2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)

Affective Computing as a Tool for Understanding Emotion Dynamics from Physiology: A Predictive Modeling Study of Arousal and Valence

Tomás A. D'Amelio
INFINA – UBA
CONICET
Buenos Aires, Argentina
tomasdamelio@psi.uba.ar

Nicolás M. Bruno
INFINA – UBA
CONICET
Buenos Aires, Argentina
nbruno@psi.uba.ar

Leandro A. Bugnon
sinc(i) - UNL
CONICET
Santa Fe, Argentina
lbugnon@sinc.unl.edu.ar

Federico Zamberlan
INFINA – UBA
Tilburg University
Tilburg, Netherlands
fxzamberlan@gmail.com

Enzo Tagliazucchi
INFINA – UBA – CONICET
BrainLat – UAI
Buenos Aires, Argentina
tagliazucchi.enzo@googlemail.com

***Abstract*— Affective computing has traditionally relied on predictive models that use summary annotations to understand emotions, an approach that often fails to capture the continuous nature of emotions. In this paper, we explore the previously unexamined possibility of understanding the temporal dynamics of emotions using the Continuously Annotated Signals of Emotion (CASE) dataset during the Emotion Physiology and Experience Collaboration (EPiC) 2023 competition. We present the first performance benchmark for predictive models using continuous annotations on this dataset, in which we achieve significantly better results than baseline models for specific scenarios. Our contributions include the development and comparison of predictive models for different affective dimensions, demonstrating that arousal models outperform valence models, a finding consistent with existing affective science literature. In addition, our analysis shows that predictions incorporating features from past data are more informative than those based on future data, suggesting that physiological activity precedes affective experience and subsequent annotation. These findings contribute to a deeper understanding of the temporal dynamics of emotion and have broad implications for both affective computing and affective science, highlighting the potential of this interdisciplinary approach.**

***Index Terms*— affective computing, affective science, temporal dynamics**

I. INTRODUCTION

Affective computing, a rapidly expanding field, combines various disciplines in its quest to develop systems and algorithms that recognize, interpret, and simulate human emotions [1]. These algorithms often use a variety of signals - text, audio, video, and physiological data - to infer affective states. Historically, a common approach within predictive models of affect has been to use summary annotation, where physiological data is collected over a period of time (e.g., while individuals are exposed to certain stimuli, such as watching movie clips) and then annotated post hoc (e.g., assigning a valence rating between 1 and 9 after the clip has ended). While this approach has been useful, its discontinuous nature often obscures the intricacies of affective dynamics. An underexplored area lies in understanding the temporal dynamics of emotions, which has the potential to enrich both affective computing and affective science [2]-[4].

The temporal dynamics of human emotions is a key area of research in affective science, as it provides insights into the mechanisms that govern the initiation, maintenance, and resolution of emotional experiences [3]. Understanding these dynamics can provide valuable information about how emotions unfold and interact over time, which is essential for developing more accurate and adaptive affective computing systems [5]-[6]. Our work takes advantage of this opportunity by using the Continuously Annotated Signals of Emotion (CASE) dataset [7] and participating in the Emotion Physiology and Experience Collaboration (EPiC) 2023 competition [8], aiming to bridge the gap between affective computing and affective science by studying the temporal dynamics of emotions. Our research is guided by two main goals:

1. Analyze the performance of predictive models of affective states using continuous annotations, marking the first attempt to establish a performance benchmark for continuous predictive models on the Continuously Annotated Signals of Emotion (CASE) dataset.
2. Investigate the relationships between arousal and valence prediction tasks, as well as the role of physiological data in affective experiences, providing insights into the complex interplay between affective dimensions, their temporal dynamics, and potential

979-8-3503-2745-8/23/$31.00 ©2023 IEEE

relationships between physiological activity and affective states [9].

This work begins by presenting our research strategy and methodology, followed by a detailed analysis of our findings, and concludes with a discussion on the impact of our research on the field.

II. Methodology

A. Dataset

The Continuously Annotated Signals of Emotion (CASE) dataset used in this study contains simultaneous physiological data and affective annotations from 30 participants [7]. Participants viewed a series of video stimuli designed to elicit different emotional responses while continuously reporting their emotions using a joystick-based interface that allowed simultaneous annotation of valence and arousal (on a scale of 0 to 10). Collected physiological measures, including electrocardiograph (ECG), blood volume pulse (BVP), electrodermal activity (EDA), respiration (RSP), skin temperature (SKT), and electromyography (EMG), provide an in-depth understanding of participants' physiological responses to the stimuli. The pioneering approach of the CASE dataset, with continuous and simultaneous annotation of valence and arousal, is of significant value to both the affective computing and affective science communities.

