CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Inter and intra signal variance in feature extraction and classification of affective state
Authors	Dair, Zachary;Dockray, Samantha;O'Reilly, Ruairi
Publication date	2023-02-23
Original Citation	Dair, Z., Dockray, S. and O'Reilly, R. (2023) 'Inter and intra signal variance in feature extraction and classification of affective state', AICS2022, in L. Longo and R. O'Reilly (eds) Communications in Computer and Information Science, vol 1662: Springer Nature Switzerland, pp. 3–17. https://doi.org/10.1007/978-3-031-26438-2_1.
Type of publication	Conference item;Book chapter
Link to publisher's version	10.1007/978-3-031-26438-2_1
Rights	© 2023 The Author(s). Open Access. This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made - http://creativecommons.org/licenses/by/4.0/
Download date	2026-10-23 05:12:31
Item downloaded from	https://hdl.handle.net/10468/14345

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Inter and Intra Signal Variance in Feature Extraction and Classification of Affective State

Zachary Dair[1](✉), Samantha Dockray[2], and Ruairi O'Reilly[1]

[1] Munster Technological University, Cork, Ireland
zachary.dair@mycit.ie, ruairi.oreilly@mtu.ie
[2] University College Cork, Cork, Ireland
s.dockray@ucc.ie

Abstract. Psychophysiology investigates the causal relationship of physiological changes resulting from psychological states. There are significant challenges with machine learning-based momentary assessments of physiology due to varying data collection methods, physiological differences, data availability and the requirement for expertly annotated data. Advances in wearable technology have significantly increased the scale, sensitivity and accuracy of devices for recording physiological signals, enabling large-scale unobtrusive physiological data gathering. This work contributes an empirical evaluation of signal variances acquired from wearables and their associated impact on the classification of affective states by (i) assessing differences occurring in features representative of affective states extracted from electrocardiograms and photoplethysmography, (ii) investigating the disparity in feature importance between signals to determine signal-specific features, and (iii) investigating the disparity in feature importance between affective states to determine affect-specific features. Results demonstrate that the degree of feature variance between ECG and PPG in a dataset is reflected in the classification performance of that dataset. Additionally, beats-per-minute, inter-beat-interval and breathing rate are identified as common best-performing features across both signals. Finally feature variance per-affective state identifies hard-to-distinguish affective states requiring one-versus-rest or additional features to enable accurate classification.

Keywords: Machine learning · Classification · Psychophysiology · Electrocardiogram · Photoplethysmography · Affective states

1 Introduction

A significant goal of Affective Computing is to improve human-to-computer interaction by providing a system with a level of emotional intelligence that

This publication has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 18/CRT/6222.

© The Author(s) 2023
L. Longo and R. O'Reilly (Eds.): AICS 2022, CCIS 1662, pp. 3–17, 2023.
https://doi.org/10.1007/978-3-031-26438-2_1

aids natural communications and is capable of including emotional components [27]. This has commonly been approached by deriving emotional states from speech, facial expressions, gestures and body posture analysis. However, utilising physiological signals to communicate psychological information is a recent exploration in the domain, likely due to the increased accessibility of signals from wearables.

A physiological signal represents an individual's biological processes derived from core aspects of human biology. These signals can enable diagnostics, for instance, analysing heart rate (HR) to detect arrhythmia [29]. Psychological analysis can also be enabled as mental states originating from unconscious effort typically present a noticeable physiological change in the relevant human system [16]. The combined analysis enables a richer understanding of individuals in terms of their mental and physical health [8].

Psychological states are complex processes comprised of several components, including feelings, cognitive reactions, behaviour and thoughts [1]. Mapping psychological states to individual experience provides valuable information regarding well-being, health (physical and mental), social contexts, experiences and emotions [7].

Electrocardiograms (ECG) are physiological signals that measure the electrical activity of the heart. Typically recorded in a clinical setting using multiple electrodes attached to the individual. Photoplethysmography (PPG) is a physiological signal used to measure heart activity through variations in the blood volume of the skin, using a light-emitting-diode and photodetector. Wearable devices predominately utilise PPG to monitor heart activity. However, recently advanced wearables have included ECG capabilities for a limited number of commercial off-the-shelf (COTS) devices.

Data variances occur when recording ECG and PPG due to differing sensor placement and signal granularity [8,22]. A lower sampling frequency is commonly used in PPG compared to ECG to reduce battery consumption in COTS devices. Such variances are under-recognised in the field of psychophysiology.

This work investigates the impact of signal variances occurring in ECG and PPG signals acquired from wearable devices for classifying affective states by addressing the following research aims: (i) To assess differences in features representative of affective states on a per signal basis, (ii) To investigate the disparity in precedence ordering of feature importance per signal, and (iii) To investigate the disparity precedence ordering of feature importance per affective state.

These aims inform the development of machine learning (ML) pipelines for classifying affective states. Utilising feature variance per signal to identify abnormal signal activity or similar affective states which are causing reduced classification accuracy. In conjunction, feature importance is utilised to provide insights into feature selection, aiding performance in tailored signal- or affect-specific approaches.

