Decoding Indoor Comfort Models through Dynamic Physiological and Environmental Data Collection on Vehicles

Name Surname[*x,+], Name Surname[†+,x], Name Surname[‡+,x], and Name Surname[§+,x]

[+]Engineering Department, University of Perugia, 06125 Perugia, Italy
[x] CIRIAF—Interuniversity Research Center, University of Perugia, 06125 Perugia, Italy

ABSTRACT

The concept of a smart environment stands for a significant shift in personalized comfort and intelligent indoors control. Key technologies, such as the Internet of Things (IoT), Artificial Intelligence (AI), cloud computing, and multisensory data analysis, can be integrated to translate the preferences of occupants. This study aims to explore the intricate correlation between individual preferences and environmental conditions, adding dynamic physiological and environmental measurements during participants commuting activities. The central focus of our experiment lies in formulating personalized comfort models for commuters by exploring the combination of physiological signals, namely electrodermal activity, skin temperature and heart rate, and thermal environmental parameters. Data from eight different participants was collected during summertime in Perugia (Italy), for at least two entire weeks of their commuting activities to describe their thermal exposure. The data from physiological signals were gathered all the times the participant was driving. The collected data was employed to set up personalized comfort profiles, highlighting the impact of environmental fluctuations on subjects physiological measurements collectively (and individually). The investigation was not reliant on comfort survey application. In parallel, dedicated lab studies on thermal comfort and individual indoor comfort for commuters were conducted in the offices in order to analyse their preferences and attitudes in controlling in-vehicle HVAC with respect to their daily routine while in office. Our study aims to pave the way for innovative solutions that prioritize individual satisfaction and well-being, recognizing the personalized nature of comfort within varying environmental contexts. Additionally, this study could contribute to the development of tailored indoor comfort solutions by addressing individual physiological and environmental needs within dynamic indoor settings

Keywords: Physiological signal processing, Wearable sensors, Multisensory data analysis, Personal comfort models, Machine learning

1 Introduction

Comfort, a pivotal facet of human well-being and productivity, exhibits significant variations among individuals within diverse environmental contexts. Consider a routine journey—a daily commute—where the concept of comfort transcends mere physical movement, becoming a dynamic interplay between the individual and their surroundings. In this intricate interaction, the fundamental nature of comfort is intricately interwoven into the fabric of our daily experiences, displaying nuanced characteristics.

Comfort is not a one-size-fits-all experience; it is a kaleidoscope that intertwines the threads of human physiology with the ever-shifting backdrop of the environment. The question arises: What if we could decipher this intricate interaction and understand the unique pattern of comfort for each individual?

This brings us to the domain of personal comfort models—an exploration unfolding at the intersection of human physiology and environmental conditions. Our goal is to unravel the stories told by the body's signals and the whispers of the surroundings. It is an expedition where you, the individual, are the protagonist, and the stage is the indoor space where your daily odyssey unfolds.

[*]Email: name.surname@example.com
[†]Email: name.surname@example.com
[‡]Email: name.surname@example.com
[§]Email: name.surname@example.com

Envision a wrist-worn companion, the Empatica E4, capturing the silent conversations within your body—heartbeat, skin's tales of arousal, temperature fluctuations, and the cadence of your movements. Now, add to this ensemble the ambient murmurs of the indoor environment—a controlled microcosm like a car, where you navigate through the ebbs and flows of temperature, humidity, and the gentle breeze.

This narrative comes to life through a symphony of data—physiological signals seamlessly blending with environmental cues. It is not just a collection of numbers; it is the brushstrokes painting a portrait of your comfort, a tapestry woven with threads of individuality.

Why embark on this journey? Understanding these tales of comfort has the power to transform realms beyond the immediate. Our exploration extends beyond the confines of this study, holding the promise of revolutionizing automotive comfort systems, shaping user-centric smart buildings, and elevating wearable technology to new heights of well-being.

In the sections that follow, we delves into the details of our methodology, witness the unveiling of patterns in physiological and meteorological data, and delve into the realms of practical applications. We narrate not just a scientific endeavor but a story of potential impact on user satisfaction and well-being.

The research aims to bridge the physiological signals within the human body with environmental conditions, laying the groundwork for personalized comfort models. The potential applications extend beyond this study, including the optimization of automotive comfort systems, the design of user-centric smart buildings, and advancements in wearable technology for enhanced well-being.

In the upcoming Sections 2 through 4, we provide a systematic presentation of our methodology, results and discussions, and conclusion, offering a detailed exploration of the scientific landscape underpinning the development of personalized comfort models.

2 Methodology

Table 1: Selected physiological features

Domain	Feature	Description	Unit
FD	hrv_vlf	Power in very low frequency range below or equal 0.04 Hz	–
FD	hrv_lf	Power in low frequency range 0.04 Hz and 0.15 Hz	–
FD	hrv_hf	Power in high frequency range 0.15 Hz and 0.4 Hz	–
TD	hrv_sdnn	SD of all NN intervals	ms
TD	hrv_rmssd	Root mean square of successive differences between adjacent NN intervals.	ms
TD	hrv_mean_nni	Average time between successive normal heartbeats	ms
TD	hrv_median_nni	Indicate central tendency in heart rate variability	ms
TD	hrv_range_nni	Difference between maximum and minimum normal-to-normal intervals	ms
TFD	hrv_entropy	Entropy of IBIs, reflect complexity of heart rate time series	–
TD	eda_tonic_mean	Mean of the tonic component of the EDA signal	μS
TD	eda_tonic_std	Standard deviation of the tonic component	μS
TD	eda_tonic_sum	Summation of the tonic component	μS
TD	eda_tonic_energy	Energy of the tonic component	μS^2
TD	eda_phasic_mean	Mean of the phasic component	μS
TD	eda_phasic_std	Standard deviation of the phasic component	μS
TD	eda_phasic_sum	Summation of the phasic component	μS
TD	eda_phasic_energy	Energy of the phasic component	μS^2
TFD	acc_acc_x_min	Minimum value of acceleration along the X-axis	–
TFD	acc_acc_y_min	Minimum value of acceleration along the Y-axis	–
TFD	acc_acc_z_min	Minimum value of acceleration along the Z-axis	–
TFD	acc_l2_min	Minimum magnitude of acceleration vector (Euclidean norm)	–
TFD	acc_acc_x_max	Maximum value of acceleration along the X-axis	–
TFD	acc_acc_y_max	Maximum value of acceleration along the Y-axis	–
TFD	acc_acc_z_max	Maximum value of acceleration along the Z-axis	–
TFD	acc_l2_max	Maximum magnitude of acceleration vector (Euclidean norm)	–
TFD	acc_acc_x_ptp	Peak-to-peak range of acceleration along the X-axis	–
TFD	acc_acc_y_ptp	Peak-to-peak range of acceleration along the Y-axis	–
TFD	acc_acc_z_ptp	Peak-to-peak range of acceleration along the Z-axis	–
TFD	acc_l2_ptp	Peak-to-peak range of magnitude of acceleration vector (Euclidean norm)	–

TD:Time domain FD:Frequency domain TFD:Time Frequency Domain μS:Microsiemens ms:Milliseconds

In this section, we provide a detailed account of the procedures and techniques employed in our research to collect and analyze physiological and environmental signals for the development of personal comfort models.

2.1 Data acquisition protocol

Eight participants, representing diverse demographics, were recruited for this study. Recorded participant details, including age, gender, and relevant health information, aimed to assess potential influences on comfort preferences. Informed consent was obtained, and participants were acquainted with the study's objectives and the use of wearable devices. To monitor physiological responses, participants wore wrist-worn Empatica E4 devices and additional strategically placed sensors inside the vehicle.

