EasyChair Preprint

№ 13098

A Feasibility Study on Improving Emotion Recognition from ECG Signals and HRV Features Through Baseline Clusterization

Elena Sajno, Alessio Rossi, Luca Pappalardo, Federico Dagostin and Giuseppe Riva

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 25, 2024

A Feasibility Study on Improving Emotion Recognition from ECG Signals and HRV Features through Baseline Clusterization*

Elena Sajno[1,2,3][0000−0002−9621−8981], Alessio Rossi[4][0000−0002−6400−5914], Luca Pappalardo[4][0000−0002−1547−6007], Federico Dagostin[5][0009−0006−9406−6343], and Giuseppe Riva[2,6][0000−0003−3657−106X]

[1] Department of Computer Science, University of Pisa, Pisa, Italy
elena.sajno@phd.unipi.it
[2] HTLAB, Università Cattolica del Sacro Cuore, Milan, Italy
[3] Responsive Environment, MIT MediaLab, Massachusetts Institute of Technology, Cambridge, USA
[4] Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR), Pisa, Italy
[5] BMI Labs "Mario Stefanelli", Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy
[6] Applied Technology for Neuro-Psychology Laboratory, IRCCS Istituto Auxologico Italiano, Milan, Italy

Abstract. Electrocardiography (ECG) has the potential for bringing Affective Computing outside laboratories, thanks to the spread of wearable and inexpensive instrumentation. Nevertheless, intra individual variability could influence Machine Learning (ML) models' accuracy. To assess this issue, we propose to group the participants according to their general cardiovascular status, through the clusterization of HRV baseline features. A specific ML model aimed at classifying emotional responses was developed for each baseline cluster. This processing will lead to cardiac-state specific classification models to mitigate ML performance issues. We experimented this data analytics framework on the Mahnob HCI database containing ECG paired with emotional self-report assessment. Baseline data was clustered using k-means, dividing the dataset into two parts. Successively, classification models were separately applied to each group to predict arousal, valence, and dominance levels from ECG features. Classifiers applied after clustering outperformed those without clustering, reaching higher scores and lower randomness. Clustering ECG baselines to create individualized classifiers may alleviate intra-individual variability and improve emotion recognition performance, making affective computing more applicable.

Keywords: Affective Computing · ECG · HRV · Machine Learning · Emotion Recognition · ECG Features · Database.

* Partially funded by the Italian Ministry of Health.

1 Introduction

Affective computing [1], which involves the computational recognition and interpretation of human emotions, has the potential to revolutionize human-computer dynamics, allowing to add the emotional layer. However, the results of emotion recognition studies using Machine Learning (ML) models are often complex, and the most accurate results are often achieved through multimodal designs (that is, integrating different signals) or with the use of complex systems such as EEG, eye tracking, and video recordings [2]. These tools are invasive, expensive, require specific settings, and could raise privacy issues, leading to difficulties in their use outside of a laboratory [3]. Electrocardiogram (ECG) and Electrodermal Activity (EDA) sensors offer more promising solutions due to the availability of non-invasive wearable devices [4] pairing user-friendliness with less expensive instruments, and efficient data management. However, work is still to be done, as performance obtained with peripheral signals is generally far from perfect[5].

One major issue in using ECG in Affective Computing can be the intra-individual variations present in a subject's heart activity, especially when computed outside the classical hospital environments [8]. This range in heart rate patterns could influence the training of ML models, leading to a less performant result.

Affective computing models typically involve a standardized workflow [6, 7], beginning with physiological data acquisition where subjects are connected to devices to record multiple physiological activations, followed by collecting baseline "ground truth" data during resting states to mitigate interpersonal differences. Subjects then engage in tasks or receive stimuli to elicit various emotional responses for data labeling. Subsequent steps involve cleaning data, calculating sub-features, and populating a labeled database. Multiple supervised classification algorithms are then trained and validation is typically performed through leave-one-out cross-validation. A test phase evaluates the performance levels and the best performer algorithm is selected.

