Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Heart Rate Variability Signatures of Standardized Emotional Expression During Professional Acting Performances

Emi Yuda

yuda.icsdf@mie-u.ac.jp

Mie University

Junichiro Hayano

Heart Beat Science Lab, Inc.

Research Article

Keywords:

Posted Date: June 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9805935/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

A major challenge in affective computing is the definition of reliable emotional ground truth. Emotional labels based on subjective self-reports may be influenced by psychological and social biases, potentially limiting reproducibility and interpretability in physiological emotion research. Acting-based paradigms, in which professional actors intentionally reproduce standardized emotional expressions, provide an alternative framework for investigating the physiological organization of emotional expression under controlled conditions. While heart rate variability (HRV) reflects autonomic nervous system activity, reliable discrimination of emotional valence remains difficult using conventional HRV indices alone. We hypothesized that recently developed HRV indices reflecting the stability and temporal organization of cardiac regulation could improve discrimination of acting-based emotional expressions.

Methods

Electrocardiographic data were obtained from an open-access dataset of 11 professional actors performing four standardized emotional expressions (calm, excited, tense, and tired). A comprehensive set of HRV indices, including autoregressive variability (ARV) and the high-frequency signal power concentration index (Hsi), was extracted. Feature selection was performed using recursive feature elimination, and multiple machine learning classifiers, including a gradient boosting machine (GBM), were evaluated using participant-level separation between training and test cohorts.

Results

The GBM model demonstrated the best classification performance in the training cohort, with area under the receiver operating characteristic curve (AUC) values of 0.87 for arousal and 0.94 for valence. In an independent participant-level test cohort, the model maintained acceptable generalization performance, with AUCs of 0.71 and 0.70 for arousal and valence, respectively. Key predictors included ARV, mean NN interval, and Hsi. In particular, Hsi and ARV contributed substantially to valence discrimination, suggesting that the temporal organization and stability of cardiac dynamics contain information relevant to standardized emotional expression beyond conventional autonomic magnitude measures.

Conclusions

The acting-based paradigm enabled investigation of reproducible and standardized emotional expressions without primarily relying on subjective self-evaluation. Our findings suggest that higher-order properties of cardiac autonomic regulation, including signal stability and temporal organization, contribute to the physiological discrimination of emotional expression. This framework may provide a useful experimental model for studying the autonomic physiology of emotional expression under controlled conditions.

1. Introduction

A central methodological challenge in affective computing is the definition of emotional “ground truth.” In many studies, emotional labels are derived from participants’ subjective self-reports following emotionally evocative stimuli. However, such self-reported emotional states may be influenced not only by internal affective experience, but also by psychological and social factors, including implicit expectations regarding “appropriate” emotional responses, interpersonal self-presentation, and demand characteristics within the experimental setting. Consequently, subjective emotional ratings may not necessarily provide a stable or unbiased reference for physiological emotion modeling.

From this perspective, acting-based paradigms offer a conceptually distinct approach. Professional actors are trained to intentionally and consistently express specific emotional states under controlled conditions. Rather than modeling private subjective feelings, such paradigms enable investigation of physiologically expressed emotional states that are socially and behaviorally standardized. This framework may provide a useful experimental model for studying the autonomic organization of emotional expression with improved reproducibility and reduced ambiguity of emotional labeling.

Physiological modalities, including electroencephalography, electrocardiography, and galvanic skin response, have been extensively employed in affective computing for emotion recognition [1, 2]. Among these, heart rate variability (HRV) derived from electrocardiographic (ECG) recordings has attracted considerable attention because of its non-invasive nature and its ability to reflect autonomic nervous system (ANS) dynamics [3, 4]. Previous studies have demonstrated the utility of various HRV indices for estimating emotional arousal and related autonomic states [5, 6].

However, a persistent challenge in HRV-based emotion recognition is the discrepancy in estimation accuracy between emotional dimensions. Conventional HRV indices often perform reasonably well in estimating emotional arousal, but their utility for estimating emotional valence has remained limited [7, 8]. This limitation may partly reflect the fact that conventional HRV measures primarily quantify the magnitude of autonomic activity, whereas emotional valence may be more strongly associated with the stability, organization, and temporal structure of autonomic regulation.

