

# A New Perspective on Stress Detection: An Automated Approach for Detecting Eustress and Distress

**Abstract**— Previous studies have solely focused on establishing Machine Learning (ML) models for automated detection of stress arousal. However, these studies do not recognize stress appraisal and presume stress is a negative mental state. Yet, stress can be classified according to its influence on individuals; the way people perceive a stressor determines whether the stress reaction is considered as eustress (positive stress) or distress (negative stress). Thus, this study aims to assess the potential of using an ML approach to determine stress appraisal and identify eustress and distress instances using physiological and behavioral features. The results indicate that distress leads to higher perceived stress arousal compared to eustress. An XGBoost model that combined physiological and behavioral features using a 30 second time window had 83.38% and 78.79%  $F_1$ -scores for predicting eustress and distress, respectively. Gender-based models resulted in an average increase of 2-4% in eustress and distress prediction accuracy. Finally, a model to predict the simultaneous assessment of eustress and distress, distinguishing between pure eustress, pure distress, eustress-distress coexistence, and the absence of stress achieved a moderate  $F_1$ -score of 65.12%. The results of this study lay the foundation for work management interventions to maximize eustress and minimize distress in the workplace.

**Index Terms**— Psychological Stress; Physiological Data; Behavioral Data; Machine Learning

## 1. INTRODUCTION

Stress, labeled the "epidemic of the 21<sup>st</sup> century," [1] affects a majority of Americans, with job pressure being the main stressor [2]. Office work, which encompasses 18.5 million people in the US [3], leads to significant stress due to long hours, heavy workload, job insecurity, conflicts, and inappropriate task assignments.

Distress refers to the overwhelming feeling of being "stressed out" when facing uncontrollable stressors [4]. It negatively impacts workers, leading to psychological effects like loss of concentration, impaired performance, insecurity, as well as physical consequences such as tension, insomnia, and headaches. This places a burden on the healthcare system, with American companies estimated to lose up to \$300 billion annually due to worker distress [5]. A survey of 17,000 American office workers revealed that 33% missed work due to distress [6], reducing overall productivity and the national gross domestic product. Thus, distress among office workers is a significant concern that requires an urgent solution.

On the other hand, eustress, or positive stress, occurs when people feel confident in handling a stressor, resulting in higher concentration, energy, motivation, confidence, engagement, and excitement [7]. It serves as a driving force for individuals to achieve success, fulfillment, and overcome challenges [8]. While the negative effects of job stress have been extensively studied, the variations in eustress and distress remain largely unexplored [9].

Work organizations typically focus on limiting stressors with the assumption that stress is negative, disregarding the potential benefits of eustress [10]. However, effective

management plans should aim to minimize distress and maximize eustress by optimizing work stressors. This can be achieved through approaches that set challenging yet attainable expectations for employees [11]. Determining workers' perception of stressors as eustress or distress is challenging but necessary. While indicators of distress are well-understood, knowledge of indicators specific to eustress is limited. Machine Learning (ML) offers the potential to examine psychophysiological responses in relation to both eustress and distress. Previous studies have mainly focused on detecting stress arousal by differentiating between "stress" and "no stress," neglecting the appraisal component [12]. In fact, only one study has attempted to detect eustress using an automated approach. Li et al. [13] utilized a small sample size ( $n=7$ ) and data from participants' computers, phones, and heart rate sensors. They achieved a moderate detection accuracy of 70% using a machine learning algorithm. This study demonstrates that automated stress detection can go beyond arousal detection and focus on the appraisal of stress as eustress or distress.

Furthermore, personal factors, such as age and gender, influence how individuals appraise stress experiences. For instance, a study of 281 office workers in technology firms found that younger females reported higher eustress, while males experienced more distress due to a lack of emotional support at work [9]. Another questionnaire-based study of 595 office workers revealed that older employees and those with higher academic qualifications perceived work overload as more distressing compared to younger counterparts [14]. To that end, any attempt to understand the

differences between eustress and distress must not ignore the impact of personal characteristics on the appraisal of stress.

Built on this background, this study aims to assess the potential of using an ML approach to determine stress appraisal and identify instances of eustress and distress among office workers. The study investigated six research questions. (1) How does stress level (i.e., arousal) change as a function of eustress and distress (i.e., valence)? (2) What ML algorithms are best suited for the prediction of eustress and distress? (3) What window size for data processing is best suited for the prediction of eustress and distress? (4) What data modality (i.e., physiological, behavioral, or combination of both) is best suited for the prediction of eustress and distress? (5) How does gender affect the prediction of eustress and distress? and (6) How can we create a stress appraisal prediction model to differentiate between eustress, distress, eustress-distress coexistence, and no-stress?

The remainder of this paper is organized as follows. Section 2 provides a comprehensive background overview of stress detection research. Section 3 explains in detail the experimental setup for data collection, the procedure for data cleaning and processing, and the training and testing of the different ML algorithms. Section 4 provides a summary of the results, while Section 5 offers a discussion, and provides insights into the feasibility of using ML for identifying positive and negative appraisals of stress. Section 6 focuses on the conclusions drawn from the results and outlines the study limitations and future research directions.

## 2. BACKGROUND

### 2.1 Psychophysiological and behavioral responses to stress

Multimodal stress detection research typically relies on analyzing three main categories of responses: psychological, physiological, and behavioral. These categories encompass different aspects of human responses to stress and are often used in combination to provide a comprehensive understanding of stress levels.

Psychological processes play a pivotal role in shaping the stress response and have been employed to establish precise labels for training machine learning algorithms in stress detection [12]. The assessment of acute stress can be achieved by examining various facets of the psychological response. Questionnaires designed to gauge perceived stress levels, emotional valence, and arousal, for example, serve as indicators of acute stress [15]. While previous research has primarily concentrated on stress arousal, stress appraisal questionnaires have not garnered widespread recognition. Nevertheless, the Valencia Eustress-Distress Appraisal Scale (VEDAS) [16], [17] offers an opportunity to

evaluate the psychological dimensions of stress appraisal, thus advancing stress detection research by incorporating appraisal in addition to arousal. This scale serves as a validated instrument for assessing stress appraisal and has undergone translation into multiple languages, as well as validation across diverse populations worldwide.

In addition to psychological responses, stress activates the autonomic nervous system, leading to variations in bodily biomarkers and physiological signals [12]. While various biomarkers have been used to measure stress, some are inconvenient to collect (such as cortisol levels from saliva or blood samples, or EEG via electrode cap) and unsuitable for continuous stress detection. Non-invasive physiological measures such as Heart Rate (HR), Heart Rate Variability (HRV), Skin Temperature (ST), ElectroDermal Activity (EDA), and Blood Volume Pulse (BVP) (i.e., volume of blood flowing through the peripheral blood vessels) are more commonly studied in stress research, as they can be collected using wearable devices [18]. For instance, HR and HRV are direct indicators of stress, with higher levels of HR and lower levels of HRV generally associated with psychological stress [19]. During periods of psychological stress, EDA typically increases due to increased sweating, BVP tends to show an increase under stress. On the other hand, ST tends to decrease during stress due to vasoconstriction, which reduces blood flow to the skin and results in cooler skin temperatures.

