



OPEN ACCESS

EDITED BY

Panagiotis Tzirakis,
Hume AI, United States

REVIEWED BY

Parimita Roy,
Thapar Institute of Engineering & Technology,
India
Yating Huang,
East China Normal University, China

*CORRESPONDENCE

Lydia Helene Rupp
✉ lydia.rupp@fau.de

RECEIVED 18 February 2025

ACCEPTED 17 April 2025

PUBLISHED 30 April 2025

CITATION

Rupp LH, Kumar A, Sadeghi M, Schindler-Gmelch L, Keinert M, Eskofier BM and Berking M (2025) Stress can be detected during emotion-evoking smartphone use: a pilot study using machine learning. *Front. Digit. Health* 7:1578917. doi: 10.3389/fdgth.2025.1578917

COPYRIGHT

© 2025 Rupp, Kumar, Sadeghi, Schindler-Gmelch, Keinert, Eskofier and Berking. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Stress can be detected during emotion-evoking smartphone use: a pilot study using machine learning

Lydia Helene Rupp^{1*}, Akash Kumar², Misha Sadeghi², Lena Schindler-Gmelch¹, Marie Keinert¹, Bjoern M. Eskofier^{2,3} and Matthias Berking¹

¹Lehrstuhl für Klinische Psychologie und Psychotherapie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, ²Machine Learning and Data Analytics Lab, Faculty of Engineering, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany, ³Translational Digital Health Group, Institute of AI for Health, Helmholtz Zentrum München - German Research Center for Environmental Health, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Introduction: The detrimental consequences of stress highlight the need for precise stress detection, as this offers a window for timely intervention. However, both objective and subjective measurements suffer from validity limitations. Contactless sensing technologies using machine learning methods present a potential alternative and could be used to estimate stress from externally visible physiological changes, such as emotional facial expressions. Although previous studies were able to classify stress from emotional expressions with accuracies of up to 88.32%, most works employed a classification approach and relied on data from contexts where stress was induced. Therefore, the primary aim of the present study was to clarify whether stress can be detected from facial expressions of six basic emotions (anxiety, anger, disgust, sadness, joy, love) and relaxation using a prediction approach.

Method: To attain this goal, we analyzed video recordings of facial emotional expressions collected from $n = 69$ participants in a secondary analysis of a dataset from an interventional study. We aimed to explore associations with stress (assessed by the PSS-10 and a one-item stress measure).

Results: Comparing two regression machine learning models [Random Forest (RF) and XGBoost], we found that facial emotional expressions were promising indicators of stress scores, with model fit being best when data from all six emotional facial expressions was used to train the model (one-item stress measure: MSE (XGB) = 2.31, MAE (XGB) = 1.32, MSE (RF) = 3.86, MAE (RF) = 1.69; PSS-10: MSE (XGB) = 25.65, MAE (XGB) = 4.16, MSE (RF) = 26.32, MAE (RF) = 4.14). XGBoost showed to be more reliable for prediction, with lower error for both training and test data.

Discussion: The findings provide further evidence that non-invasive video recordings can complement standard objective and subjective markers of stress.

KEYWORDS

stress, emotion, machine learning, emotion expression, automated stress recognition

1 Theoretical background

It is late and you are preparing to leave the office. In the hallway, you meet your supervisor, and she asks you for an important report, which, as you suddenly remember, was due today. Your heart starts to race, you start to sweat, your stomach tightens: You are feeling stressed. Encountering a challenging situation such as this—experiencing a threatening situation while subjective coping resources are deemed insufficient—may prompt a stress response (1). This reaction is adaptive and necessary to ready an individual to either fight or flee the threat (2, 3). The stress response can be subdivided into the subjective experience of stress, typically assessed via self-report (e.g., the Perceived Stress Scale, PSS-10, 4), and the physiological stress response of nervous, endocrine, and immune mechanisms. Notably, both self-report and psychophysiological stress assessment suffer from important drawbacks. The assessment of subjective stress via questionnaire measures might be skewed due to biases such as, e.g., social desirability (5) or extreme responding biases (6). Assessment of physiological markers, such as cortisol levels or heart rate, can be confounded by a plethora of factors, including sampling time, smoking, alcohol consumption, medication, physical activity levels, and use of hormonal contraception (7–9) and markers may not be specific to stress (10). Lastly, obtaining physiological measures might be perceived as obtrusive and their assessment and analysis requires considerable time and resources. Thus, they may not be feasible in every research context, which often limits physiological stress research to controlled (laboratory) situations such as the Trier Social Stress Test (TSST, 11). Thus, stress research could be advanced with scalable and accessible assessment tools.

Recently, novel technological developments have advanced non-invasive assessment of voluntary or involuntary behavioral changes occurring under stress and allow passive sensing of stress in various daily applications. In this context, tools already integrated into the everyday lives of many people, such as smartphones, could be used as sensors that are unobtrusive and easy-to-disseminate. Smartphones are typically equipped with a plethora of sensors that could potentially be used to measure stress (e.g., camera, depth sensors, gyroscope). Externally visible physiological changes under stress include both “macro” changes involving larger muscle groups (e.g., facial muscles, body posture) and “micro” changes, which are caused by physiological processes (e.g., Kurz et al., in preparation). As users typically face their smartphone when using it, and as facial expressions seem to be a promising target of stress research (12, 13), the assessment of facial expressions via a smartphone front camera might be a promising avenue for stress assessment. Empirically, previous research on the relationship between stress and facial expressions found a link between arousal and visible facial expressions, as activity in several facial action units (AUs), which correspond to facial expressions, correlated with markers of the psychophysiological stress response (14). Another study found that reporting fear vs. indignation after a stress induction task was related to differential patterns in cortisol and cardiovascular activity (15). Lastly, one study found that not only confrontation

with a stressor, but also anticipatory appraisal of a potentially stressful situation induced a cardiovascular stress response, highlighting how not only the presence of a “real” stressor but also our psychological appraisal can influence physiological responding (16). In a recent study (12, 13, 17, 18), we developed and evaluated a novel smartphone-based training to reduce stress by reacting to stress-related cognitions with facial expressions of positive and negative emotions, providing data on both emotional facial expressions during smartphone use and self-reported stress levels, which could be used to explore various emotional facial expressions as a marker for subjective stress.

