

Using Electrocardiogram and Photoplethysmography Data to Assess Human Emotions

Aakash Soni^{*1,2}, Ian Daly^{*2}, Keming Zhou^{*1}, Huosheng Hu^{†2}, Dongbing Gu^{‡2}, Hongtao Zhang^{*1}

¹August International Limited, Hoddesdon, Hertfordshire, EN11 0EE, UK

²School of Computer Science and Electronic Engineering, University of Essex, Colchester, Essex, CO4 3SQ, UK

*aakash@augustint.com, *i.daly@essex.ac.uk, *kmzhou@augustint.com, †hhu@essex.ac.uk,

‡dgu@essex.ac.uk, ‡hzhang@augustint.com

Abstract—Human emotions are linked to mental well-being and physical health, making emotion recognition via physiological signals increasingly important. Although recent studies show promise, the combined use of electrocardiogram (ECG) and photoplethysmography (PPG) data for emotion assessment remains underexplored. This study examines the feasibility of using joint ECG and PPG signals for emotion evaluation within the Affective Dimensional Model (ADM) framework. Morphological features extracted from these signals are used to classify felt arousal and valence with Support Vector Machines (SVM) and Neural Networks (NN). On a per-participant basis, SVM achieved average valence and arousal accuracies of 72.69% ($p < 0.05$) and 72.30% ($p < 0.05$), while NN reached 72.48% ($p < 0.05$) and 73.01% ($p < 0.05$). The findings suggest that the morphological features of ECG and PPG encode emotion-dependent information, enabling accurate prediction of emotional states.

Index Terms—affective computing, physiological sensors, signal processing, machine learning

I. INTRODUCTION

Recent advances in AI and semiconductor technology have enabled the development of powerful computing devices for eHealth and IoT-enabled wearables. Affective computing—a field combining psychology, biomedical engineering, and computer science—aims to enhance decision-making through emotional intelligence, as first introduced by Rosalind W. Picard [1]. Physiological sensors are vital in these systems, using signals like electrocardiogram (ECG), photoplethysmography (PPG), electroencephalography (EEG), electromyography (EMG), galvanic skin response (GSR) or electrodermal activity (EDA), skin temperature (SKT), and respiration patterns (RSP) [2]. Emotion assessment mainly follows two models: Discrete Emotional Models (DEM), based on Darwin’s work [3], which classifies emotions like joy, sadness, and anger [4]; and Affective Dimensional Models (ADM), rooted in Wundt’s research [5] and expanded by Russell [6], which uses arousal and valence to map emotional states.

This study analyzes short-duration ECG and PPG signals to extract morphological features, applying classifiers on a per-participant due to signal variability. To our knowledge, no prior work has combined ECG and PPG morphology within the

ADM framework. Our contributions are: (1) combining ECG-PPG features with ADM; (2) using a continuously annotated emotion dataset; (3) ensuring computational efficiency by focusing on signal shape; and (4) evaluating classification performance per participant using SVM and NN with optimization techniques.

II. RELATED WORK

Various techniques exist for emotion recognition via speech, facial expressions, text, and physiological signals. This review focuses specifically on physiological signal-based classification using ADM and DEM approaches.

Interbeat Interval (IBI) and Heart Rate Variability (HRV) are key features derived from heartbeats. HRV has been used to classify emotions like happiness, disgust, fear, sadness, and neutral emotions using DEM, achieving up to 69.75% accuracy with KNN and 67.81% with LDA [7]. Another study used EEG, ECG, PPG, GSR, and RSP signals with 15 features, reaching an accuracy of up to 86% for joy, 91% for happiness, 79% for fear, 87% for anger, 76% for despair, and 94% for sadness with a NN classifier and DEM approach [8].

Comparative studies show ECG and PPG perform differently in emotion classification. One found ECG achieved 58.81% (valence) and 68.75% (arousal), while PPG scored 64.94% and 64.61%, respectively, using an SVM and ADM approach [10]. A deep learning approach reported slightly lower ECG and PPG accuracies, with PPG dropping to 34.63% for arousal [11]. Another study compared HRV (from ECG) and PRV (from PPG), reporting up to 62% valence and 65% arousal accuracy using an SVM and ADM approach [12], suggesting PPG is a viable alternative for emotion assessment.

III. METHODS

A. ECG/PPG Signals

ECG signals are measured as the voltage difference between two electrodes—one active and one reference—placed on the heart or body, reflecting the time-varying amplitude of the electrical activity from the heart. In contrast, the PPG signal is obtained by an infrared LED sensor that emits light to penetrate the skin and blood vessels, with the reflected light capturing blood flow and oxygenation changes during each heartbeat. Both signals are typically noisy and require filtering to extract meaningful information.