To address the inter and intra-variability issue, we propose to add a preliminary step to the classic Affective Computing ML classification training: we suggest clustering of baseline data for similar characteristics. Clustering is a technique that automatically groups similar data points into distinct clusters based on their inherent patterns or characteristics: in this case, we want to target how people differ before the experimental conditions. Each task immediately follows a baseline recording, in which the subject remains in a neutral/resting state. The baseline values are specifically conditioned both on the subject's cardiac profile, and on any residual activation, and offer the key for evaluating both inter-personal and intra-personal differences. The proposed additional step clusters the baseline recording, grouping the ECG data connected to the emotional activation, before applying algorithms to these different groups. The starting dataset will be divided according to clusterization results. To each group will be applied a standard classification model, using the physiological data collected during the emotional tasks as input. We hypothesize that the model perfomance will improve, in the clustered groups; in this feasibility study, we experiment with the application of this model to the Mahnob HCI open database, achieving an improvement in the performance of the final classifiers. This approach can potentially alleviate the burden of intra-individual variability, improve the accuracy of ML models for emotion recognition, and allow their use in a more accessible, efficient, and effective way outside a laboratory setting.

2 Methods

2.1 Manhob-Hci Database

This study was performed on the Mahnob HCI multimodal database [2], which is openly licensable for scientific purposes [9]. The database contains EEG, peripheral physiological signals, face and body videos, gaze, and audio of 27 subjects who were watching 20 emotionally connotated videos. Subjects were also requested to evaluate their emotional state after every video and rate their affect for arousal, valence, and dominance level, following Russell et al., and Fontaine et al. frameworks [10, 11].

Soleymani et al. also present results in the classification of emotion and valence, arousal, and dominance levels [2]; their results show that the use of aggregated peripheral physiology for predicting low, medium, and high levels of Arousal and Valence obtained a performance of 46% and 45.5% respectively.

In [12], the authors show an increase to 51% in performance for both Arousal (with Random Forest and K-Nearest Neighbors) and Valence (with Decision Tree), using HRV temporal and frequency features, calculated from the sole ECG signal. HRV is recognized as a reliable physiological data of choice to discriminate a person's mental and emotional state [13, 14]; better performance in Valence and Arousal detection with HRV features, rather than ECG, was obtained also on the RECOLA dataset [15].

2.2 Data analytics Framework

Following these premises, our first analysis will be focused on the HRV features extracted from the ECG signals, which will be used to train and test multiple classification algorithms. The prediction will be aimed at three different levels of valence, arousal, and dominance (low, medium, and high).

As the second step, a preliminary clusterization will be presented, as a way to overcome inter-individual differences in ECGs, which could make algorithms less generalizable. Baselines will be clustered to create different starting states; after that, the classification process will be repeated for each group. This will create a model in which every new subject, at a specific moment, will be associated with a specific classificator, trained with a heart pattern starting with a similar baseline (Fig. 1).

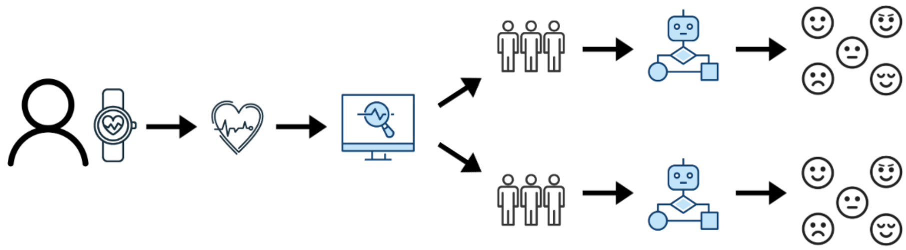

Fig. 1. ECG from each subject is acquired and processed. A first analysis of the ECG pattern will select to which cluster it belongs; after that, the recorded data will flow through the classifier trained on the data group it pertains to. The classifier will consequently select a more precise affective state.

