Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Facial video based stress detection for enhancing ecological validity

Dang Ding[a,b], Weiwei Xu[a,b], Xiaoqian Liu[a,b], Tingshao Zhu[a,b,*]

[a] *Chinese Academy Sciences Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China*
[b] *Department of Psychology, University of Chinese Academy of Sciences, Beijing, China*

ARTICLE INFO

Keywords:
Stress detection
Machine learning
Ecological validity
Stress scale

ABSTRACT

In contemporary society, high-level stress poses significant detrimental effects on mental and physical well-being, impacting performance in various aspects of life including work, studies, and social interactions. Previous research efforts have primarily relied on the induction of stress through diverse mental tasks under artificial experimental conditions, which may lack ecological validity. This study aimed to address this limitation by collecting facial data without additional contextual interventions during self-introductions from participants. A regression model was developed to evaluate an individual's stress level based solely on their facial expressions captured via video. Utilizing a dataset of 240 participants, the model incorporated both facial videos and perceived stress levels for analysis. Our findings revealed that specific facial areas and features were strongly correlated with perceptions of stress, offering insights into how facial cues can mirror subjective experiences of stress. The regression model achieved impressive performance metrics, attaining a Pearson correlation efficient of 0.539 and internal consistency reliability of 0.70. These results suggest that the model possesses high applicability for early detection and management of stress, particularly by demonstrating an elevated level of ecological validity compared to previous methodologies. The positive outcomes of this study highlight considerable potential for utilizing facial analysis as a tool in identifying stress at an early stage, enabling proactive interventions for stress alleviation. Future research is encouraged to refine the concept further and enhance its accuracy, thereby maximizing its utility in real-world scenarios.

1. Introduction

Stress plays a significant role in exacerbating existing mental health conditions such as phobias, major depression, and bipolar disorder (Fink, 2016). Stress also poses serious risks to physical health by increasing the likelihood of cardiovascular diseases (Yaribeygi et al., 2017). Given these concerns, it's essential to accurately detect stress levels in order to manage stress effectively. Utilizing effective strategies for detecting stress can act as a preemptive system akin to fire alarms, safeguarding against potential harm to mental health. By promptly identifying signs of increased stress levels, individuals are better equipped to implement preventive measures or seek professional assistance, thereby facilitating disease prevention and management. This proactive approach not only enhances personal well-being but also contributes to improvements in the quality of life and workplace safety (Pluntke et al., 2019).

Modern research has shown that machine learning algorithms can classify stress with high accuracy (often over 90 % binary classification (Almeida & Rodrigues, 2021)) using various types of data including physiological and behavioral indicators like electrocardiography (ECG) readings (Hemakom et al., 2023) and facial movements (Jeon et al., 2021). Facial information is increasingly being utilized for stress detection studies, thanks to advancements in face recognition technology. However, the effectiveness of these models can be influenced by several factors, such as the type of facial data used, methods of stress induction, preprocessing techniques applied to collected data, how much of the face area is recognized by the algorithm, and the model-building process. Stress inductions typically involve tasks like the Social Stress Test (TSST) (Greco et al., 2023; Kirschbaum et al., 1993) or Stroop Color-Word Test (SCWT) (Giannakakis et al., 2021; Stroop, 1935), with real-life scenarios being less common (Can et al., 2019; Healey & Picard, 2005; Sharma & Gedeon, 2014). This issue was acknowledged in Giannakakis' review on psychological stress detection (Giannakakis et al., 2019), which noted that existing studies have a significant problem with their experimental setups for inducing stress, which can lead to a lack of effectiveness and ecological validity.

[*] Corresponding author.
E-mail address: tszhu@psych.ac.cn (T. Zhu).

https://doi.org/10.1016/j.actpsy.2025.104877
Received 11 November 2024; Received in revised form 26 February 2025; Accepted 4 March 2025
Available online 11 March 2025

The effectiveness of stress induction is often inconsistent due to the variety of tasks used and other influencing factors. For example, labelling a brief, 30-s text reading task within a social exposure task as indicative of anxiety or stress can be problematic (Giannakakis et al., 2017, 2021). Furthermore, focusing solely on maximizing stress recognition accuracy by employing experimental settings that elicit intense perceived stress can compromise ecological validity. This approach often results in a disproportionately high percentage of stress labels in the collected data, misrepresenting real-life stress experiences. While a universally standardized experimental protocol for stress simulation remains elusive, prioritizing the elicitation of natural stress states with minimal intervention, rather than solely pursuing methodological improvements, offers a viable alternative. Inspired by the TSST, a well-established laboratory procedure for stress induction, we propose a modified approach. By retaining the oral self-presentation component of the TSST while mitigating other stressors, such as mental arithmetic, the presence of interviewers, and the emphasis on demonstrating competence, we aim to enhance ecological validity.

Ecological validity is crucial as it ensures the results apply to real-world situations. Researchers employ various strategies like virtual reality environments (Parsons, 2015) or understanding creators' intentions behind stimuli (Susino, 2023) to enhance this aspect. However, there's no standardized method to evaluate ecological validity across studies (Holleman et al., 2020). To address these concerns in our research, we aim to apply oral introduction as a natural stressor instead of relying on pre-defined tasks for stress induction. By using this approach and avoiding inconsistent experimental settings, we seek to improve both the ecological validity and effectiveness of our study results. Our goal is to build an automatic system capable of detecting stress levels that are reflective of everyday life conditions rather than induced stressors found in controlled environments.

In conclusion, by focusing on improving the ecological validity through naturalistic stress inductions like oral introductions, we strive to contribute to the development of stress detection systems that can better serve people's mental health needs and improve our understanding of how stress impacts both mental and physical well-being.

2. Materials and methods

The following Fig. 1 displays the procedures of the method part of this paper.

2.1. Data collection

2.1.1. Participant

In August 2018 and January 2019, we conducted two experimental phases, each supported by separate ethical approvals (H15010 for the first phase and H18022 for the second). We enrolled a total of 240 adult volunteers without any mental or physical disabilities.

The participants were predominantly graduate students, comprising 130 females (54.2 %) and 110 males (45.8 %). The average age was 22.8 years, with a standard deviation of 0.18 years. All participants provided written informed consent, which included a statement ensuring their privacy rights would be observed. They received appropriate compensation both before and after their participation in the experiment.

2.1.2. Collection process

Participants were given five minutes to prepare and then introduced themselves to topics such as their hometown, major, and work plans for at least one and a half minutes, recorded by a camera positioned parallel to their heads (as shown in Fig. 2). After completing the video recording, participants filled out the Perceived Stress Scale (PSS) (Cohen et al., 1983), which contains 14 items. Each item on the PSS offers five options indicating different frequencies of occurrence, ranging from 'never' to 'always.' Responses are evaluated with scores from 0 to 4, depending on the relationship between the frequency of occurrence and the stress level. The PSS result is the sum of the scores from the 14 items, with higher scores indicating a higher degree of perceived stress.