2 Related Work

2.1 Heart-Related Physiological Signals

The prevalence of heart-related data in wearable devices stems from a desire to monitor health through arrhythmia detection and HR as a measure of fitness [29]. As the heart is controlled involuntarily through the autonomic nervous system (ANS), it facilitates identifying relationships between involuntary physiological changes in heart activity and psychological states such as emotions or behaviour. Multiple psychophysiological theories aim to explain this relationship, such as Polyvagal Theory [30], which proposes that the ANS provides the neurophysiological substrates for adaptive behavioural strategies [28].

Heart activity is complex to capture. In medicine, the gold standard utilises a 12-lead ECG, resulting in comprehensive data recorded from multiple electrodes on the human body. However, in ambulatory research and daily life, this method is not feasible. Typically research-grade (RG) equipment uses several electrodes, commonly 3-lead ECG, and occasionally includes PPG as an additional measure. COTS devices tend to rely solely on PPG to monitor heart activity. However, with recent advances, top-of-the-range smart-watches (Apple Watch 4–9, Galaxy Active 2, Fitbit) include a 1-lead ECG, which is promising for portable ECG analysis [25].

Additional physiological signals such as electrodermal activity (EDA), respiration, skin temperature, electromyogram (EMG), and electrooculogram (EOG) have demonstrated potential for affective state detection [8,20]; however, due to additional sensor requirements they are excluded from this work.

Numerous studies of affective states conduct custom data collection, providing precise control over the psychological domain explored. Varied stimuli have aided the elicitation of psychological states, for example, images, movie clips, music, and dedicated tasks to elucidate stress, such as the Trier Social Stress Test [2,33]. As denoted in Table 1, several open-access or on-request datasets containing ECG and PPG are available. The distinct lack of emotionally labelled ECG signals from COTS devices is likely due to the recent inclusion of ECG monitoring capabilities [25].

2.2 Affective ECG Analysis

ECG signals contain noise introduced by motion artefacts, biological differences and sensor de-attachment. Signal processing techniques such as Butterworth Bandpass, Notch filters and Empirical Mode Decomposition (EMD) are utilised to reduce the signal noise levels [1]. Subsequently, features suitable for affective state classification can be extracted from the pre-processed signals.

An overview of features derivable from ECG and PPG is denoted in Table 2, grouped by extraction method. Performant ECG-based approaches typically utilise handcrafted features, particularly time-based HRV features, such as R-R intervals (RR) which are the intervals between heartbeats, successive dif-

Table 1. Datasets containing affectively labelled ECG, PPG or both

Dataset	ECG	PPG	Participants	Label
CASE [34]	✓ (1000 Hz)	✓ (1000 Hz)	30	Aro./Val.
WESAD [33]	✓ (700 Hz)	✓ (64 Hz)	15	B, S, A, M
DREAMER [15]	✓ (256 Hz)	x	23	Aro./Val./Dom.
SWELL [18]	✓ (2048 Hz)	x	25	S, Aro./Val./Dom.
DEAP [17]	x	✓ (256 Hz)	32	Aro./Val

B: Baseline, S: Stress, A: Amusement, M: Meditation Aro: Arousal, Val: Valence, Dom: Dominance

ferences (SD) and frequency-based features, such as relative, peak and absolute power of various frequency bands. Automated feature extraction is less frequently adopted, with only three of the reviewed approaches utilising deep learning or signal-processing feature extraction methods.

Recent approaches have favoured deep learning methodologies [31], achieving significant accuracies on multi-class classifications. However, older studies focusing on linear and quadratic discriminant analysis (LDA, QDA) [1,5,26] and support vector machines (SVM) [12] remain highly relevant, achieving high accuracy for their respective classifications. Combinations of ML classifiers forming ensembles have demonstrated potential for binary classifications in emotion detection [6]. In comparison to other studies, [31] achieved the highest accuracy for multiple emotion detection from ECG data utilising a CNN and reported setting the new state of the art for ECG emotion detection. Despite the high performance of deep learning approaches in the literature, this work focuses on classifiers using handcrafted features.

2.3 Affective PPG Analysis

PPG analysis provided by COTS devices has typically focused on tracking medical conditions, physical activity, and stress. The detrimental effects of stress on human health are a significant motivator for physiological analysis and preventative healthcare research [3]. However, instances of PPG have demonstrated similar noise levels to ECG, with the addition of skin tone and environmental light effects impacting signal quality, requiring signal cleaning techniques.

There is no consensus on the most frequently used features from the reviewed PPG-based approaches, see Table 2. The most performant approach [11] leverages handcrafted non-linear entropy features, followed by [24] using an autoencoder method for automatic feature extraction. Importantly, both handcrafted and automatically extracted features aid in achieving a high classification accuracy above 90% [10,11,14,24].

Table 2. Handcrafted, automated and statistical features utilised for affective state classification. Note the divergence between features used by PPG and ECG.