2.3 HRV Data Preprocessing

We received access to the Mahnob HCI multimodal database [2] and downloaded from the web-portal [16] the ECG data collected for the experiment called "emotion elicitation", alongside the subject and the session ID, and if the ECG track was recorded during an emotional stimulus or not. Each session recorded during an emotional stimulus was paired with the precedent not stimulated recording (i.e. its baseline). We also collected the assessment that the participant provided on the levels of arousal, valence, and dominance (evaluated on a 9-point Likert Scale). These data were used to populate our dataset.

The ECG signal was processed using the Python libraries Numpy [17], Pandas [18], HeartPy [19], and hrvanalysis [20]; the signal was cleaned from outliers and ectopic beats, and the following missing values substituted through linear interpolation. HRV features in the domains of time and frequency were calculated (Time domain features: Mean NNI, SDNN, SDSD, NN50, pNN50, NN20, pNN20, RMSSD, Median NN, Range NN, CVSD, CV NNI, Mean HR, Max HR, Min HR, STD HR; Frequency domain features: LF, HF, VLF, LH/HF ratio, LFnu, HFnu, Total Power. See Table 1 for acronyms and definitions) and the baseline was subtracted.

2.4 ML Prediction Models Training and Validation

First, we apply a typical classification model to have benchmark performance values. Decision Tree (DT), Random Forest (RF), KNeighbors (KNn), Support Vector Machine (SVM) [24], and XGBoost (XGB) [25] algorithms were trained to predict arousal, valence, and dominance in a 3-step layout, as in Soleymani et

Table 1. Description of acronyms used in the article, regarding heart rate variability (HRV, the variation in the time intervals between adjacent heartbeats), divided for time and frequency domain features [21–23].

Time domain	Definition
RR intervals	Interbeat intervals between successive R points
NN intervals	Interbeat intervals after artifacts removal
Mean NNI	Mean NN Intervals
SDNN	NN intervals standard deviation
SDSD	Standard deviation of differences between adjacent NN intervals
NN50	Number of NN interval differences more than 50ms
pNN50	Percentage of successive RR intervals that differ by more than 50 ms
NN20	Number of NN interval differences more than 20ms
pNN20	Percentage of successive RR intervals that differ by more than 20 ms
RMSSD	Root mean square of successive RR interval differences
Median NN	Median of NN values
Range NN	Range of NN values
CVSD	RMSSD divided MeanNN
CV NNI	SDNN divided by MeanNN
HR	Heart Rate
Mean HR	Mean of HR values
Max HR	Maximum HR value
Min HR	Minimum HR value
STD HR	Standard deviation
Frequency domain	**Definition**
LF	Low-frequency band (0.04–0.15 Hz) in absolute power
HF	High-frequency band (0.15–0.4 Hz) in absolute power
VLF	Very-low-frequency band (0.0033–0.04 Hz) in absolute power
LH/HF ratio	LF/HF power ratio
LFnu	Relative power of LF in normal units
HFnu	Relative power of HF in normal units
Total Power	Absolute power of the sum of the frequency bands

al., and Ferdinando, et al. (1-3 low, 4-6 medium, 7-9 high) [2, 26]. Each classification was evaluated through a Leave-One-Subject-Out Cross-Validation (LOSO CV), in which the trained algorithm is repeatedly tested on all the subject data -1, until all subjects are "left out". A further Features Selection through a Recursive Feature Elimination with Cross-Validation (RFECV) [24] was also applied to all the conditions, to select which features probably increase the functionality of the model. The "dummy value" was calculated for each classification through a Dummy Classifier, that acts as a benchmark to assess how well a machine learning model performs compared to random chance or simple rules, and helps establish a baseline for randomness in the evaluation of more complex models. [24]. Each model was evaluated by estimating precision (the ratio of correctly predicted positive observations to total predicted positives), recall (the ratio of correctly predicted positive observations to all actual positives), and F1-score (the harmonic mean of precision and recall, balancing their values). Similar models were also applied in [12] to check the validity of HRV features as input for the prediction of the affective state.