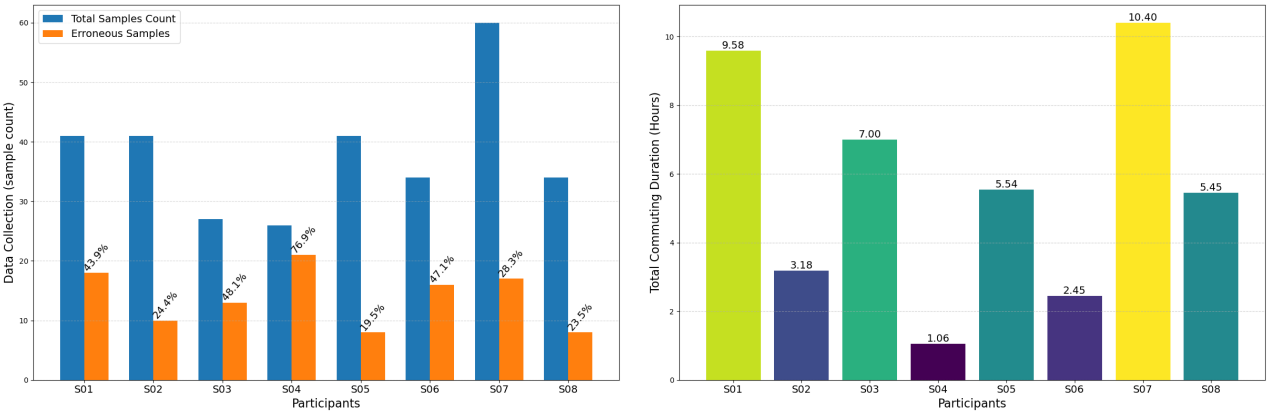

Figure 1: *(a) Erroneous samples in collected physiological data per participants. (b) Total commuting duration of considered samples for each participants*

The Empatica E4 wristwatch recorded heart rate, skin conductance, temperature, and accelerometer data. Detailed technical information is provided in appendix section on empatica E4. Simultaneously, four tinytag sensors within the vehicle (left front seat, right front seat, back of the left front seat, and left rear seat) captured environmental data, encompassing air temperature, humidity, and wind speed. This combined approach allowed the synchronized monitoring of individual physiological responses and environmental conditions, providing a comprehensive assessment of an individual's physiological state.

The Empatica E4 wristwatch served as the core element of the data collection setup, continuously recording physiological data. These parameters offered insights into cardiovascular responses, stress levels, thermal comfort, and physical activity. The simultaneous use of tinytag sensors further enriched the dataset by capturing environmental parameters, providing a holistic understanding of comfort levels. Data collection occurred during the summer of 2023 in Perugia, Italy, focusing on the daily commuting activities of eight participants—two men and six women.

Participants were instructed on the operation of Empatica E4 devices and adjusted controlled comfort parameters, such as ventilation and car wind, aligning them with personal preferences during commuting activities. Physiological signals were exclusively collected during participants' commuting activities, ensuring a targeted focus. Simultaneously, strategically placed tinytag sensors within the vehicle recorded environmental parameters, providing a comprehensive snapshot of the indoor environment.

This synchronized data collection facilitated a holistic understanding of the factors influencing personalized comfort preferences in indoor vehicle settings, forming the foundation for subsequent analyses and the development of personalized comfort models. In the next subsection, we would discuss the environmental and physiological data processing.

2.2 Environmental data

In the meticulous curation of environmental data, our methodology began with the thorough cleaning of the raw dataset acquired from the four *tinytags* sensors, denoted as *N1, N2, N3, N4,*. This rigorous process aimed to eliminate extraneous values and information, ensuring a refined dataset suitable for in-depth analysis. Subsequently, the data underwent structuring into a standardized format encompassing DateTime, Temperature, Humidity, and Dew-Point columns.

To address temporal considerations, the original environmental data samples were initially recorded at a frequency of 45 seconds. To align with experimental requirements and ensure consistent analysis, we applied resampling to the time series data, adjusting the frequency to 1 second. This meticulous operation resulted in a uniform temporal resolution, a crucial step for maintaining consistency across sensors and facilitating meaningful comparisons.

Following resampling, the environmental data underwent interpolation to address missing values arising from the shift in frequency. During this process, certain seconds within the time series, initially unrecorded, were filled through estimation, ensuring a continuous dataset. Interpolation is paramount for creating a seamless dataset, enabling a thorough exploration of environmental patterns and trends over time.

The final structured environmental dataset is a time-series featuring DateTime, Temperature, Humidity, and Dew-Point columns, now stands as a comprehensive and consistent representation of the conditions recorded by sensors *N1, N2, N3, N4*. This refined dataset forms the cornerstone for subsequent analyses of physiological signals, playing a pivotal role in understanding the consensus related to indoor comfort, climate monitoring, and the development of personalized comfort models.

2.3 Physiological data

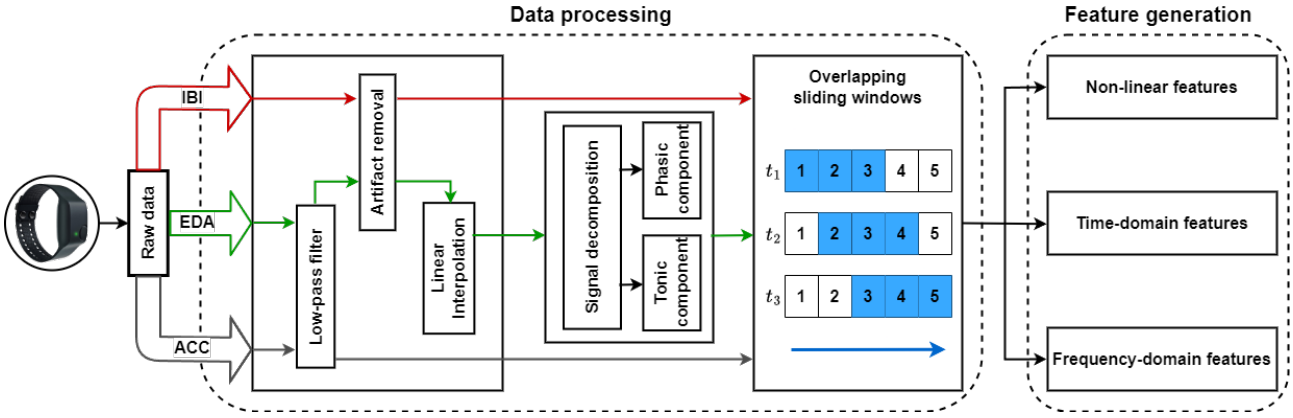

Figure 2: *Empatica features processing pipeline*

As illustrated in fig-1(a), our study involved 304 commuting instances by eight participants, with 112 instances containing erroneous data. To address this, we extracted the remaining 192 instances for final processing, shown in fig-1(b) displays the total hours each participant spent commuting after excluding the flawed data.

Wearable data recordings often encounter challenges such as artifacts, gaps, or deviations from the expected measurement regime [1]–[3]. We adopted a data segmentation approach, breaking down the time-series data into 180-second windows and extracting features for each window as shown in data proocessing pipeline fig-2. The current literature emphasizes three feature categories for physiological aspects: statistical, time-domain, and frequency-domain features [4]–[8]. Features for HRV, EDA, and ACC were chosen based on these categories, as outlined in table-1.

Empatica E4 wearables utilize photoplethysmography (PPG) sensors to measure Inter-Beat Intervals (IBIs), representing the time between consecutive heartbeats. Due to the unavailability of raw PPG recordings, we focused on processing the more accessible IBI (or normal-to-normal (NN) interval) data. Using this IBI data, we computed HRV measures to capture the variations between successive heartbeats. In the initial preprocessing step, we removed evident artifacts within the IBI series. Additionally, IBIs outside the physiologically feasible range of 250–2000 ms (equivalent to a heart rate (HR) of 30–240) were discarded [9]. The cleaned IBIs were then partitioned into overlapping time windows for feature generation.

