Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 59-24 (2025) 179–184

Impact of Sitting vs Standing Baselines on Performance Parameters of Stress Classification Models in Assembly Tasks

Danish Abbas Syed*, Walter Quadrini*, Marta Pinzone*

* *Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133, Milano, Italy (danishabbas.syed@polimi.it; walter.quadrini@polimi.it; marta.pinzone@polimi.it)*

Abstract: Within the domain of Industry 5.0, the "Healthy operator" concept aims at using biometric sensors to enhance the physical and mental wellbeing of the operators. However, a sizable portion of the tools and the methodologies involved are borrowed from fields such as neuroscience and medical science where experimental context and conditions are significantly different from the typical manufacturing plant. In the current work we explored if the method for recording baselines of biometric signals used in clinical studies should be directly transferred to the shop floor. We measured baselines in sitting and standing positions and compared these two methods by measuring the performance of Stress state prediction model during standing assembly tasks.

Keywords: Industry 5.0, Baseline, Stress Classification, Cognitive workload, Assembly, Human centric, Machine Learning

1. INTRODUCTION

Human-centricity and the crucial role of operators in manufacturing have recently garnered significant attention from academia, industry, and policymakers (European Commission, 2024)(Beducci et al., 2024)(Pinzone et al., 2024) . In this context, the "Healthy Operator" concept, introduced by (David et al., 2016), focuses on operators' wellbeing using biometric sensors such as Heart Rate (HR), Electro Dermal Activity (EDA), Electro Encephalogram (EEG), Electro Myogram (EMG), and skin temperature (ST) to assess their physical, cognitive, and psychological states(Arsalan et al., 2023). This data is used to adjust the work system to enhance operator wellbeing and performance(Arsalan et al., 2023)(Rosemeyer et al., 2024) (David et al., 2016) (Blandino et al., 2024)

However, the integration and proper usage of such sensors under industrial conditions is quite a challenging task. A part of this challenge comes from the purely informatics related aspects of biosensor integration (Syed et al., 2025). Another pressing facet of this challenge is the legal consequence of using and processing biometric data(Cutrona et al., 2024). Yet another aspect of this challenge comes from the fact that the sensors and the associated techniques and methodologies are often borrowed from fields like medical science, sports science, neuroscience, and other associated fields. The experimental settings in such fields can be significantly different from what is encountered on a manufacturing shop floor. One such issue that is noteworthy is the method of recording baselines (Tran et al., 2023).

Normalization of individual physiological signals with baselines is quite common and recommended (Mathôt et al., 2018). The main reason for such normalization is the fact that physiological signals are highly individual in nature. As such a direct comparison of absolute signal values across subjects cannot yield meaningful outcomes (NUNAN et al., 2010).

In contemporary applications involving tasks with screen-based stimuli(Androutsou et al., 2023) (Karthikeyan et al., 2012) most of the experiments are usually conducted in seated or resting positions. In this context, taking baselines in resting conditions seems quite a logical choice. However, in a manufacturing scenario, the operators often perform tasks in standing positions. Furthermore, the values of physiological signals are often affected by posture (Rajendra Acharya et al., 2005) (Cardenas, 2021). As such the use of resting or sitting baselines may not be the most effective method for normalizing data related to standing tasks.

Despite this difference, our review of sixty papers on manufacturing experiments revealed that most baselines either involve a sitting posture or the authors do not explicitly specify the posture conditions. Only the paper by (Brunzini, Peruzzini, et al., 2021) clearly specified that the baselines were recorded while standing. However, the paper does not offer any justification for deviating from the norm.

In this paper we aim to contribute to filling this gap by evaluating the impact of using sitting vs standing baselines on the performance of stress classification models trained on biometric data gathered while performing standing assembly operations.

According to its objective, the rest of the paper is divided into five sections. Section 2 covers the methodology, which describes the experiment protocol, sensors used and the data set. Section 3 briefly covers the Data processing pipeline. Section 4 describes the results obtained. The conclusion and future work are covered in Section 5.

