

Wearable devices and Machine Learning algorithms to assess indoor thermal sensation: metrological analysis

Gloria Cosoli¹, Silvia Angela Mansi², Gian Marco Revel¹, Marco Arnesano²

¹ Department of Industrial Engineering and mathematical Sciences, v. Breccie Bianche snc, 60131 Ancona, Italy

² Faculty of Engineering, Università Telematica eCampus, v. Isimbardi 10, 22060 Novedrate, Italy

ABSTRACT

Personal comfort modeling is considered the most promising solution for indoor thermal comfort management in buildings. The use of wearable sensors is investigated for the real-time measurement of physiological signals to train comfort models for buildings monitoring and control. To achieve the required reliability, different uncertainty sources should be considered and weighted in the measurement results evaluation. This study presents an example of personal comfort model (PCM) development based on wearable sensors (i.e., Empatica E4 smartband and MUSE headband) acquiring multimodal signals (i.e., photoplethysmographic – PPG, electrodermal activity – EDA, skin temperature – SKT, and electroencephalographic – EEG ones), together with a metrological characterization of the modeling procedure. Starting from the data collected within an experimental campaign on 76 subjects, different Machine Learning (ML) algorithms were exploited to create comfort models capable of predicting the human thermal sensation (TS). The most accurate model was considered to investigate the impact of sensors uncertainty through a Monte Carlo simulation. Results showed that the Random Forest model is the best performing one (accuracy: 0.86). Monte Carlo simulation method proved that the model is very robust towards measurement uncertainties of input features (expanded uncertainty of the model accuracy: ± 0.04 , $k = 2$). This confirms the possibility to derive the subject's TS exploiting only physiological signals; measurement uncertainty is influenced mostly by PPG and EDA signals. This kind of investigation could lead to the development of PCMs, exploitable within control systems to optimize subjects' well-being and building energy efficiency.

Section: RESEARCH PAPER

Keywords: machine learning; measurement uncertainty; Monte Carlo simulation; personal comfort model; physiological signals; thermal sensation; uncertainty propagation; wearable sensors

Citation: Gloria Cosoli, Silvia Angela Mansi, Gian Marco Revel, Marco Arnesano, Wearable devices and Machine Learning algorithms to assess indoor thermal sensation: metrological analysis, Acta IMEKO, vol. 12, no. 3, article 57, September 2023, identifier: IMEKO-ACTA-12 (2023)-03-57

Section Editor: Francesco Lamonaca, University of Calabria, Italy

Received May 12, 2023; **In final form** July 10, 2023; **Published** September 2023

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.

Funding: This research was co-funded by the Italian Ministry of Research.

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

1. INTRODUCTION

The World Health Organization (WHO) states that physical, mental, and social domains influence a subject's well-being [1]. Several studies report that thermal comfort, defined as “the condition of mind that expresses satisfaction with the thermal environment” [2], is correlated to well-being, satisfaction, and productivity in indoor environments [3], [4]. Recently, research on thermal comfort modelling has focused on the possible correlations among environmental parameters (e.g., air temperature, relative humidity, etc.), occupants' physiological and psychological data (e.g., heart rate, skin temperature, etc.), and their subjective feedback (e.g., thermal sensation and thermal

preference, commonly collected by questionnaires). The subject's thermal sensation (and perception) can be considered as feedback towards the building environmental regulation, in particular in the context of Personal Comfort Models (PCMs), which can be exploited also for human-centred control purposes. This is pivotal when designing a living environment, process with a twofold aim: improving human well-being, health status, and productivity, as well as streamlining the building energy consumption, in a view of energy efficiency [5], [6]. Indeed, both the energy consumptions optimization and the guarantee of the occupants' thermal comfort can be achieved through the adoption of proper control strategies, which are based on the accurate measurement of thermal sensation [7], thus considering

the subjectivity within the framework of environmental control systems, acting on both physiological state and comfort. Nevertheless, the most challenging aspect is related to the need of simple acquisition systems, for a greater spreading of the assessment procedure. A possible solution could be the restriction of the analysis to physiological parameters (preferably acquired through wearable sensors, not requiring specific installation in the built environment), monitored with appropriate accuracy. Several physiological parameters in literature have been demonstrated to be related to thermal comfort [8], such as ElectroCardioGram (ECG – Heart Rate, HR, and its variability, HRV, are expected to vary with thermal comfort [9]), ElectroEncephaloGram (EEG – whose frequency bands, related to diverse neural activities, vary along with thermal conditions [10], [11]), ElectroDermal Activity (EDA – quantifying the arousal of the SNS and mirroring the activity of the sweat glands, hence carrying information dealing with the subject's perceived emotions [12] and experienced stress [13]), and SKin Temperature (SKT – reflecting the perception of the environmental temperature).

More and more frequently wearables are used in combination with ML or Artificial Intelligence (AI) techniques for the classification/prediction of individuals' thermal comfort or thermal sensation in indoor environments. Park and Park [14] exploited the ensemble transfer learning method on data acquired through wearable and environmental sensors to predict individual thermal comfort, reaching accuracy values of 0.85-0.95. A moderate prediction power was reached also by Liu and colleagues [15], who developed PCMs through lab-grade wearable sensors; they observed better results when not in thermal neutrality conditions. Also, non-conventional wearable sensors have been investigated in the literature, such as sweat rate sensors (with a measurement error < 10 %), which were demonstrated to be able to discriminate among 4 different thermal states (even if no AI techniques were employed for classification) [16]. Chaudhuri et al. [17] acquired physiological signals through wearable sensors, finding relevant differences between females and males also in terms of subjective responses to thermal conditions. Using the Random Forest classifier, they reached prediction accuracy values of approximately 0.93 and 0.94 for males and females, respectively.

Models for classification/prediction purposes based on physiological signals can support the development of relatively compact, non-intrusive, nor invasive monitoring systems, decoding the subjective perception of thermal conditions, without inevitably counting on environmental sensors. In this way, personalized comfort systems could be developed and adapted based on the subject's physiological state, which can be monitored through wearable sensors, in an occupant-centric perspective. This paves the way towards sustainability in terms of buildings energy consumption, leading to the development of advanced Heating, Ventilation, and Air Conditioning (HVAC) and lighting control systems taking advantage of the interaction with PCMs.

