Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 148 (2018) 798–805

Energy Procedia

www.elsevier.com/locate/procedia

73rd Conference of the Italian Thermal Machines Engineering Association (ATI 2018), 12–14 September 2018, Pisa, Italy

Application of IoT and Machine Learning techniques for the assessment of thermal comfort perception.

Francesco Salamone[a,*], Lorenzo Belussi[a], Cristian Currò[b], Ludovico Danza[a], Matteo Ghellere[a], Giulia Guazzi[a], Bruno Lenzi[b], Valentino Megale[b], Italo Meroni[a]

[a]*Construction Technologies Institute, National Research Council of Italy (ITC-CNR), Via Lombardia, 49, 20098 San Giuliano Milanese (MI)*
[b]*SCS, SoftCare Studios Srls, Franco Sacchetti St., 52-00137 Roma, Italy*

Abstract

Thermal comfort is traditionally assessed by using the PMV index defined according to the EN ISO 7730:2005 where the user passively interacts with the surrounding environment considering a physic-based model built on a steady-state thermal energy balance equation. The thermal comfort satisfaction is a holistic concept comprising behavioral, physiological and psychological aspects. This article describes a workflow for the assessment of the thermal conditions of users through the analysis of their specific psychophysical conditions overcoming the limitation of the physic-based model in order to investigate and consider other possible relations between the subjective and objective variables.

Keywords: indoor thermal comfort; wearable; nearable; IoT; machine learning; parametric models.

1. Introduction

Recent studies show how the Internet of Things (IoT) approach has been applied to the physical environment aimed at improving the user's satisfaction [1] and innovative projects were born in order to create smart environment

* Corresponding author. Tel.: +39 3890429919
E-mail address: salamone@itc.cnr.it

10.1016/j.egypro.2018.08.130

based on open source and hardware devices for the optimization of different aspects of the Indoor Environmental Quality, such as: Thermal Comfort (TC) [2], illuminance level and air quality [3]. The so reported example of ubiquitous sensing has allowed to store an enormous amount of data [4]. A branch of Artificial Intelligence (AI), the Machine Learning (ML) [5], can automatically detect patterns among the collected data, predict future trend or define decision-making rules [6]. The Trends service [7] by Google allows to verify the rising interest of the previous key topics (IoT, ML and AI) starting from 2015 (Figure 1).

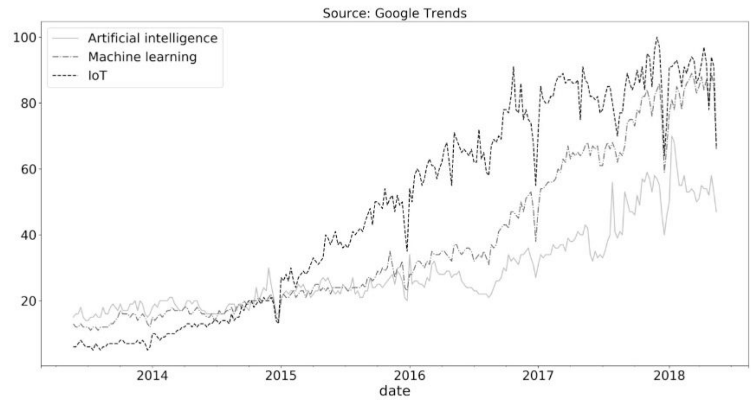

Fig. 1. Google trends for Artificial Intelligence, Machine Learning and IoT.

The article describes a new workflow that connects the theme of ML and IoT with that of TC through an in-field assessment campaign.

1.1. Case study

The system has been installed on the desktop of eight workstations of a two-stories office building located in San Giuliano Milanese, near Milan (Italy) and 8 individuals are involved in the survey. The workstations are placed in 5 offices, 3 on the ground floor and 2 on the first floor of the building (Figure 2).