The above protocols were performed with permission from the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences on 11 June 2015 and 29 December 2018 separately, with approval numbers: H15010 and H18022.

2.2. Data processing

Before delving deeply into the facial data, we first need to process it. This includes key points extraction, data cleaning, data smoothing, and calculating inter-frame differences.

2.2.1. Key-points extraction

We utilized OpenPose (Cao et al., 2017) to extract key points from facial videos. OpenPose is an open-source system that detects multi-person 2D poses in images and videos by identifying and tracking locations on the body, feet, hands and face. It achieves high accuracy with fast processing speed. In this study, OpenPose tracked the 2D

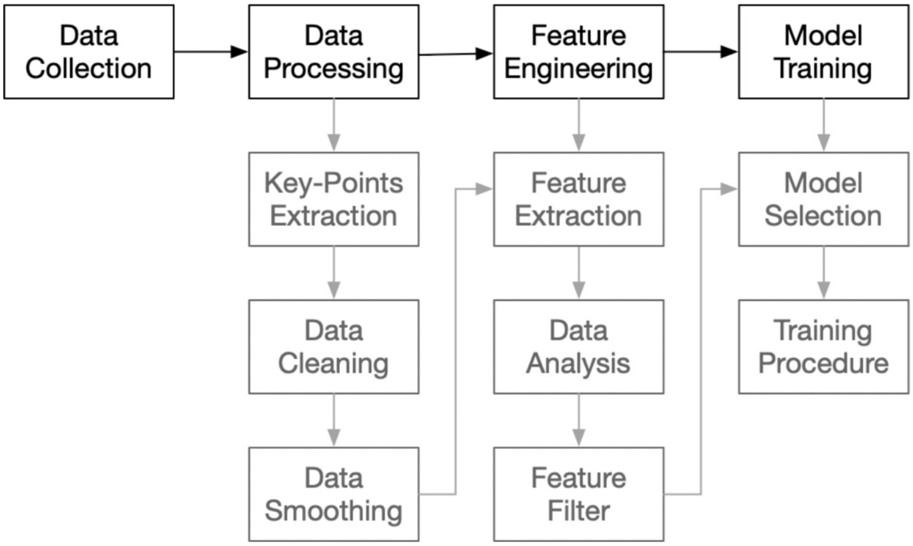

Fig. 1. The procedures of method.

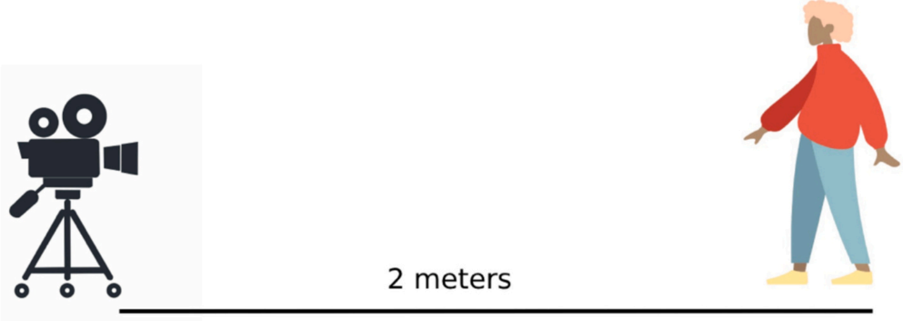

Fig. 2. The locations of the camera and participant during the facial video record.

coordinates of 70 facial key points (see Table 1). The specific locations of these key points on the face are illustrated in Fig. 3, based on the work by (Cao et al., 2017) for locations and by (Penton-Voak et al., 2006) for face figure. Following extraction, we obtained the x and y coordinates of each point for every frame.

2.2.2. Data cleaning

Considering that participants needed some time to familiarize themselves with the experimental environment, we removed the initial portion of the video data. To ensure that the facial data we processed accurately reflected participants' actual feelings after they had adapted to the unfamiliar and instructed environment and focused on the topics they were told to express, we selected the data starting from the 150th frame. The camera sampling rate was 25 Hz.

2.2.3. Data smoothing

During video recording, random noise from participants, the environment, and the recording device could have affected data quality. To mitigate this noise, we applied a mean filter, given the low-frequency nature of facial movements. This filter functions as a sliding window that averages every three frames, as illustrated in Fig. 4. The smoothing equation is as follows:

$$P_n' = \frac{P_n + P_{n+1} + P_{n+2}}{2}, \tag{1}$$

2.3. Feature engineering

2.3.1. Feature extraction

Movements between the same points on the face can provide useful information for revealing our stress and anxiety levels (Giannakakis et al., 2017). Therefore, we calculated the inter-frame difference for every x and y coordinate (as shown in Fig. 5), reducing redundant information processing. The equations for inter-frame difference are given by

Table 1
Descriptions key-points on face.

Name of the facial region	Numbers	Serial number of points
Cheek silhouette of right face	6	0–5
Jaw	5	6–10
Cheek silhouette of left face	6	11–16
Right eyebrow	5	17–21
Left eyebrow	5	22–26
Nose bridge	4	27–30
Nose base	5	31–35
Right eye	6	36–41
Left eye	6	42–47
Top mouth	10	48–53 + 60–63
Bottom mouth	10	54–59 + 64–67
Right pupil	1	68
Left pupil	1	69

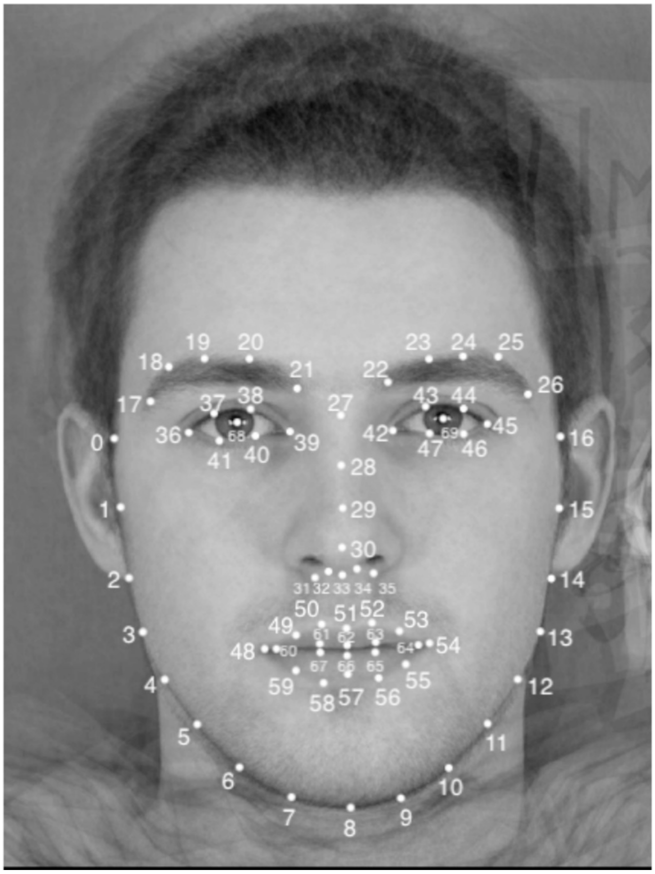

Fig. 3. Specific location of each key-point on face.