2. METHODOLOGY

2.1 Experiment protocol

Figure 1 shows the overall experimental protocol. At the start of each experiment the participants were asked to fill in an

10.1016/j.ifacol.2025.11.861

online survey form which explains the benefits of the study. The same survey form was used to collect informed consent and some demographic information of the participants. This was followed by filling a Numerical Analog Scale 0 (NAS 0)(Brunzini, Papetti, et al., 2021) survey to collect the participants self-reported stress levels before the start of the experiment. The participants were then instructed to wear biometric sensors. The sensors were then connected to their respective recording applications.

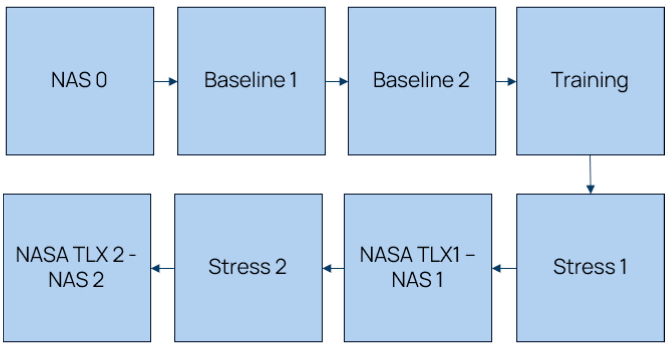

Figure 1:Experiment protocol.

Once sensor connections were verified, two consecutive baseline measurements were taken, one of which was taken while the participants were sitting and another one while standing. In both these phases headphones equipped with active noise cancellation were used to ensure a noise-free calm environment.

Figure 2: Assembled component and the workstation.

The participants then went to the actual assembly station shown in Figure 2 where they underwent a training phase in which each participant read a set of instructions and physically assembled the valve seen in Figure 2. This was done to mitigate learning effects. Based on our experience with the same assembly, the assembly time tends to stabilize after 3 assembly trials. The training phase was followed by two timed assembly phases named Stress 1 and Stress 2 in Figure 1. In Stress 1 phase the participant simply assembled the valve while listening to Industrial noises of up to 60dB max. In the Stress 2 phase the participants were required to perform mental subtraction tasks in addition to the valve assembly and the noise exposure up to 80dB. The limit of 80dB was chosen in accordance with Italian Legislative decree no. 195 dated 10th April 2006. Above this limit a company is obligated to provide personal hearing protection equipment to the employees. With regards to the 60dB limit, it was introduced to uniform basal state to circumvent the intra and inter-day variations in ambient noise. An iOS Application named SoundLab was used to communicate directly to the participants through headphones while being consecutively exposed to the Industrial noise. The second task was specifically designed to induce higher cognitive workload in the participants. To counteract the learning effects the order of the two assembly tasks was reversed in half of the participants. After each assembly phase the participants filled in three surveys a NASA – Task Load Index (NASA-TLX), NAS and a question about perceived complexity level of the task answered on 7-point Likert scale, to capture self-reported measures.

2.2 Experiment apparatus

The study used three devices to measure biometric signals. An Empatica E4 to capture EDA and ST, Neon eye tracker for capturing pupil data and a Polar H10 for capturing cardiac activity. Although Empatica E4 also reports cardiac activity, the Polar H10 was preferred due to its higher accuracy and less likelihood of motion artefacts (Hinde et al., 2021)(Giggins et al., 2021). Sony WH-1000XM5 headphones were used for communicating with the participants and administering industrial noise.

Figure 3:Sensors used.

The Sampling frequency of the sensors were as follows: for EDA and ST it was 4 Hz and for pupil data it was 200 Hz. For Cardiac activity we relied on the processed R-R interval data from the Polar H10 which internally samples ECG data at a sampling frequency of 1000 Hz to calculate R-R intervals.