	ECG								PPG							Combined	
	[31]	[1]	[13]	[5]	[26]	[12]	[6]	[32]	[11]	[24]	[10]	[14]	[21]	[19]	[32]	[4]	[9]
HRV Time:																	
HR	x	x	x	✓	x	✓	x	✓	x	x	x	✓	x	x	✓	x	x
IBI	x	x	x	x	x	x	x	✓	x	x	x	x	x	x	✓	✓	x
RR	x	x	x	✓	✓	✓	✓	✓	x	x	x	x	x	x	✓	x	✓
SD	x	x	✓	✓	✓	✓	✓	✓	x	x	x	x	x	x	✓	x	✓
P-QRS-T	x	x	✓	x	x	x	✓	✓	x	x	x	x	x	x	✓	x	x
HRV Frequency:																	
Low Freq	x	x	✓	✓	✓	✓	✓	✓	x	x	x	✓	x	x	✓	x	✓
High Freq	x	x	x	✓	✓	✓	✓	✓	x	x	x	✓	x	x	✓	x	✓
Freq. Ratios	x	x	x	✓	✓	✓	✓	✓	x	x	x	✓	x	x	✓	✓	✓
Non-Linear:																	
Poin	x	x	x	x	✓	✓	✓	x	x	x	✓	x	x	x	x	x	x
Entr	x	x	x	x	✓	✓	x	✓	✓	x	x	x	x	x	✓	✓	✓
Deep Learning:																	
Model	✓	x	x	x	x	x	x	x	x	x	x	x	✓	✓	x	x	x
AE	x	x	x	x	x	x	x	x	x	✓	x	x	x	x	x	x	x
Additional Features:																	
BR	x	x	x	x	x	✓	x	x	x	x	x	x	x	x	x	x	✓
Sig. Amp	x	x	x	x	x	x	x	✓	x	x	x	x	x	x	✓	x	✓
EMD	x	✓	x	x	x	x	✓	x	x	x	x	x	x	x	x	✓	x
Classifier	CNN	LDA	KNN	LDA	QDC	SVM[b]	Ens	SVM	PNN	SVM	SVM	DNN	CNN	CNN	SVM	SVM[b]	FNN
No. Classes	4	2	2	3	4	2	4	2	14[a]	4[a]	2	5	2	2	2	2	4
Accuracy	95%	89%	85%	85%	84%	82%	80%	69%	100%	99%	96%	91%	83%	76%	65%	93%	70%
Datasets	[33]	Priv	Priv	Priv	Priv	Priv	Priv	Priv	[17]	[33]	[17]	Priv	[33]	[17]	Priv	Priv	Priv

HRV: Heart Rate Variability, IBI: Inter-beat Interval, RR: R-R Intervals, SD: Successive Differences, Freq: Frequency, Poin: Poincare, Entr: Entropy, AE: Auto-Encoder, BR: Breathing Rate, Ens: Ensemble, Sig. Amp: Signal Amplitude, EMD: Empirical Mode Decomposition, LDA: Linear Discriminant Analysis, (C/D/F)NN: Convolutional/Deep/Feed-Forward Neural Network, QDC: Quadratic Discriminant Classifier, SVM: Support Vector Machine, KNN: K-Nearest Neighbours, [a]One Vs Rest, [b]Least-Squares SVM

These affective state classifications are conducted by variations of neural networks [11,14,19,21] and SVMs [10,24,32], which demonstrates great potential for both binary and multi-class affective state detection using PPG solely. Notably, these approaches leverage extensive signal processing to reduce signal noise and contribute to the high performances achieved.

3 Methodology

The proposed methodology provides an approach for investigating ECG and PPG variances and the subsequent impact on affective state classification. The baseline performance of affective state classification is achieved using multiple ML classifiers per signal. The inter-signal performance variances are investigated by analysing the disparity in features between temporally aligned ECG and PPG, where the degree of feature variance is an indicator of signal quality. Inter-affective state feature variance is analysed using statistical measures to

provide insights into the distribution and similarity of affective states. Feature importance is employed to identify commonalities among the best-performing features across both signals and evaluate each feature's utility for affect-specific approaches. Finally, a one-versus-rest (OVR) classification is adopted to improve performance when classifying similar affective states.

3.1 Datasets

For the purposes of this work, the focus was narrowed to RG physiological signals due to a lack of publicly available data for COTS devices. "The Dataset of Continuous Affect Annotations and Physical Signals for Emotion Analysis" (CASE) [34] and "The Wearable Stress and Affect Detection Dataset" (WESAD) [33], see Table 1, were utilised in this work. The datasets were selected due to their inclusion of temporally aligned ECG and PPG with psychological annotations. Additionally, these signals were recorded using RG devices in a laboratory environment. CASE incorporates Arousal and Valence annotations, achieved by collecting joystick movement resulting from emotionally stimulating video clips. WESAD focuses on stress detection with limited affective states: a baseline state elicited from "neutral reading", amusement caused by comedic video clips, a Trier Social Stress Test [2] to provoke stress, and a meditation stage aimed at "de-exciting" the individual following the amusement and stress stages.