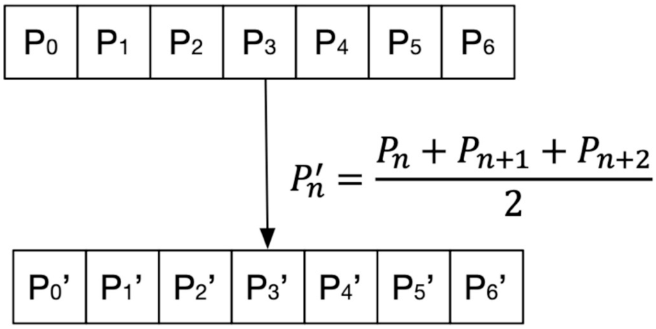
P0 P1 P2 P3 P4 P5 P6
$P_n' = \frac{P_n + P_{n+1} + P_{n+2}}{2}$
P0' P1' P2' P3' P4' P5' P6'

Fig. 4. The calculation of data smoothing. P_n means one point in one frame and P_n' stands for the same point in the next frame.

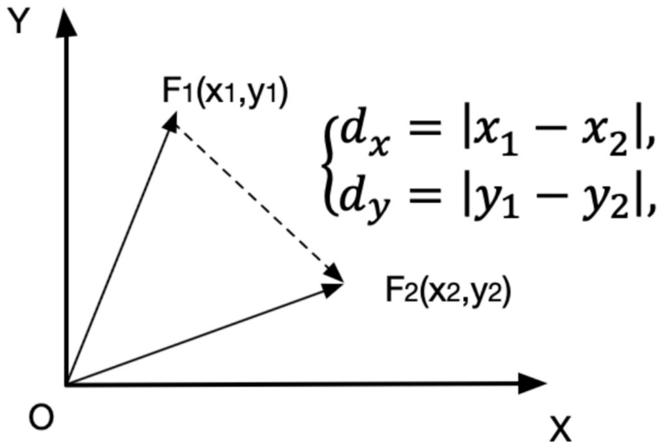
Y
F1(x1,y1)
F2(x2,y2)
O
X
$\begin{cases} d_x = |x_1 - x_2|, \\ d_y = |y_1 - y_2|, \end{cases}$

Fig. 5. The formula of difference between two frames. F_1 represents one point with x_1 and y_1 coordinates in one frame and F_2 represents the same point with x_2 and y_2 in the next frame.

$$\begin{cases} d_x = |x_1 - x_2|, \\ d_y = |y_1 - y_2|, \end{cases} \tag{2}$$

After the above procedures, the data contained at least 2250 units for each location (90 s at 25 frames per second). To simplify the modelling process, prevent overfitting, and enhance the model's robustness, we extracted representative features before analyzing the relationship between these data and PSS scores and building the best machine learning model.

We selected 30 types of items in the time domain and obtained 10 features by applying the Fast Fourier Transform (FFT) to convert the data from the time domain to the frequency domain. The specific 40 features are briefly introduced in Table 2.

2.3.2. Correlation

The initial step in our analysis was to determine if the extracted features correlated with PSS scores. We used the Pearson correlation coefficient to identify relationships between facial movements and subjective stress. Features exhibiting statistically significant correlations ($p < 0.05$) were selected for subsequent feature filtering.

Table 2
Descriptions of extracted features.

Feature's name	Description
f0-f4	Maximum/minimum/mean/variance/standard variance
f5-f8	Skewness/kurtosis/median/absolute of energy
f9	The sum of absolute values of the differences between adjacent data
f10	Variance > standard deviation (bool)
f11-f12	The number above/below mean value
f13-f16	First/last location of maximum/minimum
f17	Whether any value occurs more than once (bool)
f18	Whether the maximum value occurs more than once (bool)
f19	Whether the minimum value occurs more than once (bool)
f20-f21	The length of the longest consecutive subsequence that is bigger/smaller than mean value
f22	Mean value of f9
f23	Mean of the sum of values of the differences between adjacent data
f24	The percentage of non-unique data points
f25	The percentage of unique data points
f26-f27	The sum of reoccurring data points/values
f28	Sum of all values
f29	The difference between the maximum and minimum
f30-f34	The five top values of the frequency domain signal amplitude
f35-f37	Maximum/minimum/mean of the frequency domain signal phase angle
f38-f39	Maximum/mean of the frequency domain signal power spectrum

2.3.3. T-test

To further validate these correlations and identify the key features most reflective of emotional state, we divided the participants into high-stress and low-stress groups. These groups comprised the top 27 % and bottom 27 % of participants, respectively, based on their PSS scores (Kelley, 1939). We then performed an Independent Samples *t*-test for each feature between these two groups. Features demonstrating statistically significant differences ($p < 0.05$) in the t-test were subsequently used in the feature filtering process.

2.3.4. Feature filter

We filtered important features using four distinct methods: based on correlation results, outcomes of the t-test, a combination of both, and no filtering as a point of contrast. Subsequently, we reduced the dimensionality of the features using common techniques in the machine learning field. These techniques included Principal Component Analysis (PCA), MinMaxScaler normalization, Step Forward feature selection, and univariate linear regression tests.

2.4. Model training

2.4.1. Model selection

As the primary objective of our research was to predict individuals' PSS scores based on facial changes indicative of their stress levels, we explored various regression models. Specifically, we experimented with twelve different regression algorithms, including Adaptive Boosting Regressor, Bagging Regressor, Category Boosting Regressor, Extremely Randomized Trees Regressor, Gaussian Process Regressor, Gradient Boosting Regressor, k-nearest Neighbors Regressor, Light Gradient Boosting Machine Regressor, Linear Regression, Random Forest Regressor, Support Vector Regression, and Extreme Gradient Boosting Regressor. These models were implemented using machine learning libraries available in the Python programming platform.

2.4.2. Training procedure

In total, we employed five distinct settings, resulting in a substantial number of combinations for model training. These settings included the feature extraction method, the number of features (either 25 or 30), the selection of the regression model, the specific parameters of the model, and the number of cross-validation folds (3 or 5). Our objective was to develop the most effective model with optimal performance outcomes.