In recent years, several researchers have successfully developed algorithms to detect stress based on video data of facial expressions, with accuracies of up to 88.32% (19–28). However, many of these studies use data collected in measurement setups in which stress was induced through performance demands (20, 23, 25, 27, 28) or in situations where individuals were prone to experience elevated stress, such as driving (21). Secondly, studies focused mostly on the role of negative emotions in stress prediction (22, 24, 26). Thirdly, the previously mentioned studies have employed classification instead of prediction approaches, which limits the applicability to real-world contexts, in which stress may change continuously and in nuanced patterns, and few studies validated their findings against a standard measure of stress. Lastly, research on the role of facial emotional expressions in the stress process has been hindered by the fact that detecting and labeling emotional expressions in video recordings by human raters, although reliable (29), requires intensive training and time (30). In recent years, software such as OpenFace2.0 (31) have harnessed machine learning (ML) approaches to allow for automatic detection of facial emotional expression in video data, extracting landmarks that can be used in further analyses using ML methods.

In summary, whereas these studies offer important insights in the role of facial expressions under stress, several gaps remain: Firstly, many studies experimentally induced high levels of stress and study its consequences. Less is known about correlates of subjectively reported stress in conditions where stress was not explicitly induced. Secondly, it is unclear which specific facial emotional expressions are associated with subjective stress and not only negative, but positive emotions should be considered. Individuals might not only experience stress in situations where they display negative emotions, but also when they display facial expressions of positive emotions. Different emotions with the same valence might be differentially associated with subjective stress levels (15). Lastly, current methods of detecting stress from facial expressions still often require extensive laboratory setups (32–36).

To address these gaps, this study analyzes videos of facial emotional expressions recorded via smartphone in a setting in which participants were prompted to display facial expressions of both negative and positive emotions (12, 13, 17, 18), but where stress was not directly induced. In the current study, we examined whether the extracted visual features processed with a Random Forest (RF) regression algorithm relate to subjectively reported stress.

2 Methods

This study is a secondary analysis of data from a randomized controlled pilot study which evaluated a smartphone-based intervention aimed to reduce stress. A detailed description of the study procedure can be found in the protocol paper (for the protocol paper, see 37). This study did not evaluate the efficacy and clinical potential of the intervention. The evaluation of the randomized controlled pilot study can be found elsewhere (13, 17, 18). The study was conducted in accordance with the declaration of Helsinki and ethical approval was obtained by the university's ethical review board. The study was preregistered in the German clinical trials register (Deutsches Register Klinischer Studien; DRKS00023007).

2.1 Participants

Individuals with elevated stress ($n = 80$) were recruited through advertisements in public places in Erlangen, a medium-sized German city. Participants were eligible to participate in the study if they (1) reported elevated levels of stress, corresponding to a score of 19 or higher in the German version of the Perceived Stress Scale-10 (PSS-10, 4, 38), (2) were ≥ 18 years, and (3) provided informed consent. Exclusion criteria were (1) acute severe psychiatric conditions/symptomatology (e.g., suicidal ideation, substance abuse, or psychotic symptoms), physical impairment of facial emotion expression (e.g., facial paralysis), and heavy smoking (due to the assessment of salivary cortisol, see protocol paper; 33). In this study, we analyzed data of $n = 69$ participants who were allocated to the experimental conditions, as only participants in these conditions were asked to display emotional facial expressions.

2.2 Procedure

Participants completed a 4-day training aimed to reduce stress. Stress was not specifically induced before the training, instead, participants were invited to work on their stress that arose from day-to-day life. The training required the display of various facial emotional expressions in response to written statements displayed on a smartphone screen. The statements included potentially stress-reducing beliefs (e.g., "It is okay to make mistakes.") and stress-increasing beliefs (e.g., "I always have to be perfect."). Participants were instructed to distance themselves from the stress-inducing beliefs by displaying negative facial emotional expressions (such as anxiety, anger, sadness, or disgust) and approach stress-reducing beliefs by displaying positive facial emotional expressions (such as joy, relaxation, confidence, or pride). Participants were randomly allocated to eight different intervention groups (six intervention and two control conditions). The six intervention groups differed in the negative facial emotional expressions participants were asked to display. Participants in the first group ($n = 10$) were asked to

display anxiety in response to stress-inducing statements, the second group ($n = 10$) was asked to display anger, the third group ($n = 10$) was asked to display sadness, the fourth group ($n = 10$) was asked to display disgust, and the fifth and sixth groups (total $n = 20$) were asked to display all four negative emotions in varying ratios with the positive emotions (1:1 vs. 1:4). The emotions were chosen from the six basic emotions (39) and relaxation. Different experimental groups performed different negative emotions as the aim of the original study was to compare variations of the training enhanced by different emotions (37). Participants in all groups were asked to display the same positive facial emotional expressions (training day 1: joy, relaxation, and love; day 2: excitement, tranquility, and gratitude; day 3: happiness, resolve, and contentment; day 4: courage, confidence, and pride). The order of the emotions across days was determined before the beginning of the study to include a variety of different positive emotions. Participants were given examples on how the emotional expression could be performed by the experimenter and viewed videos of an actor displaying the different emotional expression on the study smartphone. The experimenter highlighted that these options should only serve as examples, and that the participant could perform the emotional expression as they would normally do (e.g., in terms of intensity, accompanying gestures, etc.). During the training session, participants' facial emotional expressions were recorded with a video camera and the experimenter provided feedback whether the expression had been performed correctly. To monitor participants' compliance with the instructions, the experimenter additionally rated the perceived quality of the facial emotional expressions in face and body. Participants in the active control condition ($n = 10$, not included in this study) were not specifically asked to display facial emotion expressions and participants in the inactive control condition ($n = 10$, not included in this study) did not participate in any laboratory intervention. A full description of the study procedure can be found in the protocol paper (12).

2.3 Measures

A full overview of all outcome assessed in the study can be found in the protocol paper (12). This study included two measures of stress, the PSS-10 (4) and a one-item stress measure.