B. Data Preprocessing and Feature Extraction

During data preprocessing, each physiological variable was filtered according to its specific characteristics, following recommendations from the physiological literature [10]. ECG data were refined using a 0.5 Hz high-pass Butterworth filter of order 5, followed by power line filtering at 50 Hz. The EDA signals underwent a low-pass filter with a cutoff frequency of 5 Hz, followed by a 4th order Butterworth filter. For the respiratory data, we implemented linear detrending and an IIR low-pass filter of order 5 with a cutoff frequency of 2 Hz [11]. The EMG data collected from the zygomatic, corrugator, and trapezius channels were subjected to a 100 Hz 4th order Butterworth high pass filter with constant detrending. No preprocessing was applied to the BVP and SKT signals; instead, the raw data were used.

Five additional continuous signals were then extracted: three from EDA (phasic component, sparse SMNA driver of the phasic component, and tonic component) [11], one from ECG (NN intervals) [13], and one from respiration (instantaneous respiratory rate) [14].

The inclusion of EDA and its decomposition into three distinct signals is essential to improve the prediction of arousal in affective states. EDA, a measure of skin electrical conductance, is widely recognized as a reliable indicator of emotional arousal due to its sensitivity to sympathetic nervous system (SNS) activity. The phasic component of EDA captures rapid fluctuations in conductance that are closely related to momentary changes in arousal. The sparse SMNA driver of the phasic component provides insight into the underlying neural activity driving these rapid conductance changes, further enriching the arousal-related information derived from the EDA signal [11]. Finally, the tonic component of EDA reflects the slower changing, baseline level of arousal [9]. By incorporating these three EDA-derived signals into the model, we aim to capture a more comprehensive representation of arousal dynamics and thereby improve the prediction of affective states.

On the other hand, both the extraction of the respiratory rate signal and the calculation of the distances between the R and R peaks in the ECG signal after removal of ectopic beats (i.e., NN intervals) were included as variables to predict affective states because of their established associations with emotional regulation and reactivity. Breathing rate has been associated with emotional arousal, with faster breathing rates typically observed during high arousal states [15]. Furthermore, breathing patterns have been shown to differ between emotional states, providing valuable information about affective experiences [16]. Similarly, NN intervals in the ECG signal allow the measurement of heart rate variability (HRV), a widely used index of autonomic nervous system (ANS) regulation [12]. HRV has been shown to be sensitive to changes in emotional state, with reduced HRV observed during stress and negative emotions [17]. Consequently, the inclusion of respiratory rate and NN intervals as predictors may contribute to a more accurate estimation of affective states, building on the established relationships between these physiological measures and emotional processes in the literature.

C. Data Aggregation

In the data aggregation phase, our primary goal was to accurately capture temporal physiological data patterns for effective prediction of affective states. We employed a strategy of variable sliding windows, the length of which was determined through a validation process. Specifically, we used a 6-second window length for scenario 1 and a 10-second window length for scenarios 2, 3, and 4. These window sizes were chosen based on their ability to capture and represent the dynamics of emotional responses while maintaining computational efficiency. We then centered each annotation within its respective sliding window, thereby capturing both past and future physiological data relative to the current affective annotation. This arrangement is intended to account for the temporal evolution of emotional experiences and provide a comprehensive snapshot of the affective context at each annotation point. To further explore the temporal dynamics of emotions, we divided each window into three equally sized segments, labeled "past," "present," and "future. For example, in scenario 1 with a 6-second window, each 2-second segment (from -3 to -1 sec, from -1 to 1 sec, and from 1 to 3 sec) was analyzed separately. The motivation behind this tripartite division is to distill meaningful features from different temporal contexts to provide a granular understanding of the affective state trajectory.

Finally, the data within each window was aggregated using various statistical measures. After a preliminary test combining different features for predicting scenario 1 (using a validation set), the mean and the minimum were chosen as aggregation measures. These two measures provide insight into both the central tendency and the dispersion (and outliers) in the data, providing a comprehensive representation of the physiological signals for predicting affective states.