3. Results

3.1. Data analysis

3.1.1. Correlation

The complete correlation outcomes are presented in Figs. 6 and 7, separated by y and x coordinates by locations, with blank areas indicating no results. By comparing Figs. 6 and 7, it is evident that the positive maximum attained along the y-axis is slightly higher than that on the x-axis. Additionally, both figures show that relatively high coefficients are distributed from f13 to f16 and from f24 to f27. Features f13 to f16 belong to one category related to the initial or final positions of maximum or minimum movements of facial points, whereas features f23 to f26 pertain to the percentage or sum of recurring data units.

Correlation analysis identified a total of 127 significant data units ($p < 0.05$) across all features. By summing the number of significant correlations for each facial key point, we can gain insights into which specific areas of the face are more strongly correlated with PSS scores. As shown in Fig. 8(a) and (b), the eye area, including the eyebrows and the base of the nose, is particularly prominent in Fig. 8(a), while the left cheek silhouette, eyebrows, and base of the nose are notable in Fig. 8(b).

3.1.2. T-test

The complete correlation outcomes are presented in Figs. 9 and 10,

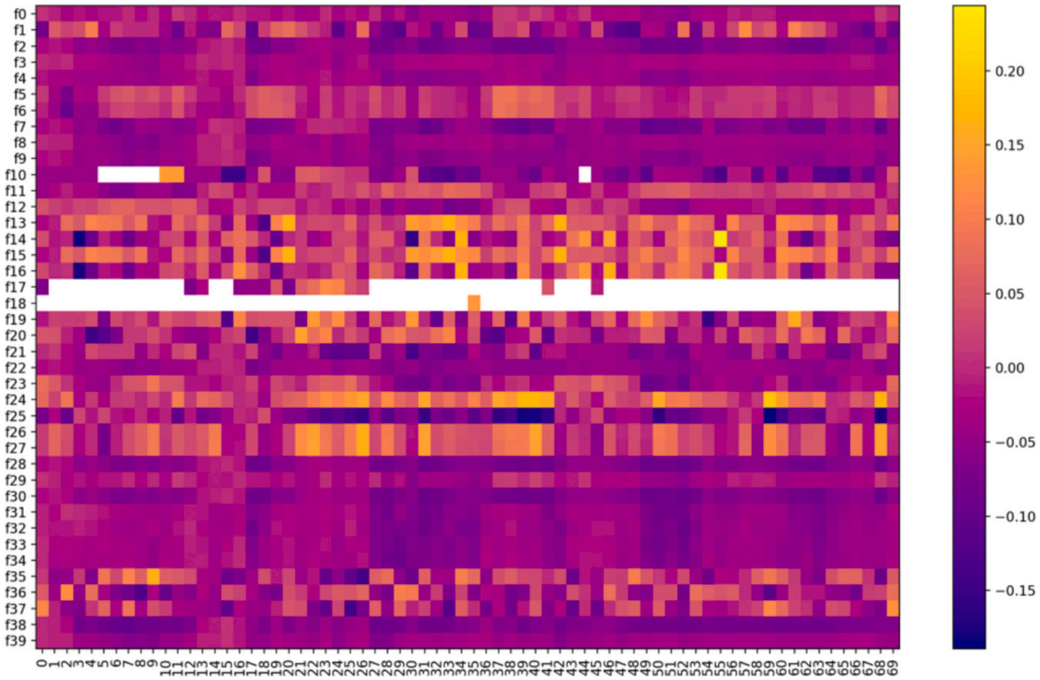

Fig. 6. The heatmap showing the coefficients between features on *y* axis and PSS scores.

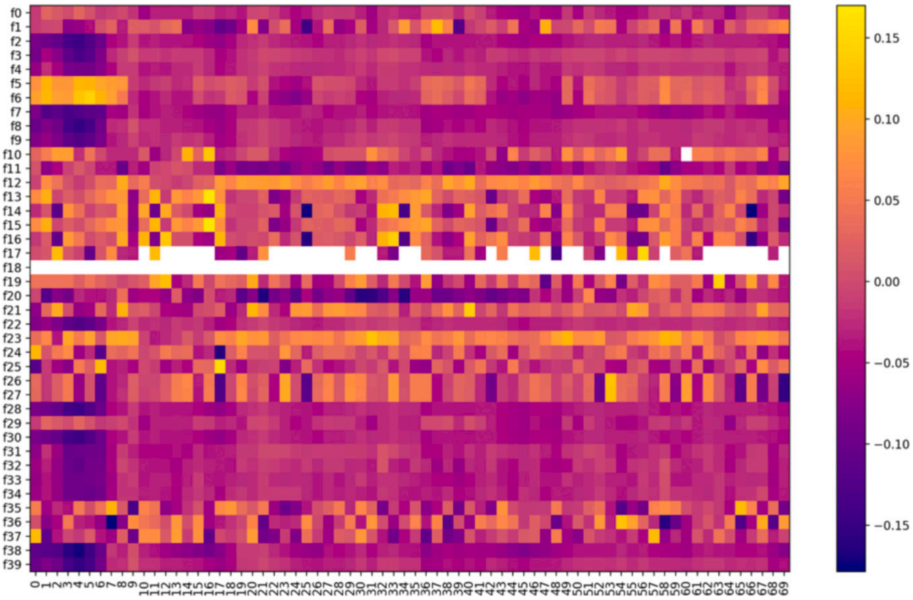

Fig. 7. The heatmap showing the coefficients between features on *x* axis and PSS scores.

separated by y and x coordinates based on locations. Similar to the previous correlation results, the positive maximum attained along the y-axis is slightly higher than that on the x-axis. Additionally, it is evident that relatively high T values are centered from f13 to f16 and from f24 to f27 in both the x and y coordinates.

The total number of significant data units ($p < 0.05$) identified under each feature through t-test analysis amounts to 129. The numbers of significant t-test results for each facial key point are shown in Fig. 11(a) and (b). The eyebrows, eye area, and bottom of the nose exhibit a high number of significant results in both figures, consistent with the findings from the correlation analysis.

3.2. Model evaluation

The parameters selected for model evaluation included the average Pearson correlation coefficient between the true and predicted values, mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). To comprehensively assess model performance, these parameters were averaged across the results obtained from the split datasets, according to the specified number of cross-validation folds.