Recent advances in HRV analysis have introduced novel indices designed to capture these dynamic regulatory properties. Autoregressive variability (ARV) quantifies the relative contribution of unpredictable fluctuations in heart rate dynamics [9], whereas the high-frequency signal power concentration index (Hsi) reflects the spectral stability of respiratory-related heart rate oscillations during vagal modulation [10]. These indices may therefore provide additional information regarding the physiological organization of emotional expression beyond conventional spectral measures.

In the present study, we investigated whether emotional arousal and valence associated with standardized acting-based emotional expressions could be estimated using HRV-derived physiological features. Using the open-access DECEiVeR dataset of professional actors [11], we evaluated the performance of multiple machine learning classifiers incorporating conventional HRV indices together

with ARV and Hsi. This approach aimed not only to assess classification performance, but also to explore whether higher-order properties of cardiac autonomic regulation contribute to the physiological discrimination of standardized emotional expression.

2. Methods

2.1 Database

This study utilized the DECEiVeR (DatasEt aCting Emotions Valence aRousal) dataset, an open-access repository provided by Aly et al. [12]. The dataset comprises ECG recordings from 11 professional Portuguese actors (5 females, 6 males; mean age 36 ± 10.82 years) [11]. Participants performed five emotional states—neutral, calm, excited, tense, and tired—in a standardized sequence. Each emotional session was separated by an intermission to facilitate a return to the baseline neutral state. The protocol was repeated after a 15-day interval period for each participant to ensure longitudinal data collection.

2.2 ECG Processing and NN Interval Generation

The present study used the ECG signals provided in the publicly available DECEiVeR dataset, in which the original recordings had already been resampled to 500 Hz by the dataset developers.

QRS complexes were detected using a fast peak-detection algorithm developed in our laboratory. Following automated detection, all recordings were visually inspected using beat interval trend plots and ECG waveforms to identify detection errors and non-sinus beats. Detection errors were manually corrected, and non-sinus rhythm beats were labeled and excluded from subsequent analyses.

Normal-to-normal (NN) interval time series were then constructed exclusively from consecutive sinus-rhythm R-R intervals. Only NN intervals free from ectopic beats and detection artifacts were used for HRV analyses.

2.3 HRV Analysis

HRV indices evaluated in the present study are summarized in Table 1.

Time-domain indices included the mean normal-to-normal interval (MNN) and the standard deviation of NN intervals (SDNN). These indices were calculated directly from NN interval time series using consecutive 30-second non-overlapping windows.

For frequency-domain and nonlinear analyses, NN interval time series were linearly interpolated and resampled at 2 Hz. Frequency-domain indices included the amplitudes and central frequencies of the very-low-frequency (VLF; 0.0033–0.04 Hz), low-frequency (LF; 0.04–0.15 Hz), and high-frequency (HF; 0.15–0.40 Hz) components. These amplitudes and frequencies were continuously estimated by complex demodulation [13].

Nonlinear and advanced HRV indices included approximate entropy (ApEn) [14], the high-frequency signal power concentration index (Hsi) [10], and autoregressive variability (ARV) [9]. ApEn was calculated using an embedding dimension of m = 2 and a tolerance of r = 20% of the standard deviation of the analyzed NN interval series. ARV was defined as the fractional variance of the residual time series obtained from autoregressive modeling of NN interval fluctuations. These indices were calculated using 90-second sliding windows with a 30-second step size.

To temporally align all HRV features, values derived from each 90-second window were assigned to the corresponding 30-second epoch according to the end time of the window.

2.4 Feature Extraction and Epoch Definition

Each emotional session lasted four minutes. Because several HRV indices, including ApEn, Hsi, and ARV, required a 90-second analysis window for stable estimation, the initial 90 seconds of each session were excluded from analysis.

The remaining 150 seconds of data were divided into five consecutive 30-second epochs. For each epoch, time-domain HRV indices were calculated directly from the corresponding 30-second NN interval segment. Frequency-domain and nonlinear indices were derived from 90-second sliding windows with a 30-second step size, and the resulting values were assigned to the corresponding epoch according to the end time of each window.

Accordingly, each emotional session yielded five temporally aligned feature sets for subsequent machine learning analyses.

2.5 Operational Definition of Arousal and Valence

Emotional states were operationally classified according to Russell's circumplex model of affect, which conceptualizes emotions within a two-dimensional space defined by arousal and valence [15].