In this perspective, the authors carried out a preliminary study [18] to demonstrate the feasibility of thermal sensation assessment through the combined use of Machine Learning (ML) classifiers and wearable sensors, measuring physiological signals in a fully controlled environment, and exposing the participants to predetermined thermal conditions. Results showed classification accuracy values up to 0.80, encouraging the expansion and deepening of this line of research. However, the final accuracy and reliability of this kind of models depend on

the measurement uncertainties linked to each step of the whole measurement chain, including both hardware and software components. In recent years, more and more studies employ wearable sensors for remote monitoring in a plethora of frameworks, also providing multimodal signals collection better depicting the subject's physiological status [19]. But it is pivotal to always characterize and validate them from a metrological point of view to provide information related to the sensor's uncertainty. In fact, despite wearable sensors have several advantages (e.g., being user-friendly, available in many quality and cost segments, providing multimodal signals, etc.), on the flip side they are currently relatively scarcely investigated in terms of metrological characteristics: measurement accuracy and precision are frequently not available, or the used test protocol is not declared, neither standard protocols are established, resulting in barely comparable data. Furthermore, the combination of wearable sensors with ML and AI algorithms introduces other uncertainty sources to the results of the measurement procedure, since these models have a probabilistic nature and, consequently, an intrinsic inaccuracy. The different sources of uncertainty (both from hardware and software points of view) should be properly considered, and their contribution should be evaluated along the whole measurement chain. The best method to do this is following the recommendations provided in the Guide to the Expression of Uncertainty in Measurement (GUM) [20].

Hence, the main objective of this paper is to test different types of ML classifiers for TS and to consider the best performing one for an uncertainty analysis according to the Monte Carlo simulation method, so as to take into account the impact of wearable sensors uncertainties on the model classification performance.

The paper is organized in this way: Section 2 presents the materials and methods used for the study; in Section 3 the authors report the results related to both classification of thermal sensation and uncertainty propagation plus sensitivity analysis. The results are discussed in Section 4, where the authors also provide possible future developments. The study schematic pipeline is reported in Figure 1.

2. MATERIALS AND METHODS

2.1. Data acquisition campaign

The experimental acquisition campaign involved 76 healthy subjects (aged 26.7 ± 2.8 years), without clinical histories possibly altering the results. It is worthy to highlight the juvenile age of the test population, since well-being perception can vary with age (e.g., older people generally have less capability to effectively respond to thermal changes), and this directly impact on the parameters of models based on physiological response. The tests were performed within the NEXT.ROOM [21] (University of Perugia), a fully controlled environment equipped with different systems able to control stimuli related to multiple domains, i.e., a radiant system on all the internal surfaces of the test room, an HVAC system, and LED panels plus RGB reflectors for lightning. The tests were performed in a period covering wintertime, springtime, and summertime (51 %, 32 %, and 17 %, respectively); the metabolic rate can be assumed equal to 1.1 met according to the standard ISO 7730 for sedentary activities (writing, desk work, using computer) [22].

All subjects gave their informed consent for inclusion before they participated in the study. The tests were carried out in accordance with the WMA Declaration of Helsinki [23] and with the statute of the Ethics Committee of the University of Perugia.

All the data were managed in accordance with the General Data Protection Regulation (GDPR) and anonymization was applied to protect the privacy and confidentiality of all the gathered information.

The subjects were also asked to fill out a survey, collecting both personal information and thermal sensation (TS) and thermal comfort votes based on 5-point scales. For TS we have -2 (cold), -1 (slightly cold), 0 (neutral), +1 (slightly warm), and +2 (warm); the scale was designed in line with the purpose of the study, which aims at exposing individuals on two opposite indoor TS, to discriminate between warm/cold sensations, without considering different degrees of intensity. These scores were used as labels for ML-based classification.

Three different thermal conditions were chosen as representative of cold (16.0 ± 0.6 °C), neutral (24.0 ± 0.9 °C), and hot (31.0 ± 2.4 °C) scenarios, in accordance with ISO 7730:2005 [22]. Lighting conditions were neutral, with a CCT equal to 4114 K and an illuminance of 500 lx (measured on the desk at 75 cm from the floor and near the subject).

The duration of a test session was equal to approximately 20 minutes; at least 15 minutes for acclimatization (literature reports a preconditioning time range of 20.7 ± 9.7 min reported in the literature [24]), 5 minutes for signals recording. If the target TS had not been reached, the acquisition was postponed until the achievement of the desired thermal condition. Tests with inconsistencies between the reported and targeted TS were removed from the dataset.

2.2. Acquisition devices

Physiological signals were acquired through two wearables sensors: 1) Empatica E4 smartband [25] (Empatica, 2020, firmware 3.2.0.8313), which is a Class IIA medical Device (93/42/EEC Directive) with 4 sensors onboard, i.e.,

- i) PPG sensor for the acquisition of the blood volume pulse (BVP) signal,
 - ii) EDA sensor for the measurement of skin,
 - iii) infrared thermometer for SKT, and
 - iv) 3-axial MEMS accelerometer, and
- 2) Interaxon MUSE headband [26] for EEG acquisition, with a reference electrode (FPz) on the forehead and 4 input electrodes (2 silver-made on the forehead, AF7 and AF8, and 2 in conductive rubber above ears, TP9 and TP10).

2.3. Data processing and ML algorithm-based analysis for the prediction of thermal sensation

The electroencephalographic signal was initially filtered (bandpass filter at 0.1-45 Hz, Notch filter at 50 Hz) to remove

motion artefacts and noise [26]. Then, the Fast Fourier Transform (FFT) was employed to compute the power spectrum densities (PSDs) of five different frequency bands [27]–[30]:

- i) Delta band (0.1-4.0 Hz), dominant in infants and in certain sleep phases and involved in motivational processes; some literature studies revealed a high power in the delta band in warm conditions [31],
- ii) Theta band (4.0-7.5 Hz), often linked to memory and emotional regulation; some studies in the literature report a decrease in this band in a warm environment [32],
- iii) Alpha band (7.5-12.0 Hz), appearing when the eyes are closed and during the relaxing state and disappearing in case of alerting,
- iv) Beta band (12-30 Hz), the normal rhythm for alert/anxious people; some literature findings report a higher content in cold/warm thermal conditions with respect to neutrality [33],
- v) Gamma band (30-45 Hz), prevalent in complex and high attention-demanding tasks, (e.g., concentration and problem-solving activities).