Table 3 presents the top five results for settings with PCA dimensions set to 25 and five-fold cross-validation. The best-performing model, using extracted features, was the Category Boosting Regressor, achieving an average Pearson correlation coefficient of 0.539. The

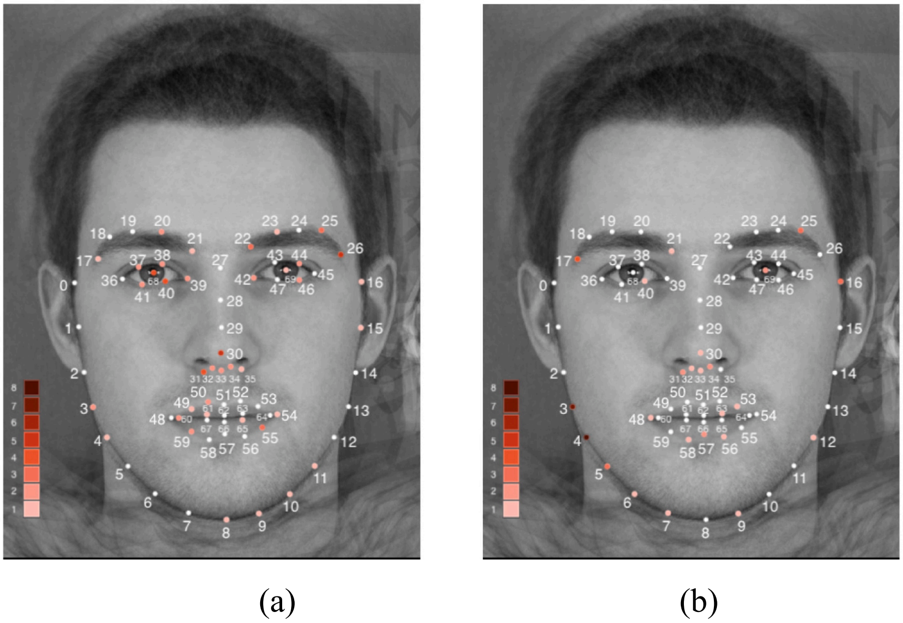

Fig. 8. Numbers of significant results accumulated on (a) y axis and (b) axis of each key-point from correlation.

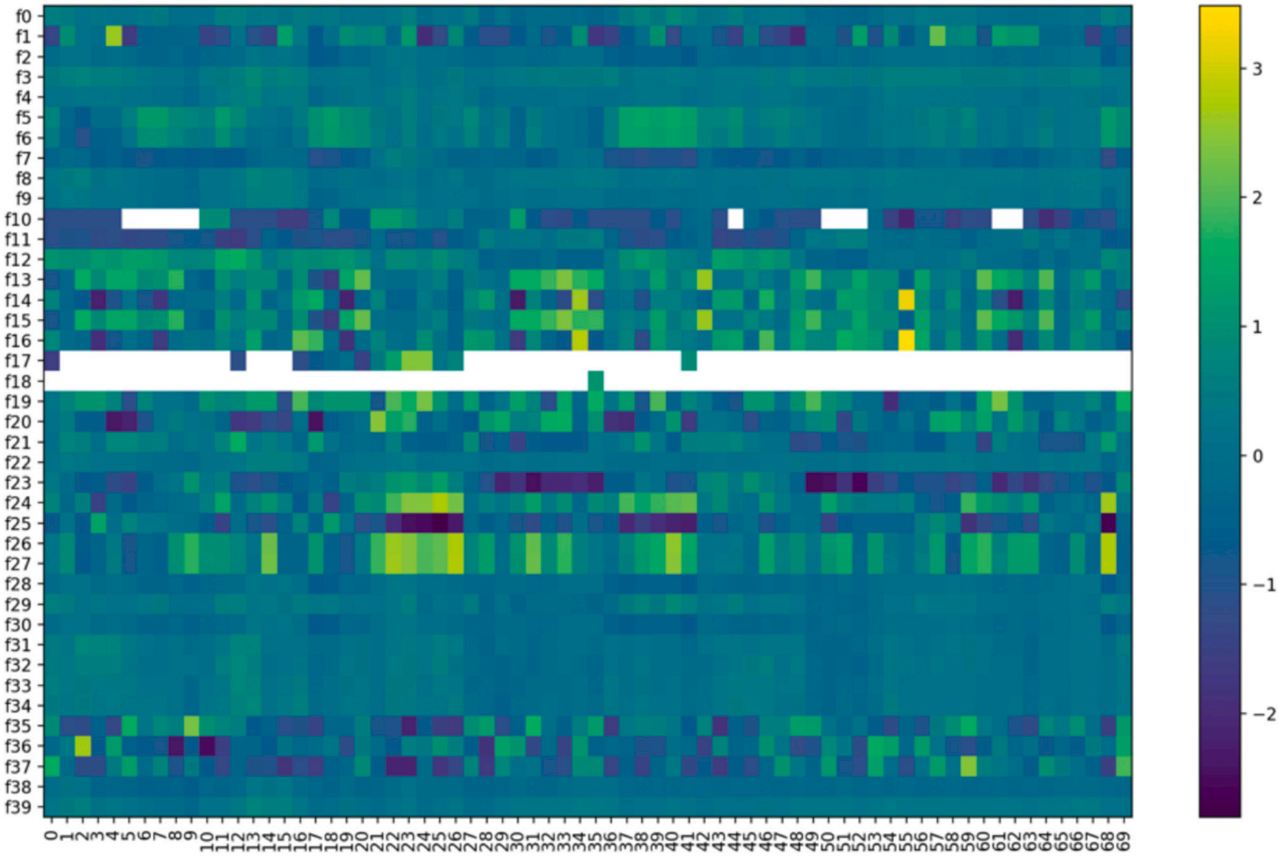

Fig. 9. The heatmap showing the T values between features on y axis and PSS scores.

corresponding evaluation metrics for this model were a MAE of 4.06, a MSE of 25.92, and a RMSE of 5.06. These results were obtained with PCA dimensions set to 25, five-fold cross-validation, and feature selection using univariate linear regression tests. Furthermore, a comparison of the top models obtained using the four different feature filtering methods revealed that the correlation and *t*-test approaches did not outperform the results achieved without filtering.

3.2.1. *Reliability and validity*

Finally, to ensure model validity, we used the cross-validation results as the final measure of performance. Model reliability was assessed by calculating half reliability, as described by Wen et al. (2022). We

