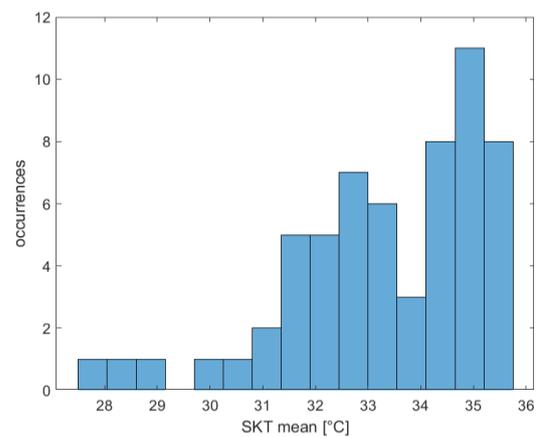


Figure 4. Histogram related to SKT_mean parameter (SKT signal).

Table 4. Variability of SKT parameters. Results are reported as $\mu \pm \sigma$ (c_v).

Subject	SKT_mean in °C	SKT_std in °C	SKT_min in °C	SKT_max in °C
1	33.20 ± 1.58 (4.8 %)	0.17 ± 0.12 (71.7 %)	32.81 ± 1.87 (5.7 %)	33.46 ± 1.53 (4.6 %)
2	31.99 ± 2.20 (6.7 %)	0.21 ± 0.15 (68.8 %)	31.58 ± 2.13 (6.8 %)	32.37 ± 2.29 (7.1 %)
3	34.02 ± 1.25 (3.7 %)	0.07 ± 0.04 (56.5 %)	33.90 ± 1.28 (3.8 %)	34.20 ± 1.25 (3.6 %)
4	32.67 ± 1.62 (4.9 %)	0.13 ± 0.07 (56.3 %)	32.37 ± 1.70 (5.3 %)	32.89 ± 1.55 (4.7 %)
5	33.05 ± 1.02 (3.1 %)	0.12 ± 0.07 (60.5 %)	32.80 ± 1.15 (3.5 %)	33.30 ± 0.98 (3.0 %)
6	32.14 ± 2.04 (6.4 %)	0.09 ± 0.04 (46.0 %)	31.88 ± 2.17 (6.8 %)	32.27 ± 2.03 (6.3 %)
7	32.02 ± 2.07 (6.5 %)	0.10 ± 0.09 (87.1 %)	31.83 ± 2.05 (6.4 %)	32.22 ± 2.21 (6.9 %)
8	35.22 ± 0.24 (0.7 %)	0.06 ± 0.04 (55.6 %)	35.09 ± 0.24 (0.7 %)	35.34 ± 0.24 (0.7 %)
9	35.32 ± 0.24 (0.7 %)	0.05 ± 0.03 (76.5 %)	35.23 ± 0.26 (0.7 %)	35.43 ± 0.27 (0.8 %)
10	33.95 ± 1.09 (3.2 %)	0.05 ± 0.02 (39.6 %)	33.86 ± 1.08 (3.2 %)	34.07 ± 1.14 (3.4 %)
Tot.	33.36 ± 1.82 (5.4 %)	0.11 ± 0.09 (83.7 %)	33.14 ± 1.91 (5.8 %)	33.55 ± 1.80 (5.4 %)

problematic in following, for example, the subject's reactions to external stimuli.

The observed variability can represent a double sword edge: on one hand, the subjective diversity can hinder a net classification by means of AI algorithms; on the other hand, considering a test population sufficiently wide to include all the characteristic variability is required to develop robust AI algorithms, not suffering from overfitting issues.

It is worthy to underline that the test population of this study is quite limited (10 subjects), therefore the normality condition could be non-optimally satisfied (verification through Shapiro-Wilk test). It would be interesting to repeat this kind of analysis on some publicly available large-scale databases (e.g. WESAD [46], k-EmoCon [47], TILES [48], etc.), in order to examine the data variability results on wider populations (possibly including also different age groups) and also considering longer acquisition intervals and different measuring devices and acquisition conditions (e.g. free-living conditions, which probably remark variability). Additionally, future studies may include one or more AI algorithms to compare the achieved performance on two datasets with different variabilities, for demonstrating the high impact of data variability on AI algorithms outputs, which can consequently impact on decision-making processes.