To mitigate the impact of measurement noise, a low-pass filter with the desired cut-off frequency was applied. This filter selectively retains frequencies below a specified threshold. Employing low-pass filtering is a conventional method for automatically reducing noise in EDA signals [10]–[12]. Detecting and eliminating artifacts utilized machine learning (ML) algorithms, which have proven effective in accurately identifying artifacts in EDA signals recorded by wrist-worn devices [13], [14].

The Electrodermal Activity (EDA) signal encompasses two fundamental Skin Conductance (SC) components: Skin Conductance Response (SCR) and Skin Conductance Level (SCL), representing the phasic and tonic components, respectively. In our analysis, the focus on EDA begins with the preprocessing of the signal to address artifacts and deviations. Subsequently, the calculation of features involves extracting relevant information from both SCR and SCL components.

To initiate the feature extraction process, the EDA signal undergoes thorough preprocessing. This involves the application of low-pass filters to minimize noise, a common approach supported by prior studies in the field [15], [16]. The resulting filtered EDA signal serves as the basis for subsequent feature computation.

Building on established methodologies [17], we compute a set of features for each ACC axis and extend this analysis to include the L2-norm of the three ACC axes. This comprehensive approach provides a detailed understanding of the multidimensional aspects of ACC data.

The data processing and feature generation pipeline are visually represented in fig-2. Following the generation of HRV, EDA and ACC features, we meticulously check for missing values and employ the data imputation technique *HistGradientBoostingRegressor*. Moreover, we eliminate the time offset (+00:00) from the index values. Subsequently, feature selection is conducted with a dual emphasis on both time and frequency domains, ensuring the retention of the most relevant information for subsequent analyses.

2.4 Data fusion

Following the preprocessing of physiological and environmental signals, it was imperative to synchronize the timestamped physiological data with the corresponding environmental time-series data. This synchronization aimed to establish meaningful relationships between physiological responses and meteorological conditions,

allowing for a precise alignment of physiological events and environmental conditions. The goal was to consolidate the two streams of data into a single, unified dataset for comprehensive analysis. During the data fusion process, careful consideration was given to labeling participant IDs and commuting instances. This step ensured the accurate association of physiological and environmental data with specific individuals and commuting events, laying the groundwork for subsequent analyses and the development of personalized comfort models. The ultimate dataframe, amalgamating physiological features and environmental data, includes columns for "Subject IDs" and "Commuting" labels. The final shape of this dataframe is (331,825 rows, 46 columns). We have appropriately labeled the combined dataframe based on "Subject IDs" and instances of "Commuting".

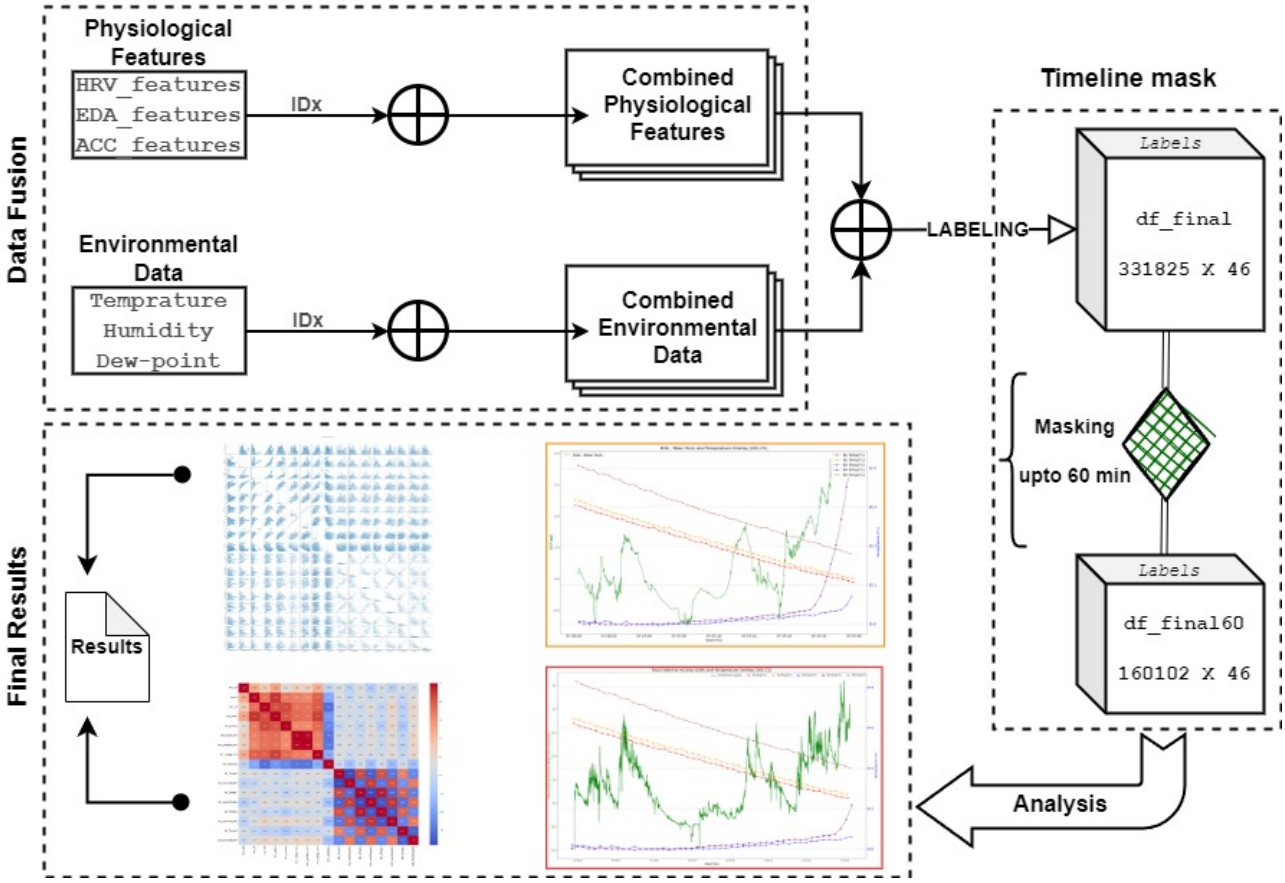

Figure 3: *Workflow of the research*

2.5 Statistical analysis

In our analysis, we focus on commuting instances lasting less than an hour. This criterion is applied due to the presence of unexpected commuting activities observed in the recorded data. By restricting our analysis to instances of commuting within this timeframe, we aim to ensure the inclusion of genuine experimental data, minimizing the likelihood of errors introduced by either human factors or devices. Consequently, the final size of the dataframe, encompassing these filtered commuting instances, is (160,102 rows, 46 columns). In this study, multiple statistical tests, including Spearman's rank correlation analysis, were conducted to achieve various objectives. Specifically, the investigation focused on establishing relationships between thermal sensation and heart rate variation, temperature variation, and skin conductance level. Additionally, the study aimed to identify individual factors influencing pulse rate normalization and explore the potential correlation between physiological features and environmental signals. To ensure the robustness of the statistical analyses, the normality of the data was assessed using the Shapiro-Wilk test. The chosen significance level for all analyses was set at 95% (p ¡ 0.05), adhering to standard practices in scientific research. This rigorous approach to statistical analysis enhances the reliability and validity of the findings in our investigation.