D. Hyperparameter Tuning and Model Training for Interpretable Affective State Prediction

The model selection and training process focused on using interpretable models with an emphasis on understanding affective states. Two different models, Random Forest and XGBoost, were individually trained and evaluated for their potential to predict valence and arousal. These models were chosen specifically for their ability to provide insight into feature importance and generalizability [18], [19]. During the model validation process, both the Random Forest and XGBoost models were trained separately on each training partition (i.e., fold) of the different scenarios. The decision to use the Random Forest or XGBoost model for the final prediction was made independently for each scenario, based on the performance metrics obtained during validation. It is important to clarify that we did not use an ensemble of both models, but rather selected one or the other based on their individual performance in each specific context.

For the Random Forest Regressor model, the following hyperparameters were examined:

- The number of trees, with values of 50 and 100.
- The maximum depth of the trees, with values of 10 and None (indicating no maximum depth constraint).
- The minimum number of samples required to split an internal node, with values of 2 and 5.
- The minimum number of samples required to be at a leaf node, with a value of 1.

For the XGB Regressor model, the grid search considered the following hyperparameters:

- The number of boosting rounds, with values of 50 and 100.
- The maximum depth of a tree, with values of 6 and 10.
- The learning rate, or step size shrinkage, with values of 0.01 and 0.1.
- The fraction of samples to be used for fitting the individual base learners (subsample), with values of 0.5 and 0.8.
- The fraction of columns to be used by each tree, with values of 0.5 and 0.8.
- The L1 regularization term on the weights, with values of 0 and 0.1.
- The L2 regularization term on the weights, with values of 0.1 and 1.

These hyperparameters were tested in different combinations, and the models were trained and evaluated using cross-validation within each training partition to avoid data leakage [20]. The optimal set of hyperparameters for each model was determined based on the performance on the validation set.

During the model training process, feature importance was evaluated, but due to time constraints during the competition, this informed feature selection process was only implemented in scenario 1. In this scenario, the feature matrix was refined by selecting only the *n* most important features, and the performance of the models trained with these selected features was evaluated on a validation set. The optimal value of *n* that maximized performance on the validation set was determined specifically for scenario 1. The full feature set was used for the remaining scenarios.

The computational resources used for model training and evaluation included a cluster of five high performance computers with 8 cores each, two Xeon workstations with 12 parallel processes and 64 GB of RAM, and an RTX 3090 GPU for enhanced processing capabilities.

E. Validation Scenarios

To thoroughly examine the different dimensions of model generalization, the competition organizers designed four different validation scenarios: across-time, across-subject, across-elicitor, and across-version. Each of these scenarios, consisting of different numbers of folds, was designed to assess a unique aspect of the model's ability to generalize:

1. Across-time (1-fold): This scenario corresponds to a chronological hold-out validation approach where each sample, representing a single person watching a single video, is divided into training and test parts based on the timeline. The earlier parts of the video contribute to the training set, while the later parts form the test set. This scenario evaluates the model's ability to apply the knowledge gained from past data to make predictions about newly collected data within the same participant group and emotional context.
2. Across-subject (5-fold): Following the leave-N-subjects-out validation approach, this scenario divides participants into random groups, ensuring that all samples of a given participant group belong exclusively to either the training or test set. Each of the 5 folds leaves out a different set of subjects. This approach tests the model's ability to apply the knowledge learned from a given group of people to a different, previously unseen group.
3. Across-elicitor (4-fold): During the data collection, each subject watched two videos per quadrant in the arousal-valence space. In this leave-one-stimuli-out validation scenario, each fold excludes both samples associated with a given quadrant from the train set, resulting in 4 folds. This scenario assesses the ability of the model to generalize from training on three arousal-valence quadrants to infer states experienced in the excluded quadrant.
4. Across-version (2-fold): Each subject has two samples (videos) per quadrant in the arousal-valence space. In this scenario, one sample is used to train the model, while the other is used for testing, resulting in two folds. In this way, the ability of the model to generalize across different annotation versions can be investigated.

Together, these four scenarios provide a robust and comprehensive assessment of model generalization, shedding light on model strengths and potential areas for improvement.