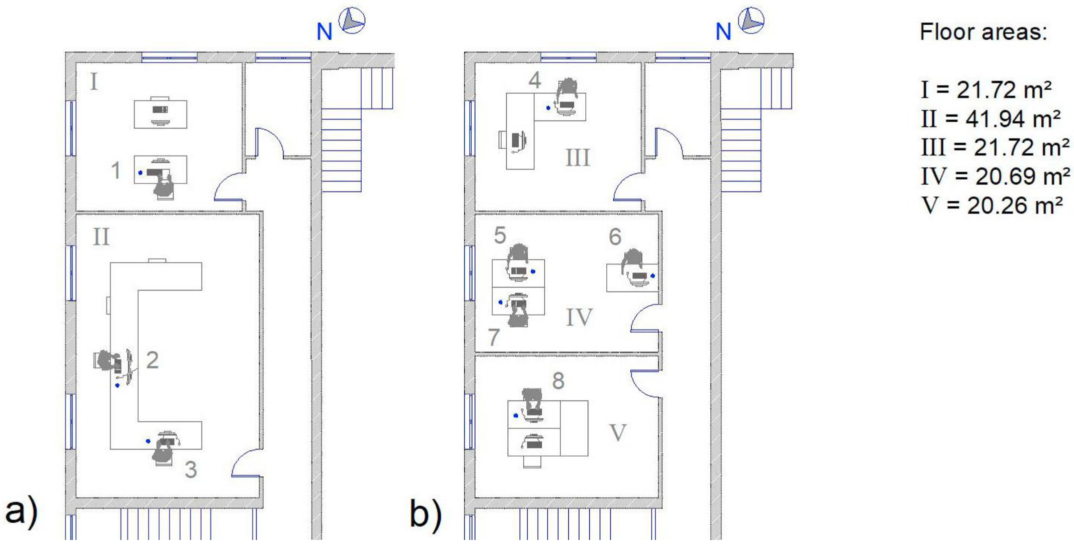

Fig. 2. Workstations: (a) first floor; (b) ground floor. With black dot the point where nearable are installed

Table 1 reports the personal data of the involved users and the periods of the tests and related information about the weather data. All subjects gave their informed consent for inclusion before they participated in the study.

Table 1. User and weather data

User	Age [y]	Weight [kg]	Height [m]	Gender [-]	Profession [-]	Test period [-]	Air temperature [°C] Avg. (St.dv)	Relative humidity [%] Avg. (St.dv)	Solar radiation [W/m^2] Avg. (St.dv)
1	61	61.4	1.75	Male	Senior researcher	Nov 13-17	6.2 (3.9)	85.7 (19.6)	222.0 (153.9)
2	39	81	1.78	Male	Researcher	Nov 6-10	9.3 (1.5)	98.1 (3.5)	102.8 (101)
3	35	85	1.79	Male	Researcher	Nov 6-10	9.3 (1.5)	98.1 (3.5)	102.8 (101)
4	43	46	1.64	Female	Researcher	Nov 20-24	7.2 (2.8)	96.2 (8.4)	143.4 (129.8)
5	29	60	1.60	Female	Junior researcher	Nov 27-30	2.5 (3.1)	88.8 (20.1)	172.0 (164)
6	37	57	1.79	Female	Researcher	Nov 20-24	7.2 (2.8)	96.2 (8.4)	143.4 (129.8)
7	33	80.2	1.91	Male	Technician	Nov 27-30	2.5 (3.1)	88.8 (20.1)	172.0 (164)
8	35	70	1.77	Male	Researcher	Nov 13-17	6.2 (3.9)	85.7 (19.6)	222.0 (153.9)

1.2. Workflow structure

Figure 3 reports the defined workflow for the assessment of TC considering a complete set of environmental and biometric data, acquired by an IoT system composed by a wearable and nearable device. The first one is the Empatica E4 wristband [8], a medical device class II according to the FDA 21 CFR Part 860.3, equipped with the following sensors:

- a Photoplethysmography (PPG) sensor for the detection of the heart rate (HR);
- an Electrodermal Activity (EDA) sensor;
- an infrared thermopile;
- a 3-Axis accelerometer.

The second one is based on low-cost sensors and open-source hardware able to monitor indoor environmental parameters (air temperature, relative humidity, radiant temperature, air velocity, CO_2 concentration, illuminance level). These variables are useful to assess different aspects of the Indoor Environmental Quality (IEQ). More details are available in [9, 10]. The nearable was installed at a distance of less than 40 cm from the worker in order to consider the environmental parameters as close as possible to the user.