3.2 Pre-processing

ECG and PPG signals recorded per subject within these datasets span the duration of the experiment resulting in approx 91/40 min for WESAD/CASE. Each signal is pre-processed into 10-second windows to facilitate analysis, accomplished using a sliding window technique with a 1-second overlap. A 10-second duration was selected due to efficient performance demonstrated in [31]; additionally, this duration enables low latency as classification occurs every 10-seconds and contains adequate data for feature computation.

A Butterworth-Bandpass filter is used to reduce signal noise, facilitating the extraction of selected features while maintaining a degree of "rawness" in the signal. This filter was adopted as it is frequently adopted in the literature and more closely aligns with COTS devices and their reduced computational power.

Once filtered and windowed, the data is aligned with the psychological annotations. For WESAD, annotations were numeric values sampled 700 Hz. Each value from 0–4 is associated with the psychological states: Transient, Baseline, Stress, Amusement and Meditation. Annotations 5–7 and Transient data are omitted as per the author's instructions [33]. Certain windows may include multiple emotive annotations; hence to identify the most pertinent emotion, the mean of all annotation values per window is calculated and rounded to the nearest annotation (1–4) using Euclidean distance. Alternative approaches [6] omit these windows and the neighbouring segments to prevent overlap.

A similar procedure is required for CASE; the raw annotation data is provided as values on an x and y-axis representing Arousal and Valence [34], these

values are normalised to a range of 0.5 to 9.5, and subsequently converted to discrete representations, resulting in low (0.5–5) and high (5.01–9.5) Arousal and Valence for each window.

Both signals provide capabilities to derive a wide array of handcrafted features useful for identifying affective information. This work utilises a python toolkit HeartPy [35] to enable extraction of HRV features from each window of data, summarised in accompanying table of Fig. 4.

3.3 HRV Feature Variance

The feature variance approach proposed is to statistically evaluate any disparity occurring in derived features from ECG and PPG under multiple conditions. Inter-signal variance is evaluated by computing the absolute difference between an ECG-derived feature and its PPG counterpart from temporally aligned signals. This is assessed using the same window of heart activity and provides a granular analysis to aid in identifying noisy, erratic or abnormal signal activity, causing unreliable computations of features. This variance is depicted by a significant absolute difference of a feature between the two signals.

Analysing the inter-affective state variance in features enables the identification of the degree of change between states, as investigated in [10,12]. The proposed methodology computes the minimum, maximum, mean and standard deviation of each feature value per affective state. Additionally outliers are identified, these are observations found in the upper and lower quartiles. This method identifies states which are complex to distinguish due to a similar feature distribution, such as meditation and relaxation. An OVR approach is adopted to convert a multi-class problem into multiple binary classifications. Using OVR, a classifier aims to identify an affective state individually from the remaining states, which increases the degree of distinction between classes.

3.4 HRV Feature Importance

This work adopts a game theory approach for feature importance known as "Shapley Additive exPlanations" [23]. This method computes SHAP values representing the degree of change on the classifier output caused by each individual feature, the magnitude of change and number of samples affected indicate the impact factor of a given feature.

Feature importance has enabled the identification of signal-specific features in [32]. However, their approach utilised different features for ECG and PPG, as such, an intra-signal comparison could not be conducted, which would provide insights into the commonality of features between ECG and PPG, motivating the intra-signal feature importance analysis provided in this work.

Feature importance can also provide insights into the variance of features per affective state, valuable for the creation of tailored emotion-specific approaches. In [9], a neural network is used for classification, and the most important features were identified from the first layer's weights. These features were then evaluated to identify a statistical difference between affective states.

3.5 ML Based Classification of Affective State

A range of classifiers was selected to provide a holistic view of the classification performance using different architectures and the suitability of ECG and PPG for automated affective state detection.

Each classifier conducts a per-signal classification on each dataset, where 20% of the data acts as a hold-out test set, which is unseen data used to evaluate the final classifier. To ensure generalisability, five-fold cross-validation is utilised, transforming the remaining 80% of data into "folds", enabling a per-fold classification. Subsequently, comparing the per-fold and average performance across the five folds enables the identification of the most robust and performant classifier. Finally, the most performant classifier is trained on the entire training set and evaluated against the hold-out set to assess expected performance in real-life classifications.