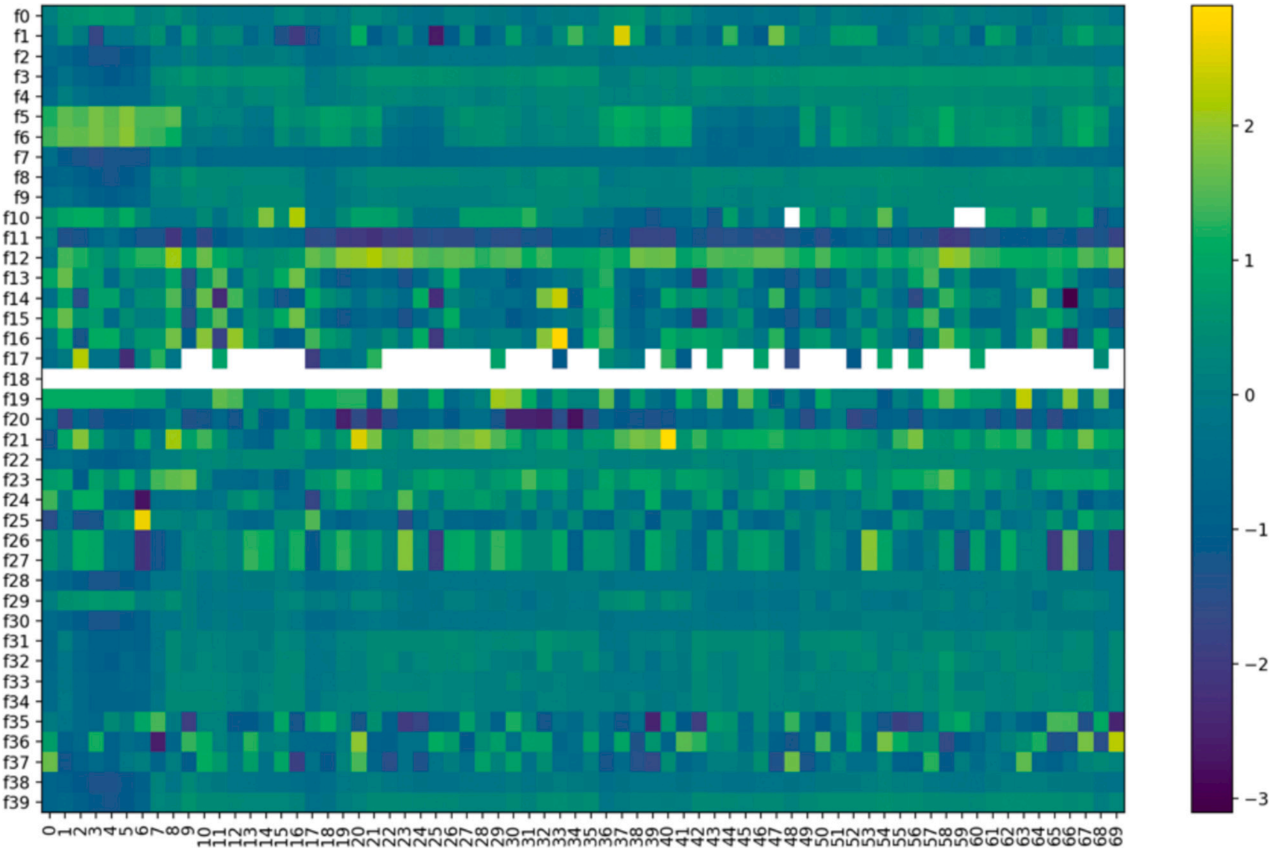

Fig. 10. The heatmap showing the T values between features on x axis and PSS scores.

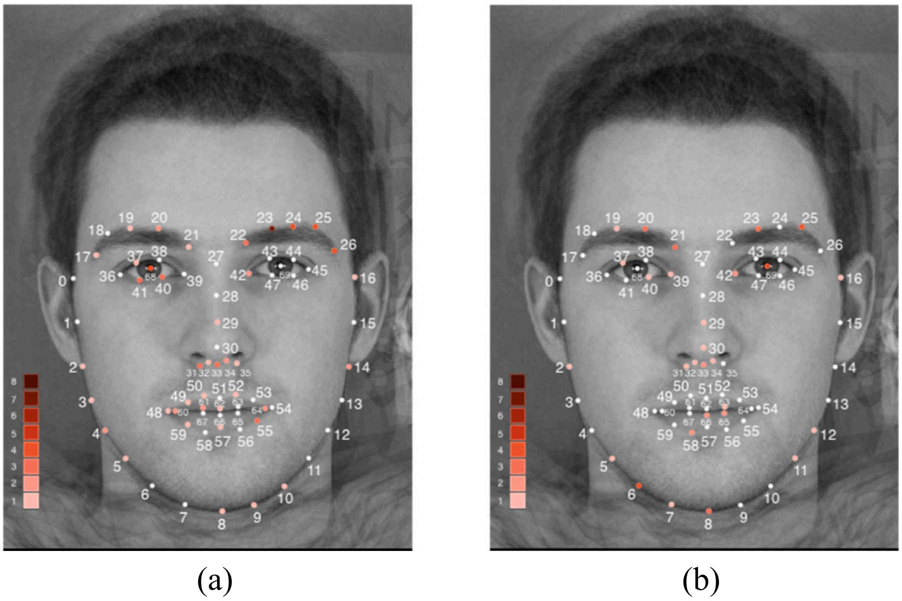

Fig. 11. Numbers of significant results accumulated on (a) y axis and (b) axis of each key-point from *t*-test.

stratified all samples into training and test datasets at a 5:1 ratio, and further dividing the test dataset into two groups: one containing even-numbered frames and the other containing odd-numbered frames. The half reliability was determined by calculating the adjusted Pearson correlation coefficient between the predicted values generated by the model trained on the training dataset and the actual values of participants from both frame groups. The detailed steps of this process are illustrated in Fig. 12.

The results of half-reliability and validity of the best model were 0.70 and 0.539, respectively, indicating that the model demonstrated both reliability and validity.

Table 3
The outcomes of predicting models.

Feature filter	Feature number	Model type	Coefficient	MAE	MSE	RMSE
Correlation	127	GradientBoostingRegressor	0.527	3.989	25.918	5.061
		RandomForestRegressor	0.492	4.238	27.816	5.262
		CatBoostRegressor	0.477	4.179	26.926	5.165
		LGBMRegressor	0.458	4.268	27.683	5.259
		XGBoostRegressor	0.418	4.273	29.974	5.463
t-test	129	LRRregressor	0.486	4.323	40.693	6.248
		CatBoostRegressor	0.483	4.031	27.017	5.188
		ExtraTreeRegressor	0.421	4.321	29.516	5.394
		GradientBoostingRegressor	0.417	4.280	29.537	5.411
		LGBMRegressor	0.407	4.338	30.867	5.529
Correlation & t-test	197	CatBoostRegressor	0.516	4.064	26.417	5.129
		GradientBoostingRegressor	0.496	4.255	27.444	5.217
		ExtraTreeRegressor	0.484	4.260	27.787	5.264
		RandomForestRegressor	0.473	4.199	27.982	5.280
		LGBMRegressor	0.400	4.434	30.567	5.508
None	5600	CatBoostRegressor	0.539	4.018	25.872	5.060
		LRRregressor	0.486	4.901	45.979	6.622
		GradientBoostingRegressor	0.430	4.251	28.766	5.375
		ExtraTreeRegressor	0.404	4.368	30.020	5.446
		LGBMRegressor	0.378	4.348	30.289	5.480