Funding: The authors thank Knowledge Transfer Partnership (KTP) and August International Limited for supporting this research under the grant, KTP11528.

B. Dataset

The CASE dataset comprises data from 30 participants (15 males and 15 females) [13]. Unlike other public datasets that focus mostly on EEG signals [9], CASE is unique in including both ECG and PPG data with real-time valence-arousal annotations. In CASE, four emotional states (amusing, boring, relaxing, and scary) were elicited using eight videos (two per state), each shown once to all participants. Responses were collected via a Joystick-based Emotion Reporting Interface (JERI) that provided simultaneous arousal and valence annotations. Additionally, the dataset records six physiological signals (ECG, PPG, GSR, RSP, EMG, and SKT) at a sampling rate of 1000 Hz, while annotations were captured at 20 Hz.

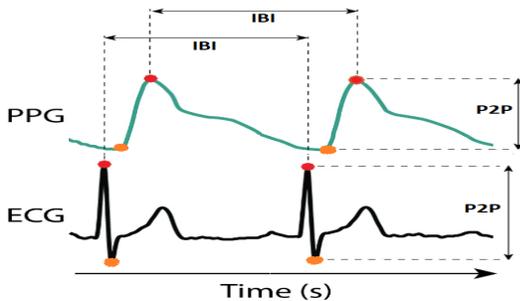


Fig. 1. Signal Morphology Features

C. Approach

In this study, we examine the impact of emotions on the cardiac cycle by extracting four features—peak amplitude, valley amplitude, peak-to-peak amplitude, and IBI—from short signal segments. These features were selected for their simplicity in identifying local maxima and minima, from which the features can be derived. A 3-second window was chosen because 1- or 2-second windows often contain only a single peak, insufficient for IBI calculation. The 3-second window consistently provides at least two peaks in both ECG and PPG signals, enabling accurate extraction of the required features. As shown in Figure 1, peaks and valleys were identified as local maxima and minima to derive peak-to-peak amplitude (y-axis difference) and IBI (x-axis difference).

For per-participant training, data from eight videos are divided in 3-second length. Therefore, a total of 418 trials are obtained for each participant. For each 3-second segment, ECG and PPG signals are extracted and matched with the corresponding annotations. Since annotations are collected every 50 milliseconds (20 Hz), each 3-second window corresponds to 60 annotations. The mean of these 60 annotations is calculated and matched with the 3-second signal.

The dataset used in this study employs four emotions from each quadrant of the circumplex model of affect. This model maps emotions in a 2D space using valence (pleasantness) and arousal (intensity) [6]. It forms four quadrants: high arousal and positive valence (e.g., amusing), high arousal and negative valence (e.g., scary), low arousal and negative valence (e.g., boring), and low arousal and positive valence (e.g., relaxed).

This model helps represent emotions as continuous states rather than fixed categories.

D. Signal Processing

Both ECG and PPG signals can drift over time due to various factors. To address this, baseline correction was performed by subtracting the mean value from a 120-second baseline period (recorded during a blue-screen video shown before the stimulus) from the subsequent signal. A band-pass filter was then applied to remove unwanted noise: for ECG, the low-pass cut-off was 40 Hz (for PPG, 4 Hz) and both used a high-pass cut-off of 0.5 Hz.

For the ECG signal, the Pan-Tompkins algorithm [14] was implemented to detect the peaks. For the PPG signal, peaks were identified by finding local maxima within a 500-sample window, which was determined empirically. Some PPG signals were found to contain outliers, and it was necessary to remove these before further processing. A linear interpolation method was applied, where outliers were replaced with values obtained by averaging the neighboring non-outlier samples. This approach prevents the loss of samples in the PPG signal.

E. Classifiers

For per-participant classification, a total of 418 trials were obtained by extracting features from 3-second windows. These were subsequently used to train SVM and NN classifiers. Two separate models were trained for arousal and valence, using MATLAB-R2020b. Arousal and valence annotations in the dataset range from 0.5 to 9.5. For binary classification, a threshold value of 5 was applied to create two distinct classes: any annotation value less than or equal to 5 was classified as Low Arousal (LA)/Negative Valence (NV), while value greater than 5 was classified as High Arousal (HA)/Positive Valence (PV). The dataset is balanced, ensuring an equal number of trials for each class. For SVM classification, 80% of the trials were used for training and 20% for testing. For NN classification, 70% of the trials were used for training, 15% for validation, and 15% for testing. A 5-fold cross validation is employed in this study.