Regarding the PPG signal, the Inter-Beat Intervals (IBIs) provided by Empatica E4 were considered for HRV analysis, which was performed through the “hrvanalysis” Python module [34] (in both time and frequency domains).

Concerning EDA, the EDA toolkit 11 allowed to remove movement artefacts, then the cvxEDA tool 13 was exploited to determine tonic and phasic components [35].

A total of 110 features (averaged on the whole 5-min recording duration) were extracted from the different physiological signals acquired in each test. A preliminary statistical analysis was carried out on the extracted features. In particular, the Shapiro test [36] was exploited to evaluate the normality of groups; then, since all the features’ distributions resulted in non-Gaussian like, their statistical significance ($p < 0.05$) in terms of differences with TS was evaluated through the Kruskal-Wallis test [37]. At this point, data were prepared for the ML-based classification procedure. At first, the correlation matrix was manually cleaned up from superfluous features, then LASSO regularization and ANOVA methods were exploited to derive the weight of the features in the classification process. To reduce the problem dimensionality, a Random Forest method, based on entropy decrease, was used to select the features with the higher impact. The classification threshold between hot and cold TS was chosen equal to 0.5 (0 equal to the cold sensation and 1 equal to the warm sensation). This procedure was executed to fine-tune the subset of features to be used as input for six

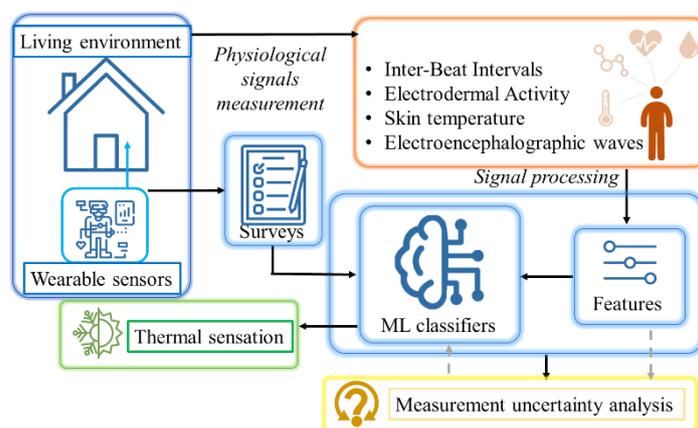


Figure 1. Schematic pipeline of the study.

supervised ML classifiers (among the most exploited in literature [38], [39], as well as recommended for classification purposes according to the scikit-learn [40] documentation), namely:

- i) K-Nearest Neighbors (kNN), making the prediction on the basis of test-training points distance,
- ii) Random Forest (RF), based on decision trees for classification/prediction purposes, considering respectively the majority and the average,
- iii) Logistic Regression (LR), a non-linear model computing the probability of a determined output given a variable of input,
- iv) Gaussian Naïve Bayes (GNB), which exploits the Gaussian distribution to model continuous features,
- v) Gradient Boosting (GB), whose cost function is based on the mean square error when the model is used for regression purposes, whereas is a logarithmic loss when exploited for classification,
- vi) Decision Tree (DT), which learns basic decision rules on the basis of the characteristics of the input data.

The thermal sensation prediction model was designed utilizing scikit-learn [40]. The acquired dataset was divided into training and testing sub-datasets, with a ratio of 7:3. Bayesian optimization and 10-fold cross-validation were employed to optimize the ML classifiers hyperparameters. Regarding the former optimization tool, the BayesSearchCV function has been exploited; a fit and a score method were implemented, and a cross-validated search was used to optimize the estimator parameters. The parameter space has been optimized according to parameters specific for each model (Table 1).

As above mentioned, TS values were used as labels in a binary classification: hot ($TS > 0$) and cold ($TS < 0$) thermal sensations. The evaluation of the classification performance of the classifiers was made through standard metrics, namely Accuracy, Sensitivity (or Recall), Precision, and Area Under Curve (AUC) [41]. The

Table 1. Hyperparameters optimization for the considered ML classifiers.

ML classifier	Parameter	Range for optimization
kNN	Leaf size	[1, 2, 3, ..., 50]
	Number of neighbours	[1, 2, 3, ..., 30]
	Weights for prediction	['uniform', 'distance']
	Metric for distance computation	['euclidean', 'manhattan', 'minkowski']
RF	Number of estimators	[10, 100, 1000]
	Maximum number of features	['log2', 'sqrt']
LR	Norm of the penalty	['L2']
	C (inverse of regularization strength)	[100, 10, 1.0, 0.1, 0.01], [10 ⁻⁴ , ..., 10 ⁴] with 20 steps
	Solver	['liblinear', 'newton-cg', 'lbfgs']
GNB	Var smoothing	[1, ..., 10 ⁻⁹] with 100 steps
	Number of estimators (boosting stages)	[10, 100, 1000]
GB	Learning rate	[0.001, 0.01, 0.1]
	Subsample for fitting the base learners	[0.5, 0.7, 1.0]
	Maximum depth of the estimators	[3, 7, 9]
DT	Maximum depth of the tree	[3, 5, 7, 9]
	Maximum features for the best split	['log2', 'sqrt']
	Minimum number of samples for a leaf node	[1, 3, 5, 7, 9]
	Criterion to measure the split quality	["gini", "entropy"]

models training was completed within the cuda toolkit, exploiting the Nvidia GeForce RTX 3080 Ti graphic processing unit.

The input features feeding the classifiers came from the multidomain physiological signals, selected according to their correlation with the class to be predicted and the simultaneous non-correlation with the other features. Then, the best feature subset was selected for the evaluation of the model classification performance.

2.4. Uncertainty propagation and Monte Carlo simulation

The authors performed an uncertainty analysis to identify the most relevant interfering sources in the context of the assessment of the indoor thermal sensation. The best performing ML algorithm with the related feature subset was considered for the analysis of the measurement uncertainty. In particular, the Monte Carlo simulation method was adopted to numerically estimate the measurement uncertainty according to the GUM. The chosen model was run 10⁴ times, taking as input the selected features extracted from the considered signals (i.e., IBI, EDA, SKT, and PSD of EEG brain waves), which were perturbed according to a Gaussian distribution.