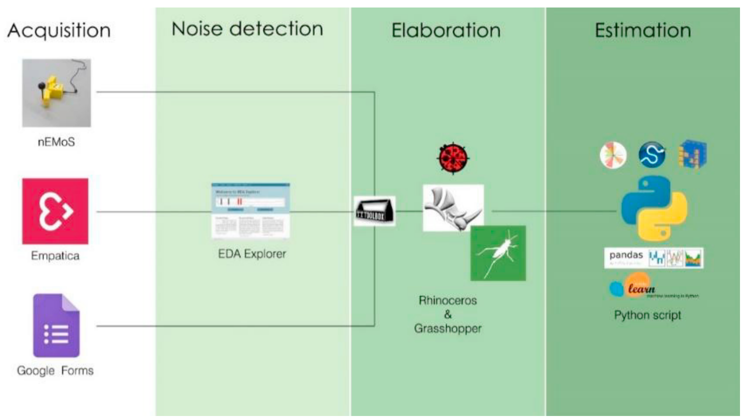

Fig. 3. Workflow of the integrated approach for the assessment of thermal comfort.

The biometric data monitored by the wearable device are processed using a noise detection ML algorithm that allows to automatically detect EDA artefacts [11, 12]. Among the possible reasons of the presence of noise in a wearable device outputs, the most probably is the variation in the contact between the skin and the recording electrode caused by excessive movement or adjustment of the device. Through ML algorithm the raw data, acquired with a sampling frequency of 4 Hz, are divided in periods of 5 seconds and then filtered considering a noise classification number equal to -1 (noise data), or 1 (clean data). Then the environmental data and users' feedback (derived from a google spreadsheet where all data inputted by users on a web multiplatform survey based on a Google Form model are automatically collected), are merged with the filtered biometric dataset considering only the clean data and related time. Besides all data are used for a parametric analysis through a set of open source plug-in [13, 14] for Grasshopper, a graphical algorithm editor tightly integrated with Rhinoceros 3-D modeling tools. This step is useful to define the optimal personal thermal comfort defined (Figure 4): the comfort zones personalized for each user are based on the environmental data averaged over a period of a minute at the time when the users have felt a neutral thermal comfort sensation.