Furthermore, the psychophysiological stress response can manifest in behavioral changes, which may be observed through alterations in body posture, facial expression, and interaction with the environment. While the exploration of behavioral measurements for stress detection is not as extensive as that of physiological measures, pioneering studies have demonstrated their potential predictive power, and further research holds the potential to strengthen these findings. Video cameras have been employed to capture and analyze facial and posture features in relation to stress development, yielding substantial improvements in stress arousal prediction. Additionally, within the context of office work, observing workers' interactions with their computer, such as mouse or keyboard usage, can provide valuable insights into work pressure and the associated increase in stress arousal. These behavioral indicators, when combined with physiological measures, contribute to a more comprehensive understanding of stress dynamics.

Hans Selye characterized stress as a state of heightened arousal and emphasized that when faced with stress, the crucial factor is how it is perceived by the individual—whether as a positive or negative experience [20]. Consequently, stress appraisal emerges as an outcome of stress arousal, providing a means to anticipate eustress and distress by employing physiological and behavioral indicators already employed for stress arousal prediction. However,

further investigation is required to determine the association between stress appraisal and physiological and behavioral changes and determine the extent of their impact during shifts in positive and negative valence.

## 2.2 Stress detection

The existing body of literature has primarily focused on the identification of stress arousal [12]. Unfortunately, this approach has largely overlooked the appraisal component inherent in the stress response. However, the limitations observed in these stress detection studies can offer valuable insights that can be utilized in the development of dependable stress appraisal models. Notably, a majority of these investigations have predominantly relied on physiological data as the foundation for constructing their machine learning prediction models [12]. In light of this, Alberdi et al. [12] contend that the integration of a multimodality stress detection approach is imperative to enhance the accuracy of detection. This viewpoint finds support in the work of Liao et al. [21], who hypothesize that physical symptoms, such as an accelerated heart rate, are not exclusively indicative of stress. Consequently, stress detection machine learning models that incorporate the fusion of information from multiple modalities are likely to exhibit increased reliability and proficiency in discerning between stressful and non-stressful situations.

Yet, while some studies have attempted to adopt a multi-modal approach, many of them have focused solely on combining various physiological features without incorporating data from other domains, such as behavioral data [12]. In contrast, Koldijk et al. [22] conducted a laboratory experiment that simulated stressors commonly experienced in office work, such as interruptions and time pressure, and collected both physiological data (heart rate and skin conductance) and behavioral data (posture, facial expressions, and human-computer interactions). The feature importance analysis of their stress detection model showed that facial expressions, head movement, and skin conductance were among the most crucial features for detecting stress arousal. This demonstrates the contribution of combining different modalities in stress detection research. However, additional research is necessary to investigate the trade-offs between physiological and behavioral features in terms of prediction accuracy, especially in the context of stress appraisal.

Finally, many stress detection studies in office-like environments rely on using psychometric tests (arithmetic calculations, Stroop tests, memory tests) or visual stimuli to induce stress [23], [24]. Although proven to induce stress effectively, these tests do not accurately mimic real office work (e.g., completing reports, writing, preparing presentations, etc.), which could lead to unreliable stress detection results when implementing the models in real office environments.

means of creating eustress and distress conditions, and thus, it is necessary to reconsider the experimental procedures, particularly when examining stress appraisal.

To this day, the investigation conducted by Li et al. [13] represents the only explicit attempt to employ machine learning techniques for the prediction of eustress. Nevertheless, the study is not without its limitations, which include a small sample size comprising merely 7 individuals, an inadequate grasp of a comprehensive methodology for distinguishing between eustress and distress, as well as limited analysis pertaining to the behavioral and physiological variations observed in instances of eustress. Additionally, the study fails to explore how eustress may vary in relation to personal characteristics. In another study, Setz et al. [25] aimed to differentiate between stress and cognitive load in a way that is similar to the distinction between distress and eustress. By focusing on this differentiation, they sought to provide a more accurate representation of the psychological experiences of individuals in office work settings. Their results showed a good prediction accuracy that reached 82%. However, their work falls short in detecting situations characterized by the absence of stress or instances where stress coexists with cognitive load.

### 3. METHODOLOGY

We conducted an experimental procedure to study the physiological and behavioral signals that are most useful for the automated detection of eustress and distress among office workers. To obtain as wide a range of signals as possible within each participant, the 70-minute experiment incorporated a phase of low-stress engagement at a computer workstation, followed by a phase of engagement that incorporated multiple stressful components. The study was approved by the Institutional Review Board of the University of Southern California.

### 3.1 Participants

A total of 48 participants voluntarily completed the experiment, of which 28 were females and 20 were males. Participants were mainly graduate and undergraduate students with a mean age of 22.6 years and a standard deviation of 2.1 years. Individuals with eye/vision problems that would prevent them from working on a computer, with psychological problems that make them sensitive to stress-inducing tasks, who were pregnant, or who were taking any medication that would affect their physiological signals were excluded.

### 3.2 Experimental Procedure

To collect physiological data such as heart rate, BVP, EDA, ST, and wrist accelerations, participants wore an E4 Empatica wristband [18] and an H10 polar chest strap [26]. To reduce motion artifacts, the E4 device was placed on the non-dominant hand, as research has shown that this hand ex-

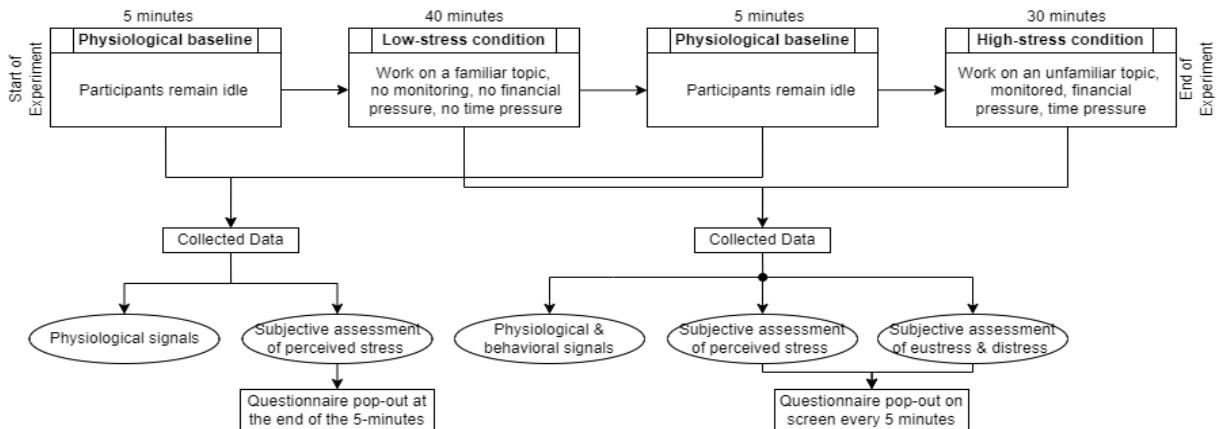


Fig. 1. Experimental design schematic

reduces instances of motion interference in the data collected by the E4 device. During the experiment, a Microsoft Azure Kinect DK camera was installed facing the participant at the top of the screen to record their faces. Additionally, a logging application called Mini Mouse Macro [27] ran in the background of the computer to record participants' interactions, such as keyboard keystrokes and mouse clicks.