The second model was applied to the same starting dataset. HRV features of baseline data were selected, considering every baseline recorded before showing to the subjects the emotive connoted videos as a single entry. The subject's baseline values were clustered through a K-means by using both time and frequency domain characteristics as input features. The hyperparameters were configured as follows: the maximum number of interactions was set to 300, the clustering algorithm used was "Lloyd". To determine the appropriate number of clusters, we calculated the silhouette score, which measures the sensibility of clustering by considering both the cohesion within clusters and the separation between clusters. A higher silhouette score indicates better clustering quality. In our case, we selected 2 clusters based on both the silhouette score and the scarcity of data when divided into more clusters. The calculated silhouette score reached 0.79, indicating that approximately 79% of the data points were confidently assigned to a specific cluster, while the remaining points were considered potential outliers or subject to some degree of uncertainty.

To understand the contribution of each feature to the clusterization tasks, a DT (found as the best performer algorithm) was trained with baseline data to predict the labels Cluster 0 or Cluster 1. We calculated permutation feature importance scores using the eli5 package, which assesses the decline in model performance when each feature is randomly permuted, to obtain insight into the significance of each feature in cluster discrimination.

The starting database was consequently split, referring to the cluster belonging to the baseline data. Therefore the new databases underwent the same classification models used in the first analysis, i.e.: DT, RF, KNn, SVM, and XGB models were trained to predict arousal, valence, and dominance in the 3-step layout with and without an RFECV, evaluated through a LOSO CV, and compared to a Dummy Classifier.

3 Results

3.1 Clusterization Results

Baseline clustering grouped 76,34% of data in Cluster 0 and 23,66% in Cluster 1. Figure 2 shows the mean of feature importance scores in the clusterization between Cluster 0 and Cluster 1, excluding features with null relevance. We calculated the DT classification with LOSO CV and reported the mean value of the resulting features. Specifically, the major influence in the clustering division is brought by the total power, followed by LF, RMSSD, and LF/HF ratio.

3.2 Affective Classification

Table 2 reports the best-performing algorithms in the aggregate, divided by arousal, valence, and dominance. The results of the first classifications are shown alongside the results obtained on clustered data.