The standard deviation (σ) of this distribution was set according to typical values taken from literature, namely:

- 2.5 bpm for HR (i.e., approximately 42 ms considering the IBIs). This value was chosen to remain in cautionary conditions (also because the signal is extracted from PPG, which is prone to many sources of uncertainty [42]), considering the mean absolute percentage error judged acceptable for an accurate HR monitor according to the ANSI/AAMI/IEC 60601-2-27:2011/(R)2016 [43],
- 0.006 μ S in terms of skin conductance for EDA [44],
- 0.5 °C for SKT (conservative range with respect to the 0.2 °C value reported by Empatica user E4 manual [45]),
- 5 % of the average PSD for EEG waves (authors' protective hypothesis).

The Gaussian distribution of the considered physiological quantities (i.e., IBI, EDA, SKT, and PSD of EEG waves) was built and the signal processing pipeline was repeated in order to extract the features from the perturbed signals. Hence, these features were exploited to train the selected best performing ML classifier, perturbing all the input variables simultaneously. The propagation of uncertainty in the whole measurement chain for TS prediction was evaluated in terms of the classifier accuracy. This was done to demonstrate the robustness of the model against the variability of physiological signals (including not only the sensors uncertainty, but also the variability of the vital signs, which plays a relevant role in the determination of uncertainty [46]). Then, perturbing only a physiological signal at a time and maintaining the others as-are, a variance-based sensitivity analysis was performed, in order to understand which signals mostly affect the results. In particular, the first-order variance-based sensitivity coefficients $S(x_i)$ were computed according to equation (1):

$$S(x_i) = \frac{u_i(\mathbf{y})^2}{u(\mathbf{y})^2} \cdot 100 \%, \quad (1)$$

where $u_i(\mathbf{y})$ is the standard uncertainty of the output due to the i -th input uncertainty, whereas $u(\mathbf{y})$ is the standard uncertainty of the output when the uncertainty of all the inputs is considered. The variance can be calculated as the square of the standard deviation in the case of normal distribution (i.e., the one selected in the presented work).

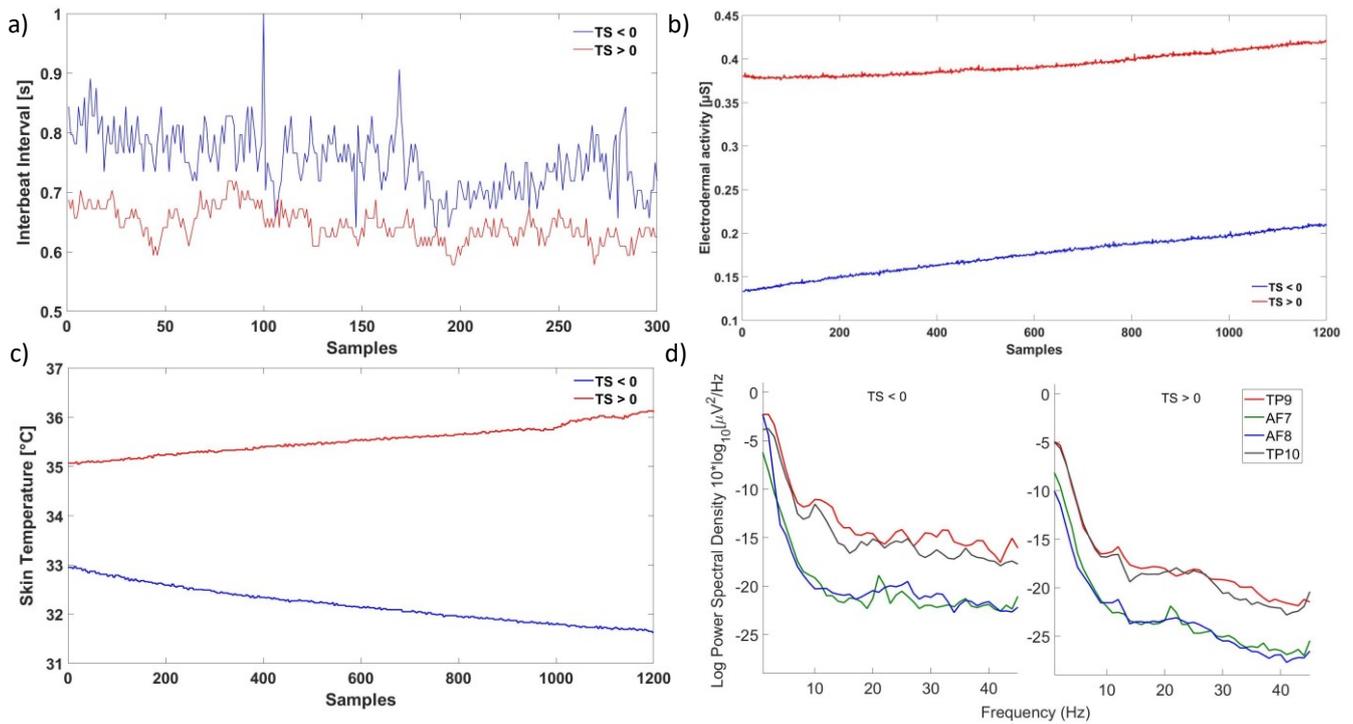


Figure 2. Examples of a) IBI (sampling frequency: 64 Hz), b) EDA (sampling frequency: 4 Hz), c) SKT (sampling frequency: 4 Hz), and d) EEG signals (PSD in logarithmic values) in case of cold and hot TS (blue and red lines, respectively, for a), b), and c) subfigures).

3. RESULTS

An example of comparison between hot and cold TS (> 0 and < 0 , respectively) is reported in Figure 2 for IBI, EDA, SKT, and EEG signals. It is possible to note that IBIs (a) have a higher duration in the case of TS < 0 , given that HR decreases in cold conditions, as expected in the case of cold defence [47]. The skin temperature (b) directly reflects the ambient temperature, hence SKT signals decrease with temperature. The EDA signal (c) shows a higher intensity in case of hot TS; in fact, the skin conductivity increases with sweat production. Indeed, the SNS controls this phenomenon, and its activation is more marked in warm condition, when heat dissipation mechanisms are actuated [48]. Concerning EEG (d), it is possible to observe that the PSD values decrease in the case of hot TS; in particular, when the subject is in a cold environment, her/his attention level tends to be higher (corresponding to an intense mental activity) and this is reflected at least in Beta and Gamma waves (frequency > 35 Hz) [49].