ACKNOWLEDGEMENT

A. P. and S. S. gratefully acknowledge the support of the Italian Ministry for Economic Development (MiSE) in implementation of the financial programme "Research and development projects for the implementation of the National Smart Specilization strategy – "DM MiSE 5 Marzo 2018", project "ChAALenge", proposal no. 493, project nr. F/180016/01-05/X43.

REFERENCES

[1] G. Cosoli, S. Spinsante, L. Scalise, Wrist-worn and chest-strap wearable devices: systematic review on accuracy and metrological characteristics, *Measurement*, Apr. 2020, p. 107789. DOI: [10.1016/j.measurement.2020.107789](https://doi.org/10.1016/j.measurement.2020.107789)

[2] S. Farivar, M. Abouzahra, M. Ghasemaghaci, Wearable device adoption among older adults: A mixed-methods study, *Int. J. Inf. Manage.*, vol. 55, 2020, p. 102209. DOI: [10.1016/j.ijinfomgt.2020.102209](https://doi.org/10.1016/j.ijinfomgt.2020.102209)

[3] N. Morresi, S. Casaccia, M. Sorcinelli, M. Arnesano, G. M. Revel, Analysing performances of Heart Rate Variability measurement through a smartwatch, 2020 IEEE International Symposium on

Medical Measurements and Applications (MeMeA), Bari, Italy, 1 June-1 July 2020, pp. 1–6. DOI: [10.1109/MeMeA49120.2020.9137211](https://doi.org/10.1109/MeMeA49120.2020.9137211)

[4] S. Levikari, A. Immonen, M. Kuisma, H. Peltonen, M. Silvennoinen, H. Kyröläinen, P. Silventoinen, Improving Energy Expenditure Estimation in Wrist-Worn Wearables by Augmenting Heart Rate Data with Heat Flux Measurement, *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, 8 pp. DOI: [10.1109/TIM.2021.3053070](https://doi.org/10.1109/TIM.2021.3053070)

[5] G. Cosoli, G. Iadarola, A. Poli, S. Spinsante, Learning classifiers for analysis of Blood Volume Pulse signals in IoT-enabled systems, in *IEEE MetroInd4.0&IoT*, Rome, Italy, 7-9 June 2021, pp. 307-312. DOI: [10.1109/MetroInd4.0IoT51437.2021.9488497](https://doi.org/10.1109/MetroInd4.0IoT51437.2021.9488497)

[6] S. Cecchi, A. Piersanti, A. Poli, S. Spinsante, Physical Stimuli and Emotions: EDA Features Analysis from a Wrist-Worn Measurement Sensor, *IEEE Int. Workshop on Computer Aided Modeling and Design of Communication Links and Networks, CAMAD*, Pisa, Italy, 14-16 September 2020, pp. 1-6. DOI: [10.1109/CAMAD50429.2020.9209307](https://doi.org/10.1109/CAMAD50429.2020.9209307)

[7] C. John, Z. Mueller, L. Prayaga, K. Devulapalli, A neural network model to identify relative movements from wearable devices, *Proc. of IEEE SoutheastCon*, Raleigh, NC, USA, 28-29 March 2020, vol. 2, pp. 1-4. DOI: [10.1109/SoutheastCon44009.2020.9368261](https://doi.org/10.1109/SoutheastCon44009.2020.9368261)

[8] N. Mahadevan, Y. Christakis, J. Di, J. Bruno, Y. Zhang, E. Ray Dorsey, W. R. Pigeon, L. A. Beck, K. Thomas, Y. Liu, M. Wicker, C. Brooks, N. Shaafi Kabiri, J. Bhangu, C. Northcott, S. Patel, Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices, *npj Digit. Med.*, vol. 4, no. 1, 2021, pp. 1–10. DOI: [10.1038/s41746-021-00402-x](https://doi.org/10.1038/s41746-021-00402-x)