F. Evaluation and Model Testing

According to the rules of the EPiC 2023 competition, participants were initially given only the training data set. To estimate the performance of the models, we were asked to provide the predicted time series for both arousal and valence scores for each individual and stimulus in each fold, across all scenarios. The performance of our models in each scenario was evaluated using the Root Mean Square Error (RMSE) across all scenarios and dimensions (i.e., valence and arousal).

It's worth noting that for each individual scenario, performance was estimated by calculating the average RMSE in each fold, with a lower value indicating better performance. After the competition, we were given access to the test sets, which allowed us to extend our analysis and gain further insight into the model's performance. These further evaluations using the test set are presented in the following sections.

III. Results

A. Evaluate Predictive Model Performance Across Scenarios

We first assessed the performance of the predictive models for both arousal and valence metrics across four different scenarios. To further ensure the robustness of our results, we compared the performance of the models against a set of dummy regressors. The dummy regressor chosen in our study consistently predicted a constant value. Specifically, the dummy predictor was programmed to consistently forecast a value of 5 for each dimension. This choice was directly influenced by the self-report scale, which ranged from 0 to 10, and was initiated with a default value of 5. This strategy was adopted after considering alternative methods for the dummy predictor selection, including one that would constantly predict the mean of the training set. However, the constant 5 strategy proved to be the most robust and provided the highest overall performance across all scenarios, making it the most rigorous benchmark for comparison.

Table I presents the comparative results between our models and the dummy predictor, and Figure 1 provides a visual representation of these results. One-tailed Wilcoxon signed-rank tests were performed to assess whether our models outperformed the dummy predictor. The *p*-values obtained from these tests were corrected for multiple comparisons using the Bonferroni method. Our models significantly outperformed the dummy predictor in scenario 1 for both affective dimensions (arousal: $W = 3140$, $p < 0.001$; valence: $W = 3482$, $p < 0.001$) and in scenario 2 for arousal ($W = 9856$, $p < 0.001$). Specifically, our model achieved lower RMSE values than the constant value dummy predictor in these scenarios, demonstrating its superior performance on unseen data. However, in scenario 2 for valence and in scenarios 3 and 4 for both dimensions, we found no significant evidence that our models outperformed the dummy predictor ($p > 0.98$).

TABLE I. AVERAGE RMSE FOR EACH SCENARIO VS. DUMMY PREDICTOR[a]

Scenario	*Arousal RMSE*	*Dummy Predictor Arousal RMSE*	*Valence RMSE*	*Dummy Predictor Valence RMSE*
Scenario 1	**0.8463**	1.5548	**0.8670**	1.5924
Scenario 2	**1.1694**	1.2864	1.3612	1.2719
Scenario 3	1.6246	1.2855	1.4879	1.2654
Scenario 4	1.5179	1.2846	1.3614	1.2602

a. The average RMSE values shown in Table 1 were computed across different scenarios and files within the hold-out set, which was inaccessible during the competition. Therefore, these values represent the predictive performance of the model on unseen data, as compared to a baseline provided by the dummy predictor.

In scenarios 1 and 2, where our models performed above chance in at least one dimension (outperforming the dummy regressors), we compared the performance of our arousal and valence predictions using the RMSE metric. We performed a one-tailed Wilcoxon signed-rank test to determine whether the arousal predictions performed significantly better than the valence predictions. The result showed that arousal predictions were indeed significantly better than valence predictions when looking at the combined performance in scenarios 1 and 2 ($W = 49898$, $p < 0.01$).