2.3 Data acquisition

iMotions software was used as the primary data acquisition platform for acquisition of data. The platform supports seamless integration of multiple sensors and management of experiments.

It is also possible to manually integrate sensors. In our case, the integration of Polar H10 with iMotions was managed with the help of a python script. Detailed instructions and scripts are available at the doi link associated with (Syed, 2025)

The Neon Companion mobile app was used for interfacing with the Neon eye tracker. The eye tracker is integrated by default in the iMotions platform which allows the user to remotely start and stop the recording within the mobile app of the eye tracker. This is quite convenient in experimental settings because the eye tracker must be connected to the phone hosting the mobile app with a USB Type-C cable.

For the Empatica E4, the E4 Realtime mobile application was used to record the signals.

2.3 Dataset description

The experiment included 28 participants: 11 females and 17 males. Most participants were university students aged 29±6 years. Participants came from seven nationalities, with Italian being the most common. They were recruited using word of mouth. None of the participants had prior experience in assembling the component that was used for the purpose of this experiment.

4. DATA ANALYSIS PIPELINE

3.1 Preprocessing

Since the signals were coming from three different devices, the data representation format was unified, this primarily involved the conversion of timestamps to a common format. Including the conversion from UTC to local time. This was done to simplify the programming effort.

An Excel file with the start and end of each experimental phase was created for all participants. This was done post-experiment by viewing the scene camera footage of the Neon eye tracker and adding the necessary markers on the Pupil Cloud platform. The markers were downloaded along with other files as a markers.csv file.

This file was used to segment the data from each of the four sensors into four phases i.e., Sitting Baseline, Standing Baseline, Stress 1, and Stress 2. This step was conducted with the help of python scripts. To ensure that the length of each data segment was consistent with the length of experiment phase, the script generated a short report summarizing expected and actual lengths of the segments.

3.2 Artefact correction

After splitting the data into segments, each sensor data was separately treated to take care of the artefacts in each of the signals.

For the cardiac data we used a combination of filters described in (Benchekroun et al., 2022) and a median filter-based data cleaning like the one used in Kubios HRV software (Tarvainen et al., 2014).

The Neon eye tracker has its own blink detection algorithm and generates a blinks.csv file containing the start, end timestamps and duration of the blinks. This was directly used to remove the blink artefacts. The missing data was then replaced with linear interpolation(Hershman et al., 2018)(Mathôt et al., 2018).

A fourth order median filter was used to eliminate anomalous ST data from Empatica E4.

For EDA data we used the preprocessing functions for EDA from the neurokit2 python library (Makowski et al., 2021).

3.3 Data Visualization

A 2-stage data visualization was performed. The first set of plots covered the raw unsegmented data. The second stage was used to visually compare clean data to raw data.

3.4 Feature extraction

For each of the rest phases, a single 3-min window was used for the extraction of baseline values of the features. However, for the Stress 1 and Stress 2 phases, 3-minute windows were used to extract three sets of feature vectors for each stress phase.

A total of 31 features were extracted from all four sensors. The guidelines provided by(Luzzani et al., 2024) were used for selecting which features to be extracted. For feature extraction a mixture of self-developed python code and existing library functions were used.

For Heart Rate Variability (HRV) features self-developed code was used for time domain features and pyHRV (Gomes et al., 2019) python library was used for frequency domain features.

ST and pupil data feature extraction was based on self-written code. While Neurokit2 Python library (Makowski et al., 2021) was used for extracting EDA features.

3.5 Normalization

Three distinct types of baseline normalization methods were used i.e., division (1), subtraction (2), subtraction with division (3). Although specific recommendations about normalizations methods exist in literature (Mathôt et al., 2018), for the sake of simplicity, each time we compared standing baseline to sitting baseline normalization, only one of these three normalization methods were used for all the 31 extracted features. Figure 4 shows a pictorial representation of the normalization pipeline. Where, “p” stands for participant number, “n” stands for feature number and “m” stands for the normalization method used each of which is described by equations (1), (2) and (3). The 6 data sets obtained from this normalization process were then used to train the machine learning models which were used for a comparative analysis summarized in Table 1.