The optimization techniques explore various hyperparameter combinations using an optimization strategy designed to minimize the model's classification error. For SVM classification, Bayesian optimization [15] is used for hyperparameter optimization. This process returns a model with optimized hyperparameters, including box constraint, kernel scale, kernel function, and polynomial order. For NN classification, the ADAM (adaptive moment estimation) optimizer [16] is used for hyperparameter optimization. This process returns a model with optimized hyperparameters, including learning rate, gradient decay factor, squared gradient decay factor, and epsilon (a small constant to ensure numerical stability).

The NN consists of a feature input layer, a fully connected layer with 50 neurons, batch normalization, and a ReLU activation layer. This is followed by another fully connected layer matching the number of target classes, a SoftMax layer for probability output, and a classification layer. The model

TABLE I
SVM TRAINING RESULTS

Participant No.	Valence				Arousal			
	Test Acc. (%)	p-Value	F1-Score (%)		Test Acc. (%)	p-Value	F1-Score (%)	
			PV	NV			HA	LA
1	75.68	$p < 0.001$	76.90	74.30	87.50	$p < 0.001$	88.20	86.70
2	77.23	$p < 0.001$	78.90	75.40	61.77	$p < 0.05$	59.40	63.90
3	80.00	$p < 0.001$	77.80	81.80	67.14	$p < 0.01$	71.60	61.01
4	73.43	$p < 0.001$	72.13	74.63	76.00	$p < 0.001$	76.92	75.00
5	79.17	$p < 0.001$	78.87	79.45	64.29	$p < 0.01$	71.91	50.98
6	71.73	$p < 0.01$	72.34	71.11	72.22	$p < 0.001$	68.75	75.00
7	62.50	$p < 0.05$	55.00	67.80	57.14	$p > 0.05$	58.62	55.56
8	66.67	$p < 0.05$	57.70	72.50	78.57	$p < 0.001$	80.00	76.92
9	67.19	$p < 0.05$	61.82	71.23	67.19	$p < 0.01$	60.38	72.00
10	63.64	$p > 0.05$	55.56	69.23	64.29	$p < 0.05$	42.30	50.00
11	82.00	$p < 0.001$	82.35	81.63	74.29	$p < 0.001$	75.00	73.53
12	69.23	$p < 0.001$	66.67	71.43	68.18	$p < 0.01$	64.41	71.23
13	77.08	$p < 0.001$	78.43	75.56	73.21	$p < 0.001$	74.58	71.70
14	76.47	$p < 0.01$	75.00	77.78	58.97	$p < 0.05$	60.00	57.90
15	63.51	$p < 0.01$	67.47	58.46	62.50	$p < 0.05$	66.67	57.14
16	61.76	$p < 0.05$	65.79	56.67	77.50	$p < 0.001$	74.29	80.00
17	71.62	$p < 0.001$	72.00	71.23	79.73	$p < 0.001$	76.92	81.93
18	71.67	$p < 0.001$	70.17	73.02	77.94	$p < 0.001$	78.26	77.61
19	70.27	$p < 0.001$	71.05	69.44	83.33	$p < 0.001$	81.48	84.84
20	83.33	$p < 0.001$	84.00	82.61	80.49	$p < 0.001$	80.00	80.95
21	68.96	$p < 0.01$	65.38	71.87	77.42	$p < 0.001$	76.67	78.12
22	70.97	$p < 0.001$	68.97	72.73	72.22	$p < 0.001$	64.29	77.27
23	71.21	$p < 0.001$	70.77	71.64	77.14	$p < 0.001$	75.00	78.95
24	76.67	$p < 0.001$	75.00	78.12	70.00	$p < 0.01$	63.41	74.58
25	77.94	$p < 0.001$	76.92	78.87	75.00	$p < 0.001$	73.68	76.19
26	72.22	$p < 0.001$	75.41	68.08	72.58	$p < 0.001$	72.13	73.02
27	68.06	$p < 0.001$	72.94	61.02	72.97	$p < 0.001$	74.36	71.43
28	62.50	$p < 0.05$	64.41	60.38	71.25	$p < 0.001$	72.29	70.13
29	98.08	$p < 0.001$	98.11	98.04	82.26	$p < 0.001$	81.97	82.54
30	69.56	$p < 0.01$	68.18	70.83	66.00	$p < 0.01$	66.67	65.31
Test Accuracy	72.69%				72.30%			

is trained with a mini-batch size of 10, with data shuffled in each epoch and validated using the validation set.