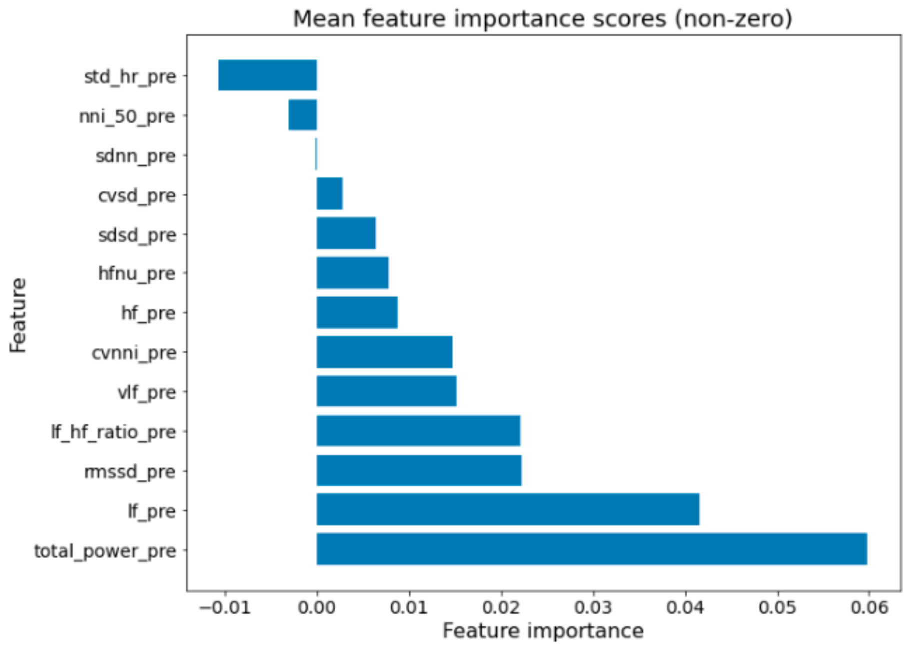

Fig. 2. The findings are charted by mean significance ratings across all leave-one-subject-out cross-validation splits. We eliminated characteristics with little relevance to the plot to increase its interpretability. Refer to table 1 for acronym definitions.

In the first round of classification, Random Forest without feature selection performs best for arousal levels, XGboost with RFECV features selection for valence, and Decision Tree with RFECV feature selection for dominance.

For both Cluster 0 and Cluster 1 data, the best-performing algorithm is the Decision Tree without feature selection for arousal. For valence, XGBoost with RFECV feature selection is the best performing in both the clusters, with Random Forest (with RFECV feature selection in Cluster 0 and without in Cluster 1). For dominance, Random Forest with RFECV feature selection performs best with Cluster 0, while Decision Tree with RFECV feature selection with Cluster 1. For valence and arousal, classifications after clusterization reach higher scores for both clusters, showing a wider distance from the dummy random values. For dominance, results without clusterization are comparable in Cluster 1 and higher in Cluster 0, while in Cluster 1 distance from randomness is more marked. For exhaustiveness, the weighted mean results of Clusters 0 and Cluster 1 are also shown in table 2.

Table 2. Performance Results: Classification results are separated by affect and clustering condition: no clusterization or pre-clusterized baseline data in Cluster 0 and Cluster 1. Each cell reports Precision, Recall, and F-1 scores of the best-performing algorithm and, in brackets, if the results were obtained with (+) or without (-) an RFECV Features Selection. The Dummy Classifier results are also reported with the same criteria. Finally, the weighted mean between Cluster 0 and Cluster 1 results is reported.

Affect	***No Cluster***	***Cluster 0***	***Cluster 1***	***Weighted Mean of Clusters 0 & 1***
Arousal	RF (-FS) Precision: 0.53 Recall: 0.44 F1-score: 0.42	DT (-FS) Precision: 0.57 Recall: 0.52 F1-score: 0.51	DT (-FS) Precision: 0.54 Recall: 0.51 F1-score: 0.51	Precision: 0.56 Recall: 0.52 F1-score: 0.51
	Dummy Precision: 0.43 Recall: 0.34 F1-score: 0.36	Dummy Precision: 0.26 Recall: 0.18 F1-score: 0.20	Dummy Precision: 0.44 Recall: 0.38 F1-score: 0.39	Dummy Precision: 0.30 Recall: 0.23 F1-score: 0.24
Valence	XGB (+FS) Precision: 0.48 Recall: 0.45 F1-score: 0.44	XGB (+FS) Precision: 0.50 Recall: 0.44 F1-score: 0.44 RF (+FS) Precision: 0.51 Recall: 0.41 F1-score: 0.42	RF (-FS) Precision: 0.58 Recall: 0.66 F1-score: 0.60 XGB (+FS) Precision: 0.60 Recall: 0.62 F1-score: 0.59	XGB Precision: 0.52 Recall: 0.48 F1-score: 0.475 RF Precision: 0.53 Recall: 0.47 F1-score: 0.46
	Dummy Precision: 0.50 Recall: 0.40 F1-score: 0.43	Dummy Precision: 0.47 Recall: 0.39 F1-score: 0.41	Dummy Precision: 0.17 Recall: 0.12 F1-score: 0.14	Dummy Precision: 0.40 Recall: 0.32 F1-score: 0.35
Dominance	DT (+FS) Precision: 0.51 Recall: 0.47 F1-score: 0.46	RF (+FS) Precision: 0.47 Recall: 0.36 F1-score: 0.38	DT (+FS) Precision: 0.51 Recall: 0.45 F1-score: 0.46	Precision: 0.48 Recall: 0.38 F1-score: 0.40
	Dummy Precision: 0.50 Recall: 0.40 F1-score: 0.43	Dummy Precision: 0.46 Recall: 0.37 F1-score: 0.39	Dummy Precision: 0.27 Recall: 0.22 F1-score: 0.24	Dummy Precision: 0.415 Recall: 0.33 F1-score: 0.35