[9] R. Dai, C. Lu, M. Avidan, T. Kannampallil, RespWatch: Robust Measurement of Respiratory Rate on Smartwatches with Photoplethysmography, *Proc. of the International Conference on Internet-of-Things Design and Implementation*, Charlottesville VA, USA, 18 - 21 May 2021, pp. 208-220. DOI: [10.1145/3450268.3453531](https://doi.org/10.1145/3450268.3453531)

[10] J. Chen, M. Abbod, J. S. Shieh, Pain and stress detection using wearable sensors and devices—a review, *Sensors (Switzerland)*, vol. 21, no. 4, 2021 MDPI AG, pp. 1–18. DOI: [10.3390/s21041030](https://doi.org/10.3390/s21041030)

[11] K. Bayoumy, M. Gaber, A. Elshafeey, O. Mhaimed, E. H. Dineen, F. A. Marvel, S. S. Martin, E. D. Muse, M. P. Turakhia, Kh. G. Tarakji, M. B. Elshazly, Smart wearable devices in cardiovascular care: where we are and how to move forward, *Nat. Rev. Cardiol.*, 18 (2021), pp. 581–599. DOI: [10.1038/s41569-021-00522-7](https://doi.org/10.1038/s41569-021-00522-7)

[12] S. Cajigal, As Consumer Sleep Trackers Gain in Popularity, Sleep Neurologists Seek More Data to Assess How to Use Them in