The four instructed acting conditions included in the DECEiVeR dataset were categorized as follows:

- Calm: low arousal / positive valence
- Excited: high arousal / positive valence
- Tense: high arousal / negative valence
- Tired: low arousal / negative valence

Separate binary classification models were independently developed for arousal (high vs. low) and valence (positive vs. negative).

The present study used predefined acting conditions as operational emotion labels rather than subjective self-ratings of emotional states. This approach was adopted because self-reported emotional ratings may be influenced by multiple psychological and social factors, including individual

preconceptions regarding “appropriate” emotional responses, interpersonal self-presentation, and implicit expectations toward the experimental context or investigators. Such factors may introduce systematic bias into subjective emotional labeling.

In contrast, professional acting paradigms provide socially standardized and reproducible emotional expressions under controlled experimental conditions. The present study therefore aimed to evaluate whether HRV-derived physiological features could discriminate intentionally expressed emotional states without relying primarily on subjective self-evaluation.

Accordingly, the present findings should be interpreted as classification of standardized acting-based emotional expressions rather than direct estimation of subjectively labeled emotional states.

2.6 Machine Learning Analysis

All machine learning analyses were conducted using R (R Foundation for Statistical Computing, Vienna, Austria) with RStudio (version 2025.09.2). The analyses were primarily implemented using the caret package together with supplementary packages including randomForest, gbm, pROC, VIM, and themis.

2.6.1 Data partitioning

To avoid information leakage arising from repeated measurements within the same participant, the dataset was partitioned at the participant level rather than at the segment level. The 11 actors were divided into a training cohort of seven actors and an independent test cohort of four actors. Because each actor completed two recording sessions, all data from both sessions of a given actor were assigned to the same cohort. Thus, no actor contributed data to both the training and test cohorts.

This participant-level separation was specifically adopted to prevent overoptimistic performance estimates caused by within-subject similarity across repeated sessions and multiple temporal segments.

2.6.2 Feature Selection

Feature selection was performed exclusively within the training cohort using recursive feature elimination (RFE). A random forest classifier was used as the base estimator for feature ranking. Feature selection was optimized using the area under the receiver operating characteristic curve (AUC) estimated by bootstrap resampling with 25 iterations.

To avoid excessively complex models in the small dataset, the maximum number of selected variables was constrained to seven. The final subset of variables was determined by selecting the feature combination that yielded the highest bootstrap-estimated AUC within the predefined variable limit.

2.6.3 Model Training and Hyperparameter Tuning

Multiple supervised classifiers were evaluated, including logistic regression, random forest, gradient boosting machine (GBM), k-nearest neighbors (KNN), and multilayer perceptron (MLP) [6, 16].

Model training and hyperparameter tuning were performed exclusively within the training cohort using bootstrap resampling with 25 iterations. The primary optimization metric was AUC. Hyperparameters were automatically optimized within the caret framework using model-specific tuning grids determined by the tuneLength parameter.

Because class imbalance was present, synthetic minority oversampling technique (SMOTE) sampling was applied within the resampling procedure using the themis package.

Missing values were imputed using k-nearest neighbor imputation (k = 5). Imputation procedures were incorporated into the machine learning pipeline to avoid inconsistencies between training and prediction stages.

All feature selection, model tuning, and classifier optimization procedures were performed exclusively within the training cohort before independent evaluation in the test cohort.

2.6.4 Model Evaluation

The classifier demonstrating the highest AUC in the training cohort was selected as the final model for evaluation in the independent test cohort.

Model discrimination performance was evaluated using AUC, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and F1 score. Receiver operating characteristic curves were generated using the pROC package. Probability thresholds for final binary classification were optimized in the training cohort by maximizing the F1 score.

Calibration performance was additionally evaluated using calibration slope and intercept derived from the val.prob function in the rms package.

2.6.5 Variable Importance

Variable importance was evaluated using the varImp function implemented in the caret framework. Relative importance scores were scaled to facilitate comparison across predictors within the final selected model.

2.7 Statistical Analysis

Statistical analyses were conducted using SAS software (SAS Institute, Cary, NC, USA).

Differences in HRV indices between arousal levels (high vs. low) and valence polarities (positive vs. negative) were evaluated using linear mixed-effects models. In these analyses, arousal level or valence

polarity was treated as a fixed effect, and participant (actor) was treated as a random effect to account for repeated measurements and inter-individual variability.

Data are presented as mean ± standard deviation unless otherwise specified. A two-sided p-value < 0.05 was considered statistically significant.