The PSS-10 (4) was assessed before the first training day (T1), after the last training day (T2), and 1 week after the last training day (T3). Participants rated statements regarding their stress level during the past week on a 5-point Likert scale (0 = *never* to 4 = *very often*). The German version of the PSS-10 has been found to have good reliability (Cronbach's $\alpha = 0.84$; 38). As the PSS-10 records stress experienced within the past week, it was included in this study to capture stress experienced over a longer period of time.

Stress at the time of assessment was assessed before the first study session and after the last study session with one self-constructed item using an 11-point Likert scale (0 = *not at all* to 10 = *very much*). The question stated (1) "Please indicate how

stressed you feel at the moment” (current stress). This measure was included to capture participants’ stress at the time of assessment. The item had a medium test-retest reliability, with a correlation of $r = .29$ between the two assessment time points.

2.4 Statistical analyses

2.4.1 Facial emotion detection

Data on facial emotional expressions collected during the training session was used to explore associations with stress experienced over the past week (assessed with the PSS-10 at the beginning of the study) and with stress experienced at the time of assessment (assessed with the one-item stress measure at the beginning of the study).

To analyze facial emotional expressions, we employed the Facial Action Coding System (FACS), which provides a standardized taxonomy for observable and anatomically grounded movements of facial muscles (40). FACS assign a unique code to each AU, thereby enabling researchers to systematically analyze and classify facial expressions.

OpenFace2.0 (31) was utilized for extracting facial behavior features. This open-source toolkit enables real-time analysis of facial videos. Among the extracted parameters are (1) AUs, (2) facial landmarks, (3) eye gaze, and (4) head pose. OpenFace2.0 accurately predicts the presence and intensity of a subset of 18 AUs, namely AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU28, and AU45). The intensity of an AU is defined on a 5-point scale (0 = *not present* to 4 = *present at maximum intensity*). The tool achieves an average Concordance Correlation Coefficient (CCC) of 0.73 on the DISFA dataset (31), indicating an agreement of 73% between the predicted AU intensities by OpenFace2.0 and manually annotated AU intensity in the DISFA dataset (41). Apart from AUs, facial landmarks also represent key facial features such as eyes, nose, mouth, and consist of 64 locations (2D and 3D) extracted per video frame. Along with these parameters, the angle of left and right eye gaze is given in radians. We analyzed the entire duration of the experiment using video recordings captured with a frame rate of 30 frames per second (fps). Additionally, the analysis incorporates the three-dimensional head position relative to the camera, in addition to rotational data: roll (rotation around the head’s front-to-back axis), pitch (rotation around the head’s side-to-side axis), and yaw (rotation around the head’s vertical axis). This ensured comprehensive coverage of all emotional expressions exhibited by participants, rather than focusing only on specific movements.

In the next step, noise in the data was reduced in a four-step approach. Firstly, with the aid of the attribute “Confidence” in OpenFace 2.0, frames where less than 80% of the face were visible (e.g., during the start and end of the video where participants sometimes moved out of the recorded frame) were excluded. Secondly, only frames where the AU was 100% present were considered. We followed this approach as AUs are the primary features considered in this work and the AUs of interest might not be visible in all frames of the video if the participant

had difficulties performing the emotional expression. Thirdly, the mean and standard deviation of the features across the video were computed to reduce noise and to capture the variability of emotional expression over the video. The missing values induced through the calculation of standard deviation were systematically eliminated $p \leq 0.05$.

2.4.2 Training of machine learning models

To develop a stable model for detecting subjective stress scores from emotional facial expressions, we employed Random Forest (RF, 68) and Extreme Gradient Boosting (XGBoost, 37) as two ML-based regression techniques. The features derived from the video data served as input. These features were considered as the independent variables, whereas the stress score was considered as the dependent variable. The methodology followed a structured pipeline comprising feature standardization, hyperparameter tuning via cross-validation, model training using optimized parameters, and performance evaluation. Prior to training, we standardized the feature set to ensure that all input variables had a mean of zero and a standard deviation of one. As we used ensemble methods in our ML models, we included standardization as a preprocessing step to mitigate the impact of varying feature scales and enhances model stability. This transformation prevents features with larger magnitudes from disproportionately influencing the model, thereby ensuring a more balanced learning process.

To optimize the predictive performance of both models, we applied a hyperparameter tuning approach using randomized grid search within a five-fold cross-validation framework. Cross-validation ensures that the model generalizes well by evaluating it across multiple subsets of the data, thereby reducing the risk of overfitting. The tuning process was carried out in the following steps: Firstly, a predefined range of hyperparameters was established for both the RF and XGBoost models. Secondly, randomized grid search was conducted to sample hyperparameter combinations from the defined space, enabling efficient exploration of the parameter landscape without an exhaustive search. Thirdly, five-fold cross-validation was employed during the search process, splitting the dataset into five subsets where, in each iteration, the model was trained on four subsets and validated on the remaining one. Lastly, the best-performing hyperparameters were selected based on the lowest mean squared error (MSE) observed during cross-validation.

Once the optimal hyperparameters were identified, both the RF and XGBoost models were trained using five-fold cross-validation to ensure consistent performance across different data partitions. Prior to the model training, feature selection was performed using the SelectKBest method with an ANOVA F -test, retaining 750 most relevant features out of the original 1,315 features. This step not only enhanced model interpretability by focusing on the most informative features but also served as a dimensionality reduction technique, mitigating potential overfitting.

We first employed a RF regression algorithm, an ensemble learning method that enhances predictive accuracy and reduces overfitting (42). RF operates by constructing multiple decision trees during training and aggregating their predictions. This ensemble approach minimizes variance and improves

generalization. The algorithm utilizes bootstrapping and bagging techniques, where bootstrapping generates diverse datasets by randomly sampling from the original data with replacement, and bagging ensures each tree is trained on a different subset of data. This de-correlation of data contributes to a more robust predictive model by mitigating overfitting. For this study, we employed the optimized RF regressor with the best-selected hyperparameters, ensuring that the model effectively learned the relationship between the facial features extracted from video data and the subjective stress scores of participants.