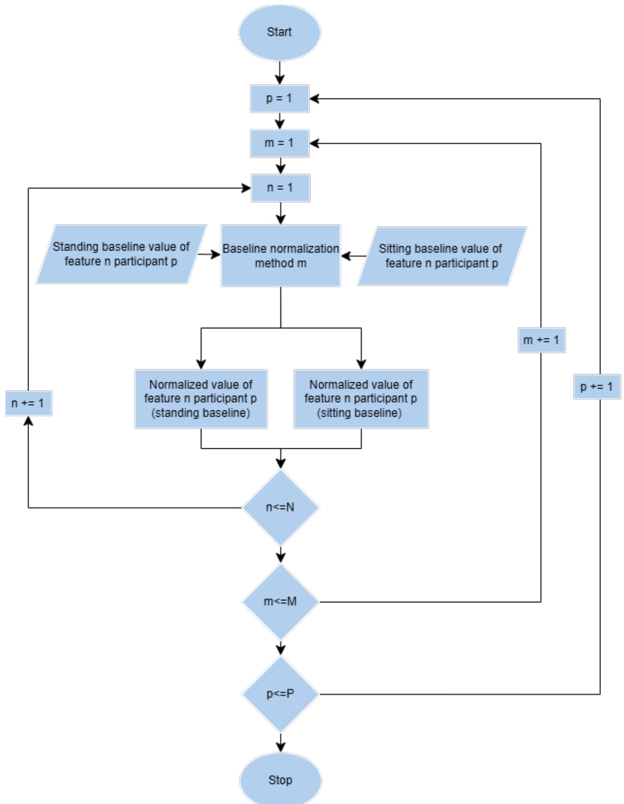

Figure 4: Baseline normalization pipeline

$$\frac{Feature\ value}{Baseline\ value} \tag{1}$$

$$Feature\ value - Baseline\ value \tag{2}$$

$$\frac{Feature\ value - Baseline\ value}{Baseline\ value} \tag{3}$$

3.6 Machine Learning

To decide whether sitting or standing baselines are better, Random Forest classifiers were trained on the six combinations of datasets. Given the small size of the dataset, 10-Fold cross validation was performed, and F1 Score was chosen as a metric to compare the effectiveness of standing and sitting baseline recording, the underlying hyper parameters of each model were tuned to obtain the best score for each case. A high-level diagram of the whole process is shown in Figure 5. The process was repeated 3 times corresponding to each different normalization method.

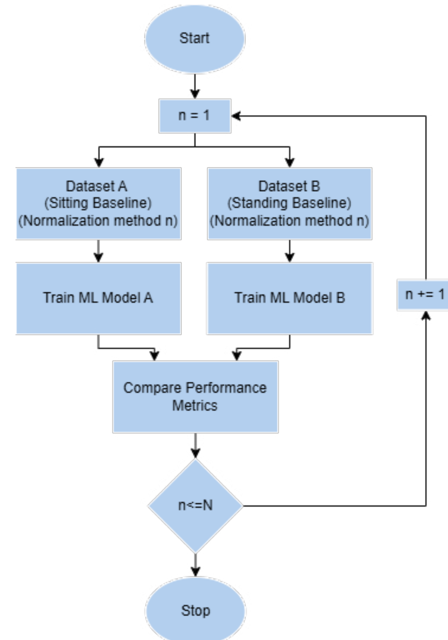

Figure 5: High level ML training and comparison pipeline

4. RESULTS

Table 1 depicts the overall results of the experiment. In general, we can notice that the F1 score obtained when using standing baselines is marginally higher than the F1 scores obtained when sitting baselines were used for the normalization process.