4 Results and Discussion

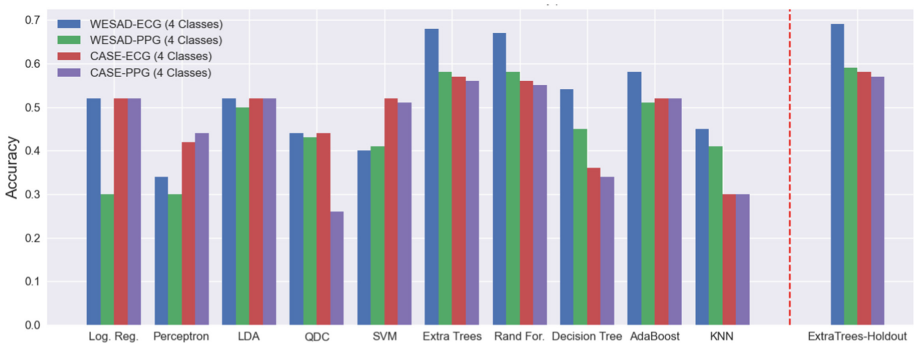


Fig. 1. Mean cross-validation accuracy classifying affective states for model selection, and the performance on the holdout test set from the best performing classifier. The performance variance in ExtraTrees classifier between ECG and PPG reflects the degree of feature variance identified per dataset.

4.1 HRV Feature Variance

The wearables' sample rate disparity (See Table 1) is evident in the inter-signal feature variance results depicted in Fig. 2. The reduced sample rates in WESAD result in slightly decreased granularity of ECG data and significantly in PPG data compared to CASE. A higher fluctuation in feature variance occurs in WESAD in terms of magnitude and frequency, stemming from the high sample rate disparity.

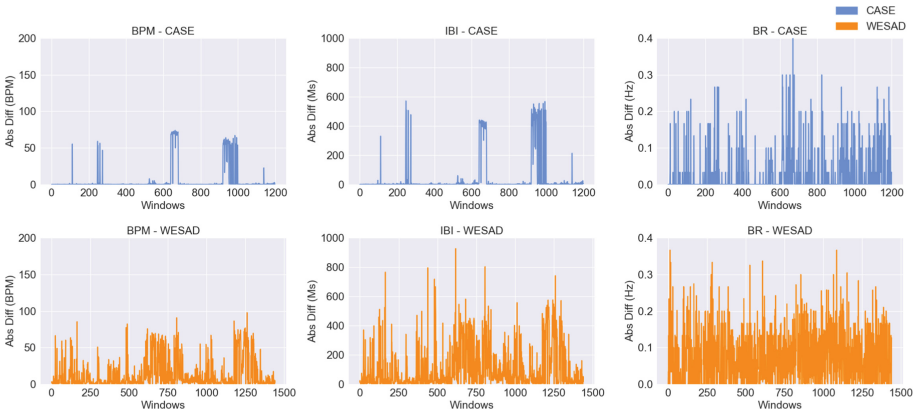


Fig. 2. Absolute difference between ECG and PPG features: BPM, IBI and BR, for CASE and WESAD. High variance demonstrates unreliable feature computation in one of the signals due to signal noise or sensor differences.

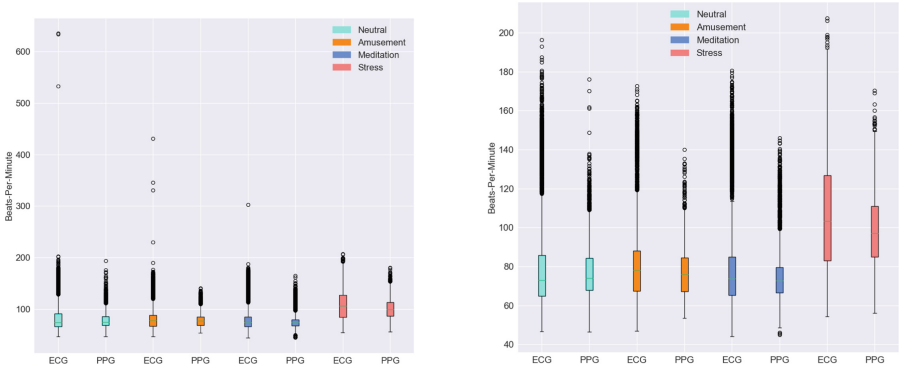
In CASE, beats per-minute (BPM) and inter-beat interval (IBI) contain a small variance with substantial spikes relative to the average. These variances occur in isolated data segments and are likely caused by electrode disconnection, movement, or subject-specific factors, visible in Fig. 2 at approximately window numbers 250, 550 and 900. Such occurrences may benefit from additional signal processing to reduce noise and improve feature computation accuracy.

Interestingly, the breathing rate (BR) feature exhibits a high deviation between signals in both datasets. This deviation indicates that at least one of the signals is unreliably computing BR, likely due to the wrist and finger placement of the PPG sensors.

A low degree of inter-affective state feature variance was identified between WESAD baseline, amusement, and meditation states for all features indicating these states are difficult to distinguish as depicted in Fig. 3. Statistically similar features negatively impact automated classification, as the classifier struggles to differentiate between the classes. This impact is demonstrated by the reduced performance in multi-class classifications (58%–69%) as compared to the OVR performance depicted by the ROC curves (ROC Area: 0.70–0.95) in Fig. 5. This performance increase validates the utility of OVR classifications when classifying affective states that are difficult to differentiate due to statistical similarities.