In addition to the RF model, we also trained an XGBoost regression model to predict subjective stress scores. XGBoost is an advanced gradient boosting framework that builds trees sequentially, where each new tree corrects the errors of its predecessors. Unlike traditional boosting methods, XGBoost incorporates advanced regularization techniques such as L1 and L2 regularization, which improve generalization and reduce the risk of overfitting. It also employs a weighted quantile sketch algorithm for efficient handling of sparse data and parallelized tree construction for computational efficiency. Moreover, XGBoost utilizes a unique split-finding algorithm that balances model complexity and predictive performance, making it well-suited for structured data applications. With these optimizations, the XGBoost model was trained using the optimized hyperparameters, following the same five-fold cross-validation process as the RF model.

2.4.3 Model evaluation

The models were trained using the facial features extracted through OpenFace as input. The stress levels based on the PSS-10 and the one-item stress measure served as labels. For model evaluation, Mean Absolute Error (MAE) and Mean Squared Error (MSE) were used. MAE quantifies the average magnitude, while MSE emphasizes larger error due to squaring. Lower values for both metrics indicate better performance.

The mathematical representation of MAE and MSE are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Where:

- n : Number of data points
- y_i : Actual value of the target variable for the i th data point
- $\hat{y}_i$: Predicted value of the target variable for the i th data point

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Where:

- n : Number of data points
- y_i : Actual value of the target variable for the i th data point
- $\hat{y}_i$: Predicted value of the target variable for the i th data point

MAE treats all errors equally due to its absolute nature and does not adequately penalize large errors or outliers. In contrast, MSE squares errors, thereby heavily penalizing large errors. In this study, both metrics were evaluated. The algorithm, however, tried to minimize MSE. The code that was used to generate and train the model is available under OSF.io (https://osf.io/ksyda/?view_only=5da173bacf4b485bb9e6e28510d3844b).

3 Results

3.1 Demographic characteristics

The demographic information for the sample is presented in Table 1. Overall, the sample was mostly female, highly educated, and young, with a mean age of 21.36 (SD = 16.65; range from 19 to 46). Descriptive statistics for PSS-10 and the one-item stress measure are presented in Table 2. Correlations between the PSS-10 and the one-item stress measure are displayed in Table 3. Neither baseline scores of the PSS-10 [$t_{(67)} = -0.41$, $p = .682$], nor of the one-item stress measure [$t_{(67)} = 0.66$, $p = .514$] differed significantly between male and female participants.

TABLE 1 Demographic characteristics of the sample ($n = 69$).

Variable	Frequency (percentage)
Gender	
Male	14 (20.29%)
Female	55 (79.71%)
Occupation	
Student	1 (1.44%)
University student	62 (89.86%)
Employee	6 (8.70%)

TABLE 2 Means (M), standard deviation (SD), minimum (min) and maximum (max) for PSS-10 scores and one-item stress measure at pre- and post-assessment ($n = 69$).

Variable	Pre			Post		
	M (SD)	Min	Max	M (SD)	Min	Max
PSS-10	20.17 (6.10)	5	33	15.72 (6.32)	0	31
Current stress (one-item measure)	4.17 (2.07)	0	8	3.51 (2.08)	0	9

TABLE 3 Pearson correlations between the PSS-10 and the one-item stress measure at pre- and post-assessment ($n = 69$).

Measures	Measures			
	PSS-10, t1	PSS-10, t2	One-item measure, t1	One-item measure, t2
PSS-10, t1	–	–	–	–
PSS-10, t2	.54	–	–	–
One-item measure, t1	.54	.17	–	–
One-item measure, t2	.34	.52	.29	–

3.2 ML model metrics

Table 4 illustrates the error metrics for the prediction of PSS-10 scores across considered emotions, while Table 5 presents the corresponding metrics for the prediction of the one-item stress measure. The MSE and MAE for PSS-10 predictions were 2.07 and 0.82, respectively, when all emotions are considered. For the one-item stress measure, the MSE and MAE were 0.41 and 0.35, respectively. The R^2 score for PSS-10 prediction was 0.94 and that of one-item stress measure was 0.9 (Tables 4, 5).

4 Discussion

The aim of this exploratory study, conducted as a secondary analysis of data from a randomized controlled pilot study on a novel emotion-based intervention to reduce stress (12), was to assess how facial emotional expressions relate to stress. Previous studies have found the emotion-based intervention to be acceptable and clinically effective (13, 17, 18). Using video data collected during emotion-evoking smartphone use, we found that facial emotional expressions can be used to predict stress, as

assessed using both the PSS-10 and a one-item stress measure, with low model error. The evaluation of XGBoost and RF algorithms across two different stress prediction tasks demonstrated marked differences their performance, generalization capability, and potential overfitting tendencies.

Overall XGBoost achieved lower MSE and MAE on the test set compared to RF, indicating higher robustness and better generalization to unseen data. This was particularly evident for the one-item stress measure, where XGBoost’s average test MSE (2.31) was notably lower than RF’s (3.86). Findings for PSS-10 were similar, as XGBoost maintained lower test errors, which indicates its advantage for handling structured tabular data with complex patterns. The performance advantage of XGBoost could be attributed to its gradient boosting approach, which sequentially corrects errors from previous iterations and enabling fine-tuned predictions. Additionally, XGBoost’s built-in L1 and L2 regularization mechanisms helped control overfitting and ensure stable generalization. In contrast, RF exhibited higher test errors (particularly for MSE) which indicated a tendency to overfit. The larger gap between training and test errors in RF suggests that its reliance on bagging, which reduces variance but does not explicitly refine errors iteratively, might lead to suboptimal generalization compared to boosting techniques.

TABLE 4 MSE and MAE error metrics for the RF and XGBoost algorithms for the training and test data sets for the PSS-10.

Emotion	Algorithm							
	RF		RF		XGBoost		XGBoost	
	MSE		MAE		MSE		MAE	
	Train	Test	Train	Test	Train	Test	Train	Test
Joy	19.17	24.94	3.65	4.14	23.73	32.77	4.01	4.76
Love	19.00	27.12	3.60	4.30	23.98	32.32	4.05	4.71
Relaxation	18.77	26.10	3.58	4.22	24.68	34.62	4.07	4.88
Anger	18.45	21.46	3.42	3.73	14.15	29.23	2.84	4.22
Disgust	19.77	21.53	3.73	3.87	13.53	15.95	2.75	2.97
Sadness	2.59	2.40	0.03	1.05	5.34	6.48	1.83	2.01
Anxiety	6.81	7.81	2.14	2.29	0.46	2.50	0.38	0.89
All emotions	25.10	25.65	4.12	4.16	25.83	26.32	4.10	4.14

MSE, mean squared error; MAE, mean absolute error.