- Practice, *Neurol. Today*, vol. 21, no. 9, 2021, pp. 8-14.
DOI: [10.1097/01.nt.0000752872.13869.cc](https://doi.org/10.1097/01.nt.0000752872.13869.cc)
- [13] C. P. Wen, J. P. M. Wai, C. H. Chen, W. Gao, Can weight loss be accelerated if we exercise smarter with wearable devices by subscribing to Personal Activity Intelligence (PAI)?, *Lancet Reg. Heal. - Eur.*, vol. 5, 2021, p. 100133 (8 pp.).
DOI: [10.1016/j.lanepc.2021.100133](https://doi.org/10.1016/j.lanepc.2021.100133)
- [14] L. Scalise, G. Cosoli, Wearables for health and fitness: Measurement characteristics and accuracy, *Proc. of the 2018 IEEE International Instrumentation and Measurement Technology Conference I2MTC: Discovering New Horizons in Instrumentation and Measurement*, Houston, TX, USA, 14-17 May 2018, pp. 1-6.
DOI: [10.1109/I2MTC.2018.8409635](https://doi.org/10.1109/I2MTC.2018.8409635)
- [15] J. Ringrose, R. Padwal, Wearable Technology to Detect Stress-Induced Blood Pressure Changes: The Next Chapter in Ambulatory Blood Pressure Monitoring?, *American journal of hypertension*, vol. 34, no. 4. NLM (Medline), 2021, pp. 330–331.
DOI: [10.1093/ajh/hpaa158](https://doi.org/10.1093/ajh/hpaa158)
- [16] G. Quer, J. M. Radin, M. Gadaleta, K. Baca-Motes, L. Ariniello, E. Ramos, V. Kheterpal, E. J. Topol, S. R. Steinhubl, Wearable sensor data and self-reported symptoms for COVID-19 detection, *Nat. Med.*, vol. 27, no. 1, 2021, pp. 73–77.
DOI: [10.1038/s41591-020-1123-x](https://doi.org/10.1038/s41591-020-1123-x)
- [17] J. Budd, B. S. Miller, E. M. Manning, V. Lampos, M. Zhuang, M. Edelstein, G. Rees, V. C. Emery, M. M. Stevens, N. Keegan, M. J. Short, D. Pillay, Ed Manley, I. J. Cox, D. Heymann, A. M. Johnson, R. A. McKendry, Digital technologies in the public-health response to COVID-19., *Nat. Med.*, vol. 26, no. 8, Aug. 2020, pp. 1183–1192.
DOI: [10.1038/s41591-020-1011-4](https://doi.org/10.1038/s41591-020-1011-4)
- [18] M. L. Millenson, J. L. Baldwin, L. Zipperer, H. Singh, Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis, *Diagnosis*, vol. 5, no. 3, 2018, pp. 95–105.
DOI: [10.1515/dx-2018-0009](https://doi.org/10.1515/dx-2018-0009)
- [19] G. Cosoli, S. Spinsante, L. Scalise, Wearable devices and diagnostic apps: beyond the borders of traditional medicine, but what about their accuracy and reliability?, *Instrum. Meas. Mag.*, vol. 24, no. 6, September 2020, pp. 89 - 94.
DOI: [10.1109/MIM.2021.9513636](https://doi.org/10.1109/MIM.2021.9513636)
- [20] M. Cukurova, C. Kent, R. Luckin, Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring, *Br. J. Educ. Technol.*, vol. 50, no. 6, 2019, pp. 3032–3046.
DOI: [10.1111/bjct.12829](https://doi.org/10.1111/bjct.12829)
- [21] A. Chang, *The Role of Artificial Intelligence in Digital Health*, Springer, Cham, 2020, pp. 71–81
DOI: [10.1007/978-3-030-12719-0_7](https://doi.org/10.1007/978-3-030-12719-0_7)
- [22] E. B. Hansen, S. Bøgh, Artificial intelligence and internet of things in small and medium-sized enterprises: A survey, *J. Manuf. Syst.*, vol. 58, 2021, pp. 362–372.
DOI: [10.1016/j.jmsy.2020.08.009](https://doi.org/10.1016/j.jmsy.2020.08.009)
- [23] M. Borghetti, P. Bellitti, N. F. Lopomo, M. Serpelloni, E. Sardini, F. Bonavolonta, Validation of a modular and wearable system for tracking fingers movements, *Acta IMEKO*, vol. 9, no. 4, 2020, pp. 157–164.
DOI: [10.21014/acta_imeko.v9i4.752](https://doi.org/10.21014/acta_imeko.v9i4.752)
- [24] A. Razzaque, A. Hamdan, Artificial intelligence based multinational corporate model for EHR interoperability on an e-health platform, *Studies in Computational Intelligence*, vol. 912. Springer, 2021, pp. 71–81.
DOI: [10.1007/978-3-030-51920-9_5](https://doi.org/10.1007/978-3-030-51920-9_5)
- [25] T. Zhang, A. El Ali, C. Wang, A. Hanjalic, P. Cesar, Cornnet: Fine-grained emotion recognition for video watching using wearable physiological sensors, *Sensors (Switzerland)*, vol. 21, no. 1, 2021, pp. 1–25.
DOI: [10.3390/s21010052](https://doi.org/10.3390/s21010052)
- [26] S. Mekruksavanich, A. Jitpattanakul, Biometric User Identification Based on Human Activity Recognition Using Wearable Sensors: An Experiment Using Deep Learning Models, *Electronics*, vol. 10, no. 3, 2021, pp. 1-21.