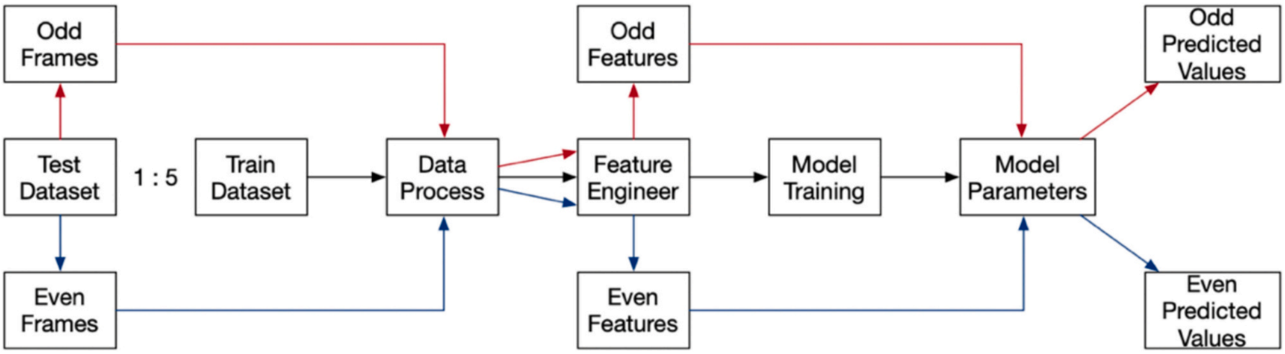

Fig. 12. The steps for half-reliability.

4. Discussion

This study analyzed the relationships between facial features and PSS scores and evaluated the model's ability to predict perceived stress levels based on facial cues and PSS scores. The data analysis results, revealing associations between facial information and stress perception, contributed to the performance of our best-performing model. More importantly, some of the discoveries from this analysis can deepen our understanding of the physical manifestation of stress, particularly in the face.

4.1. Principal findings

During the development of our target model, the feature filtering process identified two crucial types of features related to stress perception. The first type comprises features representing the initial and final positions of extreme inter-frame differences. The second type consists of features representing the repetition rate of key point shifts. Correlation analysis suggests that location features reflect the association between stress levels and the timing of extreme facial movements, while repetition ratio features indicate a correlation between stress levels and the repetitive motion of specific facial areas over a given distance.

Taking the x_16 key point under feature f13 as a specific example, we observe that the later the maximum occurs, the higher the perceived stress level, suggesting that increased stress may lead to more constrained movement. This finding contrasts with some studies (Dinges et al., 2005; Liao, 2005), indicating that stress can result in greater amplitude of head movements; however, the relationship between stress and the intensity of head movements remains disputed. Additionally, considering the overlap between stress and depression in feelings of sadness, and findings suggesting that depression slows and reduces head movements (Anis et al., 2018), it is plausible that stress may also inhibit spontaneous body language during verbal expression.

Regarding feature f24, a higher proportion of repetitions of the distance moved at the lower lip is associated with elevated stress levels, potentially indicating increased frequency of movement. This association aligns with previous literature (Liao, 2005) demonstrating that a higher frequency of mouth opening and closing correlates with heightened stress levels.

In summary, the two types of features related to stress levels suggest that as perceived stress levels increase, certain facial movements during speech tend to exhibit a pattern characterized by more restricted amplitude and higher frequency.

Among the features mentioned above, the key facial points highly correlated with the perception of stress include the left facial contour where it joins the ear, the lower lip, the left eyebrow, and the pupil of the right eye. Facial expressions serve as a means of conveying emotions, and the negative emotions associated with stress can be expressed through facial cues. It is plausible that facial information contains cues indicative of stress. Numerous studies have established a connection between facial movements and stress, which can be categorized into three main areas based on facial regions: head movements, eye movements, and mouth movements.

Head movements tend to exhibit greater amplitude and speed under stress (Giannakakis et al., 2018), with nodding and head bobbing movements being utilized to recognize complex emotional expressions (Adams et al., 2015). In the eye region, the frequency of blinks (Haak et al., 2009; Korda et al., 2021) and eyebrow movements (Bevilacqua et al., 2018) have been identified as useful indicators in predictive models that effectively assess stress levels. Mouth movements have also been consistently associated with stress levels in many studies (Pampouchidou et al., 2016). For instance, Gavrilescu's study (Gavrilescu & Vizireanu, 2019), demonstrated that a model built using four Action Units (AUs) could predict stress levels measured by the Depression Anxiety Stress Scale (DASS) (Lovibond & Lovibond, 1995) with high accuracy (88.4 %). These AUs include AU1 (inner brow raiser), AU6 (cheek raiser), AU12 (lip corner puller), and AU15 (lip corner depressor), which correspond to the cheek contour, lower lip, and left eyebrow highlighted in our findings.

Feature engineering is crucial for enhancing model performance. In this study, we employed a combination of 30 temporal statistical features and 10 frequency-domain features. This approach incorporates dynamic changes in facial expressions, capturing the variability of facial movements over time through frequency-domain analysis. By integrating frequency-domain features with temporal statistical features, the model gains a more comprehensive understanding of regular patterns in facial movements, thus improving its ability to accurately reflect stress-related information encoded in facial expressions.

4.2. Strength and shortcoming with future work

This study boasts several strengths, primarily revolving around the non-invasive, objective, and real-time monitoring capabilities of the model. By utilizing data acquired through a camera, the study benefits from a more ecologically valid dataset compared to other methods that require embedded sensors. Moreover, the model is developed based on a sizable sample of 240 individuals, ensuring a high degree of generalizability and transferability. The stress scale used in this study accurately measures the real stress state of subjects without inducing other emotions or states, thereby enhancing ecological validity and meeting the requirements for real-world application.

However, a notable limitation of this study is the homogeneity of the sample, consisting mainly of young individuals around 23 years old. As a result, the sample may not be strongly representative of other age groups. Future research endeavors could address this limitation by including participants from diverse age groups, thereby enhancing the robustness and representativeness of the model.

5. Conclusions

This study developed a non-invasive methodology for detecting stress levels with high ecological validity by analyzing facial cues. Our research revealed that specific facial features, particularly mouth and eyebrow movements, play a crucial role in reflecting subjective stress experiences, serving as strong predictors. The achieved correlation coefficient of up to 0.539 and a half-reliability value of 0.70 validate the effectiveness of this approach.

The significance of these findings extends beyond scientific circles, bridging the gap between theory and practical application. Our model offers a promising prototype for comprehensive stress monitoring systems that could be integrated into various fields, including healthcare, psychology, workplace management, and education.

Future research could expand the model's capabilities by collecting data from a more diverse participant pool, encompassing different demographics, cultures, age groups, and occupations. This would enhance the robustness and versatility of our methodology, ensuring its broader applicability.