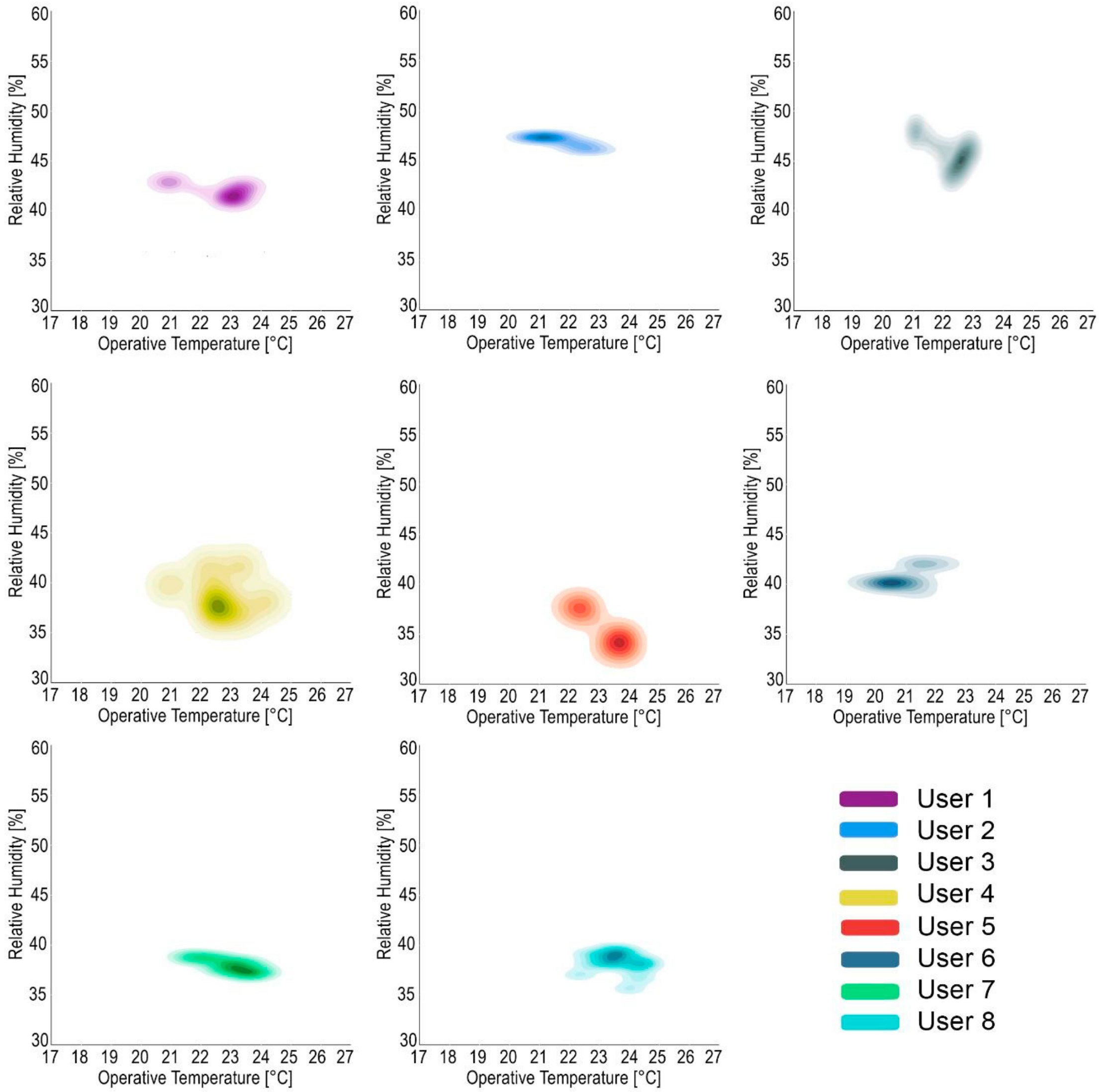

Fig. 4. Personalized Graphic Comfort Zone for all the users

Finally all data are used by a Python script for ML purpose. The following case study allows to verify the application of the workflow considering an in-field assessment.

2. Results and discussion

The workflow reported above allows to filter the dataset considering the output of the EDA explorer algorithm. The heatmap reported in Figure 5 allows to verify the consistencies of the defined dataset considering all filtered data of the users 1, 2, 3, 5, 6, 7 and 8. The data of user 4 have not been considered due to a very high noise level in the biometric monitored data.

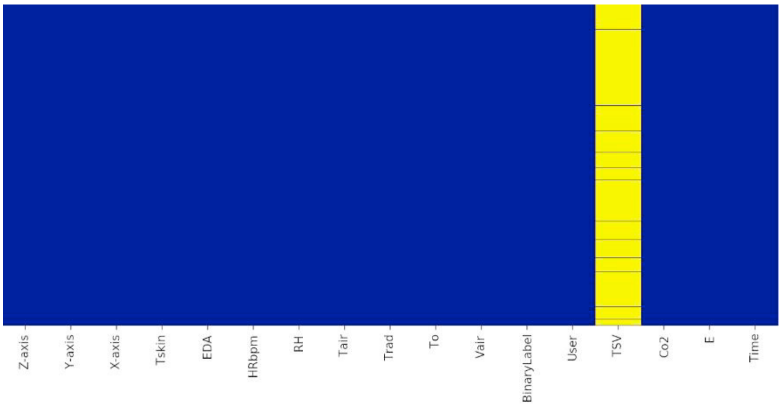

Fig. 5. Output dataset and related parameters: in yellow – null data, in blue – non-null data.

It is possible to highlight how the dataset is completely imbalanced: it is not possible to use the TSV as a target value in a ML approach. To overcome the limit of an imbalanced dataset due to a small number of users' feedback in relation to the environmental and biometric data, the ML is applied to identify the users and consequently their TC perception. Starting from the basic dataset it is possible to define a new one with the characteristics reported in Table 2.