3. Results

3.1 HRV Characteristics of Acting-Based Emotional Expressions

Linear mixed-effects analyses revealed significant differences in multiple HRV indices between acting-based emotional expression categories, as summarized in Table 2. These findings provided the physiological basis for subsequent feature selection and machine learning analyses.

High-arousal emotional expressions were associated with significantly shorter MNN, increased VLF amplitude, lower LF frequency, and higher HF frequency compared with low-arousal expressions. In addition, high-arousal expressions showed lower Hsi and higher ARV values, indicating altered autonomic and cardiorespiratory dynamics during standardized emotional expression.

Positive-valence emotional expressions exhibited longer MNN, greater SDNN, higher LF frequency, and higher HF amplitude than negative-valence expressions. Positive-valence expressions were also associated with higher values of nonlinear indices including ApEn, ApEnNorm, CGApEnNorm, and Hsi, suggesting more stable and organized cardiac dynamics during positive emotional expression.

3.2 Feature Selection

Recursive feature elimination (RFE) performed exclusively within the training cohort identified a parsimonious subset of HRV features associated with acting-based emotional expression classification.

For arousal classification, the selected feature subset included MNN, ApEnNorm, CGApEnNorm, Hsi, and ARV. These features reflected multiple aspects of cardiac dynamics, including autonomic variability, temporal organization, and signal stability.

For valence classification, the optimal feature subset consisted of MNN, Hsi, and ARV. Notably, Hsi and ARV were selected as informative predictors for valence discrimination, suggesting that higher-order characteristics of cardiac regulation contributed to classification performance beyond conventional spectral HRV measures.

The selected feature subsets were subsequently used for training and evaluation of the machine learning classifiers.

3.3 Classification Performance in the Training Cohort

The classification performance of the evaluated machine learning models in the training cohort is summarized in Table 3.

Among the evaluated classifiers, the GBM demonstrated the highest AUC and F1 score for both arousal and valence classification. For arousal classification, the GBM achieved an AUC of 0.87 with an F1 score of 0.78. For valence classification, the GBM achieved an AUC of 0.94 with an F1 score of 0.85.

Random forest models also demonstrated relatively high discrimination performance, whereas logistic regression, k-nearest neighbors, and multilayer perceptron models showed comparatively lower AUC and F1-score values.

Based on the highest bootstrap-estimated AUC in the training cohort, the GBM classifier was selected as the final model for subsequent evaluation in the independent participant-level test cohort.

3.4 Performance in the Independent Test Cohort

The generalization performance of the final GBM classifier was evaluated in an independent participant-level test cohort consisting of actors not included in the training cohort (Tables 4 and 5).

For arousal classification, the model achieved an AUC of 0.71, with a sensitivity of 0.70, specificity of 0.73, and F1 score of 0.71. For valence classification, the model achieved an AUC of 0.70, with a sensitivity of 0.74, specificity of 0.64, and F1 score of 0.70.

Although classification performance decreased compared with the training cohort, the model maintained moderate discrimination performance in previously unseen participants. These findings indicate that HRV-derived physiological features contained information associated with standardized acting-based emotional expression that generalized beyond the participants used for model training and feature selection.

3.5 Variable Importance

Figure 1 illustrates the relative contribution of each feature within the GBM models. ARV, MNN, and Hsi emerged as the most influential predictors, further validating the hypothesis that both the magnitude and the qualitative organization of autonomic activity are critical for physiological discrimination of standardized emotional expression.

4. Discussion

In the present study, we investigated whether standardized acting-based emotional expressions could be discriminated using HRV-derived physiological features. By combining advanced HRV indices, including ARV and Hsi, with nonlinear machine learning approaches, we identified physiological patterns associated with emotional expression that generalized to independent participants.

A central conceptual aspect of this study is the use of acting-based emotional expression as an alternative framework to conventional subjective emotion labeling. In many affective computing studies, emotional ground truth is defined by participants' self-reported feelings following emotionally evocative stimuli. However, subjective emotional ratings may be influenced by psychological and social biases, including implicit expectations regarding "appropriate" emotional responses, interpersonal self-presentation, and demand characteristics within the experimental setting. From this perspective, acting-based paradigms provide a distinct experimental framework in which emotional expressions are intentionally and reproducibly standardized by trained actors under controlled conditions.