3.1. ML-based classification

The feature selection process led to the definition of a peculiar feature subset (Table 2); this was used as input for the ML classifiers to be tested. These features are from all the considered physiological signals domains: 8 from PPG, 4 from EDA signal, 4 from EEG, and 1 feature from SKT signal.

The performance metrics for all the considered ML classifiers are reported in Table 3. In the configuration with the input features coming from all the recorded physiological signals, the best results are achieved by the RF classifier (Accuracy: 0.86; Sensitivity: 0.87; Precision: 0.80; AUC: 0.90), with a very good ability in distinguishing between hot and cold TS. The second-best performing classifiers are GB and GNB (Accuracy: 0.80). However, GNB reports a slightly higher Sensitivity (0.88 against

0.86 of GB), proving a better performance in retrieving the instances of interest. All the other tested classifiers, namely kNN, LR, and DT, provide an Accuracy < 0.80 , but still good (i.e., 0.77, 0.73, and 0.70, respectively). It is worthy to underline the optimal Sensitivity of kNN algorithm, which is even the best one among

Table 2. Feature subset selected as input to the ML classifiers.

Feature and description	Signal
RMSSD : root mean square of successive RR interval differences	PPG
pnni_20 : percentage of successive RR intervals exceeding 20 ms	PPG
mean_temp: mean skin temperature	SKT
lfnu : Relative power of low-frequency band (0.04–0.15 Hz) in normal units	PPG
Cvstd : Coefficient of variation of successive differences between RR intervals	PPG
Tonic_STD : Standard deviation of tonic component	EDA
gamma_tp9: Power of gamma band	EEG
pnni_50 : Percentage of successive RR intervals exceeding 50 ms	PPG
Sdsd : Standard deviation of differences between adjacent NN intervals	PPG
lf_hf_ratio : Ratio of LF-to-HF power	PPG
relative_gamma_tp10 : Relative power of gamma band, TP10 electrode	EEG
Tonic_perc25 : 25th percentile of tonic component	EDA
Tonic_quartdev : Quartile deviation of tonic component	EDA
relative_gamma_tp9 : Relative power of gamma band, TP9 electrode	EEG
Hfnu : Relative power of high- frequency band (0.15–0.4 Hz) in normal units	PPG
Temporal_asym_delta : Difference between left and right hemisphere in delta band	EEG
Tonic_mean : Mean value of tonic component	EDA

Table 3. Accuracy, Sensitivity, Precision, and AUC achieved by the tested ML classifiers with the identified optimal feature subset.

ML classifier	Accuracy	Sensitivity	Precision	AUC
kNN	0.77	0.96	0.65	0.86
RF	0.86	0.87	0.80	0.90
LR	0.73	0.70	0.67	0.78
GNB	0.80	0.91	0.70	0.86
GB	0.80	0.83	0.73	0.88
DT	0.70	0.83	0.59	0.73

the considered models, highlighting its high capability of identifying the relevant instances.

PPG signal provides 8 out of 17 of the selected features used as input; if it is true that it delivers a lot of information, it should be reminded that it also brings many variability sources, which are intrinsic of the signal nature itself. Hence, an important part of uncertainty could be attributable to the multiple interfering sources to which it is prone: motion artefacts, illuminating conditions, skin tone, etc. Hence, particular attention to the features extracted from PPG signal (HRV analysis) was then paid in the uncertainty analysis, considering a cautionary uncertainty range for its values.

3.2. Monte Carlo simulation results

Regarding the uncertainty analysis, it can be stated that the model appears to be quite robust against the uncertainty estimated for the input metrics. In particular, from the results reported in Figure 3, it is possible to observe that the expanded uncertainty on the model performance in terms of accuracy turned out to be equal to ± 0.04 ($k = 2$). This means that, despite the significant perturbations on the model inputs, its performance remains very good and stable, i.e., > 0.82 (lower bound of 95 % of the coverage interval). The probability distribution of the model accuracy obtained with the Monte Carlo simulation is reported in Figure 2.

Concerning the sensitivity analysis, the results are reported in Table IV; it is possible to note that the PPG signal is the one mostly influencing the variability of the results, together with the EDA signal. Less influence is given by the uncertainty of the EEG and SKT signals. The sum of the sensitivity indexes $S(x_i)$ is 99.5 %, near to the unity, so the model is additive, and its inputs can be considered as uncorrelated.

4. DISCUSSION AND CONCLUSIONS

In the present manuscript, the authors used features extracted from different types of physiological signals acquired through wearable devices to train ML-based models for the classification of thermal sensation (TS). In particular, EEG, PPG, EDA, and SKT signals were considered, since a previous preliminary study by the same authors [18] verified the possibility to exploit them for TS assessment. The post-processing of these signals, carried out in both time and frequency domains, allows to obtain features that can be ingested by ML classifiers, after a proper evaluation and selection based on statistical significance with respect to TS. K-nearest neighbor (kNN), Random Forest (RF), Logistic Regression (LR), Gaussian Naïve Bayes (GNB), Gradient Boosting (GB), and Decision Tree (DT) were tested with the optimal feature subset. The results showed that the RF model is the best performing one when all the signals are exploited as input, achieving an Accuracy of 0.86 (with an AUC of 0.90). Indeed, the best Sensitivity was achieved by the kNN

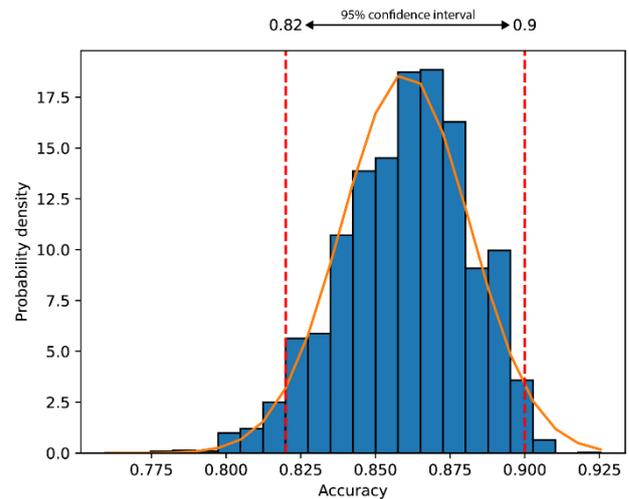


Figure 3. Probability distribution of the Accuracy obtained with the Monte Carlo simulation applied to the Random Forest model fed with all the physiological signals (i.e., PPG, EDA, SKT, and EEG) – with the optimal features subset.

classifier, probably due to its different nature (kNN makes prediction according to the distance between test and training data, whereas RF is based on the majority vote).