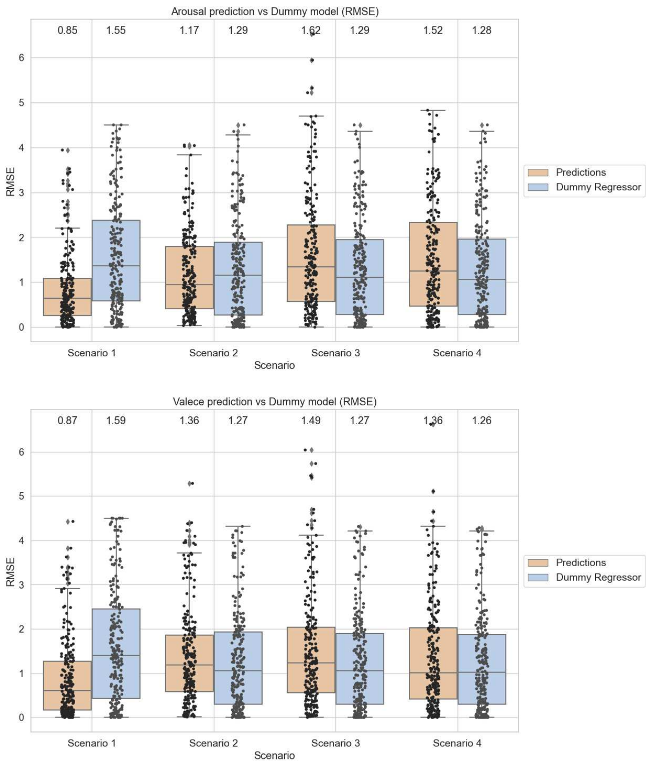

Fig. 1. Boxplots with scatter plot overlaid comparing the distribution of RMSE values across four different scenarios (scenario 1 to 4) for both the arousal and valence dimensions, with predictions in each scenario also compared to dummy predictions. Each subplot represents one dimension: arousal at the top and valence at the bottom. Each boxplot represents the interquartile range (IQR) with a line at the median. Overlaid scatter plots show the performance of each model.

B. Feature Importance in Arousal and Valence Models

To illustrate the importance of the features in the arousal and valence models, we focus primarily on scenario 1. This is because this scenario outperformed the dummy regression model, suggesting a more robust relationship between physiological signals and participants' affective annotations. Scenario 1 also had the largest number of trained models. With one model for each subject (30) and stimulus (8), a total of 240 different models were available for testing. This allows us to demonstrate a broader distribution of feature importance.

Feature importance was calculated based on the mean decrease in impurity and grouped according to the different physiological measures, with a particular focus on the partitioning of the window into past (pre-annotation), present (coinciding with the annotation period), and future (post-annotation). Figure 2 shows the distribution of feature importance for each model, revealing a trend in which the pre-annotation features are more relevant for explaining affective states than the post-annotation physiological data.

A nonparametric Kruskal-Wallis test was performed to analyze statistical differences in feature importance between the three partitions (past, present, and future), and the results indicated a significant difference between the groups ($H(2) = 199.57$, $p < 0.001$). Subsequent post hoc analyses revealed that all pairwise comparisons between the three partitions were statistically significant ($p < 0.001$). These results underscore the strong relevance of pre-annotation physiological features in explaining affective states, and highlight the significant differences in feature importance across temporal partitions. The statistical evidence provides a robust basis for the observation that pre-annotation features are more predictive than those obtained during or after the annotation period.

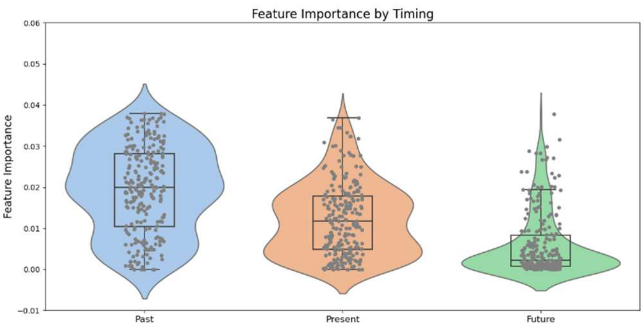

Fig. 2. Distribution of feature importance for the 240 models trained in scenario 1, It illustrates the distribution of feature importance by partitioning the window into past (pre-annotation), present (during annotation), and future (post-annotation) information. Boxplots, violin plots and scatterplots are used to visualize the feature importance distributions.

IV. Discussion

In this study, our exploration was driven by two primary goals: first, to analyze the performance of predictive models of affective states using continuous annotations, marking an unprecedented attempt to establish a performance benchmark for continuous predictive models on the Continuously Annotated Signals of Emotion (CASE) dataset; and second, to explore the complex relationships between arousal and valence prediction tasks and the role of physiological data in understanding the temporal dynamics of affective experiences.

In scenario 1, our models, trained with both physiological signals and annotations from the same participant and video segment we wanted to predict, outperformed the constant-value dummy predictor for both arousal and valence predictions. This suggests that maintaining participant and context continuity in the training data can potentially improve prediction performance. In contrast, our models only outperformed the dummy predictor for arousal prediction in scenario 2. In the remaining scenarios, the models failed to outperform the dummy predictor, suggesting possible limitations in the generalizability of our models or the complexity of the task in these conditions.