In this paper, all computations, encompassing both the aforementioned statistical tests and machine learning algorithms, were executed within the miniconda computing software environment. The analyses were carried out using JupyterLab (version 4.0.9) in conjunction with the Python programming language (version 3.11.7). Data manipulation and processing were facilitated by the employment of Pandas (version 2.1.4)[18], while the implementation of machine learning algorithms relied on the scikit-learn library (version 1.2.2). Furthermore, the visual representation of results was achieved using the Matplotlib library (version 3.8.2) [19] and Seaborn

(version 0.13)[20]. This utilization of robust computational tools and libraries ensures the accuracy, efficiency, and reproducibility of the analyses conducted throughout the study.

3 Results and Discussion

With the processed and integrated dataset in hand, we conducted a comprehensive correlation analysis to unveil patterns and relationships between physiological responses and environmental conditions. The primary goal was to identify significant associations that could be used to develop personalized comfort models.

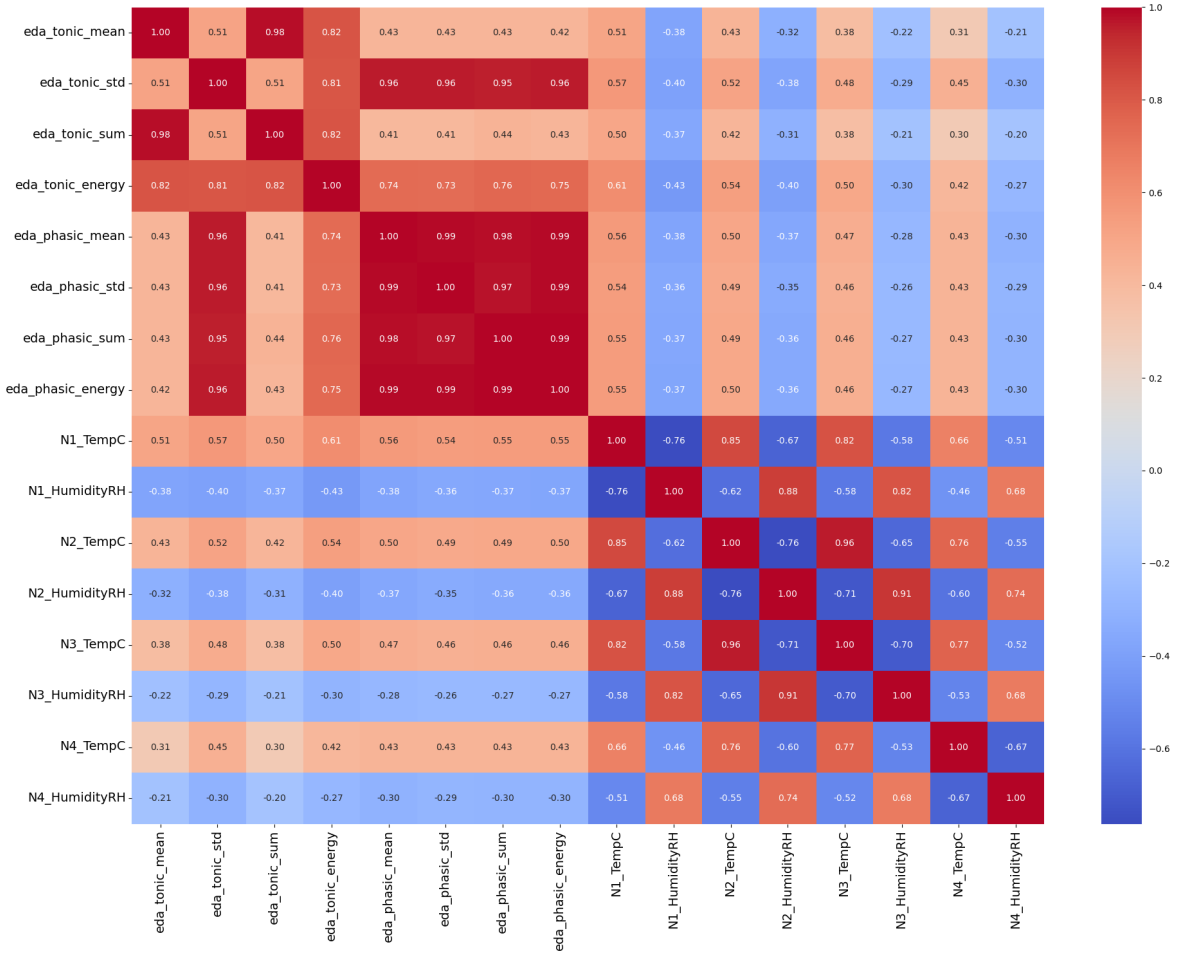

Figure 4: *Electrodermal activity features and environmental signals correlation heatmap*

Spearman Correlation Analysis

We delved into the intricate relationship between physiological measures of electrodermal activity (EDA) and environmental factors using Spearman correlation analysis. The results unveiled compelling insights into the associations between tonic electrodermal activity (`EDA_tonic_mean`) and various climate parameters. Notably, a substantial positive correlation (Spearman's $\rho = 0.372$, $p < 0.001$) was identified between `EDA_tonic_mean` and ambient temperature (`N1_TempC`), emphasizing the influence of temperature on tonic EDA levels. Conversely, a negative correlation was observed with humidity (`N1_HumidityRH`) (Spearman's $\rho = -0.265$, $p < 0.001$), signifying a potential inverse relationship between humidity and `EDA_tonic_mean`. These findings were consistently replicated with the inclusion of additional temperature and humidity variables (`N2_TempC, N2_HumidityRH, N3_TempC`), further substantiating the nuanced interplay between environmental conditions and tonic electrodermal activity. The significance of these correlations, coupled with the robust p-values ($p < 0.001$), underscores the relevance of climatic factors in shaping autonomic nervous system responses as reflected in electrodermal activity. This nuanced understanding has implications for research in psychophysiology and could inform the development of interventions tailored to specific environmental contexts. However, it is essential to acknowledge the multifaceted nature of these relationships, and further investigations, including potential moderating factors, are warranted to enrich our comprehension of the intricate dynamics between environmental conditions and electrodermal activity.

Regression Analysis

In our comprehensive regression analysis, we employed Ordinary Least Squares (OLS) regression to examine the relationship between the dependent variable, `eda_tonic_mean`, and the independent variable `N1_TempC`. The model demonstrated statistical significance, with an overall R^2 of 0.162, indicating that approximately 16.2% of the variance in `eda_tonic_mean` can be explained by the predictor variable. The F-statistic (3.090×10^4) was highly significant (Prob (F-statistic) < 0.001), supporting the validity of the model. The coefficient for `N1_TempC` was estimated at 0.5608 ($p < 0.001$), suggesting a positive relationship between the mean electrodermal activity and temperature. The 95% confidence interval for the coefficient (0.555 to 0.567) further supported the precision of this estimate. Notably, the intercept was -12.3328 ($p < 0.001$), representing the estimated value of `eda_tonic_mean` when `N1_TempC` is zero. The diagnostic tests, including the Omnibus, Durbin-Watson, and Jarque-Bera statistics, were conducted to assess the model's assumptions. While the results provide valuable insights, further exploration of residuals and potential sensitivity analyses will be essential for a comprehensive understanding of the findings.

Further, we investigated the relationship between the mean electrodermal activity (`eda_tonic_mean`) and various temperature variables (`N1_TempC, N2_TempC, N3_TempC, N4_TempC, N5_TempC`). Initially, we observed individual regressions for `N1_TempC` and `N2_TempC`. For `N1_TempC`, the model yielded an R^2 of 0.135, with a coefficient of 0.5296 ($p < 0.001$), indicating a positive association. The second model for `N2_TempC` exhibited an R^2 of 0.162, with a coefficient of 0.5608 ($p < 0.001$), also suggesting a positive relationship. Subsequently, a more complex model including `N1_TempC, N2_TempC, N3_TempC, N4_TempC, and N5_TempC` was considered. This extended model resulted in an R^2 of 0.310, indicating that the inclusion of additional temperature variables substantially improved the explanatory power of the model. Individual coefficients for each temperature variable were estimated, and their significance varied, with `N2_TempC` and `N5_TempC` exhibiting strong positive associations with `eda_tonic_mean`, while `N3_TempC` showed a negative association. Additionally, similar analyses were performed for the dependent variable `eda_phasic_mean`, where temperature variables exhibited varying degrees of influence. These findings underscore the nuanced relationship between temperature and electrodermal activity, highlighting the importance of considering multiple factors in understanding physiological responses. Further examination of residuals and diagnostic tests will be conducted to ensure the robustness of these results.