IV. RESULTS

Table I summarizes the results obtained from the SVM classifier. The average test accuracy achieved by the SVM classifier is 72.69% for valence and 72.30% for arousal. Similarly, Table II provides the results of the NN classifier, with average test accuracies of 72.48% for valence and 73.01% for arousal. In addition to the accuracy of the classification, the statistical significance (p -value) and the F1 score for both classes were calculated. The F1 score measures a model's accuracy on a given dataset by calculating the harmonic mean of precision and recall. A higher F1 score indicates strong performance, while a lower score suggests lower precision and recall. Both classifiers, optimized using relevant algorithms, produced promising results. Most results were statistically significant ($p \leq 0.05$), except for a few participants in both classifiers. Overall, the statistical evaluation validates the performance of both classifiers, with only a few exceptions.

In addition to the statistical significance test and average accuracies, the standard deviation of the performance of each model over trials is reported. This is 7.48% for valence and 7.32% for arousal using the SVM classifier and 7.62% for valence and 6.06% for arousal using the NN classifier. This level of variability suggests moderate fluctuations in

the model performance across different participants. This is due to differences in signal morphology between trials and participants (i.e. Non-stationarity of the signals, noise etc.), as well as differences in participants emotional responses (different people emotionally react differently) [17].

Per-participant based emotion recognition offers several advantages, particularly in improving the accuracy and reliability of emotion classification. By training and testing models on data from the same individual, these systems can better capture the unique physiological and emotional response patterns of each person. This personalized approach reduces inter-participant variability and leads to more consistent performance. Moreover, participant-dependent models are well-suited for integration into wearable devices, where continuous monitoring and adaptation to an individual's emotional state are essential for delivering meaningful, real-time insights.

V. CONCLUSION

This study introduces a method for assessing human emotions using short physiological signal segments, specifically ECG and PPG data. The approach comprises three steps: signal processing, feature extraction, and classification. Emotional states were evaluated using 3-second segments of ECG and PPG data, showing that the morphological features effectively capture emotion-related changes. Classifiers were individually implemented for 30 participants, accounting for

TABLE II
NN TRAINING RESULTS

Participant No.	Valence				Arousal			
	Test Acc. (%)	p-Value	F1-Score (%)	F1-Score (%)	Test Acc. (%)	p-Value	F1-Score (%)	F1-Score (%)
			PV	NV			HA	LA
1	72.22	$p < 0.001$	73.68	70.59	71.43	$p < 0.05$	71.42	71.42
2	75.00	$p < 0.001$	77.19	72.34	65.45	$p < 0.01$	66.67	66.67
3	76.19	$p < 0.001$	78.26	73.68	66.67	$p < 0.01$	65.22	68.00
4	80.95	$p < 0.001$	80.00	81.82	76.32	$p < 0.001$	74.29	78.05
5	68.00	$p < 0.01$	70.37	65.22	69.05	$p < 0.01$	66.67	71.11
6	70.00	$p < 0.01$	70.00	70.00	70.83	$p < 0.01$	69.56	72.00
7	61.90	$p < 0.05$	66.67	64.86	63.83	$p < 0.05$	65.22	57.89
8	64.29	$p < 0.05$	57.14	69.39	82.50	$p < 0.001$	82.05	82.93
9	67.39	$p < 0.01$	65.12	69.39	71.43	$p < 0.01$	72.73	70.00
10	62.50	$p > 0.5$	66.67	57.14	75.00	$p < 0.001$	73.68	76.19
11	85.29	$p < 0.001$	84.85	85.71	73.08	$p < 0.001$	74.04	72.00
12	72.41	$p < 0.001$	65.22	77.14	68.00	$p < 0.01$	69.23	66.67
13	75.00	$p < 0.001$	72.22	77.27	75.00	$p < 0.01$	72.73	76.93
14	73.08	$p < 0.01$	75.86	69.56	58.62	$p < 0.05$	60.00	57.14
15	70.00	$p < 0.001$	66.67	72.73	69.05	$p < 0.01$	68.29	69.77
16	63.04	$p < 0.05$	69.09	54.05	87.50	$p < 0.001$	88.89	85.71
17	73.21	$p < 0.001$	75.41	70.59	73.21	$p < 0.001$	68.08	76.92
18	68.52	$p < 0.01$	66.67	70.17	68.97	$p < 0.01$	65.38	71.87
19	74.00	$p < 0.001$	64.70	79.36	81.82	$p < 0.001$	81.82	81.82
20	73.33	$p < 0.01$	75.00	71.43	81.25	$p < 0.001$	79.07	83.02
21	72.73	$p < 0.01$	73.91	71.43	79.17	$p < 0.001$	78.26	80.00
22	76.19	$p < 0.001$	76.19	76.19	69.23	$p < 0.01$	60.00	75.00
23	73.08	$p < 0.001$	73.08	73.08	75.86	$p < 0.001$	74.07	77.42
24	75.00	$p < 0.001$	75.56	74.42	75.00	$p < 0.01$	72.73	76.92
25	70.37	$p < 0.01$	71.43	69.23	76.67	$p < 0.01$	74.07	78.79
26	79.17	$p < 0.001$	81.48	76.19	81.03	$p < 0.001$	80.70	81.36
27	73.21	$p < 0.001$	75.41	70.59	71.15	$p < 0.001$	73.68	68.08
28	58.33	$p > 0.05$	47.37	65.52	68.52	$p < 0.01$	66.67	70.17
29	100.00	$p < 0.001$	100.00	100.00	72.50	$p < 0.01$	68.57	75.56
30	70.00	$p < 0.05$	70.97	68.97	72.22	$p < 0.01$	68.75	75.00
Test Accuracy	72.48%				73.01%			