Consistent with our predictions and the literature [8], our arousal models outperformed valence models in scenarios 1 and 2, suggesting a more robust representation of sal changes by physiological data in this affective dimension . Future research could extend these observations by examining the performance differentiation between arousal and valence models, incorporating data from other participants in the EPiC 2023 competition. By integrating a broader range of modeling approaches and signal processing techniques, we aim to increase the robustness and generalizability of our findings and further elucidate the physiological correlates of affective states and the distinctions between arousal and valence.

In our study, we specifically focused on the timing of physiological features for emotion prediction and found that peripheral signals preceding the annotation were more predictive than those following the annotation. This finding underscores the critical importance of temporal aspects and greatly enriches our understanding of emotional dynamics. Such findings are consistent with broader theories that debate the complex relationship between the autonomic nervous system (ANS) and emotion. While some theories propose a bottom-up process in which the ANS triggers emotions, others contend that emotions drive ANS responses in a top-down manner. This complex interplay defies easy categorization and remains the subject of vigorous debate [21]. In the context of this debate, our work offers a unique perspective by highlighting the primacy of peripheral information in emotional experience. Our findings shed light on the potential importance of bottom-up processes in emotion generation, thus contributing nuanced insights to this multifaceted and ongoing dialogue.

As a final point, to further improve affective computing models and to better understand the temporal dynamics of affective states, it is critical to develop better databases. These improvements should focus on expanding databases to include countries that are typically underrepresented in current studies, increasing the ecological validity of data through more immersive methods of emotion elicitation, such as virtual reality [22], and incorporating both central (e.g., electroencephalography) and peripheral (e.g., EDA, ECG, and EMG) measures of neural activity [23]. It is worth noting that the overall poor performance of predictive models may be due to low emotional induction resulting from data collection in the database. By addressing these areas, researchers can gain a more complete understanding of the complex interplay between physiological and neural processes underlying emotional experiences. As a result, this would lead to the development of more robust and explicable affective computational models that better reflect the complexity of real-world emotional experiences, ultimately benefiting the broader understanding of emotion dynamics within affective science.

V. Conclusion

Our work serves as an example of the potential analyses that can be conducted using affective computing and predictive modeling to shed light on affective science. This is paper that aims to benchmark continuous models of affective states, with positive results under constrained conditions. The insights gained from our study can inspire further research to better understand the temporal dynamics of affective states and their neural correlates. By bridging the gap between affective computing and affective science, our work encourages the development of more effective and explainable affective computing applications, fostering a closer connection between these two fields, and ultimately benefiting the broader understanding of emotion dynamics within affective science.

Ethical Impact Statement

In this paper, we investigate the temporal dynamics of emotions using the Continuously Annotated Signals of Emotion (CASE) dataset during the EPiC 2023 competition. As we delve into the opportunities offered by affective computing research, it is crucial to address potential ethical concerns, especially when predicting and modeling human emotions [24], [25]. There is a possibility that the research could be used deceptively, for instance, in targeted advertising based on emotional states. To mitigate this risk, regulatory measures should be established to prevent the exploitation of affective computing technologies for non ethical purposes. Moreover, affective computing technologies have the potential to be employed in applications that limit human rights or impact people's livelihoods, such as surveillance or access to jobs and schools [26]. Addressing these concerns requires ongoing dialogue and collaboration among researchers, practitioners, and stakeholders to ensure the responsible development and implementation of affective computing technologies [10]. While acknowledging the potential for bias introduced by the WEIRD (Western, Educated, Industrialized, Rich, and Democratic) participant pool of 30, future studies should extend datasets to non-WEIRD countries for better cultural and demographic representation, given the known cross-cultural differences in emotional perception [27, 28] Furthermore, the research examines the temporal dynamics of emotions in the range of seconds to minutes. To better understand affective phenomena, ecological contexts should be studied over longer periods, such as days, weeks, or months [1]. This would help reveal the complex interplay of emotions over extended durations. In conclusion, acknowledging and addressing ethical considerations and limitations are vital in affective computing research. By doing so, we can advance our understanding of human emotions while mitigating potential risks and ethical concerns, thereby fostering responsible development and deployment of affective computing technologies.