4.2 HRV Feature Importance

Analysing the SHAP values per feature indicates that BPM, IBI, and BR have the most significant impact on classification for both signals, as demonstrated in Fig. 4. The remaining features exhibit inconsistent influence between the signals. Most notably, standard deviation 1 divided by standard deviation 2 (SD1/SD2) and room-mean-square of successive differences (RMSSD) exhibit higher impact



(a) Outliers Included (b) Outliers Excluded

Fig. 3. Inter-signal and inter-affective state variance for BPM in WESAD, including and excluding outliers. Note in (a) the presence of outliers with a BPM of over 300 occurring in ECG indicating abnormal signal activity. Additionally, in (a, b), a visible overlap in neutral, amusement, and meditation occur, demonstrating the degree of similarity in these states.

in PPG as opposed to ECG. This demonstrates the need for assessing feature importance on a per-signal basis to identify which features are most informative for use in tailored signal-specific classification approaches.

Certain features demonstrate varying impacts across affective states, indicating the presence of affect-specific features. For example, BPM and IBI exhibit high impacts on the class "stress", indicating their suitability for stress detection approaches. Assessing feature importance per-affective state provides an informative analysis of feature utility for affect-specific approaches.

The high feature importance of BPM for "stress" is due to statistical distinction to the other affective states in the inter-affective state feature variance, as depicted in Fig. 3. This demonstrates the benefit of assessing inter-affective state feature variance and feature importance to gain insights to aid the creation of affect-specific approaches.

4.3 Automated Affective State Classification Variance

Finally, the selected classifier is trained on the initial 80% of data and classifies the hold-out set to assess expected performance in real-life classifications. The ExtraTrees classifier (ET) was selected as the most performant classifier from the model selection, where it was trained on 80% of the training data and evaluated on the remaining 20%. Notably, ET exhibits an increased performance when evaluated on the hold-out set as it was trained on all available training data. The full model comparison and ET hold-out performance is depicted in Fig. 1. Interestingly, the classifier performance variance between ECG and PPG is similar to the degree of the inter-signal feature variance identified per dataset.

Feature	Abbrv.
Beats Per Minute	BPM
Interbeat Interval	IBI
Std dev. of RR Intervals	SDNN
Std dev. of successive diff.	SDSD
RMSE of successive diff.	RMSSD
Proportion of diff. < 20ms	pNN20
Proportion of diff. < 50ms	pNN50
Median absolute dev. of RR intervals	MAD
Estimated breathing rate	BR
Poincare analysis	SD1, SD2, S, SD1/SD2

Dev: Deviation, Diff: Difference

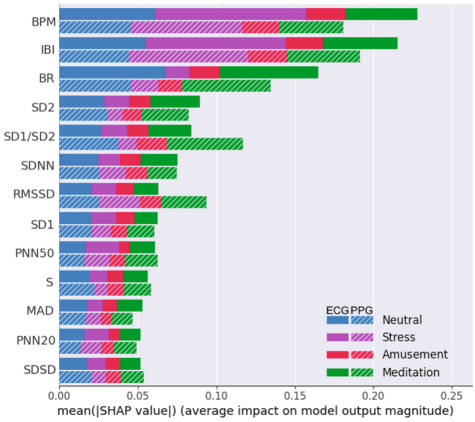
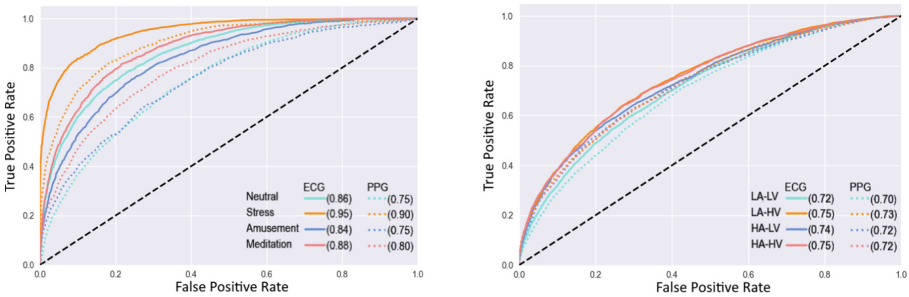


Fig. 4. Shapley **A**dditive ex**P**lanations (SHAP) Feature Importance from ExtraTrees classifying WESAD signals.

In contrast with the state-of-the-art [11,31], the performance achieved is lower for ECG and PPG; however, this work focuses on the analysis and understanding of variances between the signals for affective analysis rather than achieving high classification accuracy. Analysing the ROC curves from ET demonstrates the true and false positive rates per signal for each affective state, see Fig. 5. On average, ECG demonstrates increased capabilities for affective classification by achieving a higher ROC area than PPG, varying with a range of 0.02–0.11. The increased performance via OVR demonstrates the benefit of identifying and overcoming the effects of similar affective states to achieve greater classification performance.