TABLE 5 MSE and MAE error metrics for the RF and XGBoost algorithms for the training and test data sets for the one-item stress measure.

Emotion	Algorithm							
	XGB		XGB		RF		RF	
	MSE		MAE		MSE		MAE	
	Train	Test	Train	Test	Train	Test	Train	Test
Joy	2.40	2.92	1.32	1.46	2.89	4.01	1.43	1.72
Love	2.01	2.84	1.22	1.45	2.78	3.79	1.41	1.66
Relaxation	1.99	2.90	1.21	1.46	2.57	3.97	1.33	1.68
Anger	2.00	2.23	1.20	1.27	2.61	3.66	1.31	1.56
Disgust	2.03	2.88	1.09	1.50	2.09	3.36	1.24	1.57
Sadness	1.35	2.54	0.97	1.34	0.70	2.24	0.67	1.23
Anxiety	0.13	2.05	0.22	1.02	0.28	1.89	0.34	1.01
All emotions	2.12	2.31	1.26	1.32	3.45	3.86	1.56	1.69

MSE, mean squared error; MAE, mean absolute error.

An important difference between the two algorithms was found for their performance on data from different emotional expressions: For positive emotions such as joy, love, and relaxation, and higher stress levels in PSS-10, both models showed similar performance trends. However, RF exhibited a larger discrepancy between training and test errors, indicating mild overfitting. The improvement for test MSE with XGBoost (e.g., 2.92 for joy vs. 4.01 for RF) indicates that it can capture emotional patterns without overfitting. For anger and disgust, XGBoost significantly outperformed RF, with the largest performance gap observed in disgust (RF test MSE: 3.36 vs. XGBoost test MSE: 2.88). This suggests that RF might struggle with generalizing lower-intensity or nuanced emotional expressions, potentially due to noise sensitivity in bagging-based approaches. For the prediction of anxiety, the overfitting issue in RF was most pronounced (train MSE: 0.28, test MSE: 1.89). Similarly, train MSE values were significantly lower than test MSE values for PSS-10, which indicates that RF memorized the training data rather than learning generalizable patterns. In contrast, XGBoost maintained a more balanced train-test error ratio, which indicates that it can capture complex stress patterns.

Another important finding was that both models had important performance differences when aiming to detect subtle changes in stress. In general, and particularly for higher PSS-10 scores, XGBoost showed consistent test errors across multiple runs, while RF exhibited fluctuating test performance. This shows that boosting techniques like XGBoost are more stable for analyzing psychological stress prediction. Similarly, RF exhibited overfitting tendencies for lower stress in PSS-10, with an increase in test errors compared to training errors. XGBoost had a more stable train-test error relationship, which might make it more suitable to detect subtle psychological patterns. Taken together, these findings highlight the potential applicability of XGBoost-based automated emotion and stress recognition systems for psychology, where reliability, stability, and generalization are paramount.

From a psychological perspective, several study design features might explain these findings. Firstly, in line with the hypothesis that both stress and emotion are linked through arousal and appraisal processes (43, 44), anger, disgust, and sadness might have been characterized by higher arousal and the activation of more AUs (45), which might have increased the number of parameters that could be extracted in OpenFace and used as the basis for model evaluation. In line, emerging literature on changes in facial emotional expressions under stress indicates an association between stress and emotions such as anger (46) and activity in the *musculus corrugator supercilii* (47), which is associated with negative facial emotional expressions (42, 48). In turn, positive emotional expressions (joy, love, relaxation) might have been characterized by weaker facial muscle activity, making them more difficult to detect with OpenFace. Secondly, it should be noted that there was a gender imbalance of the sample, with 80% of participants being female. However, males and females have been only found to differ with regard to the frequency with which different emotions are displayed and not how they are displayed (49–52). As all participants in the current study were instructed to perform the same emotional expressions and there

were no significant differences in baseline stress, it is unlikely that the gender imbalance may have impacted the results. Lastly, despite being routinely used as a measure of stress in many studies (53), the accuracy for models including stress assessed with the PSS-10 (4) was slightly lower than that for the one-item stress measure. Whereas the PSS-10 asked participants to rate their stress experienced during the past week, the one-item measure asked participants to indicate their current stress at the time of assessment. Thus, the response in the one-item stress measure might have more closely reflected participants' stress at the time of video data collection (54–57).

This study yields several important implications: Firstly, this study is one of the first to use a prediction instead of a classification algorithm. Previous studies (21–23, 25–27) have used classification approaches to distinguish stressed from non-stressed individuals using ML approaches. However, instruments such as the PSS-10 (4) capture subjective stress on a continuous scale, making prediction approaches more suitable for capturing nuanced changes in stress. Secondly, this study shows that stress detection using ML-based analysis of video data is also possible in contexts in which stress was not directly induced (albeit individuals may have experienced some degree of stress due to the study context), contrasting previous studies that used data from contexts in which stress was directly induced through performance demands or individuals were likely to experience high levels of subjective stress (see e.g., 21, 22) or have used standardized laboratory paradigms such as the TSST (11).

Although this study has shed light on potential new methods of stress detection, there are several important limitations to be considered: Firstly, we predicted stress based on a pre-specified set of positive and negative facial emotional expressions instead of spontaneously expressed emotions and subjective emotion intensity was not assessed. Furthermore, we did not ask participants how strongly they experienced the emotion in question. Additionally, individuals may differ in their facial expressiveness (e.g., due to differences in cultural background, emotion regulation, and personality structure), which might impair prediction accuracy for individuals with low expressiveness and overall model generalizability. Fourthly, we used a one-item rating to assess momentary stress over the course of the study. While this is a valid and frequent procedure in longitudinal psychological studies (54), the measure was not previously validated. Lastly, the sample size was relatively small, and the sample was not representative of the general population, limiting the generalizability of the findings.