4 Discussion and Conclusions

As stated also in [12], valence, arousal, and dominance levels can be characterized by differences in HRV values. The results we obtained without clusterization are comparable to [12] but a further Cross-Validation was added, to increase their generalizability. Specifically, HRV features alone outperform the aggregate use of peripheral physiological data, if compared with classifications trained on the same dataset [2].

Our final model proposes to cluster the sole baselines, in order to decide which classification algorithm needs to be applied to the subsequent activation phase. The results suggest that clustering subjects based on HRV computed from baseline ECG can improve affect recognition performance, especially regarding arousal and valence. Classifiers applied after clustering outperformed those without clustering, indicating that creating individualized models based on groups of subjects' heart rate patterns may alleviate intra-individual variability and noise.

As shown in Table 2, regarding precision, arousal scores exhibited an improvement from 53% to 56% and valence from 48% to 52%, while dominance scores decreased from 51% to 48%. In terms of recall, arousal scores moved from 44% to 52%, valence from 45% to 47%, and dominance scores from 47% to 38%. The overall accuracy, as measured by the F-1 score, showed improvements in arousal (from 42% to 51%) and in valence (from 44% to 46%), while decreases in dominance from 46% to 40%. Clustering the baseline data resulted in enhanced precision, recall, and accuracy for both arousal and valence dimensions. However, the dominance dimension performed better when the data was not clusterized. These results can be analyzed in light of inter and intraindividual differences in cardiovascular responses. Additionally, individuals may exhibit different sensitivities to arousal and valence compared to dominance: this could impact the performance of the model.

Positive and negative mean feature importance (see Figure 1) provide an indication of which characteristics determine the cluster of belonging, and may offer insights into the underlying cardiac functioning. Total power, followed by LF, RMSSD, and LF/HF ratio have proven the major (positive) influence on the clusterization.

However, the study can have some limitations. The Mahnob-Hci Database is relatively small, and the results require validation on larger datasets. At the same time, intrinsic characteristics of the specific database, such as characteristics of the participants and the techniques used in the data collection and database creation may limit the generalizability of the results.

Regarding adding the clustering step, the optimal number of clusters is unclear and may vary among different data collections: in this study, the choice of two clusters was led by data scarcity. Having more data available would allow one to test if a larger number of clusters could lead to a more precise division of physiological patterns. Additionally, including other physiological signals like EDA and respiration in the clustering approach could potentially further improve results and offer insight into how inter and intra-personal differences can influence physiological reactions to affective and emotive states.

In conclusion, clustering subjects based on HRV features calculated from ECG baselines appears to be a promising approach to deal with intra-individual variability in ML emotion recognition. With optimization and an enlargement of the data pool, this method could make Affective Computing models more applicable outside laboratories by improving accuracy for individual users, on data that can be easily collected in the wild. Future work will aim to validate and expand this preliminary finding.