4 Conclusion

References

[1] E. Jovanov,
"Preliminary analysis of the use of smartwatches for longitudinal health monitoring,"
In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*,
IEEE, Aug. 2015.
DOI: 10.1109/embc.2015.7318499.

[2] F. Larradet, R. Niewiadomski, G. Barresi, D. G. Caldwell, and L. S. Mattos, "Toward emotion recognition from physiological signals in the wild: Approaching the methodological issues in real-life data collection,"
Frontiers in Psychology, vol. 11, Jul. 2020,
ISSN: 1664-1078.
DOI: 10.3389/fpsyg.2020.01111.

[3] L. Smital, C. R. Haider, M. Vitek, *et al.*, "Real-time quality assessment of long-term ecg signals recorded by wearables in free-living conditions,"
IEEE Transactions on Biomedical Engineering, vol. 67, no. 10, pp. 2721–2734, Oct. 2020,
ISSN: 1558-2531.
DOI: 10.1109/tbme.2020.2969719.

[4] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review,"
Medical and Biological Engineering Computing, vol. 44, no. 12, pp. 1031–1051, Nov. 2006,
ISSN: 1741-0444.
DOI: 10.1007/s11517-006-0119-0.

[5] M. Malik, J. T. Bigger, A. J. Camm, *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use,"
European Heart Journal, vol. 17, no. 3, pp. 354–381, Mar. 1996,
ISSN: 1522-9645.
DOI: 10.1093/oxfordjournals.eurheartj.a014868.

[6] J. E. Mietus, "The pnnx files: Re-examining a widely used heart rate variability measure,"
Heart, vol. 88, no. 4, pp. 378–380, Oct. 2002,
ISSN: 0007-0769.
DOI: 10.1136/heart.88.4.378.

[7] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms,"
Frontiers in Public Health, vol. 5, Sep. 2017,
ISSN: 2296-2565.
DOI: 10.3389/fpubh.2017.00258.

[8] J. P. Spiers, B. Silke, U. McDermott, R. G. Shanks, and D. W. G. Harron, "Time and frequency domain assessment of heart rate variability: A theoretical and clinical appreciation,"
Clinical Autonomic Research, vol. 3, no. 2, pp. 145–158, Apr. 1993,
ISSN: 1619-1560.
DOI: 10.1007/bf01819000.

[9] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, and L. Tarassenko, "Signal quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring,"
IEEE Journal of Biomedical and Health Informatics, pp. 1–1, 2014,
ISSN: 2168-2208.
DOI: 10.1109/jbhi.2014.2338351.

[10] S. Gashi, E. Di Lascio, B. Stancu, *et al.*, "Detection of artifacts in ambulatory electrodermal activity data,"
Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 2, pp. 1–31, Jun. 2020,
ISSN: 2474-9567.
DOI: 10.1145/3397316.

[11] I. R. Kleckner, R. M. Jones, O. Wilder-Smith, *et al.*, "Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data,"
IEEE Transactions on Biomedical Engineering, vol. 65, no. 7, pp. 1460–1467, Jul. 2018,
ISSN: 1558-2531.
DOI: 10.1109/tbme.2017.2758643.

[12] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review,"
Sensors, vol. 20, no. 2, p. 479, Jan. 2020,
ISSN: 1424-8220.
DOI: 10.3390/s20020479.

[13] S. Taylor, N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard,
"Automatic identification of artifacts in electrodermal activity data,"
In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*,
IEEE, Aug. 2015.
DOI: 10.1109/embc.2015.7318762.

[14] Y. Zhang, M. Haghdan, and K. S. Xu, "Unsupervised motion artifact detection in wrist-measured electrodermal activity data,"
Jul. 2017.
DOI: 10.48550/ARXIV.1707.08287. arXiv: 1707.08287 [cs.HC].

[15] J. Fridolfsson, M. Börjesson, C. Buck, *et al.*, "Effects of frequency filtering on intensity and noise in accelerometer-based physical activity measurements,"
Sensors, vol. 19, no. 9, p. 2186, May 2019,
ISSN: 1424-8220.
DOI: 10.3390/s19092186.

[16] M. Huang, G. Zhao, L. Wang, and F. Yang,
"A pervasive simplified method for human movement pattern assessing,"
In *2010 IEEE 16th International Conference on Parallel and Distributed Systems*,
IEEE, Dec. 2010.
DOI: 10.1109/icpads.2010.65.

[17] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven,
"Introducing wesad, a multimodal dataset for wearable stress and affect detection,"
In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*,
Ser. ICMI '18,
ACM, Oct. 2018.
DOI: 10.1145/3242969.3242985.

[18] W. McKinney,
"Data structures for statistical computing in python,"
In *Proceedings of the 9th Python in Science Conference*,
Ser. SciPy,
SciPy, 2010.
DOI: 10.25080/majora-92bf1922-00a.

[19] J. D. Hunter, "Matplotlib: A 2d graphics environment,"
Computing in Science and Engineering, vol. 9, no. 3, pp. 90–95, 2007,
ISSN: 1521-9615.
DOI: 10.1109/mcse.2007.55.

[20] M. Waskom, O. Botvinnik, D. O'Kane, *et al.*,
Mwaskom/seaborn: V0.8.1 (september 2017),
2017.
DOI: 10.5281/ZENODO.883859.

Nomenclature

- **Acc Acc X Max:** Acceleration X-axis Maximum
- **Acc Acc X Min:** Acceleration X-axis Minimum
- **Acc Acc X PTP:** Acceleration X-axis Peak-to-Peak
- **Acc Acc Y Max:** Acceleration Y-axis Maximum
- **Acc Acc Y Min:** Acceleration Y-axis Minimum
- **Acc Acc Y PTP:** Acceleration Y-axis Peak-to-Peak
- **Acc Acc Z Max:** Acceleration Z-axis Maximum
- **Acc Acc Z Min:** Acceleration Z-axis Minimum
- **Acc Acc Z PTP:** Acceleration Z-axis Peak-to-Peak
- **Acc L2 Max:** Acceleration Magnitude (L2 norm) Maximum
- **Acc L2 Min:** Acceleration Magnitude (L2 norm) Minimum
- **Acc L2 PTP:** Acceleration Magnitude (L2 norm) Peak-to-Peak
- **EDA Phasic Energy:** Electrodermal Activity Phasic Energy
- **EDA Phasic Mean:** Electrodermal Activity Phasic Mean
- **EDA Phasic Std:** Electrodermal Activity Phasic Standard Deviation
- **EDA Phasic Sum:** Electrodermal Activity Phasic Sum
- **EDA Tonic Energy:** Electrodermal Activity Tonic Energy
- **EDA Tonic Mean:** Electrodermal Activity Tonic Mean
- **EDA Tonic Std:** Electrodermal Activity Tonic Standard Deviation
- **EDA Tonic Sum:** Electrodermal Activity Tonic Sum
- **HRV Entropy:** Entropy
- **HRV HF:** Heart Rate Variability High Frequency
- **HRV LF:** Heart Rate Variability Low Frequency
- **HRV Mean NNI:** Mean Normal-to-Normal Interval
- **HRV Median NNI:** Median Normal-to-Normal Interval
- **HRV Range NNI:** Range of Normal-to-Normal Intervals
- **HRV RMSSD:** Root Mean Square of Successive Differences
- **HRV SDNN:** Standard Deviation of NN intervals
- **HRV VLF:** Heart Rate Variability Very Low Frequency
- IoT: Internet of Things..
- HVAC: Heating, Ventilation, and Air Conditioning.
- EDA: Electrodermal Activity.
- BVP: Blood Volume Pulse.
- ACC: Accelerometer.
- IBI: Inter Beat Interval.
- HR: Heart Rate.
- N1, N2, N3, N4, N5: Environmental Sensors.
- Participants IDs: S01, S02, S03, S04, S05, S06, S07, S08.