From these limitations, several directions for future research can be derived: Firstly, further investigation into hybrid models combining RF and XGBoost could be beneficial, leveraging the variance reduction of bagging (RF) and sequential refinement of boosting (XGBoost). Additionally, feature engineering techniques such as interaction terms, domain-specific transformations, and deep feature selection could further enhance model accuracy. Secondly, the unobtrusive, video-based measurement setup used in the current study could complement widely used, but often intrusive and distracting measurement paradigms employed in clinical psychological research and practice. A next step in

research could be to connect these existing methods (55, 56) to develop minimally invasive measurement set-ups. Sensor fusion approaches could leverage input from various sensors, collecting data on different physiological systems to develop multi-modal stress detection systems and explore whether including physiological parameters such as heart rate, voice, and respiratory activity can improve prediction accuracy. These approaches should be evaluated in real-time scenarios, as in-the-moment stress detection using smartphone data could allow the delivery of just-in-time interventions (57–59) to target stress as it emerges. Samples should be representative of the general population to avoid biases and allow for higher generalizability of the results. In this context, researchers should specifically explore potential differences between genders (49, 50, 60) and individuals from various cultural backgrounds. Additionally, studies should also consider the role of positive emotions, as researchers have long argued that positive emotions may also serve important functions in the stress process (61) preliminary evidence suggests that deliberately showing expressions of positive emotions can decrease the detrimental effects of stress (62). Using ML models, a previous study was able to successfully distinguish self-reported distress from eustress (63). Future works should consider the impact of positive stress-related emotional experiences and distinguish between eustress and distress states. Finally, as these methods allow for unobtrusive measurement of inner states, research in this context should be accompanied by ethical considerations to ensure that measurement setups comply with ethical standards (64, 65). In ML contexts, special care should be taken to ensure that the resulting model is free of biases (66, 67), highlighting the need for appropriate data samples (in terms of gender, ethnicity, and socioeconomic background). Additionally, stress is a sensitive psychological state, so stress detection models should comply with data privacy laws and protect the healthcare information of their users.

As an interdisciplinary study connecting biopsychological stress research and ML methods, this study highlights how an this perspective can advance psychophysiological stress research. Using contactless approaches to detect indicators of inner states (such as macro- or even micro movements) might offer novel ways of measuring psychological processes.

5 Conclusion

This study aimed to explore whether macro movements in the form of emotional facial expressions recorded on video can be correlated with subjectively experienced stress. Using both a RF and XGBoost algorithm, we found that overall model accuracy for stress scores was good, with model accuracy being better for negative facial emotional expressions. XGBoost demonstrated better generalization, particularly for subtle patterns in emotional intensity and stress levels. This makes it a more reliable choice for real-world applications where unseen data distributions must be handled effectively. RF exhibited higher overfitting tendencies, particularly in lower-intensity emotions (e.g., disgust) and complex psychological states (e.g., anxiety, PSS-10 scores).

Additional regularization strategies or hybrid approaches may be required to enhance its generalization. By demonstrating that stress can be inferred from facial emotional expressions, this study further contributes to the emerging field of research on non-invasive methods to detect inner states and offers important opportunities for further research to improve diagnostic methods in psychology.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethikkommission der Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LR: Conceptualization, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. AK: Data curation, Formal analysis, Writing – original draft, Writing – review & editing. MS: Conceptualization, Writing – review & editing. LS-G: Writing – review & editing. MK: Writing – review & editing, Conceptualization, Data curation, Investigation. BE: Funding acquisition, Writing – review & editing. MB: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The study presented is part of the research project “Optimierung von Apps zur Stärkung der psychischen Gesundheit [Optimizing Apps to Strengthen Mental Health]” that is part of the Bavarian Research Association on Healthy Use of Digital Technologies and Media (ForDigitHealth) and is funded by the Bavarian Ministry of Science and Arts. This study was (partly) funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) - SFB 1483 – Project-ID 442419336, EmpkinS.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. AI technology was used to improve the language, grammar, and readability of the manuscript.