Accordingly, the present findings should not be interpreted as direct estimation of private subjective emotional states, but rather as discrimination of physiologically expressed emotional states associated with standardized emotional expression. This distinction may be important when considering the methodological foundations of physiological affective computing.

4.1 The Acting-Based Paradigm

A distinctive feature of this study is the use of standardized acting-based emotional expression rather than spontaneously elicited affective states. Professional actors are specifically trained to intentionally reproduce emotional expressions with consistency across repeated performances [11]. This characteristic provides a highly controlled experimental framework for investigating physiological correlates of emotional expression while reducing variability associated with spontaneous emotional experience.

Importantly, this approach does not necessarily assume that the actors internally experienced identical subjective emotions during every performance. Rather, the acting paradigm enabled investigation of reproducible physiological organization associated with socially expressed emotional states. Such a framework may complement conventional self-report-based paradigms by providing a more standardized reference structure for physiological emotion research.

4.2 Challenges in Valence Estimation

Consistent with previous literature, our results show that emotional arousal was more accurately classified than valence, especially in the independent test dataset. This aligns with the prevailing view that peripheral physiological signals, including HRV, more directly reflect autonomic activation (arousal) than affective polarity [3, 17]. While arousal is deeply rooted in the sympathetic–parasympathetic balance, valence involves higher-order cognitive and contextual processes that are often not fully captured by cardiac autonomic indices alone [8].

However, the incorporation of advanced indices—Hsi and ARV—enhanced the classification of emotional valence. Conventional spectral indices (e.g., LF and HF) primarily reflect the tonic or phasic "amount" of autonomic activity. In contrast, Hsi reflects the spectral concentration of high-frequency components, serving as an indicator of the stability and efficiency of parasympathetic modulation. Our findings of

higher Hsi in positive valence states suggest that organized and stable cardiac-respiratory coupling is a characteristic of positive-valence emotional expression, potentially reflecting more stable autonomic organization [4]. Similarly, ARV provides insights into the short-term temporal instability of cardiac control. The significant contribution of ARV to valence classification suggests that negative emotions may disrupt the regularity of heart rate intervals—a nuance often overlooked by traditional linear features. Collectively, our results suggest that emotional valence-related information may be reflected not merely in the magnitude of autonomic activity, but in the regulatory quality and temporal organization of the cardiac rhythm.

4.3 Modeling Nonlinear Dynamics

The consistent identification of ARV and Hsi as key predictors, combined with the superior performance of the GBM model, underscores the importance of nonlinear organization of cardiac autonomic dynamics associated with emotional expression. While linear models often struggle to capture the complex, non-monotonic relationships between autonomic stability and affect, the GBM effectively modeled these intricate patterns. Furthermore, the use of RFE ensured model parsimony, focusing on a subset of physiologically meaningful features to maintain robustness and prevent overfitting.

4.4 Limitations and Future Directions

Several limitations warrant consideration. First, the sample size was relatively small and limited to professional Portuguese actors, which may restrict the generalizability of the findings to broader or more diverse populations. Second, although acting provides high controllability, it may not fully represent the spontaneous affective experiences of daily life; physiological signatures in naturalistic contexts may differ from those observed in a controlled acting paradigm. Third, this study focused exclusively on HRV. Integrating multimodal signals—such as electrodermal activity, respiration, and facial electromyography—could further enhance recognition accuracy, particularly for the valence dimension.

Future research should expand this framework to include larger, non-actor cohorts and clinical populations. Bridging the gap between experimental controllability and ecological validity remains a key challenge. Longitudinal investigations over extended periods would also provide deeper insights into the temporal stability of these physiological markers.

4.5 Applied Implications

From an applied perspective, the ability to reliably estimate arousal and valence from HRV has significant implications for affective computing and human–computer interaction. These findings support the development of adaptive systems in healthcare, mental health monitoring, and immersive media. Acting-based datasets, in particular, serve as essential benchmarks for validating physiological modeling of emotional expression under standardized conditions.

5. Conclusion

This study demonstrates that HRV-derived physiological features contain information capable of discriminating standardized acting-based emotional expressions associated with arousal and valence. By incorporating advanced indices reflecting the temporal organization and stability of cardiac dynamics, together with nonlinear machine learning approaches, we identified reproducible physiological patterns that generalized to independent participants.

Beyond classification performance itself, the present findings highlight the potential methodological value of acting-based paradigms for physiological affective computing. Standardized emotional expression may provide a complementary framework to subjective self-report-based emotion labeling for investigating the autonomic physiology of emotional expression under controlled conditions.