Given that the majority of commercial wrist-worn wearable devices is based on a PPG sensor, it is worthy to properly consider such signal for thermal sensation prediction. Likewise, it is fundamental to optimize both the hardware and the whole acquisition procedure (e.g., sensor-skin contact pressure) to enhance the signal quality and, thus, the reliability of the derived parameters. However, all the tested ML models, trained with features coming from multimodal signals (i.e., PPG, EEG, EDA, and SKT), provided a good Accuracy (always > 0.70). This result underlines again the high correlation between physiological signals and thermal sensation, hence introducing objective measurements in the thermal comfort assessment.

The classification performance of the ML algorithms was undoubtedly influenced by the hyperparameters optimization procedure, which was performed through Bayesian optimization and 10-fold cross-validation. Then, considering all the physiological signals as input and found the best configuration for the RF classifier (i.e., the best performing one in these conditions), the model was subjected to a Monte Carlo simulation to evaluate how the input uncertainties propagate and reflect on the classifier Accuracy. Such an analysis considered Gaussian distributions for all the input features, setting plausible uncertainty values, hence evaluated the robustness of the model against this variability. The results showed that the Accuracy of the prediction model can be obtained with an expanded uncertainty of ± 0.04 ($k = 2$). Such uncertainty is mostly related to the uncertainty of the PPG and EDA signals, counting for 58.7 % and 23.6 % of the output variance, respectively. The relatively low perturbation of the model performance with respect to the input signals uncertainty leads to the conclusion that Personal Comfort Models performance presents a contribution of uncertainty due to the inter- and intra-subject variabilities. Such variability must be deeply investigated for a wider application of these models in real contexts.

The results from this study can be relevant for the development of human-centred indoor environmental control systems in the perspective of achieving a sustainable built environment within the context of a quite complex climate

change. In the future, it would be interesting to widen the database collecting data on older subjects, in order to include more variability in terms of thermal comfort perception. Indeed, physiological response mechanisms vary with age, hence the model parameters are different; a study performed on subjects in a wide age range would give the possibility to consider also age as input parameter for classification models. The experimental campaign data could be also further analysed (and also widened, in terms of both numerosity and physiological variability) in order to develop Personalized Comfort Models, considering not only hot and warm thermal sensation, but also the neutral condition. Indeed, some of the considered features (e.g., from EDA and SKT signals) have been already proved to be suitable for the identification of neutral TS [49]. These models could cover pivotal roles in control and actuation perspectives within living environments, focusing on the real users' needs. This could lead to the definition of a sustainable ecosystem with automated control abilities, where the environmental parameters of a building would be controlled with the support of data recorded by wearable sensors. The use of wearables can be combined with ML and AI algorithms, appropriately trained with datasets including proper physiological variability and, hence, being able to easily adapt to specific users' needs. In this context, the occupants' physiological conditions, linked to their comfort and overall well-being level, cover a twofold role: on the one hand, they determine the subject's well-being (i.e., the target of the control system), on the other hand, they represent feedback for the built environment control loop. This enables the evaluation of the control process effects on the living environment occupants.

ACKNOWLEDGEMENT

This research was co-funded by the Italian Ministry of Research through the NEXT:COM (Prot. 20172FSC4) "Towards the NEXT generation of Multiphysics and multidomain environmental COMfort models: theory elaboration and validation experiment" project, within the PRIN 2017 program. The authors want to thank the colleagues from Università di Perugia for the availability of the NEXT.ROOM and the support for the experimental campaign.

REFERENCES

[1] L. Breslow, A Quantitative Approach to the World Health Organization Definition of Health: Physical, Mental and Social Well-being, *Int. J. Epidemiol.*, vol. 1, no. 4, 1972, pp. 347–355. DOI: [10.1093/ije/1.4.347](https://doi.org/10.1093/ije/1.4.347)

[2] ANSI/ASHRAE, Thermal Environmental Conditions for Human Occupancy, 2017, Online [Accessed 18 August 2022]. <https://hogiaphat.vn/upload/docs/ASHRAE55-version2017.pdf>

[3] T. Williamson, V. Soebarto, H. Bennetts, L. Arakawa Martins, D. Pisaniello, A. Hansen, R. Visvanathan, A. Carre, J. van Hoof, Assessing human resilience A study of thermal comfort, well-being and health of older people, in *Routledge Handbook of Resilient Thermal Comfort*, Taylor & Francis Group, 2022, eBook ISBN 9781003244929, p. 20.

[4] A. Lipczynska, S. Schiavon, L. T. Graham, Thermal comfort and self-reported productivity in an office with ceiling fans in the tropics, *Build. Environ.*, vol. 135, 2018, pp. 202–212. DOI: [10.1016/j.buildenv.2018.03.013](https://doi.org/10.1016/j.buildenv.2018.03.013)

[5] S. B. Yazyeva, I. A. Mayatskaya, Eco-sustainable architecture and comfortable living environment, *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1083, no. 1, Feb. 2021, p. 012018. DOI: [10.1088/1757-899X/1083/1/012018](https://doi.org/10.1088/1757-899X/1083/1/012018)

[6] Q. Li, L. Zhang, L. Zhang, X. Wu, Optimizing energy efficiency and thermal comfort in building green retrofit, *Energy*, vol. 237, 2021, p. 121509. DOI: [10.1016/j.energy.2021.121509](https://doi.org/10.1016/j.energy.2021.121509)

[7] A. Calvaresi, M. Arnesano, F. Pietroni, G. M. Revel, Measuring Metabolic Rate to Improve Comfort Management in Buildings, *Environ. Eng. Manag. J.*, vol. 17, 2018, pp. 2287–2296.