References

- Lazarus RS, Folkman S. *Stress, Appraisal, and Coping*. New York: Springer Publishing Company (1984).
- Allen AP, Kennedy PJ, Dockray S, Cryan JF, Dinan TG, Clarke G. The trier social stress test: principles and practice. *Neurobiol Stress*. (2017) 6:113–26. doi: 10.1016/j.ynstr.2016.11.001
- Rohleder N. Stress and inflammation – the need to address the gap in the transition between acute and chronic stress effects. *Psychoneuroendocrinology*. (2019) 105:164–71. doi: 10.1016/j.psyneuen.2019.02.021
- Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. (1983) 24(4):385. doi: 10.2307/2136404
- Nederhof AJ. Methods of coping with social desirability bias: a review. *Eur J Soc Psychol*. (1985) 15(3):263–80. doi: 10.1002/ejsp.2420150303
- Greenleaf EA. Measuring extreme response style. *Public Opin Q*. (1992) 56(3):328. doi: 10.1086/269326
- Adam EK, Kumari M. Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*. (2009) 34(10):1423–36. doi: 10.1016/j.psyneuen.2009.06.011
- Hellhammer DH, Wüst S, Kudielka BM. Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*. (2009) 34(2):163–71. doi: 10.1016/j.psyneuen.2008.10.026
- Stalder T, Kirschbaum C, Kudielka BM, Adam EK, Pruessner JC, Wüst S, et al. Assessment of the cortisol awakening response: expert consensus guidelines. *Psychoneuroendocrinology*. (2016) 63:414–32. doi: 10.1016/j.psyneuen.2015.10.010
- Mishra V, Sen S, Chen G, Hao T, Rogers J, Chen CH, et al. Evaluating the reproducibility of physiological stress detection models. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. (2020) 4(4):1–29. doi: 10.1145/3432220
- Kirschbaum C, Pirke KM, Hellhammer D. The ‘trier social stress test’ – a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*. (1993) 28:76–81. doi: 10.1159/000119004
- Keinert M, Eskofier BM, Schuller BW, Böhme S, Berking M. Evaluating the feasibility and exploring the efficacy of an emotion-based approach-avoidance modification training (eAAMT) in the context of perceived stress in an adult sample—protocol of a parallel randomized controlled pilot study. *Pilot Feasibility Stud*. (2023) 9(1):155. doi: 10.1186/s40814-023-01386-z
- Streit H, Keinert M, Schindler-Gmelch L, Eskofier BM, Berking M. Disgust-based approach-avoidance modification training for individuals suffering from elevated stress: a randomized controlled pilot study. *Stress Health*. (2024) 40(4):e3384. doi: 10.1002/smi.3384
- Blasberg JU, Gallistl M, Degering M, Baierlein F, Engert V. You look stressed: a pilot study on facial action unit activity in the context of psychosocial stress. *Compr Psychoneuroendocrinol*. (2023) 15:100187. doi: 10.1016/j.cpnec.2023.100187
- Lerner JS, Dahl RE, Hariri AR, Taylor SE. Facial expressions of emotion reveal neuroendocrine and cardiovascular stress responses. *Biol Psychiatry*. (2007) 61(2):253–60. doi: 10.1016/j.biopsych.2006.08.016
- Feldman PJ, Cohen S, Lepore SJ, Matthews KA, Kamarck TW, Marsland AL. Negative emotions and acute physiological responses to stress. *Ann Behav Med*. (1999) 21(3):216–22. doi: 10.1007/BF02884836
- Keinert M, Schindler-Gmelch L, Eskofier BM, Berking M. An anger-based approach-avoidance modification training targeting dysfunctional beliefs in adults with elevated stress – results from a randomized controlled pilot study. *J Cogn Ther*. (2024) 17(4):700–24. doi: 10.1007/s41811-024-00218-z
- Rupp LH, Keinert M, Böhme S, Schindler-Gmelch L, Eskofier B, Schuller B, et al. Sadness-based approach-avoidance modification training for subjective stress: pilot randomized controlled trial. *JMIR Form Res*. (2023) 7(1):e50324. doi: 10.2196/50324
- Almeida J, Rodrigues F. Facial expression recognition system for stress detection with deep learning. *Proceedings of the 23rd International Conference on Enterprise Information Systems. Online Streaming—Select a Country—: SCITEPRESS - Science and Technology Publications* (2021). p. 256–63
- Dinges DF, Rider RL, Dorrian J, McGlinchey EL, Rogers NL, Cizman Z, et al. Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviat Space Environ Med*. (2005) 76(6):B172–82.
- Gao H, Yüce A, Thiran JP. Detecting emotional stress from facial expressions for driving safety. *2014 IEEE International Conference on Image Processing (ICIP)* (2014). p. 5961–5
- Giannakakis G, Pedititis M, Manousos D, Kazantzaki E, Chiarugi F, Simos PG, et al. Stress and anxiety detection using facial cues from videos. *Biomed Signal Process Control*. (2017) 31:89–101. doi: 10.1016/j.bspc.2016.06.020
- Giannakakis G, Koujan MR, Roussos A, Marias K. Automatic stress detection evaluating models of facial action units. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)* (2020). p. 728–33
- Pedititis M, Giannakakis G, Chiarugi F, Manousos D, Pampouchidou A, Christinaki E, et al. Extraction of facial features as indicators of stress and anxiety. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2015). p. 3711–4
- Viegas C, Lau SH, Moxion R, Hauptmann A. Towards independent stress detection: a dependent model using facial action units. *2018 International Conference on Content-Based Multimedia Indexing (CBMI)* (2018). p. 1–6
- Zhang J, Mei X, Liu H, Yuan S, Qian T. Detecting negative emotional stress based on facial expression in real time. *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)* (2019). p. 430–4
- Zhang H, Feng L, Li N, Jin Z, Cao L. Video-based stress detection through deep learning. *Sensors*. (2020) 20(19):5552. doi: 10.3390/s20195552
- Drimalla H, Norden M, Langer K, Wolf OT. Video-based measurement of subjective, observable and biological stress responses. *Psychoneuroendocrinology*. (2024) 160:106753. doi: 10.1016/j.psyneuen.2023.106753
- Cohn JF, Ambadar Z, Ekman P. Observer-based measurement of facial expression with the facial action coding system. In: Coan JA, Allen JJB, editors. *Handbook of Emotion Elicitation and Assessment*. New York, NY: Oxford University Press (2007). p. 203–21.
- Lewinski P, den Uyl TM, Butler C. Automated facial coding: validation of basic emotions and FACS AUs in FaceReader. *J Neurosci Psychol Econ*. (2014) 7(4):227–36. doi: 10.1037/npe0000028
- Baltrusaitis T, Zadeh A, Lim YC, Morency LP. Openface 2.0: facial behavior analysis toolkit. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (2018). p. 59–66
- Carpenter MG, Frank JS, Silcher CP. Surface height effects on postural control: a hypothesis for a stiffness strategy for stance. *J Vestib Res*. (1999) 9(4):277–86. doi: 10.3233/VES-1999-9405
- Hagenaars MA, Stins JF, Roelofs K. Aversive life events enhance human freezing responses. *J Exp Psychol Gen*. (2012) 141(1):98–105. doi: 10.1037/a0024211
- Hashemi MM, Zhang W, Kaldewaij R, Koch SBJ, Smit A, Figner B, et al. Human defensive freezing: associations with hair cortisol and trait anxiety. *Psychoneuroendocrinology*. (2021) 133:105417. doi: 10.1016/j.psyneuen.2021.105417
- Ly V, Huys QJM, Stins JF, Roelofs K, Cools R. Individual differences in bodily freezing predict emotional biases in decision making. *Front Behav Neurosci*. (2014) 8:237. doi: 10.3389/fnbeh.2014.00237
- Stins JF, Beek PJ. Effects of affective picture viewing on postural control. *BMC Neurosci*. (2007) 8(1):83. doi: 10.1186/1471-2202-8-83
- Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016). p. 785–94