Table 1: F1 Score comparison of standing and sitting baseline.

	F1 Score	
	Standing Baseline	**Sitting Baseline**
Division	0.8395	0.8208
Subtraction	0.8466	0.8364
Subtraction with division	0.8395	0.8208

Furthermore, we also notice that this trend holds true regardless of the method used for normalization. Additionally subtractive baseline produces the best performing model. While there is no difference between the division-based method (1) and the method with division and subtraction (3). The confusion matrices corresponding to the best performing model set from Table 1 i.e. subtraction based baseline normalization are reported in Figure 6 and Figure 7. From the confusion matrices we observe that the use of standing baseline marginally improves the prediction accuracy of Stress

1 phase without affecting the Stress 2 phase prediction.

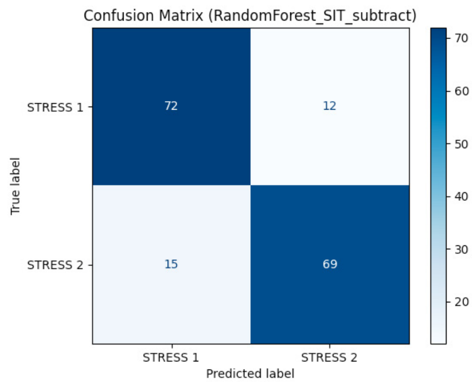

Figure 6: Confusion matrix subtraction method, sitting baseline

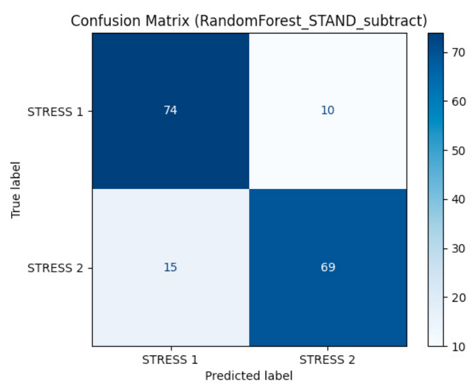

Figure 7: Confusion matrix subtraction method, standing baseline

5. CONCLUSION AND FUTURE WORK

The results of the analysis seem to indicate that using standing baselines does indeed improve the performance of a stress prediction ML model. Even though the differences are not significantly large, the performance improvements are consistent across all baseline calculation methodologies. Which proves our initial hypothesis that performance of stress prediction models is indeed affected by whether baselines are measured while standing or sitting. Therefore, for standing assembly tasks, it is preferable to take baseline measurements while standing rather than sitting.

Our findings should be interpreted considering the limitations of the current study, which we recommend future research should address. First, the study only deals with simplified versions of the analysis, e.g., only one of the three normalization methods (1), (2), (3) was used at a time for all signal features. In future analysis the normalization method can be optimized to the specific type of physiological signal or a signal feature. Additionally, other more complex forms of normalization like z-score normalization can be explored. Second, the analysis of other self-reported measures, like NASA-TLX, and assembly performance measures, like errors and number of assembled pieces, can be included in future studies. Third, a larger dataset with a more diverse participant pool can help increase the generalizability and robustness of the results. Finally, the experiment design for the current work involved standing assembly tasks. As a further step it would be interesting to verify if a similar argument can be made for sitting tasks.

In conclusion, despite its limitations, the present study is one of the first to investigate the impact of sitting versus standing baselines on stress classification during manufacturing assembly tasks. In doing so, it contributes to paving the way for better assessment and management of operators' wellbeing in Industry 5.0.

ACKNOWLEDGEMENTS

This work has been supported by the projects Programming Platform for intelligent COlaborative DEployements (P2CODE), and European Lighthouse to Manifest Trustworthy and Green AI (ENFIELD) which have respectively received funding from the European Union's Horizon Europe research and innovation program under grant agreements 101093069 and 101120657, as well as by the HumanTech Project. The HumanTech Project is financed by the Italian Ministry of University and Research (MUR) for the 2023-2027 period as part of the ministerial initiative "Departments of Excellence" (L. 232/2016). The initiative rewards departments that stand out for the quality of the research produced and funds specific development projects.