Declarations

Funding This work was supported by the New Energy and Industrial Technology Development Organization (NEDO) [Grant Principal Investigator: Emi Yuda

Ethical Considerations

The present study involved secondary analysis of the anonymized open-access DECEiVeR dataset. Ethical approval procedures and informed consent for the original data collection are described in Aly et al. [11]. No additional ethical approval was required for the present secondary analysis of publicly available anonymized data.

References

1. Picard, R. W., *Affective Computing*. MIT Press: Cambridge, MA, 1997.
2. Calvo, R. A.; D□fMello, S., Affect detection: An interdisciplinary review of models, methods, and applications. *IEEE Transactions on Affective Computing* **2010,** 1, (1), 18–37.
3. Kreibig, S. D., Autonomic nervous system activity in emotion: a review. *Biol. Psychol.* **2010,** 84, (3), 394–421.
4. Thayer, J. F.; Ahs, F.; Fredrikson, M.; Sollers, J. J., 3rd; Wager, T. D., A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* **2012,** 36, (2), 747–56.
5. Shaffer, F.; Ginsberg, J. P., An overview of heart rate variability metrics and norms. *Frontiers in public health* **2017,** 5, 258.
6. Chen, Y. C.; Hsiao, C. C.; Zheng, W. D.; Lee, R. G.; Lin, R., Artificial neural networks-based classification of emotions using wristband heart rate monitor data. *Medicine (Baltimore).* **2019,** 98, (33), e16863.
7. Posner, J.; Russell, J. A.; Peterson, B. S., The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005,** 17,

(3), 715–34.

8. Appelhans, B. M.; Luecken, L. J., Heart rate variability as an index of regulated emotional responding. *Biol. Psychol.* **2006,** 10, (3), 229–240.
9. Hayano, J.; Ohashi, K.; Yoshida, Y.; Yuda, E.; Nakamura, T.; Kiyono, K.; Yamamoto, Y., Increase in random component of heart rate variability coinciding with developmental and degenerative stages of life. *Physiol. Meas.* **2018,** 39, (5), 054004.
10. Hayano, J.; Ueda, N.; Kisohara, M.; Yoshida, Y.; Tanaka, H.; Yuda, E., Non-REM Sleep Marker for Wearable Monitoring: Power Concentration of Respiratory Heart Rate Fluctuation. *Appl Sci* **2020,** 10, (9), 3336.
11. Aly, L.; Godinho, L.; Bota, P.; Bernardes, G.; da Silva, H. P., Acting Emotions: a comprehensive dataset of elicited emotions. *Sci Data* **2024,** 11, (1), 147.
12. Aly, L.; Godinho, L.; Bota, P.; Bernardes, G.; da Silva, H. P. Deceiver a comprehensive dataset of elicited emotions. https:// doi.org/10.6084/m9.figshare.23579862.v3 (February 14),
13. Hayano, J.; Taylor, J. A.; Yamada, A.; Mukai, S.; Hori, R.; Asakawa, T.; Yokoyama, K.; Watanabe, Y.; Takata, K.; Fujinami, T., Continuous assessment of hemodynamic control by complex demodulation of cardiovascular variability. *Am. J. Physiol.* **1993,** 264, (4 Pt 2), H1229–38.
14. Pincus, S. M.; Goldberger, A. L., Physiological time-series analysis: What does regularity quantify. *Am. J. Physiol.* **1994,** 266, H1643–H1656.
15. Russell, J. A., A circumplex model of affect. *J. Pers. Soc. Psychol.* **1980,** 39, (6), 1161–1178.
16. Friedman, J. H., Greedy function approximation: A gradient boosting machine. *Annals of Statistics* **2001,** 29, (5), 1189–1232.
17. Mauss, I. B.; Robinson, M. D., Measures of emotion: A review. *Cogn Emot* **2009,** 23, (2), 209–237.