As presented in Fig. 1, the experiment consisted of two phases: low-stress work and high-stress work. At the start of each phase, participants remained still for 5 minutes to collect resting physiological data. Participants then rated their stress level on a 0-100 scale. Throughout both phases, every 5 minutes, participants completed a pop-up questionnaire on the computer screen to rate their perceived stress level on the 0-100 scale and to appraise the work as eustress and distress using the VEDAS [16], [17]. Eustress was rated as a source of opportunity/challenge using a 6-point scale (with 1 being "very definitely is not" and 6 being "very definitely is"), while distress was rated as a source of pressure using the same scale.

In the low-stress task, participants were given 40 minutes to prepare a slide deck for a presentation about their favorite movie, TV series, or book, which was a familiar topic that allowed them agency over the task. The allotted time and topic had been previously piloted, ensuring that participants had ample time to complete the assignment with no time constraints or pressure. After a break, participants were given 30 minutes to prepare a new presentation on an unfamiliar topic - the scientific and philosophical achievements of two ancient Greek philosophers. The high-stress task was carefully designed to create time pressure and an unfamiliar workload. Participants were informed that they would present their work to a committee at the end of the experiment to encourage them to take the task seriously.

Additional external stressors were added during the high-stress task. Participants turned on their video cameras and shared their screens via Zoom with a confederate posing as

a professor with expertise in optimizing work settings for office workers. The confederate informed participants that he would monitor their work and reduce their score whenever he noticed suboptimal performance. Participants were told that their final score would be compared to others in the study, with the highest scorers receiving the highest compensation (\$50) and the lowest scorers receiving minimal compensation (\$5). However, at the end of the high-stress task, participants were debriefed and informed that the confederate was not a professor, and their score had no impact on their compensation. All participants received the maximum compensation. Participants were also informed that they would not actually present their work to a committee, and the task was designed to push them to perform to the best of their abilities.

### 3.3 Feature Extraction

To analyze the HRV data, we used the Kubios software package [28], which provides accurate and detailed HRV analysis and extracts the time and frequency-domain indices of the heart rate signal for every time window. We applied a medium level of artifact correction that identifies R-R intervals varying above or below 0.25 seconds compared with the average. This method helps to preserve the variability of the data while addressing the presence of any artifacts. Kubios also uses a piecewise cubic spline interpolation method to generate corrupted or missing values, resulting in a cleaner and more accurate HRV signal. It is noteworthy that the RR-interval, which represents the time between successive R-peaks in heart rate analysis, was excluded from the feature set. This decision was made to prevent feature duplication, given the direct relationship between RR-interval and heart rate. Generally, the RR-interval and heart rate are inversely proportional, with their product being a constant value of 60,000 (HR x RR interval = 60,000). This relationship was further confirmed in our dataset, as there was a strong 94% correlation between these two features.

Before feature extraction, BVP and ST signals were filtered

using winsorization [29], a statistical technique to remove outlier values by replacing extreme values beyond the 2th and 98th percentiles. We used this method to clean the noisy BVP and ST signals collected from the Empatica E4, as done in [30]. For EDA data, we utilized the MATLAB Le-dalab toolbox [31], which provides various functions to clean and process EDA data. We applied a series of signal processing techniques, including a Butterworth low-pass filter, Hanning smoothing with a window size of 4 samples, and manual artifact correction to remove any noise introduced by movement or other sources of interference.

Following this cleaning procedure, we computed various statistical parameters including the mean, standard deviation, median, minimum, maximum, 25th and 75th percentiles, and the slope of BVP, EDA, and ST. Our analysis focuses on these specific statistical parameters as they have been used in previous studies related to stress detection, demonstrating their relevance and effectiveness in detecting stress levels, thus providing a comprehensive evaluation of the different aspects of stress appraisal [12]. All physiological related-features were subtracted from the corresponding experimental phase's 5 minutes baseline for each participant. Also, the x, y, and z wrist accelerations were calculated for every time window.

We used OpenFace [32] to extract participants' mean and standard deviation of facial action unit (AU) intensities from the RGB video recorded by the Kinect camera. AUs are pre-defined facial muscle movements that correspond to emotions and are categorized as main AUs, head movement AUs, and eye movement AUs. Facial expressions are an excellent indicator of stress, making them suitable for stress detection research [22]. We excluded the head translation vector in the x, y, and z planes from the analysis because it was dependent on the participant's height and position in the camera frame. We also dropped head rotation in the x and y planes due to high interdependence with the gaze vector, resulting in redundant features. A correlation analysis supported this finding, indicating a close relationship between these variables (Pearson correlation between 89% and 94%). By removing these features, we avoided duplicating information in our analysis.

Finally, keyboard strokes and left and right mouse clicks were aggregated for the predefined time windows. While these measures may not be directly related to physiological changes associated with stress, they are known to be affected by cognitive and emotional states and can reflect changes in work-related stress levels. The inclusion of keyboard strokes and left and right mouse clicks as features in a dataset aimed at predicting stress in an office setting is a relatively novel approach that has shown promising results in recent studies [33], [34].

### 3.4 Data Processing

Due to technical errors, some sensors failed to collect data

resulting in missing data for some participants. Keyboard and mouse files were missing for three participants during the low-stress condition, and RGB video files were missing for two others during the high-stress condition. To impute the missing data, we trained an XGBoost model using data from 43 participants with complete data. We optimized the model by tuning hyperparameters such as learning rate, maximum depth of trees, and number of trees through cross-validation. Using the optimized XGBoost model, we predicted the missing data points for keyboard, mouse, and RGB video files. This method is accurate and preserves the standard deviation and shape of the feature distribution, avoiding data loss due to deletion of rows with missing entries [35].

Depending on the window size, the physiological and behavioral dataset comprised of 48 participants  $\times$  70 minutes per participant  $\times$  1/window size. For instance, considering a 30 second time window, the total number of datapoints would be:  $48 \times 70 \times 1/(0.5\text{min}) = 6720$  datapoints. The final dataset included 83 features including 34 physiological features, 48 behavioral features including 3 human-computer interactions, 39 facial-related features and 6 features for the hand wrist acceleration, and 1 feature indicating the participant's gender. All features were normalized using min-max scaling, which involved a linear transformation of the original data to a range between 0 and 1. Table 1 presents a summary of all the features included in our analysis.