Table 2. Dataset info

Number	Data description	U.M.	Number	Null count	Data type	Model feature importance
0	Z-axis	[g]	8959	0	Double precision float	0.01897298
1	Y-axis	[g]	8959	0	Double precision float	0.03128233
2	X-axis	[g]	8959	0	Double precision float	0.04648461
3	Tskin	[°C]	8959	0	Double precision float	0.11264711
4	EDA	[µS]	8959	0	Double precision float	0.1108617
5	HR	[bpm]	8959	0	Double precision float	0.01466359
6	RH	[%]	8959	0	Double precision float	0.17893774
7	To	[°C]	8959	0	Double precision float	0.16441891
8	CO_2	[ppm]	8959	0	Double precision float	0.06317967
9	E	[lx]	8959	0	Double precision float	0.25855138
10	User	[-]	8959	0	Categorical object	-

The features of the dataset that contribute most to the prediction of user profile are automatically selected in order to: reduce overfitting and training time and improve accuracy. For this purpose the instances from 0 to 9 according to the previous Table 2 are considered. The importance of each feature useful to define the target, number 10, user profile, is defined considering an Extra Trees Classifier that identify an importance score for each attribute. Only the variables with a score higher than 0.11 have been considered: Tskin, EDA, RH, To and E. Figure 6 reports a parallel coordinates plot that display a qualitative correlation among all reference variables [15].