Appendix

Participant's commuting instances

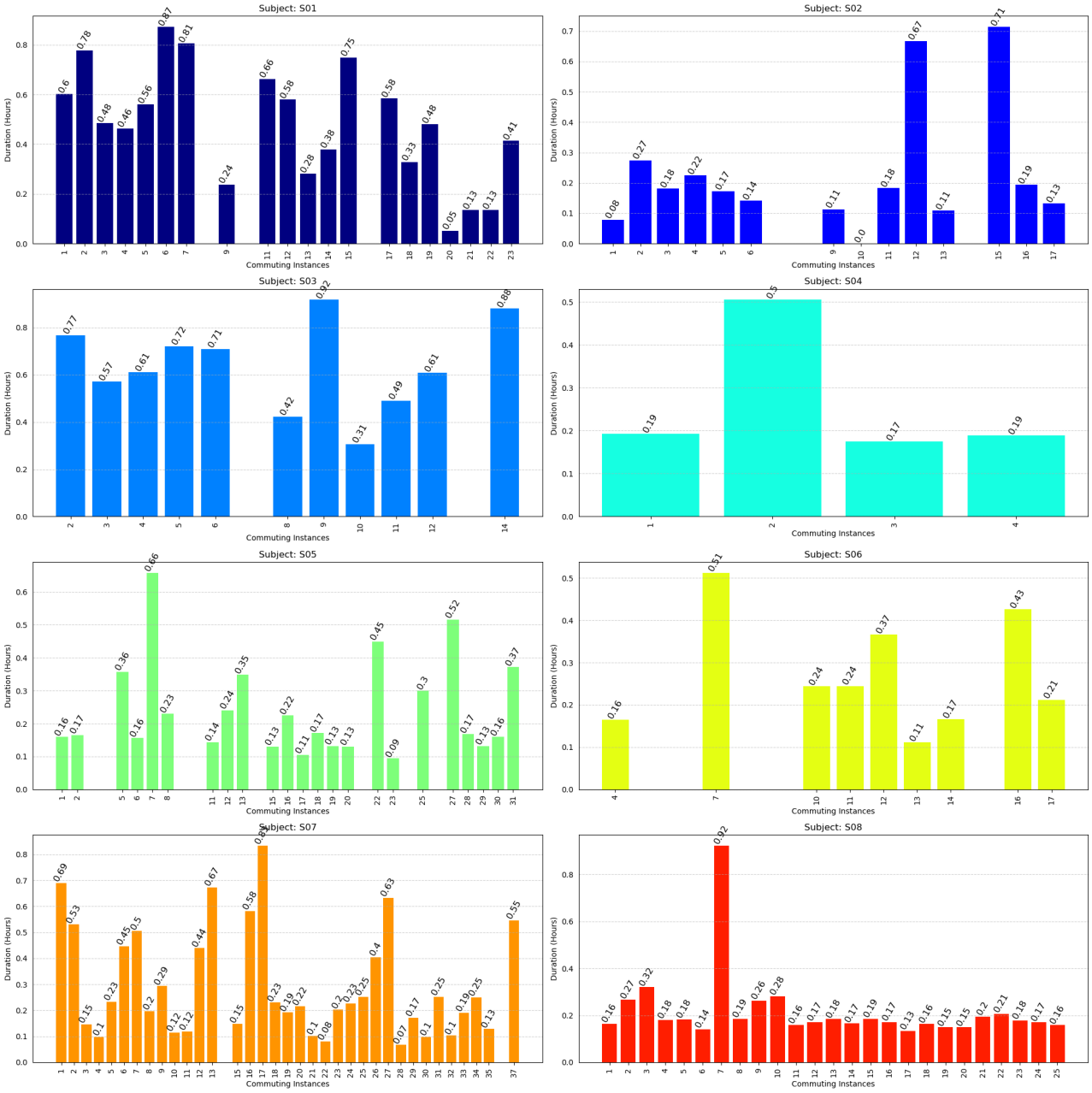

Figure 5: *Samples of each participants and duration of each sample instance*

Empatica Technical Specification

PPG Sensor

- Sampling frequency 64 Hz (Non customizabl
- LEDs: Green (2 LEDs), Red (2 LEDs) Photodiodes: 2units, total 15.5 mm2 sensitive area.
- Sensor output: Blood Volume Pulse (BVP) (variationof volume of arterial blood under the skin resultingfrom the heart cycle)
- Sensor output resolution 0.9 nW / Digit.
- Motion artifact removal algorithm:
- Combines different light wavelengths.
- Tolerates external lighting conditions

Infrared thermopile

- Sampling frequency: 4 Hz (Non customizable)
- Range:
- -40...85°C for ambient temperature.
- -40...115°C for skin temperature.
- Resolution: 0.02°C.
- Accuracy ±0.2°C within 36-39°C

EDA Sensor

- Sampling frequency: 4 Hz (Non customizable).
- Resolution: 1 digit 900 pSiemens.
- Range: 0.01 μSiemens – 100 μSiemens.
- Alternating current (8Hz frequency) with amax peak to peak value of 100 μ-Amps (at 100μ-Siemens).
- Electrodes: • Placement on the ventral (inner) wrist.
- Snap-on silver (Ag) plated with metallic core.
- Electrode longevity: • 4–6 months

3-Axis accelerometer

- Sampling frequency: 32 Hz (Non customizable).
- High sensitivity motion detection across 3 axes:X,Y, and Z.
- Default range ±2g.
- Ranges of ±4g or ±8g are selectable with custom.
- Resolution: 8 bits of the selected range.