(a) WESAD (b) CASE

Fig. 5. ROC Curves from ExtraTrees representing the OVR classification variance between ECG and PPG

5 Conclusions

The inter-signal classification performance disparity mirrors the degree of feature variance between signals from both datasets. Specifically, WESAD exhibited a high feature variance, which explains the higher disparity in classification accuracy and ROC area per signal. Conversely, a lower inter-signal feature variance and a lower disparity in the performance measures occurred for CASE. This demonstrates the utility of inter-signal feature variance in identifying inconsistent computations of features stemming from sensor differences or abnormal signal activity, which negatively impact classification performance. These occurrences are likely to be more frequent in the ambulatory analysis due to motion artefacts and uncontrolled usage of wearables.

Furthermore, inter-affective state feature variance enables the identification of affective states that contain a similar distribution of features, which causes classification confusion. To counter this, the similar states are aggregated into an OVR classification problem, leading to increased performance, demonstrated by the ROC area per affective state.

Feature importance identifies BPM, IBI, and BR as the most impactful features for affective classification across ECG and PPG. Notably, the remaining features exhibit inconsistent impacts, specifically SD1/SD2 and RMSSD, which demonstrate a greater impact in PPG, warranting the exploration of signal-specific features. Analysing statistical measures to understand the inter-affective state feature variance indicates that certain features provide a greater degree of affect-specific information beneficial for tailored applications.

This work contributes an empirical analysis of data variances in ECG and PPG acquired using wearables and the impact on affective state classification. Therefore, enabling practitioners to make informed decisions when creating ML pipelines for affective state classification. The code-base will be made open access on Github (https://github.com/ZacDair/Emo-Phys-Eval), enabling automated feature variance analysis from each of these perspectives in a combined manner, regardless of data acquisition methods. While this approach analyses handcrafted features, it can also be utilised with automatically extracted features.

Future work will expand the analysis by utilising additional datasets to provide greater insights into the variances stemming from data collection devices, affective states, and population differences. In addition, an extended analysis will be conducted using additional features and methods to further inform the development of ML pipelines for affective state detection.

References

1. Agrafioti, F., et al.: ECG pattern analysis for emotion detection. IEEE Trans. Affect. Comput. **3**(1), 102–115 (2012). https://doi.org/10.1109/T-AFFC.2011.28
2. Birkett, M.A.: The trier social stress test protocol for inducing psychological stress. J. Vis. Exp. (2011). https://doi.org/10.3791/3238
3. Can, Y.S., Chalabianloo, N., Ekiz, D., Ersoy, C.: Continuous stress detection using wearable sensors in real life: algorithmic programming contest case study. Sensors **19**(8), 1849 (2019). https://doi.org/10.3390/s19081849

4. Cheema, A., Singh, M.: Psychological stress detection using phonocardiography signal: an empirical mode decomposition approach. Biomed. Signal Process. Control **49**, 493–505 (2019). https://doi.org/10.1016/j.bspc.2018.12.028
5. Cinaz, B., Arnrich, B., La Marca, R., Tröster, G.: Monitoring of mental workload levels during an everyday life office-work scenario. Pers. Ubiquit. Comput. **17** (2013). https://doi.org/10.1007/s00779-011-0466-1
6. Dissanayake, T., Rajapaksha, Y., Ragel, R., Nawinne, I.: An ensemble learning approach for electrocardiogram sensor based human emotion recognition. Sensors **19**(20) (2019). https://doi.org/10.3390/s19204495
7. Dockray, S., O'Neill, S., Jump, O.: Measuring the psychobiological correlates of daily experience in adolescents. J. Res. Adolesc. **29**(3), 595–612 (2019). https://doi.org/10.1111/jora.12473
8. Dzedzickis, A., Kaklauskas, A., Bucinskas, V.: Human emotion recognition: review of sensors and methods. Sensors **20**(3) (2020). https://doi.org/10.3390/s20030592
9. Filippini, C., et al.: Automated affective computing based on bio-signals analysis and deep learning approach. Sensors **22**(5) (2022). https://doi.org/10.3390/s22051789
10. Goshvarpour, A., Goshvarpour, A.: Poincaré's section analysis for PPG-based automatic emotion recognition. Chaos Solitons Fractals **114**, 400–407 (2018). https://doi.org/10.1016/j.chaos.2018.07.035
11. Goshvarpour, A., Goshvarpour, A.: Evaluation of novel entropy-based complex wavelet sub-bands measures of PPG in an emotion recognition system. J. Med. Biol. Eng. **40**(3), 451–461 (2020). https://doi.org/10.1007/s40846-020-00526-7
12. Hsu, Y.L., Wang, J.S., Chiang, W.C., Hung, C.H.: Automatic ECG-based emotion recognition in music listening. IEEE Trans. Affect. Comput. **11**(1), 85–99 (2020). https://doi.org/10.1109/TAFFC.2017.2781732
13. Jing, C., Liu, G., Hao, M.: The research on emotion recognition from ECG signal. In: 2009 International Conference on Information Technology and Computer Science, vol. 1, pp. 497–500 (2009). https://doi.org/10.1109/ITCS.2009.108
14. Kalra, P., Sharma, V.: Mental stress assessment using PPG signal a deep neural network approach. IETE J. Res. 1–7 (2020). https://doi.org/10.1080/03772063.2020.1844068
15. Katsigiannis, S., Ramzan, N.: Dreamer: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE J. Biomed. Health inform. (2018). https://doi.org/10.1109/JBHI.2017.2688239
16. Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE PAMI **30**(12), 2067–2083 (2008). https://doi.org/10.1109/TPAMI.2008.26
17. Koelstra, S., et al.: Deap: a database for emotion analysis; using physiological signals. IEEE Trans. Affect. Comput. **3**(1), 18–31 (2012). https://doi.org/10.1109/T-AFFC.2011.15
18. Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M.A., Kraaij, W.: The SWELL knowledge work dataset for stress and user modeling research. In: ICMI, pp. 291–298. ACM (2014). https://doi.org/10.1145/2663204.2663257
19. Lee, M.S., Lee, Y.K., Pae, D.S., Lim, M.T., Kim, D.W., Kang, T.K.: Fast emotion recognition based on single pulse PPG signal with convolutional neural network. Appl. Sci. **9**(16) (2019). https://doi.org/10.3390/app9163355
20. Lin, S., et al.: A review of emotion recognition using physiological signals. Sensors **18**(7), 2074 (2018). https://doi.org/10.3390/s18072074