The authors express their gratitude towards the contributions of Moreno Biroli and Federico Miselli for conducting the experiments and collecting the data.

REFERENCES

Androutsou, T., Angelopoulos, S., Hristoforou, E., Matsopoulos, G. K., & Koutsouris, D. D. (2023). Automated Multimodal Stress Detection in Computer Office Workspace. *Electronics (Switzerland)*, *12*(11). https://doi.org/10.3390/electronics12112528

Arsalan, A., Majid, M., Nizami, I. F., Manzoor, W., Anwar, S. M., & Ryu, J. (2023). *Human Stress Assessment: A Comprehensive Review of Methods Using Wearable Sensors and Non-wearable Techniques*. https://arxiv.org/abs/2202.03033

Beducci, E., Acerbi, F., Pinzone, M., & Taisch, M. (2024). Unleashing the role of skills and job profiles in circular manufacturing. *Journal of Cleaner Production*, *449*, 141456. https://doi.org/https://doi.org/10.1016/j.jclepro.2024.141456

Benchekroun, M., Chevallier, B., Istrate, D., Zalc, V., & Lenne, D. (2022). Preprocessing Methods for Ambulatory HRV Analysis Based on HRV Distribution, Variability and Characteristics (DVC). *Sensors*, *22*(5), 1984. https://doi.org/10.3390/s22051984

Blandino, G., Colombo, S., & Montagna, F. (2024). On-site analysis of work-related stress to design workers-friendly manufacturing systems. *Proceedings of the Design Society*, *4*, 205–214. https://doi.org/DOI: 10.1017/pds.2024.23

Brunzini, A., Papetti, A., Germani, M., Barbadoro, P., Messi, D., & Adrario, E. (2021). *Mixed Reality Simulation for Medical Training: How It Affects Learners' Cognitive*

State (pp. 339–347). https://doi.org/10.1007/978-3-030-79763-8_41

Brunzini, A., Peruzzini, M., Grandi, F., Khamaisi, R. K., & Pellicciari, M. (2021). A Preliminary Experimental Study on the Workers' Workload Assessment to Design Industrial Products and Processes. *Applied Sciences*, *11*(24), 12066. https://doi.org/10.3390/app112412066

Cardenas, A. K. (2021). *A comparison of computer workstation postures: active sitting, traditional sitting and standing*. https://unbscholar.lib.unb.ca/handle/1882/14284

Cutrona, V., Niko, B., Elias, M., Tamas, R., Giacomo, D., & and Pedrazzoli, P. (2024). Extending factory digital Twins through human characterisation in Asset Administration Shell. *International Journal of Computer Integrated Manufacturing*, *37*(10–11), 1214–1231. https://doi.org/10.1080/0951192X.2023.2278108

David, R., Stahre, J., Wuest, T., Noran, O., Bernus, P., Berglund, Å. F., & Gorecky, D. (2016). *TOWARDS AN OPERATOR 4.0 TYPOLOGY: A HUMAN-CENTRIC PERSPECTIVE ON THE FOURTH INDUSTRIAL REVOLUTION TECHNOLOGIES*. https://api.semanticscholar.org/CorpusID:125890216

Giggins, O. M., Doyle, J., Sojan, N., Moran, O., Crabtree, D. R., Fraser, M., & Muggeridge, D. J. (2021). Accuracy of Wrist-Worn Photoplethysmography Devices at Measuring Heart Rate in the Laboratory and During Free-Living Activities. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 6970–6973. https://doi.org/10.1109/EMBC46164.2021.9629522

Gomes, P., Margaritoff, P., & da Silva, H. (2019, February). *pyHRV: Development and Evaluation of an Open-Source Python Toolbox for Heart Rate Variability (HRV)*.