Table 1  
Features Dataset

Type (Number of features)	Signal	Features Included
Physiological (34)	EDA Blood Volume Pulse Skin Temperature	Mean, Standard deviation, Median, Minimum, Maximum, 25 <sup>th</sup> & 75 <sup>th</sup> percentile, slope fitted through the data.
	Heart Rate and HRV	Mean, Standard deviation, Minimum, Maximum, rmsdd, LF peak, HF peak, LF power, HF power, LF/HF
Behavioral (48)	Facial action units Head Rotation Eye gaze direction	Mean, Standard deviation
	Blink	Count
	Hand wrist acceleration	Mean, Standard deviation
	Mouse right clicks Mouse left clicks Keyboard keystrokes	Count
	Gender (1)	Female vs Male
		Binary

Participants' ratings of stress level were each subtracted from the rating provided at the end of the corresponding resting period resulting in stress arousal ratings ranging from -100 to 100. Appraisals of eustress and distress were transformed into a binary outcome. "Stress is not appraised

response from the 3 categories of "very definitely is not a source of," "definitely is not a source of," and "generally is not a source of." Similarly, the 3 categories of "very definitely is a source of," "definitely is a source of," and "generally is a source of" were grouped into "stress appraised as" eustress or distress.

### 3.5 Metrics for prediction assessment

Eustress and distress were transformed into binary outcomes for ML analysis, a classification problem. Metrics used to evaluate prediction performance included accuracy, unweighted (average) precision, recall, and  $F_1$  score, which accounts for class imbalance. Each model presented in the results section was assessed using the leave-one-person-out cross validation method, a technique commonly used in machine learning to build models that are robust and generalizable.

## 4. RESULTS

To provide a foundation for our findings, we first present a variety of descriptive data related to stress appraisal across and within the two experimental conditions. While the overall perceived stress level among all participants for both work conditions was relatively low ( $M=13.96$ ,  $SD=20.24$ ), our experiment induced stress given that the perceived stress level was on average 13.96 points higher than the baseline. In addition, the low-stress work condition induced more eustress ( $N=2,110$ ) among participants than distress ( $N=1,380$ ). On the other hand, the high-stress work condition resulted in almost equal eustress ( $N=2,390$ ) and distress ( $N=2,220$ ) datapoints. Despite the low-stress condition being 10 minutes longer than the high-stress condition, the distress datapoints in the latter ( $N=2,220$ ) were significantly higher than the former ( $N=1,380$ ), as evidenced by the significant results of the chi-squared analysis ( $\chi^2(df=1, N=6,720) = 1,120$ ,  $p<0.001$ ). Fig. 2 provides a summary of the eustress and distress datapoints distribution across both conditions.

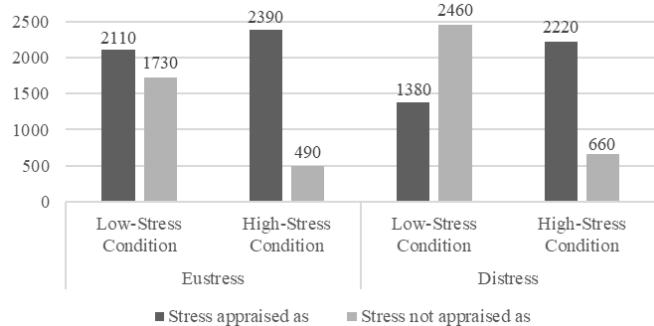


Fig. 2. Distribution of eustress and distress across the high- and low-stress conditions

### 4.1 Perceived stress levels variation across eustress and distress conditions

To answer our first research question, we conducted two

Authorized licensed use limited to: University of Southern California. Downloaded on October 23, 2023 at 16:05:33 UTC from IEEE Xplore. Restrictions apply.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

independent t-tests that examined how perceived stress level (i.e., arousal) changed as a function of eustress and distress (i.e., valence). The first test investigated the effect of eustress appraisal on stress arousal. The results show a significant effect of eustress appraisal on the stress arousal ( $t(6718)=-17.44$ ,  $p<0.001$ ); the stress arousal was significantly higher when datapoints were labeled as "stress appraised as eustress" ( $M=16.92$ ,  $SD=21.74$ ) in comparison to the data points labeled as "stress not appraised as eustress" ( $M=7.96$ ,  $SD=15.11$ ). The second test examined the effect of distress appraisal on stress arousal. The results show a significant effect of distress appraisal on stress arousal ( $t(6718)=-28.05$ ,  $p<0.001$ ), more specifically stress arousal was significantly higher with datapoints labeled as "stress appraised as distress" ( $M=20.06$ ,  $SD=22.59$ ) compared to the stress arousal associated with datapoints labeled as "stress not appraised as distress" ( $M=6.92$ ,  $SD=14.19$ ).

### 4.2 Comparison between different ML models

We answered our second question by investigating which ML model is best suited for predicting eustress and distress using the 83 features in our dataset. We examined ten models, including Naïve Bayes (NB), K-nearest neighbor (K-NN) (K-values between 3 and 15), Support Vector Machine (SVM) with different kernels, Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Multi-Layer Perceptron (MLP), and Logistic Regression (LR) as well Long Short-Term Memory (LSTM) and a combination of Convolutional Neural Network (CNN) and LSTM models. Both LSTM and CNN-LSTM were implemented with the Keras Sequential API. The LSTM model consisted of a single LSTM layer with 64 units, followed by a dense layer with sigmoid activation for binary classification. The input shape was determined by the time steps in the training data and a single feature dimension. The model was optimized using binary cross-entropy loss and Adam optimizer. The CNN-LSTM model included one-dimensional convolutional layers, followed by an LSTM layer and a dense output layer. It employed 64 filters in the convolutional layers with a kernel size of 3 and ReLU activation. A max-pooling layer and dropout were applied to reduce overfitting.

Fig. 2 reveals a somewhat unbalanced distribution of the eustress and distress classes. Although unbalanced datasets can pose challenges for classification problems, the degree of imbalance in this binary distribution is not severe enough to require statistical intervention. To confirm our assumption, we conducted ML analysis both with and without data augmentation. The results showed comparable performance between the datasets, with the augmented dataset exhibiting only a small increase of 2% in accuracy and approximately 3% for the  $F_1$ -score across all ML models. Thus, the accuracy, precision, recall, and  $F_1$ -score reported in Table 2 are based on the actual dataset without any augmentation.

Table 2  
Comparison of binary classifiers' performance in the prediction of eustress and distress

	NB	LR	SVM	KNN	MLP	LSTM	CNN-LSTM	DT	RF	XGB oost
Eustress										
Accuracy(%)	68.14	70.32	79.25	83.55	79.29	79.86	80.03	82.21	84.29	<b>85.65</b>
Precision(%)	62.74	65.52	78.07	82.13	79.30	78.42	78.07	81.59	84.47	<b>85.24</b>
Recall(%)	61.74	63.41	72.92	80.79	71.81	73.10	72.92	78.10	80.42	<b>81.60</b>
F <sub>1</sub> -score(%)	62.23	64.44	75.40	81.45	75.37	75.66	75.40	79.81	82.39	<b>83.38</b>
Distress										
Accuracy(%)	54.03	65.14	70.16	70.73	70.89	75.45	78.35	72.24	74.92	<b>78.90</b>
Precision(%)	53.53	65.20	71.56	71.92	72.29	76.06	78.37	72.12	75.31	<b>79.21</b>
Recall(%)	53.23	64.70	69.29	69.50	69.92	74.60	77.72	72.20	74.03	<b>78.38</b>
F <sub>1</sub> -score(%)	53.38	64.94	70.40	70.68	71.08	75.32	78.04	72.16	74.66	<b>78.79</b>