21. Lisowska, A., Wilk, S., Peleg, M.: Catching patient's attention at the right time to help them undergo behavioural change: stress classification experiment from blood volume pulse. In: Tucker, A., Henriques Abreu, P., Cardoso, J., Pereira Rodrigues, P., Riaño, D. (eds.) AIME 2021. LNCS (LNAI), vol. 12721, pp. 72–82. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77211-6_8
22. Mahdiani, S., et al.: Is 50 Hz high enough ECG sampling frequency for accurate HRV analysis? In: EMBC, pp. 5948–5951 (2015). https://doi.org/10.1109/EMBC.2015.7319746
23. Molnar, C.: Interpretable Machine Learning. Lulu.com (2022). https://christophm.github.io/interpretable-ml-book/
24. Mukherjee, N., et al.: Real-time mental stress detection technique using neural networks towards a wearable health monitor. Meas. Sci. Technol. **33**(4), 044003 (2022). https://doi.org/10.1088/1361-6501/ac3aae
25. Nabeel, S., et al.: A comparison of manual electrocardiographic interval and waveform analysis in lead 1 of 12-lead ECG and apple watch ECG: a validation study. Cardiovasc. Digit. Health J. (2020). https://doi.org/10.1016/j.cvdhj.2020.07.002
26. Nardelli, M., Valenza, G., Greco, A., Lanata, A., Scilingo, P.: Recognizing emotions induced by affective sounds through heart rate variability. IEEE Trans. Affect. Comput. **6**(4), 385–394 (2015). https://doi.org/10.1109/TAFFC.2015.2432810
27. Picard, R.W.: Affective computing: challenges. Int. J. Hum Comput Stud. **59**(1), 55–64 (2003). https://doi.org/10.1016/S1071-5819(03)00052-1
28. Porges, S.W.: The polyvagal theory: new insights into adaptive reactions of the autonomic nervous system. Clevel. Clin. J. Med. **76**(Suppl. 2), S86–S90 (2009). https://doi.org/10.3949/ccjm.76.s2.17
29. da José, S., Luz, E., et al.: ECG-based heartbeat classification for arrhythmia detection: a survey. CMPB **127**, 144–164 (2016). https://doi.org/10.1016/j.cmpb.2015.12.008
30. Porges, S.W., et al.: Vagal tone and the physiological regulation of emotion. Monogr. Soc. Res. Child Dev. **59**(2–3), 167–186 (1994)
31. Sarkar, P., Etemad, A.: Self-supervised ECG representation learning for emotion recognition. IEEE Trans. Affect. Comput. (2021). https://doi.org/10.1109/TAFFC.2020.3014842
32. Sayed Ismail, S.N.M., Ab. Aziz, N.A., Ibrahim, S.Z.: A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG). J. King Saud Univ. - Comput. Inf. Sci. **34**(6, Part B), 3539–3558 (2022). https://doi.org/10.1016/j.jksuci.2022.04.012
33. Schmidt, P., et al.: Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: ICMI 20. ACM (2018). https://doi.org/10.1145/3242969.3242985
34. Sharma, K., et al.: A dataset of continuous affect annotations and physiological signals for emotion analysis (2018). https://doi.org/10.48550/ARXIV.1812.02782
35. van Gent, P., et al.: HeartPy: a novel heart rate algorithm for the analysis of noisy signals. Transp. Res. F: Traffic Psychol. Behav. **66**, 368–378 (2019). https://doi.org/10.1016/j.trf.2019.09.015

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.