Tables

Table 1. Heart rate variability (HRV) indices evaluated in the present study and their definitions

Indices	Definition
MNN	Mean normal-to-normal (NN) R-R interval interval of electrocardiogram
SDNN	Standard deviation of NN interval
VLF amp	Amplitude of very-low-frequency (0.0033-0.04 Hz) component
LF amp	Amplitude of low-frequency (LF, 0.04-0.15 Hz) component
LF freq	Frequency of LF component
HF amp	Amplitude of high-frequency (HF, 0.15-0.40 Hz) component
HF freq	Frequency of HF component
ApEn	Approximate entropy [14]
ApEnNorm	Normalized ApEn [14]
CGApEn	Coarse-grained ApEn: ApEn of 10-sec moving averaged NN intervals
CGApEnNorm	Normalized CGApEn
Hsi	High frequency signal power concentration index: The degree of HF power concentration into the main spectral peak [10].
ARV	Autoregressive (AR) variability: The ratio of random component defined as the fractional variance of the autoregressive residual time series [9].

Table 2. Comparison of standardized HRV indices between high and low arousal and between positive and negative valence during acting emotions (all participants)

	Arousal			Valence		
	Low	High	*P*	Negative	Positive	*P*
MNN	-0.59 ± 2.64	-2.35 ± 2.79	<0.0001	-2 ± 2.77	-0.94 ± 2.84	<0.0001
SDNN	0.8 ± 2.36	0.55 ± 2.49	0.2	0.42 ± 2.52	0.93 ± 2.31	0.01
VLF amp	0.55 ± 1.84	0.94 ± 2.03	0.02	0.78 ± 2.02	0.71 ± 1.87	0.6
LF amp	1 ± 2.22	0.99 ± 2.23	0.9	0.91 ± 2.31	1.07 ± 2.13	0.4
LF freq	-0.09 ± 1.67	-0.42 ± 1.65	0.02	-0.54 ± 1.62	0.03 ± 1.66	<0.0001
HF amp	0.19 ± 2.56	-0.25 ± 2.72	0.05	-0.45 ± 2.72	0.39 ± 2.51	0.0002
HF freq	-0.57 ± 1.88	-0.21 ± 2.01	0.01	-0.29 ± 2.02	-0.49 ± 1.87	0.1
ApEn	-0.24 ± 2.45	-0.64 ± 2.94	0.07	-0.75 ± 2.94	-0.13 ± 2.42	0.005
ApEnNorm	-0.94 ± 2.09	-0.87 ± 2.05	0.6	-1.07 ± 2.15	-0.73 ± 1.97	0.04
CGApEn	1.39 ± 2.54	1.18 ± 2.62	0.3	1.1 ± 2.75	1.47 ± 2.39	0.09
CGApEnNorm	-0.01 ± 1.59	-0.18 ± 1.88	0.2	-0.25 ± 1.78	0.05 ± 1.69	0.03
Hsi	-1.18 ± 1.87	-1.59 ± 1.92	0.01	-1.72 ± 1.92	-1.04 ± 1.83	<0.0001
ARV	-0.25 ± 0.65	0.03 ± 1.08	<0.0001	-0.13 ± 0.86	-0.09 ± 0.94	0.5

Data are presented as mean ± standard deviation.

All values were standardized (z-scores) prior to analysis.

Abbreviations are explained in Table 1.

Table 3. Performance of machine learning classifiers for arousal and valence classification in the training dataset

Classifier	AUC	Sensitivity	Specificity	PPV	NPV	F1 score
Arousal						
LogReg	0.66	0.62	0.63	0.62	0.62	0.62
RF	0.86	0.78	0.75	0.80	0.79	0.76
GBM	0.87	0.79	0.78	0.81	0.80	0.78
KNN	0.75	0.67	0.68	0.67	0.67	0.67
MLP	0.71	0.67	0.66	0.68	0.68	0.67
Valence						
LogReg	0.64	0.61	0.57	0.66	0.63	0.61
RF	0.91	0.84	0.79	0.88	0.87	0.81
GBM	0.94	0.88	0.84	0.91	0.90	0.85
KNN	0.73	0.66	0.71	0.62	0.65	0.68
MLP	0.68	0.63	0.62	0.63	0.63	0.63

LogReg, logistic regression; RF, random forest; GBM, gradient boosting machine; KNN, k-nearest neighbor; and MLP , multi-layer perceptron.

Table 4. Confusion matrix of arousal classification in the test dataset using the GBM model

		Ground truth		
		Low	High	
Prediction	Low	58	24	NPV = 0.70
	High	22	56	PPV = 0.72
		Specificity = 0.73	Sensitivity = 0.70	Accuracy = 0.71
				F1 score = 0.71

NPV, negative predictive value; PPV, positive predictive value.