Figure 6: *E4 wristband, overview of sensors (adapted from Empatica.com)*

List of Physiological Features

HRV Features

num_ibis: The number of interbeat intervals (IBIs) calculated within a specified window or dataset. It represents the count of heartbeats analyzed. **hrv_mean_nni:** Mean of the NN intervals (IBIs), reflecting average heart rate variability. **hrv_median_nni:** Median of the NN intervals, less sensitive to outliers compared to the mean. **hrv_range_nni:** Range of NN intervals, showing the difference between the maximum and minimum values. **hrv_sdsd:** Standard deviation of successive differences between NN intervals, also known as SDSD. **hrv_rmssd:** Root Mean Square of Successive Differences, emphasizing short-term variations. **hrv_nni_50:** Percentage of differences between consecutive NN intervals greater than 50 ms, indicative of parasympathetic activity. **hrv_pnni_50:** Proportion of NN intervals differing by more than 50 ms. **hrv_nni_20:** Percentage of differences between consecutive NN intervals greater than 20 ms. **hrv_pnni_20:** Proportion of NN intervals differing by more than 20 ms. **hrv_cvsd:** Coefficient of Variation of Successive Differences. **hrv_sdnn:** Standard deviation of NN intervals, reflecting overall HRV. **hrv_cvnni:** Coefficient of Variation of NN intervals. **hrv_mean_hr:** Mean heart rate. **hrv_min_hr:** Minimum heart rate. **hrv_max_hr:** Maximum heart rate. **hrv_std_hr:** Standard deviation of heart rate. **hrv_total_power:** Total power of the HRV spectrum. **hrv_vlf, hrv_lf, hrv_hf:** Power in Very Low Frequency, Low Frequency, and High-Frequency bands, respectively, representing different autonomic nervous system activities. **hrv_lf_hf_ratio:** Ratio of LF to HF power, indicating sympathovagal balance. **hrv_lfnu, hrv_hfnu:** Normalized LF and HF power, respectively. **hrv_lf_hf_ratio:** It's the ratio of low-frequency (LF) to high-frequency (HF) power in the HRV spectrum. This ratio indicates the balance between sympathetic (LF) and parasympathetic (HF) nervous system activities. Higher values often suggest sympathetic dominance, while lower values suggest parasympathetic dominance. **hrv_lfnu, hrv_hfnu:** These are normalized LF (LFnu) and HF (HFnu) powers, respectively. They represent the relative contribution of LF and HF components to the total power after normalizing their values. **hrv_mean:** Mean value of the HRV signal. **hrv_std:** Standard deviation of the HRV signal, showing the dispersion of values around the mean. **hrv_min:** Minimum value observed in the HRV signal. **hrv_max:** Maximum value observed in the HRV signal. **hrv_ptp:** Peak-to-peak value, the difference between the maximum and minimum values in the HRV signal. **hrv_sum:** Sum of all values in the HRV signal. **hrv_energy:** Energy of the HRV signal, calculated as the sum of squares of the signal values. **hrv_skewness:** Measure of asymmetry in the HRV signal distribution. Positive values suggest a longer tail on the right side of the distribution. **hrv_kurtosis:** Measure of the "peakedness" of the HRV signal distribution. Higher values indicate more data in the tails and a sharper peak. **hrv_peaks:** Count of local maxima or peaks in the HRV signal. **hrv_rms:** Root Mean Square of the HRV signal. **hrv_lineintegral:**

Integral of the HRV signal, often calculated over a specific period. **hrv_n_above_mean, hrv_n_below_mean:** Number of data points above and below the mean HRV value. **hrv_n_sign_changes:** Number of times the HRV signal changes sign from positive to negative or vice versa. **hrv_iqr:** Interquartile Range of the HRV signal, measuring the spread of the middle 50**hrv_iqr_5_95:** Interquartile Range between the 5th and 95th percentiles of the HRV signal. **hrv_pct_5, hrv_pct_95:** Values at the 5th and 95th percentiles, respectively, of the HRV signal. **hrv_entropy:** Entropy of the HRV signal, a measure of randomness or disorder in the signal. **hrv_perm_entropy:** Permutation entropy, measuring complexity in the HRV signal using ordinal patterns. **hrv_svd_entropy:** Singular Value Decomposition (SVD) entropy, another measure of signal complexity using SVD.

EDA Features –Tonic Component Metrics

eda_tonic_mean: Mean value of the tonic component of EDA, representing the baseline skin conductance level. **eda_tonic_std:** Standard deviation of the tonic component, indicating the variability around the mean tonic level. **eda_tonic_min:** Minimum value observed in the tonic component. **eda_tonic_max:** Maximum value observed in the tonic component. **eda_tonic_ptp:** Peak-to-peak amplitude of the tonic component, the difference between the maximum and minimum values. **eda_tonic_sum:** Sum of all values in the tonic component. **eda_tonic_energy:** Energy of the tonic component, computed as the sum of squares of the signal values. **eda_tonic_skewness:** Skewness of the tonic component's distribution. Positive values indicate a longer tail on the right side. **eda_tonic_kurtosis:** Kurtosis of the tonic component's distribution. Higher values suggest a sharper peak and heavier tails. **eda_tonic_peaks:** Count of local maxima or peaks in the tonic component. **eda_tonic_rms:** Root Mean Square of the tonic component. **eda_tonic_lineintegral:** Integral of the tonic component, often calculated over a specific period. **eda_tonic_n_above_mean, eda_tonic_n_below_mean:** Number of data points above and below the mean tonic value. eda_tonic_n_sign_changes: Number of times the tonic component changes sign from positive to negative or vice versa. **eda_tonic_iqr: Interquarti**le Range of the tonic component, measuring the spread of the middle 50**eda_tonic_iqr_5_95:** Interquartile Range between the 5th and 95th percentiles of the tonic component. **eda_tonic_pct_5, eda_tonic_pct_95:** Values at the 5th and 95th percentiles, respectively, of the tonic component. **eda_tonic_entropy:** Entropy of the tonic component, a measure of its randomness or disorder. **eda_tonic_perm_entropy:** Permutation entropy, measuring complexity in the tonic component using ordinal patterns. **eda_tonic_svd_entropy:** Singular Value Decomposition (SVD) entropy, another measure of signal complexity using SVD.

EDA Features –Phasic Component Metrics

eda_phasic_mean: Mean value of the phasic component of EDA, representing the fast-changing fluctuations beyond the tonic level. **eda_phasic_std:** Standard deviation of the phasic component, indicating variability around the mean phasic level. **eda_phasic_min:** Minimum value observed in the phasic component. **eda_phasic_max:** Maximum value observed in the phasic component. **eda_phasic_ptp:** Peak-to-peak amplitude of the phasic component, the difference between the maximum and minimum values. **eda_phasic_sum:** Sum of all values in the phasic component. **eda_phasic_energy:** Energy of the phasic component, computed as the sum of squares of the signal values. **eda_phasic_skewness:** Skewness of the phasic component's distribution. Positive values indicate a longer tail on the right side. **eda_phasic_kurtosis:** Kurtosis of the phasic component's distribution. Higher values suggest a sharper peak and heavier tails. **eda_phasic_peaks:** Count of local maxima or peaks in the phasic component. **eda_phasic_rms:** Root Mean Square of the phasic component. **eda_phasic_lineintegral:** Integral of the phasic component, often calculated over a specific period. **eda_phasic_n_above_mean, eda_phasic_n_below_mean:** Number of data points above and below the mean phasic value. **eda_phasic_n_sign_changes:** Number of times the phasic component changes sign from positive to negative or vice versa. **eda_phasic_iqr:** Interquartile Range of the phasic component, measuring the spread of the middle 50**eda_phasic_iqr_5_95:** Interquartile Range between the 5th and 95th percentiles of the phasic component. **eda_phasic_pct_5, eda_phasic_pct_95:** Values at the 5th and 95th percentiles, respectively, of the phasic component. **eda_phasic_entropy:** Entropy of the phasic component, a measure of its randomness or disorder. **eda_phasic_perm_entropy:** Permutation entropy, measuring complexity in the phasic component using ordinal patterns. **eda_phasic_svd_entropy:** Singular Value Decomposition (SVD) entropy, another measure of signal complexity using SVD.

Accelerometer Features

Mean: The average value of the acceleration in the respective axis. **Standard_Deviation (Std):** Measures the dispersion or spread of values in the data. **Minimum:** The smallest observed value in the dataset. **Maximum:** The largest observed value in the dataset. **Peak-to-Peak (PtP):** The difference between the maximum and minimum values. **Sum:** The sum of all values in the dataset. **Energy:** The sum of the squares of the values. **Skewness:** Measures the asymmetry of the distribution of values. **Kurtosis:** Measures the "tailedness" or sharpness of the distribution. **Peaks:** The count of local maxima or minima. **Root_Mean_Square (RMS):** The square root of the mean of the squares of the values. **Line_Integral:** Integral of the absolute values of the acceleration. **Count_of_Points_Above_Mean:** Number of data points above the mean value. **Count_of_Points_Below_Mean:** Number of data points below the mean value. **Number_of_Sign_Changes:** Count of times the signal changes direction. **Interquartile_Range (IQR):** Range between the first and third quartiles. **Percentiles_(5th_and_95th):** Values below which a certain percentage of data fall. **Entropy:** The measure of disorder or uncertainty in the data. **Permutation_Entropy:** A complexity measure derived from time series data. **Singular_Value_Decomposition (SVD) Entropy:** Another complexity measure based on SVD.