variability in resting heart rates and signal patterns. The results indicate that the four extracted features reliably determine emotional states, with statistical measures (including p -values) confirming the method's significance.

REFERENCES

- [1] R. Picard, *Affective computing*, 1st ed. Cambridge, MA, USA: MIT Press, 1997.
- [2] A. Dzedzickis, A. Kaklauskas and V. Bucinskas, "Human Emotion Recognition: Review of Sensors and Methods", *Sensors*, vol. 20, no. 3, p. 592, 2020. Available: 10.3390/s20030592.
- [3] C. Darwin, *The Expression of the Emotions in Man and Animals*. London, UK: John Murray, 1872.
- [4] P. Ekman, "Darwin's contributions to our understanding of emotional expressions", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3449-3451, 2009. Available: 10.1098/rstb.2009.0189.
- [5] W. Wundt, *Grundriss der Psychologie*. Leipzig, Germany: Engelmann, 1896.
- [6] J. Russell, "A circumplex model of affect.", *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161-1178, 1980. Available: 10.1037/h0077714.
- [7] M. Murugappan, S. Murugappan and B. Zheng, "Frequency Band Analysis of Electrocardiogram (ECG) Signals for Human Emotional State Classification Using Discrete Wavelet Transform (DWT)", *Journal of Physical Therapy Science*, vol. 25, no. 7, pp. 753-759, 2013. Available: 10.1589/jpts.25.753.
- [8] G. Yoo, S. Seo, S. Hong and H. Kim, "Emotion extraction based on multi bio-signal using back-propagation neural network", *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4925-4937, 2016. Available: 10.1007/s11042-016-4213-5.
- [9] P. Samal and MF. Hashmi, "Role of machine learning and deep learning techniques in EEG-based BCI emotion recognition system: a review", *Artif Intell Rev*, vol. 57, no. 50, 2024. Available: 10.1007/s10462-023-10690-2
- [10] S. Ismail, N. Aziz, S. Ibrahim, "A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG)", *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 3539-3558, 2022. Available: 10.1016/j.jksuci.2022.04.012
- [11] N. Aziz, K. Tawsif, S. Ismail, M. Hasnul, K. Aziz, S. Ibrahim, A. Aziz, J. Raja, "Asian Affective and Emotional State (A2ES) Dataset of ECG and PPG for Affective Computing Research", *Algorithms*, vol. 16, no. 3, pp. 130, 2023. Available: 10.3390/a16030130
- [12] S. Rinella, S. Massimino, P.G. Fallica, A. Giacobbe, N. Donato, M. Coco, G. Neri, R. Parenti, V. Perciavalle, S. Conoci, "Emotion Recognition: Photoplethysmography and Electrocardiography in Comparison", *Biosensors*, vol. 12, no. 10, pp. 811, 2022. Available: 10.3390/bios12100811
- [13] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, F. Schwenker, A dataset of continuous affect annotations and physiological signals for emotion analysis. vol. 6, *Nature: Scientific Data*, 2019. [Dataset]. Available: 10.1038/s41597-019-0209-0.
- [14] J. Pan, W.J. Tompkins, "A Real-Time QRS Detection Algorithm", *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230-236, 1985. Available: 10.1109/TBME.1985.325532.
- [15] A. Bull, "Convergence rates of efficient global optimization algorithms", *arXiv - stat.ML*, 2011. Available: arXiv:1101.3501.
- [16] D.P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", *arXiv - cs.LG*, 2017. Available: arXiv:1412.6980.
- [17] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review", *Biological Psychology*, vol. 84, no. 3, pp. 394-421, 2010. Available: https://doi.org/10.1016/j.biopsycho.2010.03.010