Table 5. Confusion matrix of valence classification in the test dataset using the GBM model

		Ground truth		
		Negative	Positive	
Prediction	Negative	51	21	NPV = 0.71
	Positive	29	59	PPV = 0.67
		Specificity = 0.64	Sensitivity = 0.74	Accuracy = 0.69
				F1 score = 0.70

NPV, negative predictive value; PPV, positive predictive value.

Figures

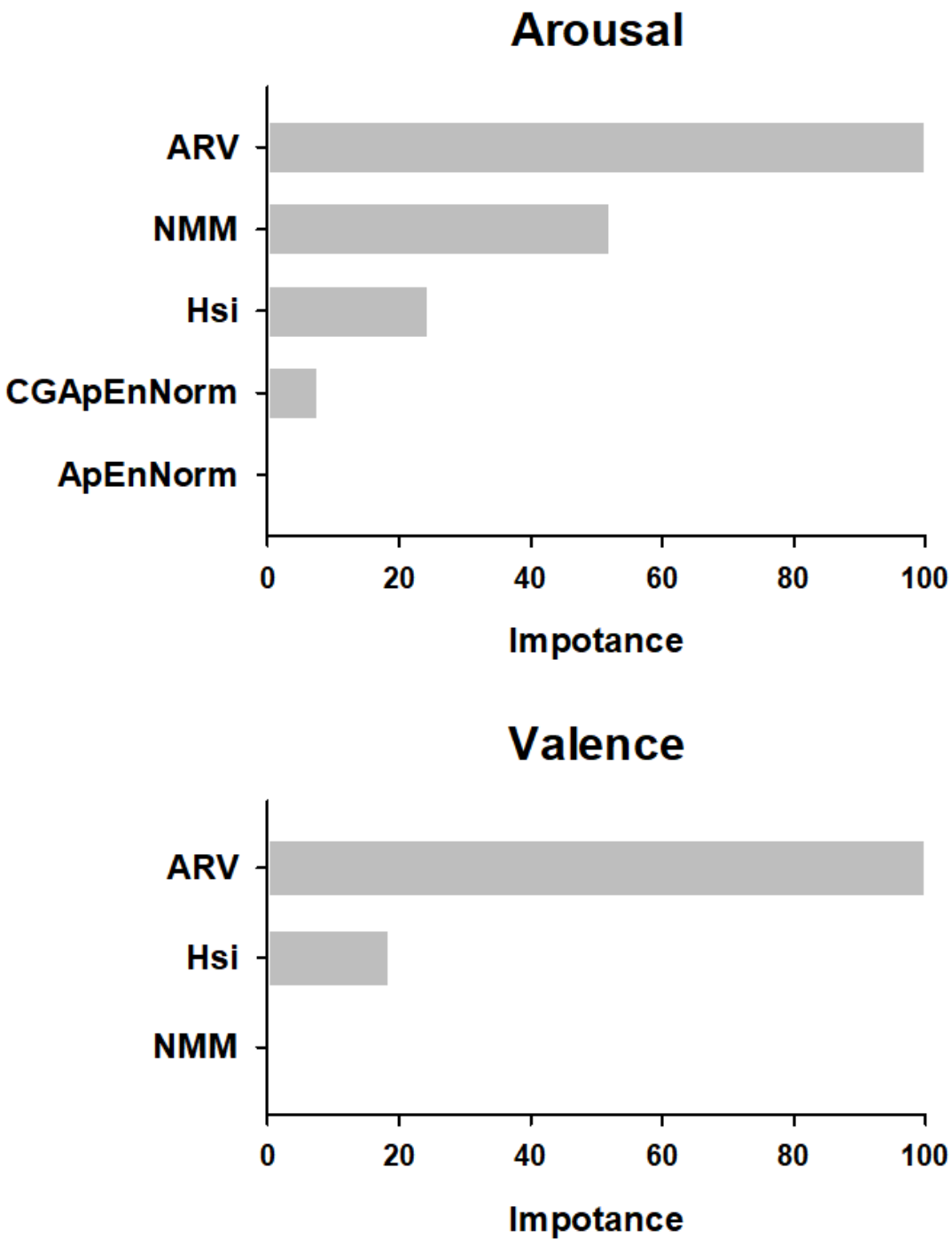

Figure 1

Relative importance of HRV features in the Gradient Boosting Machine (GBM) models for classifying acting-based emotional expressions associated with arousal (upper panel) and valence (lower panel).