Heatmaps

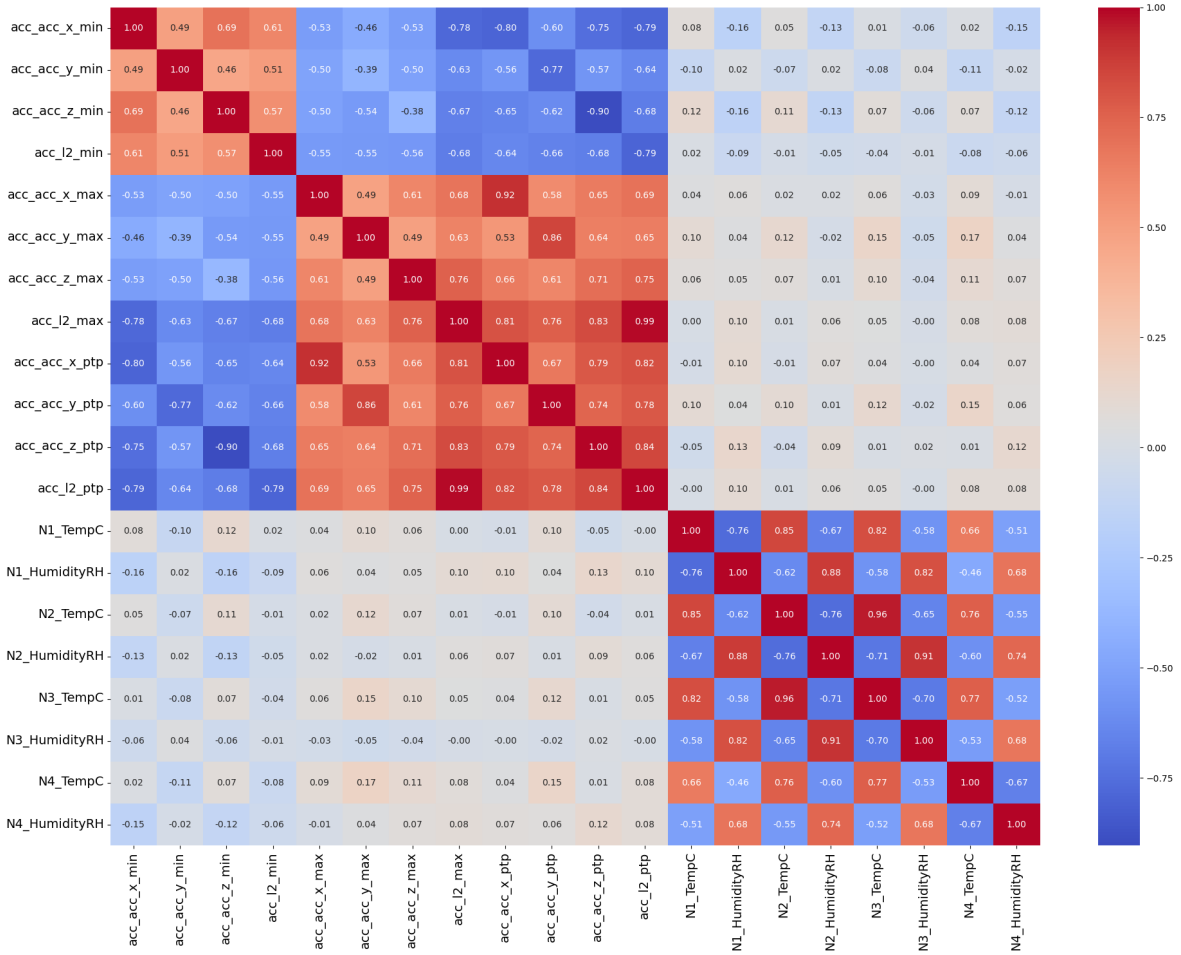

Figure 7: *Accelerometer features and environmental signals correlation heatmap*

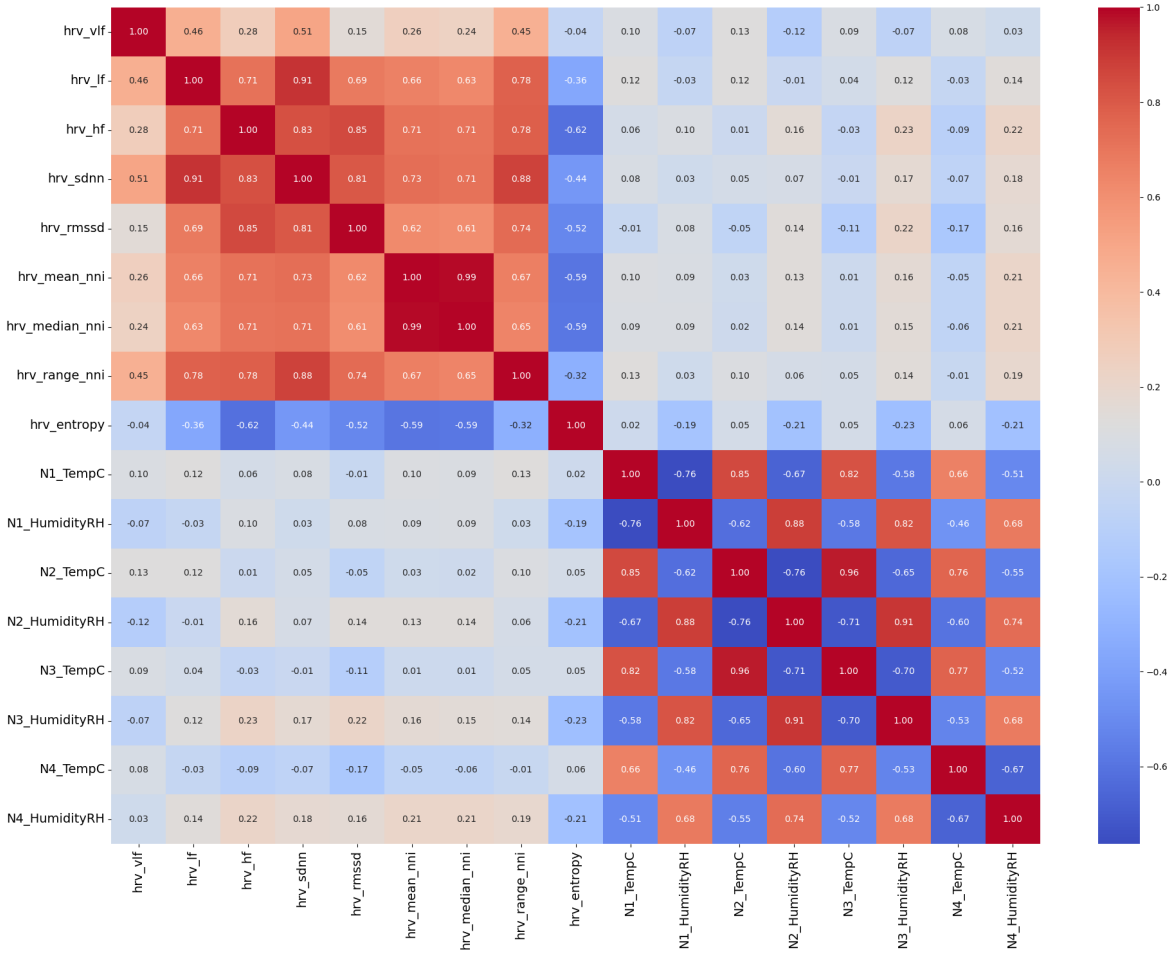

Figure 8: *Heart rate variation features and environmental signals correlation heatmap*