Acknowledgment

We would like to extend our sincere gratitude to Dr. Denis A. Engemann for the insightful conversations we had about this project. His valuable input and guidance significantly contributed to the refinement of our research direction, ultimately resulting in a more robust and comprehensive work.

References

[1] R. W. Picard, *Affective Computing.* MIT press, 2000.

[2] S. D'Mello, A. Kappas, and J. Gratch, "The affective computing approach to affect measurement," *Emotion Review,* vol. 10, no. 2, pp. 174-183, 2018.

[3] P. Kuppens, Z. Oravecz, and F. Tuerlinckx, "Feelings change: accounting for individual differences in the temporal dynamics of affect," *Journal of Personality and Social Psychology*, vol. 99, no. 6, pp. 1042-1058, 2010.

[4] L. A. Bugnon, R. A. Calvo, and D. H. Milone, "Dimensional affect recognition from HRV: An approach based on supervised SOM and ELM," *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 32-44, 2017.

[5] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and Emotion*, vol. 23, no. 2, pp. 209-237, 2009.

[6] R. A. Calvo, S. K. D'Mello, J. Gratch, and A. Kappas, *The Oxford Handbook of Affective Computing*. Oxford University Press, New York, 2015, p. 169.

[7] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Scientific Data*, vol. 6, no. 1, p. 196, 2019.

[8] EPiC Collaboration, "EPiC 2023 competition," 2023. [Online]. Available: https://epic-collab.github.io/

[9] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological Psychology*, vol. 84, no. 3, pp. 394-421, 2010.

[10] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, and S. A. Chen, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behavior research methods*, pp. 1-8, 2021.

[11] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 797-804, 2015.

[12] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.

[13] D. Khodadad, S. Nordebo, B. Mueller, A. Waldmann, R. Yerworth, T. Becher, et al., "Optimized breath detection algorithm in electrical impedance tomography," Physiological Measurement, vol. 39, no. 9, p. 094001, 2018.

[14] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological measurement*, vol. 37, no. 4, pp. 610, 2016.

[15] M. Y. Balban, E. Neri, M. M. Kogon, L. Weed, B. N ouriani, B. Jo, and A. D. Huberman, "Brief structured respiration practices enhance mood and reduce physiological arousal," Cell Reports Medicine, vol. 100895, 2023.

[16] P. Grossman and E. W. Taylor, "Toward understanding respiratory sinus arrhythmia: Relations to cardiac vagal tone, evolution and biobehavioral functions," Biological psychology, vol. 74, no. 2, pp. 263-285, 2007.

[17] B. M. Appelhans and L. J. Luecken, "Heart rate variability as an index of regulated emotional responding," Review of general psychology, vol. 10, no. 3, pp. 229-240, 2006.

[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.

[19] L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001.

[20] Z. Wang, Y. Wang, J. Zhang, C. Hu, Z. Yin, and Y. Song, "Spatial-temporal feature fusion neural network for EEG-based emotion recognition," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-12, 2022s

[21] E. F. Pace-Schott, M. C. Amole, T. Aue, M. Balconi, L. M. Bylsma, H. Critchley, et al., "Physiological feelings," *Neuroscience & Biobehavioral Reviews*, vol. 103, pp. 267-304, 2019.

[22] T. D. Parsons, "Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences," *Frontiers in human neuroscience*, vol. 9, p. 660, 2015.

[23] L. Nummenmaa and H. Saarimäki, "Emotions as discrete patterns of systemic activity," *Neuroscience letters*, vol. 693, pp. 3-8, 2019.

[24] L. Tian, S. Oviatt, M. Muszynski, B. Chamberlain, J. Healey, and A. Sano, "Applied Affective Computing," 2022.

[25] T. D. Parsons, "Ethical challenges in digital psychology and cyberpsychology," Cambridge University Press, 2019.

[26] K. Roemmich, F. Schaub, and N. Andalibi, "Emotion AI at Work: Implications for Workplace Surveillance, Emotional Labor, and Emotional Privacy," in Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pp. 1-20, 2023.

[27] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," Psychological Science in the Public Interest, vol. 20, no. 1, pp. 1-68, 2019.

[28] N. Lim, "Cultural differences in emotion: differences in emotional arousal level between the East and the West," Integrative medicine research, vol. 5, no. 2, pp. 105-109, 2016.