DOI: [10.3390/electronics10030308](https://doi.org/10.3390/electronics10030308)
- [27] Kelvin Tsoi, Karen Yiu, Helen Lee, Hao-Min Cheng, Tzung-Dau Wang, Jam-Chin Tay, Boon Wee Teo, Yuda Turana, Ariessa Ann Soenarta, Guru Prasad Sogunuru, Saulat Siddique, Yook-Chin Chia, Jinho Shin, Chen-Huan Chen, Ji-Guang Wang, Kazuomi Kario, the HOPE Asia Network, Applications of artificial intelligence for hypertension management, *Journal of Clinical Hypertension*, vol. 23, no. 3. Blackwell Publishing Inc., 2021, pp. 568–574.
DOI: [10.1111/jch.14180](https://doi.org/10.1111/jch.14180)
- [28] E. Anceschi, G. Bonifazi, M. C. De Donato, E. Corradini, D. Ursino, L. Virgili, Savemenow.ai: A machine learning based wearable device for fall detection in a workplace, *Studies in Computational Intelligence*, vol. 911. Springer Science and Business Media Deutschland GmbH, 2021, pp. 493–514.
DOI: [10.1007/978-3-030-52067-0_22](https://doi.org/10.1007/978-3-030-52067-0_22)
- [29] S. Casaccia, G. M. Revel, G. Cosoli, L. Scalise, Assessment of domestic well-being: from perception to measurement, *Instrum. Meas. Mag.*, vol. 24, no. 6, 2021, pp. 58-67.
DOI: [10.1109/MIM.2021.9513641](https://doi.org/10.1109/MIM.2021.9513641)
- [30] A. Poli, G. Cosoli, L. Scalise, S. Spinsante, Impact of Wearable Measurement Properties and Data Quality on ADLs Classification Accuracy, *IEEE Sens. J.*, Volume: 21, Issue: 13, July 2021, pp. 14221-14231.
DOI: [10.1109/JSEN.2020.3009368](https://doi.org/10.1109/JSEN.2020.3009368)
- [31] C. Sáez, N. Romero, J. A. Conejero, J. M. García-Gómez, Potential limitations in COVID-19 machine learning due to data source variability: A case study in the nCov2019 dataset, *J. Am. Med. Informatics Assoc.*, vol. 28, no. 2, 2021, pp. 360–364.
DOI: [10.1093/jamia/ocaa258](https://doi.org/10.1093/jamia/ocaa258)
- [32] M. Garbarino, M. Lai, D. Bender, R. W. Picard, S. Tognetti, Empatica E3 — A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition, *Proc. of the 4th Int. Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, Athens, Greece, 3-5 November 2014, pp. 39–42.
DOI: [10.1109/MOBIHEALTH.2014.7015904](https://doi.org/10.1109/MOBIHEALTH.2014.7015904)
- [33] F. Scardulla, L. D'acquisto, R. Colombarini, S. Hu, S. Pasta, D. Bellavia, A study on the effect of contact pressure during physical activity on photoplethysmographic heart rate measurements, *Sensors (Switzerland)*, vol. 20, no. 18, 2020, pp. 1–15.
DOI: [10.3390/s20185052](https://doi.org/10.3390/s20185052)
- [34] E. Yuda, M. Shibata, Y. Ogata, N. Ueda, T. Yambe, M. Yoshizawa, J. Hayano, Pulse rate variability: A new biomarker, not a surrogate for heart rate variability, *J. Physiol. Anthropol.* (2020), pp. 1-4.
DOI: [10.1186/s40101-020-00233-x](https://doi.org/10.1186/s40101-020-00233-x)
- [35] N. Pinheiro, R. Couceiro, J. Henriques, J. Muehlsteff, I. Quintal, L. Goncalves, P. Carvalho, Can PPG be used for HRV analysis?, *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, Orlando, FL, USA, 16-20 August 2016, pp. 2945–2949.
DOI: [10.1109/EMBC.2016.7591347](https://doi.org/10.1109/EMBC.2016.7591347)
- [36] G. Cosoli, A. Poli, L. Scalise, S. Spinsante, Heart Rate Variability Analysis With Wearable Devices: Influence of Artifact Correction Method on Classification Accuracy for Emotion Recognition, *Proc. of the 2021 IEEE Int. Instrumentation and Measurement Technology Conference I2MTC: Discovering New Horizons in Instrumentation and Measurement*, Glasgow, United Kingdom, 17-20 May 2021, pp. 1-6.
DOI: [10.1109/I2MTC50364.2021.9459828](https://doi.org/10.1109/I2MTC50364.2021.9459828)
- [37] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-aho, P. A. Karjalainen, Kubios HRV – Heart rate variability analysis software, *Comput. Methods Programs Biomed.*, vol. 113, no. 1, 2014, pp. 210–220.
DOI: [10.1016/j.cmpb.2013.07.024](https://doi.org/10.1016/j.cmpb.2013.07.024)
- [38] J. Lee, M. Kim, H. K. Park, I. Y. Kim, Motion artifact reduction in wearable photoplethysmography based on multi-channel sensors with multiple wavelengths, *Sensors (Switzerland)*, vol. 20, no. 5,