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

38. Klein EM, Brähler E, Dreier M, Reinecke L, Müller KW, Schmutz G, et al. The German version of the perceived stress scale – psychometric characteristics in a representative German community sample. *BMC Psychiatry*. (2016) 16(1):159. doi: 10.1186/s12888-016-0875-9
39. Ekman P. Are there basic emotions? *Psychol Rev*. (1992) 99(3):550–3. doi: 10.1037/0033-295X.99.3.550
40. Ekman P, Friesen WV. Facial action coding system (1978).
41. Mavadati SM, Mahoor MH, Bartlett K, Trinh P, Cohn JF. DISFA: a spontaneous facial action intensity database. *IEEE Trans Affect Comput*. (2013) 4(2):151–60. doi: 10.1109/T-AFFC.2013.4
42. Tan JW, Walter S, Scheck A, Hrabal D, Hoffmann H, Kessler H, et al. Repeatability of facial electromyography (EMG) activity over corrugator supercilii and zygomaticus major on differentiating various emotions. *J Ambient Intell Human Comput*. (2012) 3(1):3–10. doi: 10.1007/s12652-011-0084-9
43. Smith CA, Lazarus RS. Emotion and adaptation. In: Pervin LA, editor. *Handbook of Personality: Theory and Research*. New York: Guilford (1990). p. 609–37.
44. Berking M. *Training Emotionaler Kompetenzen*. Berlin, Heidelberg: Springer Berlin Heidelberg (2015).
45. Messinger DS, Cassel TD, Acosta SI, Ambadar Z, Cohn JF. Infant smiling dynamics and perceived positive emotion. *J Nonverbal Behav*. (2008) 32(3):133–55. doi: 10.1007/s10919-008-0048-8
46. Lupis SB, Lerman M, Wolf JM. Anger responses to psychosocial stress predict heart rate and cortisol stress responses in men but not women. *Psychoneuroendocrinology*. (2014) 49:84–95. doi: 10.1016/j.psyneuen.2014.07.004
47. Mayo LM, Heilig M. In the face of stress: interpreting individual differences in stress-induced facial expressions. *Neurobiol Stress*. (2019) 10:100166. doi: 10.1016/j.ynstr.2019.100166
48. Fridlund AJ, Cacioppo JT. Guidelines for human electromyographic research. *Psychophysiology*. (1986) 23(5):567–89. doi: 10.1111/j.1469-8986.1986.tb00676.x
49. Chaplin TM, Aldao A. Gender differences in emotion expression in children: a meta-analytic review. *Psychol Bull*. (2013) 139(4):735–65. doi: 10.1037/a0030737
50. Hess U, Adams R, Kleck R. Facial appearance, gender, and emotion expression. *Emotion*. (2004) 4:378–88. doi: 10.1037/1528-3542.4.4.378
51. Simpson PA, Strohm LK. Gender differences: emotional expression and feelings of personal inauthenticity. *J Appl Psychol*. (2004) 89:715–21. doi: 10.1037/0021-9010.89.4.715
52. Deng Y, Chang L, Yang M, Huo M, Zhou R. Gender differences in emotional response: inconsistency between experience and expressivity. *PLoS One*. (2016) 11(6):e0158666. doi: 10.1371/journal.pone.0158666
53. Smith KJ, Emerson DJ. An assessment of the psychometric properties of the perceived stress scale-10 (PSS10) with a U.S. public accounting sample. *Adv Account*. (2014) 30(2):309–14. doi: 10.1016/j.adiaac.2014.09.005
54. Allen MS, Iliescu D, Greiff S. Single item measures in psychological science. *Eur J Psychol Assess*. (2022) 38:1–5. doi: 10.1027/1015-5759/a000699
55. Skiendziel T, Rösch AG, Schultheiss OC. Assessing the convergent validity between the automated emotion recognition software Noldus FaceReader 7 and facial action coding system scoring. *PLoS One*. (2019) 14(10):e0223905. doi: 10.1371/journal.pone.0223905
56. Will C, Shi K, Schellenberger S, Steigleder T, Michler F, Fuchs J, et al. Radar-based heart sound detection. *Sci Rep*. (2018) 8(1):11551. doi: 10.1038/s41598-018-29984-5
57. Sarker H, Hovsepian K, Chatterjee S, Nahum-Shani I, Murphy SA, Spring B, et al. From markers to interventions: the case of just-in-time stress intervention. In: Rehag JM, Murphy SA, Kumar S, editors. *Mobile Health*. Cham: Springer International Publishing (2017). p. 411–33.
58. Jaimes L, Llofriu M, Raji A. A stress-free life: just-in-time interventions for stress via real-time forecasting and intervention adaptation. *Proceedings of the 9th International Conference on Body Area Networks*. London, Great Britain: ICST (2014).
59. Smyth JM, Heron KE. Is providing mobile interventions “just-in-time” helpful? An experimental proof of concept study of just-in-time intervention for stress management. 2016 *IEEE Wireless Health (WH)*. Bethesda, MD, USA: IEEE (2016). p. 1–7
60. Fischer AH. Gender differences in nonverbal communication of emotion. In: Fischer AH, editor. *Gender and Emotion: Social Psychological Perspectives*. Cambridge, New York, Paris: Cambridge University Press (2000); Editions de la Maison des Sciences de l’Homme (Studies in emotion and social interaction). p. 97–117.
61. Folkman S. The case for positive emotions in the stress process. *Anxiety Stress Coping*. (2008) 21(1):3–14. doi: 10.1080/10615800701740457
62. Kraft TL, Pressman SD. Grin and bear it: the influence of manipulated facial expression on the stress response. *Psychol Sci*. (2012) 23(11):1372–8. doi: 10.1177/0956797612445312
63. Singh A, Singh K, Kumar A, Shrivastava A, Kumar S. Optimizing well-being: unveiling eustress and distress through machine learning. 2024 *International Conference on Inventive Computation Technologies (ICICT)* (2024). p. 150–4
64. McCradden MD, Anderson JA, Stephenson EA, Drysdale E, Erdman L, Goldenberg A, et al. A research ethics framework for the clinical translation of healthcare machine learning. *Am J Bioeth*. (2022) 22(5):8–22. doi: 10.1080/15265161.2021.2013977
65. McCradden MD, Anderson JA, Shaul RZ. Accountability in the machine learning pipeline: the critical role of research ethics oversight open peer commentaries. *Am J Bioethics*. (2020) 20(11):40–2. doi: 10.1080/15265161.2020.1820111
66. Chakraborty J, Majumder S, Menzies T. Bias in machine learning software: why? How? What to do? *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. Athens, Greece: ACM (2021). p. 429–40
67. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *arXiv* (2022).
68. Breiman L. Random forests. *Mach Learn*. (2001) 45(1):5–32. doi: 10.1023/A:1010933404324