[8] S. A. Mansi et al., Measuring human physiological indices for thermal comfort assessment through wearable devices: A review, *Measurement*, vol. 183, Oct. 2021, p. 109872. DOI: [10.1016/j.measurement.2021.109872](https://doi.org/10.1016/j.measurement.2021.109872)

[9] N. Morresi, S. Casaccia, M. Arnesano, G.M. Revel, Impact of the measurement uncertainty on the monitoring of thermal comfort through AI predictive algorithms, *Acta IMEKO*, 10 (2021), pp. 221–229. DOI: [10.21014/acta_imeko.v10i4.1181](https://doi.org/10.21014/acta_imeko.v10i4.1181)

[10] J. Han, C. Chun, Differences between EEG during thermal discomfort and thermal displeasure, *Build. Environ.*, vol. 204, no. March 2021, p. 108220. DOI: [10.1016/j.buildenv.2021.108220](https://doi.org/10.1016/j.buildenv.2021.108220)

[11] X. Lang, P. Wargocki, W. Liu, Investigating the relation between electroencephalogram, thermal comfort, and cognitive performance in neutral to hot indoor environment, *Indoor Air*, vol. 32, no. 1, Jan. 2022, p. e12941. DOI: [10.1111/ina.12941](https://doi.org/10.1111/ina.12941)

[12] A. Horvers, N. Tombeng, T. Bosse, A. W. Lazonder, I. Molenaar, Detecting Emotions through Electrodermal Activity in Learning Contexts: A Systematic Review, *Sensors*, vol. 21, 2021, no. 23. DOI: [10.3390/s21237869](https://doi.org/10.3390/s21237869)

[13] A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, P. Zontone, Driver's stress detection using Skin Potential Response signals, *Meas. J. Int. Meas. Confed.*, vol. 122, Jul. 2018, pp. 264–274. DOI: [10.1016/j.measurement.2018.03.040](https://doi.org/10.1016/j.measurement.2018.03.040)

[14] H. Park, D. Y. Park, Prediction of individual thermal comfort based on ensemble transfer learning method using wearable and environmental sensors, *Build. Environ.*, vol. 207, 2022, p. 108492. DOI: [10.1016/j.buildenv.2021.108492](https://doi.org/10.1016/j.buildenv.2021.108492)

[15] S. Liu, S. Schiavon, H. P. Das, M. Jin, C. J. Spanos, Personal thermal comfort models with wearable sensors, *Build. Environ.*, vol. 162, 2019, p. 106281. DOI: [10.1016/j.buildenv.2019.106281](https://doi.org/10.1016/j.buildenv.2019.106281)

[16] J. K. Sim, S. Yoon, Y. H. Cho, Wearable Sweat Rate Sensors for Human Thermal Comfort Monitoring, *Sci. Reports* 2018 81, vol. 8, no. 1, Jan. 2018, pp. 1–11. DOI: [10.1038/s41598-018-19239-8](https://doi.org/10.1038/s41598-018-19239-8)

[17] T. Chaudhuri, D. Zhai, Y. C. Soh, H. Li, L. Xie, Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology, *Energy Build.*, vol. 166, 2018, pp. 391–406. DOI: [10.1016/j.enbuild.2018.02.035](https://doi.org/10.1016/j.enbuild.2018.02.035)

[18] G. Cosoli, S. A. Mansi, M. Arnesano, Combined use of wearable devices and Machine Learning for the measurement of thermal sensation in indoor environments, in *IEEE Int. Workshop on Metrology for Living Environment (MetroLivEn)*, 2022, pp. 1–6. DOI: [10.1109/MetroLivEnv54405.2022.9826956](https://doi.org/10.1109/MetroLivEnv54405.2022.9826956)

[19] G. Cosoli, A. Poli, L. Scalise, S. Spinsante, Measurement of multimodal physiological signals for stimulation detection by wearable devices, *Measurement*, vol. 184, Nov. 2021, p. 109966. DOI: [10.1016/j.measurement.2021.109966](https://doi.org/10.1016/j.measurement.2021.109966)

[20] BIPM, JCGM 100, Evaluation of measurement data - Guide to the expression of uncertainty in measurement Évaluation des données de mesure-Guide pour l'expression de l'incertitude de mesure, 2008. Online [Accessed 2 September 2022] https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf

[21] F. Vittori, C. Chiatti, I. Pigliatù, A. L. Pisello, The NEXT.ROOM: Design principles and systems trials of a novel test room aimed at deepening our knowledge on human comfort, *Build. Environ.*, vol. 211, 2022, p. 108744. DOI: [10.1016/j.buildenv.2021.108744](https://doi.org/10.1016/j.buildenv.2021.108744)