### 4.3 Comparison between different window sizes

This section answers the third research question and presents our findings on the optimal window size for training eustress and distress prediction models. In our analysis, we considered four different window sizes, namely 30 seconds, 1 minute, 2.5 minutes, and 5 minutes, with corresponding datasets of 6720 datapoints, 3360, 1344, and 672 for each window size respectively. We chose a 30-second time window based on the recommendation of Bernardes et al. [36], who found that this is the smallest time frame that can reliably capture HRV features that accurately assess psychological stress. Furthermore, we chose 1 minute as it is a commonly used window size in previous studies on stress prediction [22]. The 2.5-minute window size was chosen to capture a longer period of signals, which may provide additional information for predicting eustress and distress. Finally, given that participants received a new questionnaire every 5 minutes, a timeframe of 5 minutes was determined to be the maximum feasible window size. The results presented in Table 3, are based on training an XGBoost model using all 83 features.

Table 3  
Window size analysis for binary classification of eustress and distress

	Eustress				Distress			
	30 sec	1 min	2.5 min	5 min	30 sec	1 min	2.5 min	5 min
Accuracy(%)	<b>85.65</b>	83.16	82.15	84.23	<b>78.90</b>	78.27	76.35	77.24
Precision(%)	<b>85.24</b>	82.75	78.88	82.96	<b>79.21</b>	77.75	75.97	77.29
Recall(%)	<b>81.60</b>	80.17	83.75	84.52	<b>78.38</b>	76.96	77.11	75.05
F <sub>1</sub> -score(%)	<b>83.38</b>	81.44	81.24	83.37	<b>78.79</b>	77.35	76.53	76.15

### 4.4 Comparison between different modalities

To answer question four, we trained different ML models to determine how different data modalities affect the prediction performance of eustress and distress. Since XGBoost resulted in the highest accuracies among the ML models, all analyses conducted from this point forward used XGBoost. The results in Table 4 show that the combination of physiological and behavioral features resulted in the highest prediction accuracy and F<sub>1</sub>-scores for both eustress and distress.

Table 4  
XGBoost results for binary classification of eustress and distress

	Eustress			Distress		
	Physio	Behavior	Combined	Physio	Behavior	Combined
Accuracy(%)	83.45	73.38	<b>85.65</b>	74.83	72.35	<b>78.90</b>
Precision(%)	81.00	69.81	<b>85.24</b>	75.48	72.32	<b>79.21</b>
Recall(%)	80.26	67.48	<b>81.60</b>	73.63	71.28	<b>78.38</b>
F <sub>1</sub> -score(%)	80.63	68.62	<b>83.38</b>	74.54	71.79	<b>78.79</b>

Next, we employed SHAP feature importance analysis [37] to identify the most influential physiological and behavioral features for predicting eustress and distress. Fig. 3 displays the feature importance analysis for the eustress and distress binary models using combined physiological and behavioral data. Only the top 15 features are shown, as including additional features showed a negligible improvement in performance. Using only the top 15 features led to a slight decline in performance when compared to the full models. The accuracy and F<sub>1</sub>-score for eustress decreased from 85.65% and 83.38% to 83.99% and 82.11% respectively, while for distress, the accuracy and F<sub>1</sub>-score decreased from 78.90% and 78.79% to 76.19% and 75.40% respectively.

### 4.5 Gender-based models

This section answers the fifth question of the study. Fig. 3 shows that gender was the second most important feature in the prediction of eustress. Therefore, we decided to build gender-based stress appraisal models. The distribution of the eustress and distress binary variables based on gender is presented in Fig. 4 below.

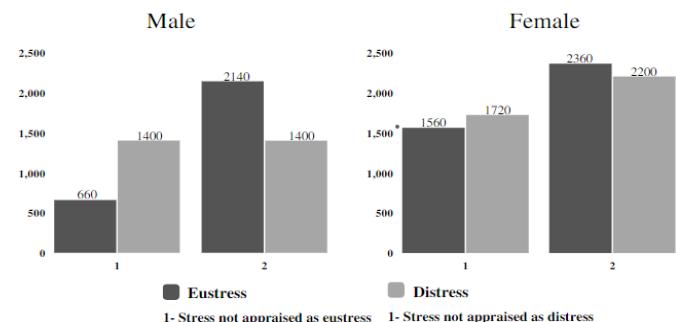


Fig. 3: Gender-based distribution of eustress and distress variables

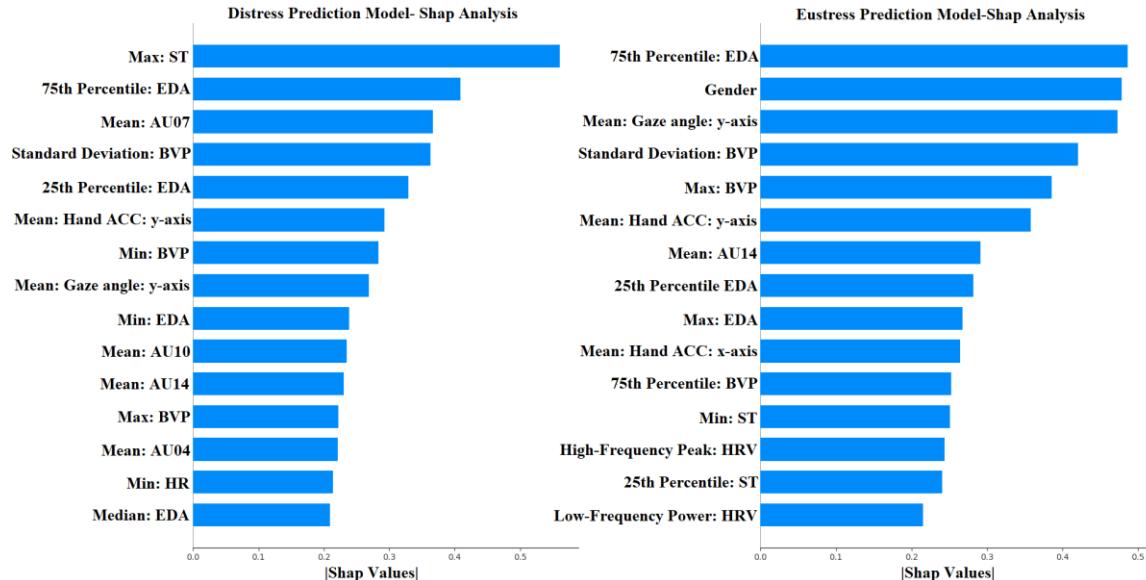


Fig. 4. Feature importance for the distress and eustress binary classification models

We created gender-based models by dividing our initial dataset of 6720 datapoints (30 second time window) into two subsets: one for males (2800 datapoints) and one for females (3920 datapoints). XGBoost models were trained for each dataset, incorporating all available features within a 30-second time window. Table 5 presents the results of these gender-based models.