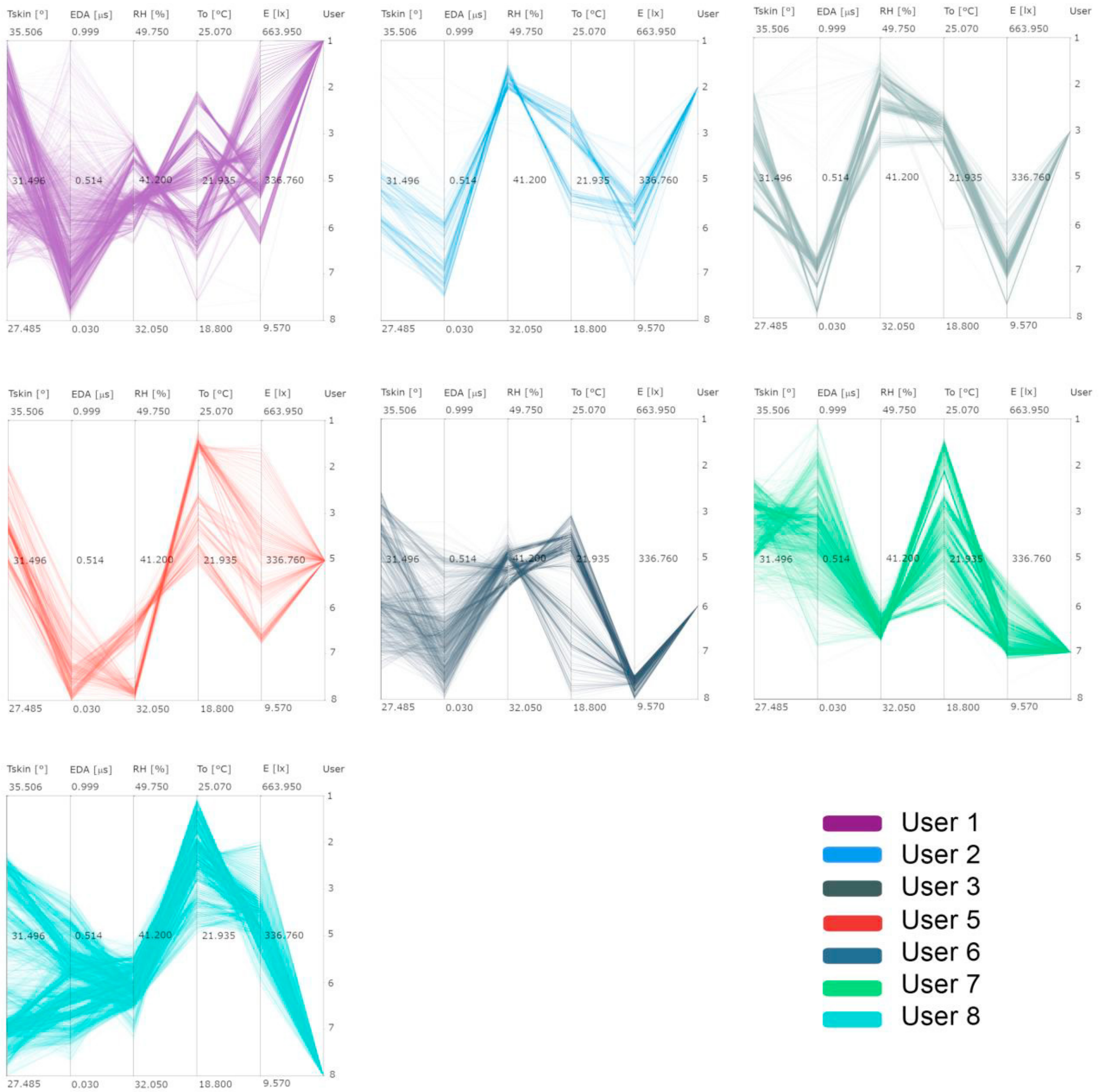

Fig. 6. Correlation among all considered parameters

It is possible to highlight how the Tskin and EDA values are strictly linked to the user profile: so as, for example, if the values of user 2 are compared with those of user 3, it is possible to verify how the human body reacts differently to external stimuli despite the relative humidity and the air temperature are quite comparable for this specific cases. The same consideration can be done if trends of users 5 and 7 are compared. The idea is that it is possible to define a personalized profile considering both biometric and environmental data. Excluding linear dependencies among the variables, four different non-linear algorithms are considered: K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Naive Bayes (NB), Support Vector Machines (SVM). The metric of 'accuracy' is used to evaluate the models defined as the ratio of the number of correctly predicted instances in divided by the total number of instances in the dataset. The k-fold cross validation (k = 10) is used to evaluate the performances of the different algorithms on the dataset (Table 3).

Table 3. Algorithms accuracy comparison

User	Avg.	St.dev.
KNN	0.959	0.006
CART	0.997	0.002
NB	0.953	0.007
SVM	0.979	0.006

Among the models, the CART algorithm, a non-parametric supervised learning method that predicts the value of a target variable by learning simple decision rules inferred from the data features, has the highest average estimation accuracy (0.997) and the lowest standard deviation (0.002).

Table 4 shows the classification report summarizing the results as a final accuracy score of the CART model directly on the validation set.

Table 4. Prediction on validation dataset with CART algorithm

User	Precision	Recall	F1score	Support
1	1.00	1.00	1.00	429
2	1.00	1.00	1.00	60
3	1.00	1.00	1.00	178
4	-	-	-	-
5	1.00	1.00	1.00	95
6	1.00	1.00	1.00	221
7	1.00	1.00	1.00	260
8	1.00	1.00	1.00	549
Avg/Tot	1.00	1.00	1.00	1792

3. Conclusions and future work

The proposed framework has allowed to detect the indoor environmental variables close to users, in addition to the biometric parameters and users' feedback. This approach has permitted to highlight TC differences among users, optimizing the control strategy and identifying the most relevant parameters for users recognition.

The future developments of this study may be focused, on one hand, on a longer detection of both environmental parameters and users' feedback in order to set up a dataset able to predict a customized TC. On the other hand, the monitoring of further variables could be implemented in order to apply the methodology in other IEQ fields [16] like ILQ, IAQ and Acoustic comfort [17, 18, 19] and their interaction with the energy performance of buildings [20, 21, 22].

References

[1] Uzelac, A.; Gligoric, N.; Krco, S. A comprehensive study of parameters in physical environment that impact students' focus during lecture using Internet of Things. Comput. Hum. Behav. 2015, 53, 427–434.