Hershman, R., Henik, A., & Cohen, N. (2018). A novel blink detection method based on pupillometry noise. *Behavior Research Methods*, *50*(1), 107–114. https://doi.org/10.3758/s13428-017-1008-1

Hinde, K., White, G., & Armstrong, N. (2021). Wearable devices suitable for monitoring twenty four hour heart rate variability in military populations. In *Sensors (Switzerland)* (Vol. 21, Issue 4, pp. 1–20). MDPI AG. https://doi.org/10.3390/s21041061

Innovation, E. C. D.-G. for R. and. (2024). *ERA industrial technologies roadmap on human-centric research and innovation for the manufacturing sector*. Publications Office of the European Union. https://doi.org/doi/10.2777/0266

Karthikeyan, P., Murugappan, M., & Yaacob, S. (2012). Analysis of stroop colorword test-based human stress detection using electrocardiography and heart rate variability signals. *Arabian Journal for Science and Engineering*, *39*(3), 1835–1847. https://doi.org/10.1007/s13369-013-0786-8

Luzzani, G., Buraioli, I., Guglieri, G., & Demarchi, D. (2024). EDA, PPG and Skin Temperature as Predictive Signals for Mental Failure by a Statistical Analysis on Stress and Mental Workload. *IEEE Open Journal of Engineering in Medicine and Biology*. https://doi.org/10.1109/OJEMB.2024.3515473

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, *50*(1), 94–106. https://doi.org/10.3758/s13428-017-1007-2

NUNAN, D., SANDERCOCK, G. R. H., & BRODIE, D. A. (2010). A Quantitative Systematic Review of Normal Values for Short-Term Heart Rate Variability in Healthy Adults. *Pacing and Clinical Electrophysiology*, *33*(11), 1407–1417. https://doi.org/10.1111/j.1540-8159.2010.02841.x

Pinzone, M., Paola, F., & and Taisch, M. (2024). Skills for Industry 4.0: a structured repository grounded on a generalized enterprise reference architecture and methodology-based framework. *International Journal of Computer Integrated Manufacturing*, *37*(8), 952–971. https://doi.org/10.1080/0951192X.2023.2278105

Rajendra Acharya, U., Kannathal, N., Mei Hua, L., & Mei Yi, L. (2005). Study of heart rate variability signals at sitting and lying postures. *Journal of Bodywork and Movement Therapies*, *9*(2), 134–141. https://doi.org/10.1016/j.jbmt.2004.04.001

Rosemeyer, J., Pinzone, M., & Metternich, J. (2024). *Digital Assistance Systems to Implement Machine Learning in Manufacturing: Current State and Future Directions*. https://doi.org/10.20944/preprints202410.0522.v1

Syed, D. A. (2025). *danish123117/iMotions_polarH10_connect: For_citations* . Zenodo. https://doi.org/10.5281/zenodo.14796915

Syed, D. A., Quadrini, W., Rahmani Choubeh, N., Pinzone, M., & Gusmeroli, S. (2025). Approaching Interoperability and Data-Related Processing Issues in a Human-Centric Industrial Scenario. In M. Presser, A. Skarmeta, S. Krco, & A. González Vidal (Eds.), *Global Internet of Things and Edge Computing Summit* (pp. 21–34). Springer Nature Switzerland.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Tran, T. A., Pentek, M., Motahari-Nezhad, H., Abonyi, J., Kovacs, L., Gulacsi, L., Eigner, G., Zrubka, Z., & Ruppert, T. (2023). Heart Rate Variability Measurement to Assess Acute Work-Content-Related Stress of Workers in Industrial Manufacturing Environment - A Systematic Scoping Review. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *53*(11), 6685–6692. https://doi.org/10.1109/TSMC.2023.3282141