Table 5  
Gender-based binary eustress and distress prediction models

	Eustress		Distress	
	Male	Female	Male	Female
Accuracy(%)	86.65	88.02	79.74	80.12
Precision(%)	81.19	87.25	78.50	79.80
Recall(%)	84.10	86.35	79.43	77.98
F1-score(%)	82.62	86.80	78.96	78.88

#### 4.6 Differentiating between eustress, distress, eustress-distress coexistence, and no-stress

Individuals in the workplace can perceive stressors differently, resulting in varying levels of eustress and distress [38]. For example, a worker may feel pure eustress when leading a successful project, but pure distress when dealing with limited work resources or a toxic work environment. While eustress and distress can be experienced separately, they can also coexist in the workplace. For instance, a worker may experience pressure to meet a deadline (eustress) while also feeling overwhelmed by workload (distress). Conversely, individuals may experience no stress at all when they're disengaged or bored at work, which can negatively impact their performance and well-being. An administrative assistant, for example, may feel no eustress or distress when performing repetitive tasks, leading to feelings of disengagement or apathy towards their work.

This coexistence of eustress and distress highlights the need for a more comprehensive understanding of workplace stress. Thus, after exploring the distinct concepts of

eustress and distress, creating prediction models for each, and identifying the physiological and behavioral characteristics that best represent them, we developed a comprehensive model to predict the simultaneous assessment of both types of stress. Our model aims to capture not only the presence of eustress and distress but also their simultaneous appraisal, enabling a more nuanced understanding of the complex experiences individuals face in the workplace. Therefore, we developed a new outcome measure that distinguishes between pure eustress, pure distress, eustress-distress coexistence, and the absence of stress. We reverted back to the binary formulation of eustress and distress, defining "Eustress" as stress appraised as eustress but not distress, "Distress" as stress appraised as distress but not eustress, "Eustress-distress coexistence" as stress appraised as both eustress and distress, and "No stress" as stress not appraised as either eustress or distress. Table 6 presents the formulation of stress appraisal states.

Table 6  
Simultaneous stress appraisal formulation and data distribution

Eustress Appraisal	Distress Appraisal	Stress Appraisal	Datapoints
Stress not appraised as eustress	Stress not appraised as distress	No stress	1890
Stress appraised as eustress	Stress not appraised as distress	Eustress	1230
Stress appraised as eustress	Stress appraised as distress	Eustress-distress coexistence	3270
Stress not appraised as eustress	Stress appraised as distress	Distress	330

The resulting dataset is imbalanced, as approximately 50% of cases reflected Eustress-distress coexistence, while Distress was only identified in 5% of cases. To address the issue of imbalanced classes, we utilized an oversampling technique by employing the synthetic minority oversampling technique (SMOTE) algorithm [39]. This technique involves

generating new synthetic samples in the minority classes by selecting a random sample from the minority class, identifying the k-nearest neighbors, and creating synthetic data points in the direction of the vector that connects the minority instance and its neighbors. It is worth noting that the SMOTE algorithm was applied solely to the training set, not the testing set. For this analysis, we utilized the XGBoost algorithm to build our predictive model and incorporated all 83 features with a 30-second time window. The XGBoost model achieved a moderate classification performance, with an accuracy of 74.42%, precision of 66.78%, recall of 63.55% and  $F_1$  score of 65.12%.

## 5. DISCUSSION

### 5.1 Perceived stress level variation across eustress and distress conditions

When participants indicated having a eustress feeling, their stress arousal was significantly higher than with a non-eustress feeling. Similarly, participants experiencing a distressing feeling showed significantly higher stress arousal compared to the non-distress feeling. However, a distressful situation was considerably more intense than situations that elicited a eustress feeling as the former led to a substantially higher level of stress arousal. Hans Selye defined stress as the body's response to a certain demand but distinguished between eustress and distress [20]. He denoted stress as arousal and explained that whenever stress arises, the question becomes about its valence and whether the stressed individual perceives it as positive or negative. The results from the t-tests are in accordance with Selye's definition of stress; both eustress and distress were associated with an increase in perceived stress arousal.

### 5.2 Comparison between different ML models

Naïve Bayes classifier had the weakest  $F_1$ -score for eustress (62.23%) and distress (53.38%), likely due to its assumption that the presence of a particular feature is independent of all other features, which is not applicable to physiological and behavioral features that are interdependent. Logistic regression also showed weak  $F_1$ -score for eustress (64.44%) and distress (64.94%) likely because the target classes have no linear correlation with the features. In contrast, decision tree, K-NN, SVM, and MLP models had fair to good prediction accuracy. The best K-NN model with a K value of 3 achieved  $F_1$ -scores of 81.45% for eustress and 70.68% for distress. Among SVM models, the polynomial kernel in the 6<sup>th</sup> degree led to the best  $F_1$ -score of 75.40% for eustress and 70.40% for distress. MLP also showed good  $F_1$ -scores of 75.37% for eustress and 71.08% for distress, despite its relatively longer training time for large datasets.

The LSTM model achieved an  $F_1$ -score of 75.66% for eustress and 75.32% for distress. Additionally, the CNN-LSTM model showed similar performance to the LSTM model in terms of  $F_1$ -scores, achieving  $\approx$ 75% for eustress and (75-

78%) for distress. These results indicate that combining convolutional and recurrent layers is an effective approach for predicting stress appraisal from physiological and behavioral signals, as it captures both local and temporal dependencies, outperforming other models in the study.

The best-performing models were XGBoost (eustress: 83.38%, distress: 78.79%) and random forest (eustress: 82.39%, distress: 74.66%), with slightly better performance for XGBoost. XGBoost is an optimized gradient boosting technique that builds decision trees sequentially and penalizes underperforming leaves. In contrast, random forest combines multiple decision trees using bagging. XGBoost, by learning from previous mistakes, can capture complex patterns and outperforms most classification algorithms. Another study by Hsieh et al. [40] ( $F_1$ -score: 89%) also identified XGBoost as the most effective algorithm to distinguish between stress and amusement states. However, it is important to note that their study employed a different experimental design from ours. In contrast, our findings highlight the importance of assessing individual appraisal of work conditions as a source of pressure. Therefore, while the results of [40] are impressive, our study provides novel insights into the context of work-related stress.

Finally, our study found that ML models can predict eustress and distress with reasonable accuracy and  $F_1$ -scores. However, both metrics were lower for distress, which may be due to its complexity and the influence of contextual factors. Response bias [41] may also have played a role, as participants may have under-reported their distress levels to appear competent, resulting in misalignment between actual and reported distress and lower performance for the distress model compared to the eustress model.