In conclusion, this study highlights the potential of using facial expressions as non-invasive indicators of subjective stress. By deepening our understanding of these patterns, we can develop more effective tools and strategies to help individuals manage stress in daily life, contributing to healthier living environments and practices.

CRediT authorship contribution statement

Dang Ding: Writing – review & editing, Writing – original draft, Conceptualization. **Weiwei Xu:** Writing – review & editing. **Xiaoqian Liu:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Methodology. **Tingshao Zhu:** Writing – review & editing, Visualization, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could appeared to influence the work reported in this article.

Acknowledgements

This research was supported by projects H15010 and H18022, both reviewed and approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences. We gratefully acknowledge the contributions on code writing by Yeye Wen and Lei Cai, and acknowledge the assistance of Lei Cai and Xiaoyang Wang during the process of collecting facial video data.

Data availability

Data will be made available on request.

References

Adams, A., Mahmoud, M., Baltrušaitis, T., & Robinson, P. (2015). *Decoupling facial expressions and head motions in complex emotions* (pp. 274–280). Xi'an, China: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII). https://doi.org/10.1109/ACII.2015.7344583

Almeida, J., & Rodrigues, F. (2021). *Facial expression recognition system for stress detection with deep learning: Proceedings of the 23rd international conference on Enterprise information systems, 256–263*. https://doi.org/10.5220/0010474202560263

Anis, K., Zakia, H., Mohamed, D., & Jeffrey, C. (2018). Detecting depression severity by interpretable representations of motion dynamics. In *2018 13th IEEE international conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 739–745). https://doi.org/10.1109/FG.2018.00116

Bevilacqua, F., Engström, H., & Backlund, P. (2018). Automated analysis of facial cues from videos as a potential method for differentiating stress and boredom of players in games. *International Journal of Computer Games Technology, 2018*, 1–14. https://doi.org/10.1155/2018/8734540

Can, Y. S., Chalabianloo, N., Ekiz, D., & Ersoy, C. (2019). Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study. *Sensors, 19*(8), 1849. https://doi.org/10.3390/s19081849

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017*, 1302–1310. https://doi.org/10.1109/CVPR.2017.143

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385. https://doi.org/10.2307/2136404

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., … Metaxas, D. N. (2005). *Optical Computer Recognition of Facial Expressions Associated with Stress Induced by Performance Demands., 76*(6).

Fink, G. (2016). *Stress: Concepts, cognition, emotion, and behavior*. Academic Press, an imprint of Elsevier.

Gavrilescu, M., & Vizireanu, N. (2019). Predicting depression, anxiety, and stress levels from videos using the facial action coding system. *Sensors, 19*(17), 3693. https://doi.org/10.3390/s19173693

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing, 13*(1), 440–460. https://doi.org/10.1109/TAFFC.2019.2927337

Giannakakis, G., Koujan, M. R., Roussos, A., & Marias, K. (2021). Automatic stress analysis from facial videos based on deep facial action units recognition. *Pattern Analysis and Applications, 25*(3), 521–535. https://doi.org/10.1007/s10044-021-01012-9

Giannakakis, G., Manousos, D., Simos, P., & Tsiknakis, M. (2018). Head movements in context of speech during stress induction. In *2018 13th IEEE international conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 710–714). https://doi.org/10.1109/FG.2018.00112

Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P. G., … Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos.

Biomedical Signal Processing and Control, 31, 89–101. https://doi.org/10.1016/j.bspc.2016.06.020

Greco, A., Valenza, G., Lázaro, J., Garzón-Rey, J. M., Aguiló, J., De La Cámara, C., … Scilingo, E. P. (2023). Acute stress state classification based on Electrodermal activity modeling. *IEEE Transactions on Affective Computing, 14*(1), 788–799. https://doi.org/10.1109/TAFFC.2021.3055294

Haak, M., Bos, S., Panic, S., & Rothkrantz, L. J. M. (2009). *Detecting stress using eye blinks and brain activity from EEG signals*.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*(2), 156–166. https://doi.org/10.1109/TITS.2005.848368

Hemakom, A., Atiwiwat, D., & Israsena, P. (2023). ECG and EEG based detection and multilevel classification of stress using machine learning for specified genders: A preliminary study. *PLoS One, 18*(9), Article e0291070. https://doi.org/10.1371/journal.pone.0291070

Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology, 11*, 721. https://doi.org/10.3389/fpsyg.2020.00721

Jeon, T., Bae, H. B., Lee, Y., Jang, S., & Lee, S. (2021). Deep-learning-based stress recognition with spatial-temporal facial information. *Sensors, 21*(22), 7498. https://doi.org/10.3390/s21227498

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*(1), 17–24. https://doi.org/10.1037/h0057123

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier social stress test' – A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology, 28*(1–2), 76–81. https://doi.org/10.1159/000119004

Korda, A. I., Giannakakis, G., Ventouras, E., Asvestas, P. A., Smyrnis, N., Marias, K., & Matsopoulos, G. K. (2021). Recognition of blinks activity patterns during stress conditions using CNN and Markovian analysis. *Signals, 2*(1), 55–71. https://doi.org/10.3390/signals2010006

Liao, W. (2005). *A decision theoretic model for stress recognition and user assistance*.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the depression anxiety stress scales (DASS) with the Beck depression and anxiety inventories. *Behaviour Research and Therapy, 33*(3), 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Pampouchidou, A., Pediaditis, M., Chiarugi, F., Marias, K., Simos, P., Yang, F., Meriaudeau, F., & Tsiknakis, M. (2016). Automated characterization of mouth activity for stress and anxiety assessment. *IEEE International Conference on Imaging Systems and Techniques (IST), 2016*, 356–361. https://doi.org/10.1109/IST.2016.7738251

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00660

Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition, 24*(5), 607–640. https://doi.org/10.1521/soco.2006.24.5.607

Pluntke, U., Gerke, S., Sridhar, A., Weiss, J., & Michel, B. (2019). Evaluation and classification of physical and psychological stress in firefighters using heart rate variability. In *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 2207–2212). https://doi.org/10.1109/EMBC.2019.8856596

Sharma, N., & Gedeon, T. (2014). Modeling observer stress for typical real environments. *Expert Systems with Applications, 41*(5), 2231–2238. https://doi.org/10.1016/j.eswa.2013.09.021

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Susino, M. (2023). Emotional expression, perception, and induction in music and dance: Considering ecologically valid intentions. *The Journal of Creative Behavior, 57*(3), 409–418. https://doi.org/10.1002/jocb.587

Wen, Y., Li, B., Chen, D., & Zhu, T. (2022). Reliability and validity analysis of personality assessment model based on gait video. *Frontiers in Behavioral Neuroscience, 16*, Article 901568. https://doi.org/10.3389/fnbeh.2022.901568

Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI Journal, 16*, 1057–1072. https://doi.org/10.17179/excli2017-480