References

1. Picard RW. Affective Computing. MIT Press; 2000.
2. Soleymani M, Lichtenauer J, Pun T, Pantic M. A Multimodal Database for Affect Recognition and Implicit Tagging. IEEE Transactions on Affective Computing. 2012;3:42–55.
3. Chen M, Zhang Y, Li Y, Hassan MM, Alamri A. AIWAC: affective interaction through wearable computing and cloud technology. IEEE Wirel Commun. 2015;22:20–7.
4. Hasnul MA, Aziz NAA, Alelyani S, Mohana M, Aziz AA. Electrocardiogram-Based Emotion Recognition Systems and Their Applications in Healthcare—A Review. Sensors . 2021;21:5015.
5. Wang Y, Song W, Tao W, Liotta A, Yang D, Li X, et al. A systematic review on affective computing: emotion models, databases, and recent advances. Inf Fusion. 2022;83-84:19–52
6. Zheng, W.-L., and Bao-Liang, L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Auton. Ment. Dev. 2015; 7:162–175. doi: 10.1109/TAMD.2015.2431497
7. Sajno E, Bartolotta S, Tuena C, Cipresso P, Pedroli E and Riva G (2023) Machine learning in biosignals processing for mental health: A narrative review. Front. Psychol. 13:1066317. doi: 10.3389/fpsyg.2022.1066317
8. Avram R, Tison GH, Aschbacher K, Kuhar P, Vittinghoff E, Butzner M, et al. Real-world heart rate norms in the Health eHeart study. NPJ Digit Med. 2019;2:58.
9. Koelstra S. HCI tagging database - home [Internet]. [cited 2023 Apr 8]. Available from: https://mahnob-db.eu/hci-tagging/
10. Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J Res Pers. 1977;11:273–94.
11. Fontaine JRJ, Scherer KR, Roesch EB, Ellsworth PC. The world of emotions is not two-dimensional. Psychol Sci. 2007;18:1050–7.
12. Sajno E, Rossi A, De Gaspari S, Sansoni M, Brizzi G, Riva G. XAI in Affective Computing: a Preliminary Study. ANNUAL REVIEW OF CYBERTHERAPY AND TELEMEDICINE 2023; 40-46
13. Rowe DW, Sibert J, Irwin D. Heart rate variability: indicator of user state as an aid to human-computer interaction. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. USA: ACM Press/Addison-Wesley Publishing Co.; 1998. p. 480–7.
14. Rakshit R, Reddy VR, Deshpande P. Emotion detection and recognition using HRV features derived from photoplethysmogram signals. Proceedings of the 2nd workshop on Emotion Representations and Modelling for Companion Systems. New York, NY, USA: Association for Computing Machinery; 2016. p. 1–6.
15. Bugnon LA, Calvo RA, Milone DH. Dimensional Affect Recognition from HRV: An Approach Based on Supervised SOM and ELM. IEEE Transactions on Affective Computing. 2020;11:32–44.
16. https://mahnob-db.eu/hci-tagging/

17. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585:357–62.
18. McKinney W. Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference [Internet]. SciPy; 2010 [cited 2023 Mar 24]. Available from: https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf
19. Van Gent P, Farah H, Van Nes N, Van Arem B. Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast Lane project. J Open Res Softw. 2019;7:32.
20. Hrv-analysis [Internet]. PyPI. [cited 2023 Mar 24]. Available from: https://pypi.org/project/hrv-analysis/
21. Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Health. 2017;5:258.
22. Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. 1996; Available from: https://air.unimi.it/handle/2434/188038
23. Ma C, Xu H, Yan M, Huang J, Yan W, Lan K, et al. Longitudinal Changes and Recovery in Heart Rate Variability of Young Healthy Subjects When Exposure to a Hypobaric Hypoxic Environment. Front Physiol. 2021;12:688921.
24. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python [Internet]. the Journal of machine Learning research, 12. 2011 [cited 2023 Mar 24]. p. 2825–30. Available from: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf
25. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–94.
26. Ferdinando H, Seppänen T, Alasaarela E. Enhancing emotion recognition from ECG signals using supervised dimensionality reduction. Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods [Internet]. SCITEPRESS - Science and Technology Publications; 2017. Available from: http://jultika.oulu.fi/files/nbnfi-fe2019081924650.pdf