### 5.3 Comparison between different window sizes

Our results suggest that shorter window sizes may capture more fine-grained fluctuations in the physiological and behavioral signals and result in a slightly more accurate prediction. However, the differences in accuracy and  $F_1$ -scores between the tested window sizes are relatively small. Within the range of window sizes that were tested, the choice of window size may not be critical for achieving good model performance. Additionally, it is important to acknowledge that emotions may not fluctuate as rapidly as within the short time frames tested in our experiment. While our study provides valuable insights, it is limited to a controlled laboratory experiment with a short duration of 70 minutes. As such, longitudinal data collection in real-world office environments is necessary to determine how eustress and distress develop over time and how well prediction models perform in such settings. This will also allow for the exploration of larger window sizes and their effectiveness in capturing changes in stress appraisal over longer time periods. Further research is needed to fully understand the complexities of stress appraisal and its

prediction in real-world settings. Nonetheless, our findings offer important validation and evidence that predicting eustress and distress is possible.

#### 5.4 Comparison between different modalities

Previous studies, focusing on detecting stress as an arousal state, demonstrated that a combination of behavioral and physiological signals leads to higher ML prediction performance. For instance, Koldijk et al. [22] showed that a combination of physiological, facial, and computer interaction features led to the highest accuracy in differentiating stressful from non-stressful work conditions for office workers. Our results show that the same conclusion holds for determining stress appraisal as distress and eustress. In the study of Li et al. [13], a composite of features derived from smartphone and computer usage, along with heart rate data, was utilized to identify occurrences of eustress in a naturalistic data setting. The authors' findings revealed a prediction accuracy of 71.33%, albeit their study population was confined to a mere seven participants. Our laboratory-based results, however, suggest that a blend of facial features and physiological measures beyond heart rate may serve as stronger indicators of eustress reaching a prediction accuracy of 85.65% and an  $F_1$ -score of 83.38%.

The combination feature set resulted in only a slight increase in performance (2-4%) compared to the physiological feature sets for eustress and distress predictions, while the behavioral feature set showed a larger increase (accuracy<sub>eustress</sub>:12%,  $F_1$ -score<sub>eustress</sub>:15% and accuracy<sub>distress</sub>:6%,  $F_1$ -score<sub>distress</sub>:7%). These findings suggest that physiological features may be more informative than behavioral features for predicting eustress and distress. However, further research is needed to fully evaluate the relative importance of each feature set. These results have practical implications for researchers interested in implementing this framework. If high prediction performance is crucial, a combination of features may be necessary, but this would require significant financial and computational resources to acquire and analyze the data. Alternatively, relying on a unimodal framework with physiological features can provide good prediction performance, comparable to the combination feature set.

Upon examining the SHAP plots, a clear contrast emerged between the dominant predictors for eustress and distress. Notably, physiological data played a prominent role in predicting eustress, as 10 out of the top 15 features were physiological, whereas only 4 were behavioral, and gender was the final feature. Conversely, in predicting distress, 6 behavioral features were among the top 15, which explains why the performance of behavioral-based models (72.35%) was relatively comparable to that of physiological-based models (74.83%). However, this trend, as shown in Table 3, did not hold for eustress prediction models (Behavioral = 73.38%, Physiological = 83.45%). These findings highlight

the importance of considering the distinct predictors for eustress and distress, particularly in developing effective prediction models.

Our study found that EDA, BVP, and ST were the most important physiological features for predicting both eustress and distress. This confirms previous research, which suggested that EDA is a strong indicator of stress but is not enough on its own to differentiate between eustress and distress [42]. Our study uniquely shows that ST and BVP are also important predictors of stress appraisal. In addition to these features, heart rate (minimum heart rate) and HRV features (high frequency bands) were among the most important features for unveiling stress appraisal. Our review of the literature showed that only one attempt has been made to determine when an office worker is feeling eustressed [13], with similar findings of importance as in our study.

The identification of brow lowering (AU04), lid tightening (AU07), and upper lip raising (AU10) as predictors of distress is consistent with prior research linking these action units with negative emotions such as anger, fear, sadness, and worry, which are commonly observed in response to threatening or stressful events [43]. Interestingly, AU14, which is not typically associated with emotional expression, has been found to predict both eustress and distress. This finding is noteworthy, as it suggests that the presence of AU14 may reflect a sense of enjoyment or pleasure, consistent with the experience of eustress. Alternatively, it may also reflect a coping mechanism or an attempt to maintain a positive mood in the face of adversity. Our study identified a unique finding of gaze angle in the y-direction, but we acknowledge that head movement and gaze cannot be interpreted separately from body posture, which we did not examine. Yang et al. [44] argue that head movements are typically associated with gaze drifting and body movements, highlighting the need for further investigation into body posture to obtain a complete understanding of the behavioral characteristics of eustress and distress.

We added productivity-related features to our model, considering the impact of eustress and distress on workers' productivity [8]. Wrist accelerations in the x and y planes emerged as significant predictors of both eustress and distress. This finding is consistent with Holder et al.'s argument that hand acceleration captured by the Empatica E4 is a crucial factor in predicting stress arousal [45], reflecting people's engagement and performance [46], which is related to the impacts of both eustress and distress on engagement and excitement during work. In contrast to previous studies, our results did not find keyboard strokes and mouse clicks as crucial features for predicting eustress or distress. Nonetheless, these metrics may be relevant in real office settings or different types of office tasks [22]. Future work could examine other HCI features like keystroke

pressure, gaze duration, and application usage to enhance the prediction of eustress and distress.

### 5.5 Gender-based models

When employing gender-based models, eustress prediction performance improved more considerably than distress prediction performance. For males and females, the eustress prediction accuracy was 86.65% and 88.02%, respectively, compared to the 85.65% prediction accuracy for the generalized eustress model. The male and female groups' distress prediction accuracy was 79.74% and 80.12%, respectively, compared to the 78.90% prediction accuracy for the generalized distress model. However, when examining the  $F_1$ -scores, the only notable improvement was observed for the eustress category in the females' group with an  $F_1$ -score of 86.80% in comparison to 83.38% for the generalized model. This larger increase in the eustress model is consistent with the SHAP results. Gender was not among the top predictors of distress, but it was the second most important feature in predicting eustress, as shown in Fig. 3.

Gender is an important factor in the way people perceive and respond to stress. Stress is a complex biological and psychological phenomenon, and research has consistently found gender differences in the physiological, cognitive, and behavioral responses to stressors. For instance, women tend to have stronger physiological responses to stress than men, including a higher heart rate and blood pressure [47], which may be attributed to hormonal factors. Additionally, societal expectations and gender roles can influence how men and women perceive and respond to stressors at work [25], leading to differences in coping strategies and outcomes. These gender differences may also extend to eustress and distress, with studies suggesting that men and women may experience different types of stressors that elicit either eustress or distress [9], [48].

In recent years, ML and predictive modeling techniques have been used to develop tools for detecting and predicting stress. These models often incorporate gender as a feature to address the effect of gender on stress detection and acquire better prediction performance [49], [34]. In the present study, separating the eustress and distress prediction models by gender led to an improvement in prediction performance, suggesting that gender-specific differences play an important role in stress appraisal.

It is important to note that stress appraisal is influenced by a variety of personal and contextual factors, and gender is just one of these factors. Other personal characteristics, such as age, personality traits, work type, and coping styles, may also play a role in stress appraisal and response. By creating more group models based on personal characteristics, we may be able to further improve the performance of machine learning techniques in stress appraisal. Building on that, future research should continue to explore the role

of gender and other personal characteristics in stress appraisal and response, to further enhance our understanding of this complex phenomenon.