[2] Salamone, F., Belussi, L., Danza, L., Ghellere, M., & Meroni, I. (2016). An open source "smart lamp" for the optimization of plant systems and thermal comfort of offices. Sensors, 16(3), 338.

[3] Salamone, F., Belussi, L., Danza, L., Galanos, T., Ghellere, M., & Meroni, I. (2017). Design and development of a nearable wireless system to control indoor air quality and indoor lighting quality. Sensors, 17(5), 1021.
[4] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.
[5] Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.
[6] Robert, C. (2014). Machine learning, a probabilistic perspective.
[7] Rech, J. (2007). Discovering trends in software engineering with google trend. ACM SIGSOFT Software Engineering Notes, 32(2), 1-2.
[8] Pietilä, J., Mehrang, S., Tolonen, J., Helander, E., Jimison, H., Pavel, M., & Korhonen, I. (2017). Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities. In EMBEC & NBC 2017 (pp. 145-148). Springer, Singapore.
[9] Salamone, F.; Belussi, L.; Danza, L.; Ghellere, M.; Meroni, I. Design and Development of nEMoS, an All-in-One, Low-Cost, Web-Connected and 3D-Printed Device for Environmental Analysis. Sensors 2015, 15, 13012-13027.
[10] Salamone, F.; Danza, L.; Meroni, I.; Pollastro, M.C. A Low-Cost Environmental Monitoring System: How to Prevent Systematic Errors in the Design Phase through the Combined Use of Additive Manufacturing and Thermographic Techniques. Sensors 2017, 17, 828.
[11] Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., and Picard, R. "Automatic Identification of Artifacts in Electrodermal Activity Data" In Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, August 2015.
[12] EDA-explorer. Available online: http://eda-explorer.media.mit.edu/ (accessed on 18 March 2018).
[13] Roudsari, M. S., Pak, M., & Smith, A. (2013, August). Ladybug: a parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design. In Proceedings of the 13th international IBPSA conference held in Lyon, France Aug.
[14] TT toolbox. Available online: http://core.thorntontomasetti.com/tt-toolbox-for-grasshopper (accessed on 18 March 2018).
[15] Inselberg, A. (2009). Parallel coordinates. In Encyclopedia of Database Systems (pp. 2018-2024). Springer US.
[16] Salamone, F., Belussi, L., Danza, L., Ghellere, M., & Meroni, I. (2017). How to control the Indoor Environmental Quality through the use of the Do-It-Yourself approach and new pervasive technologies. Energy Procedia, 140, 351-360.
[17] Torchia, F.; Ricciardi, P.; Scrosati, C.; Scamoni, F. Improvement of Façades' Sound Insulation of Schools near the Bergamo—Orio al Serio International Airport: Case Study. Build. Acoust. 2015; 22(2):123-142.
[18] Casini, D.; Cellai, G.; Fogola, J.; Scamoni, F.; Secchi, S. Correlation between facade sound insulation and urban noise: A contribution to the acoustic classification of existing buildings. Build. Acoust. 2016, 23, 145–158.
[19] Secchi, S.; Astolfi, A.; Calosso, G.; Casini, D.; Cellai, G.; Scamoni, F.; Scrosati, C.; Shtrepi, L. Effect of outdoor noise and façade sound insulation on indoor acoustic environment of Italian schools. Appl. Acoust. 2017, 126, 120–130.
[20] Belussi, L.; Danza, L.; Meroni, I.; Salamone, F.; Ragazzi, F.; Mililli, M. Energy performance of buildings: A study of the differences between assessment methods. In Energy Consumption: Impacts of Human Activity, Current and Future Challenges, Environmental and Socio-Economic Effects; Nova Science Publishers: New York, NY, USA, 2014; pp. 53–75.
[21] Danza, L., Belussi, L., Meroni, I., Salamone, F., Floreani, F., Piccinini, A., & Dabusti, A. (2016). A simplified thermal model to control the energy fluxes and to improve the performance of buildings. Energy Procedia, 101, 97-104.
[22] Danza, L., Belussi, L., Meroni, I., Mililli, M., & Salamone, F. (2016). Hourly calculation method of air source heat pump behavior. Buildings, 6(2), 16.