- [22] ISO, ISO 7730 Ergonomics of the thermal environment — Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria, Third edition, 2005. Online [Accessed 9 December 2022] <https://www.iso.org/obp/ui/en/#iso:std:iso:7730:dis:ed-4:v1:en> <https://www.iso.org/home.html>
- [23] WMA – The World Medical Association, WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. Online [Accessed 9 December 2022] <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>
- [24] S. Van Craenendonck, L. Lauriks, C. Vuye, J. Kampen, A review of human thermal comfort experiments in controlled and semi-controlled environments, *Renew. Sustain. Energy Rev.*, vol. 82, Feb. 2018, pp. 3365–3378. DOI: [10.1016/j.rser.2017.10.053](https://doi.org/10.1016/j.rser.2017.10.053)
- [25] 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, Aug. 2020. DOI: [10.3389/fnbeh.2020.00148](https://doi.org/10.3389/fnbeh.2020.00148)
- [26] O. E. Krigolson, C. C. Williams, A. Norton, C. D. Hassall, F. L. Colino, Choosing MUSE: Validation of a Low-Cost, Portable EEG System for ERP Research, *Front. Neurosci.*, vol. 0, no. MAR, March 2017, p. 109. DOI: [10.3389/fnins.2017.00109](https://doi.org/10.3389/fnins.2017.00109)
- [27] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, J. O. McNamara, S. M. Williams (editors), *Stages of Sleep*, in *Neuroscience*, 2nd edition, Sunderland (MA): Sinauer Associates, 2001. Online [Accessed 12 December 2022] <https://www.ncbi.nlm.nih.gov/books/NBK10996/>
- [28] S. U. C. Mehrotra, *Introduction To EEG- and Emotion Recognition*, 2018.
- [29] E. Bourel-Ponchel et al., Normal EEG during the neonatal period: maturational aspects from premature to full-term newborns, *Neurophysiol. Clin.*, vol. 51, no. 1, Jan. 2021, pp. 61–88. DOI: [10.1016/j.neucli.2020.10.004](https://doi.org/10.1016/j.neucli.2020.10.004)
- [30] I. A. Zapata, Y. Li, P. Wen, Rules-Based and SVM-Q Methods with Multitapers and Convolution for Sleep EEG Stages Classification, *IEEE Access*, vol. 10, 2022, pp. 71299–71310. DOI: [10.1109/access.2022.3188286](https://doi.org/10.1109/access.2022.3188286)
- [31] B. Lv, C. Su, L. Yang, T. Wu, Effects of stimulus mode and ambient temperature on cerebral responses to local thermal stimulation: An EEG study, *Int. J. Psychophysiol. Off. J. Int. Organ. Psychophysiol.*, vol. 113, Mar. 2017, pp. 17–22. DOI: [10.1016/j.ijpsycho.2017.01.003](https://doi.org/10.1016/j.ijpsycho.2017.01.003)
- [32] M. Zhu, W. Liu, P. Wargoeki, Changes in EEG signals during the cognitive activity at varying air temperature and relative humidity, *J. Expo. Sci. Environ. Epidemiol.*, vol. 30, no. 2, 2020, pp. 285–298. DOI: [10.1038/s41370-019-0154-1](https://doi.org/10.1038/s41370-019-0154-1)
- [33] U. Herwig, P. Satrapi, C. Schönfeldt-Lecuona, Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation, *Brain Topogr.*, vol. 16, no. 2, 2003, pp. 95–99. DOI: [10.1023/b:brat.0000006333.93597.9d](https://doi.org/10.1023/b:brat.0000006333.93597.9d)
- [34] PyPi.org, Heart Rate Variability analysis, 2021. Online [Accessed 25 September 2023] <https://pypi.org/project/hrv-analysis/>
- [35] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, L. Citi, CvxEDA: A convex optimization approach to electrodermal activity processing, *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, 2016, pp. 797–804. DOI: [10.1109/tbme.2015.2474131](https://doi.org/10.1109/tbme.2015.2474131)
- [36] M. Brzezinski, The Chen-Shapiro test for normality, *Stata J.*, vol. 12, no. 3, 2012, pp. 368–374. DOI: [10.1177/1536867X1201200302](https://doi.org/10.1177/1536867X1201200302)
- [37] E. Ostertagová, O. Ostertag, J. Kováč, Methodology and application of the Kruskal-Wallis test, *Appl. Mech. Mater.*, vol. 611, 2014, pp. 115–120. DOI: [10.4028/www.scientific.net/amm.611.115](https://doi.org/10.4028/www.scientific.net/amm.611.115)
- [38] F. Tartarini, S. Schiavon, M. Quintana, C. Miller, Personal comfort models based on a 6-month experiment using environmental parameters and data from wearables, *Indoor Air*, vol. 32, no. 11, Nov. 2022, p. e13160. DOI: [10.1111/ina.13160](https://doi.org/10.1111/ina.13160)
- [39] J. Kim, Y. Zhou, S. Schiavon, P. Raftery, G. Brager, Personal comfort models: Predicting individuals’ thermal preference using occupant heating and cooling behavior and machine learning, *Build. Environ.*, vol. 129, Feb. 2018, pp. 96–106. DOI: [10.1016/j.buildenv.2017.12.011](https://doi.org/10.1016/j.buildenv.2017.12.011)
- [40] F. Pedregosa, G. Varoquaux, A. Gramfort (+ 13 more authors), Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, vol. 12, 2011, pp. 2825–2830. Online [Accessed 15 February 2022] <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- [41] D. J. Cook, Narayanan C. Krishnan, *Activity learning: Discovering, recognizing, and predicting human behavior from sensor data*, ISBN: 978-1-118-89376-0, April 2015, 288 pp.
- [42] J. Fine, K. L. Branan, A. J. Rodriguez (+ 5 more authors), Sources of Inaccuracy in Photoplethysmography for Continuous Cardiovascular Monitoring, *Biosensors*, vol. 11, 2021, no. 4. DOI: [10.3390/bios11040126](https://doi.org/10.3390/bios11040126)
- [43] ANSI, ANSI/AAMI/IEC 60601-2-27:2011 (R2016) - Medical electrical equipment - Part 2-27: Particular requirements for the basic safety and essential performance of electrocardiographic monitoring equipment. Online [Accessed 12 December 2022] <https://webstore.ansi.org/standards/aami/ansiamiiicc60601272011r2016>
- [44] J. Ogorevc, G. Geršak, D. Novak, J. Drnovšek, Metrological evaluation of skin conductance measurements, *Measurement*, vol. 46, Nov. 2013, no. 9, pp. 2993–3001. DOI: [10.1016/j.measurement.2013.06.024](https://doi.org/10.1016/j.measurement.2013.06.024)
- [45] Empatica, E4 Wristband: User Manual, 2015, pp. 1–32. Online [Accessed 25 September 2023] https://archive.arch.ethz.ch/esum/downloads/manuals/emaptic_s.pdf
- [46] G. Cosoli, A. Poli, S. Spinsante, L. Scalise, The importance of physiological data variability in wearable devices for digital health applications, *Acta IMEKO*, vol. 11, no. 2, May 2022, pp. 1–8. DOI: [10.21014/acta_imeko.v11i2.1135](https://doi.org/10.21014/acta_imeko.v11i2.1135)
- [47] R. A. Piñol, A. S. Mogul, C. K. Hadley, A. Saha, Ch. Li, V. Škop, H. S. Province, C. Xiao, O. Gavrilova, M. J. Krashes, M. L. Reitman, Preoptic BRS3 neurons increase body temperature and heart rate via multiple pathways, *Cell Metab.*, vol. 33, no. 7, Jul. 2021, pp. 1389-1403.e6. DOI: [10.1016/j.cmet.2021.05.001](https://doi.org/10.1016/j.cmet.2021.05.001)
- [48] S. D. Kreibig, Autonomic nervous system activity in emotion: a review., *Biol. Psychol.*, vol. 84, no. 3, Jul. 2010, pp. 394–421. DOI: [10.1016/j.biopsycho.2010.03.010](https://doi.org/10.1016/j.biopsycho.2010.03.010)
- [49] S. A. Mansi, I. Pigliautile, M. Arnesano, A. L. Pisello, A novel methodology for human thermal comfort decoding via physiological signals measurement and analysis, *Build. Environ.*, vol. 222, 2022, p. 109385. DOI: [10.1016/j.buildenv.2022.109385](https://doi.org/10.1016/j.buildenv.2022.109385)