- 2020, 1493, pp. 1-14.
DOI: [10.3390/s20051493](https://doi.org/10.3390/s20051493)
- [39] H. Lee, H. Chung, J. W. Kim, J. Lee, Motion Artifact Identification and Removal from Wearable Reflectance Photoplethysmography Using Piezoelectric Transducer, *IEEE Sens. J.*, vol. 19, no. 10, 2019, pp. 3861–3870.
DOI: [10.1109/JSEN.2019.2894640](https://doi.org/10.1109/JSEN.2019.2894640)
- [40] M. Nabian, Y. Yin, J. Wormwood, K. S. Quigley, L. F. Barrett, S. Ostadabbas, An open-source feature extraction tool for the analysis of peripheral physiological data, *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, 2018, pp. 1-11.
DOI: [10.1109/JTEHM.2018.2878000](https://doi.org/10.1109/JTEHM.2018.2878000)
- [41] A. Greco, G. Valenza, E. P. Scilingo, Electrodermal Phenomena and Recording Techniques, *Advances in Electrodermal Activity Processing with Applications for Mental Health*. Springer International Publishing, 2016, pp. 1–17.
DOI: [10.1007/978-3-319-46705-4_1](https://doi.org/10.1007/978-3-319-46705-4_1)
- [42] S. S. Shapiro, M. B. Wilk, An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, vol. 52, no. 3/4, 1965, pp. 591-611.
DOI: [10.2307/2333709](https://doi.org/10.2307/2333709)
- [43] K. Kasos, Z. Kekecs, L. Csirmaz, S. Zimonyi, F. Viktor, E. Kasos, A. Veres, E. Kotyuk, A. Szekely, Bilateral comparison of traditional and alternate electrodermal measurement sites, *Psychophysiology*, vol. 57, no. 11, 2020, e13645, pp. 1-15.
DOI: [10.1111/psyp.13645](https://doi.org/10.1111/psyp.13645)
- [44] N. Milstein, I. Gordon, Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States, *Front. Behav. Neurosci.*, vol. 14, 2020, 13 pp.
DOI: [10.3389/fnbeh.2020.00148](https://doi.org/10.3389/fnbeh.2020.00148)
- [45] L. Menghini, E. Gianfranchi, N. Cellini, E. Patron, M. Tagliabue, M. Sarlo, Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions, *Psychophysiology*, vol. 56, no. 11, 2019, e13441, 15 pp.
DOI: [10.1111/psyp.13441](https://doi.org/10.1111/psyp.13441)
- [46] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, K. Van Laerhoven, Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection, *Proc. of the 20th ACM International Conference on Multimodal Interaction*, Boulder, CO, USA, 16 – 20 October 2018, pp. 400–408.
DOI: [10.1145/3242969.3242985](https://doi.org/10.1145/3242969.3242985)
- [47] C. Y. Park, N. Cha, S. Kang, A. Kim, A. Habib Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, U. Lee, K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations, *Sci. Data*, vol. 7, 2020, no. 1, 293, pp. 1–16.
DOI: [10.1038/s41597-020-00630-y](https://doi.org/10.1038/s41597-020-00630-y)
- [48] K. Mundnich, B. M. Booth, M. L'Hommedieu, T. Feng, B. Girault, J. L'Hommedieu, M. Wildman, S. Skaaden, A. Nadarajan, J. L. Villatte, T. H. Falk, K. Lerman, E. Ferrara, S. Narayanan, TILES-2018, a longitudinal physiologic and behavioral data set of hospital workers, *Sci. Data*, vol. 7, no. 1, 2020, pp. 354.
DOI: [10.1038/s41597-020-00655-3](https://doi.org/10.1038/s41597-020-00655-3)