### 5.6 Differentiating between eustress, distress, eustress-distress coexistence, and no-stress

The results of the classification problem involving four stress appraisal classes indicate an overall decline in performance (accuracy=74.42%,  $F_1$ -score=65.12%) when compared to the binary classification problems for eustress (accuracy=85.65%,  $F_1$ -score=83.38%) and distress (accuracy=78.90%,  $F_1$ -score=78.79%). These observed differences in performance can be attributed to various factors inherent to the nature of the classification tasks. Firstly, the binary classification task inherently possesses a simpler structure compared to the multi-class classification problem, as it involves distinguishing between only two classes. In contrast, the 4-class classifier is burdened with the intricate task of differentiating among four distinct classes. This increased complexity of the multi-class problem poses greater challenges for the classifier in accurately classifying instances. Secondly, the presence of class imbalance can significantly impact classifier performance. While the balanced distribution of classes in the binary classification problem may contribute to higher accuracy and  $F_1$  score, imbalanced class distributions in the 4-class classification problem, particularly when certain classes have significantly fewer instances, can adversely affect overall classifier performance. The minority classes (i.e., distress class), being underrepresented, may prove more difficult to accurately classify, leading to lower scores. Lastly, the overlapping features among classes in a multi-class classification scenario introduce inherent ambiguity and elevate the difficulty in correctly classifying instances. In our case, there might be a potential overlap between the "Eustress-distress coexistence" class and the "Eustress" and "Distress" classes, introducing some classification errors. Conversely, binary classification problems often exhibit more distinct boundaries between the two classes, facilitating the classifier's discrimination process.

The study conducted by Setz et al. [25] aimed to differentiate between stress and cognitive load using electrodermal activity (EDA) data collected from 33 subjects in a laboratory experiment. Although not directly related to eustress and distress, their work is similar to our study's objective of distinguishing between different stress states. Multiple ML models, including linear discriminant analysis, SVM, and nearest class center, were tested, with the highest accuracy of 82.8% achieved. This research presents a comparable analysis between cognitive load and eustress and stress and distress. However, our study contributes to the literature by identifying pure eustress, pure distress, eustress-distress coexistence, and the absence of stress. We have expanded the classification beyond the binary categorization of stress and cognitive load to include four different stress

appraisal states.

Our approach has important implications for workplace settings where stress is prevalent. By distinguishing between eustress and distress, managers and supervisors can intervene early to prevent negative emotions from escalating. Additionally, our ability to detect eustress-distress coexistence is valuable in identifying mixed emotional states that are difficult to discern through self-report measures. This information can facilitate targeted interventions that help individuals develop coping strategies and reframe negative emotions. These findings have practical significance for the development of affective computing systems that can accurately detect and differentiate between different emotional states in real-time. By using a combination of physiological and behavioral features, our approach represents a significant step forward in the field of affective computing. It has the potential to be applied in a variety of contexts, including workplace stress management, mental health monitoring, and personalized healthcare.

## 6. LIMITATIONS & FUTURE RESEARCH

While this study presents the first attempt to employ ML for differentiating eustress and distress, it also has some limitations. First, although the experimental procedure was designed to simulate stressful office work, participants were assigned predesigned tasks and were put under extreme work conditions (i.e., zoom monitoring, compensation withhold). Hence, this experiment falls short of mimicking the dynamics and complexity of office work. To that end, future research directions should examine office workers' eustress and distress in their naturalistic work environments. Second, the proposed ML models presented in this paper did not consider the full personalized experience of stress and only accounted for gender as a moderating prediction feature. Eustress and distress appraisal are affected by various personal characteristics; what is considered as eustress for one person can be distress for another. For that, future research should incorporate personal characteristics (e.g., age, personality traits) while building automated prediction frameworks for eustress and distress or establishing personalized and unique ML models for groups of workers following their personal characteristics. Finally, our results showed that both head movement and gaze were important predictors of eustress and distress, which hints at the importance of incorporating body posture in future research studies to differentiate between eustress and distress appraisal.

## 7. CONCLUSIONS

This study represents the first attempt to employ an ML framework to predict eustress and distress in an experimental setting. The study mimicked different work settings with two stress conditions: low-stress and high-stress work.

Physiological and behavioral signals were used in establishing the prediction models. Results show that the perception of distress is associated with a higher level of subjective stress arousal than the perception of eustress. The XGBoost classifier had the best prediction performance for both eustress and distress compared to nine other classifiers. Using this ML model along with a window size of 30 seconds, the combination of physiological and behavioral features led to 85.65% and 78.90% accuracy in predicting eustress and distress, respectively. Additionally, the results indicate that gender plays a role in predicting eustress and distress conditions, with a potentially higher influence in predicting eustress than distress. Finally, we developed a model to predict the simultaneous assessment of eustress and distress, distinguishing between pure eustress, pure distress, eustress-distress coexistence, and the absence of stress. The developed model achieved a moderate accuracy of 74.42% and F<sub>1</sub>-score of 65.12%.

This study presents promising findings that can be integrated with work management practices to minimize work distress and promote eustress among office workers. Personal factors play a major role in how workers perceive the stress associated with their work tasks. Thus, eustress-distress prediction models could help work managers effectively design, tailor, and assign work duties among office workers with the aim of maximizing eustress at the expense of distress. Also, implementing this framework may be useful for promoting self-awareness among workers about their negative stress levels and the specific work conditions that increase their distress. Finally, such a framework could be coupled with a notification system to alert workers about prolonged distress experiences and provide them with appropriate intervention suggestions that limit unhealthy distress exposure.

## ACKNOWLEDGMENT

This publication was supported by the Pilot Project Research Training Program of the Southern California NIOSH Education and Research Center (SCERC), Grant Agreement Number T42 OH008412 from the Centers for Disease Control and Prevention (CDC), the National Science Foundation under Grants No. 1763134 & 2204942, and by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-20-2-0053. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC, the Army Research Office or the National Science Foundation. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## REFERENCES

[1] G. Fink, "Stress: Concepts, Definition and History★" in

*Reference Module in Neuroscience and Biobehavioral Psychology*, Elsevier, 2017.

[2] American Psychological Association, "Stress in America: The state of our nation. Stress in AmericaTM Survey," 2017. [Online]. Available: <https://www.apa.org/news/press/releases/stress/2017/state-nation.pdf>.

[3] U.S. Bureau of Labor Statistics, "Occupational Employment and Wages, May 2020," 2021. <https://www.bls.gov/oes/current/oes430000.htm#nat>.

[4] S. B. Oswalt and C. C. Riddock, "What to do about being overwhelmed: Graduate students, stress and university services," *Coll. Student Aff. J.*, vol. 27, no. 1, pp. 24–44, 2007.

[5] American Psychological Association, "STRESS AND HEALTH DISPARITIES," 2017. [Online]. Available: <https://www.apa.org/pi/health-disparities/resources/stress-report.pdf>.

[6] Mental Health America, "Mind the Workplace," 2017. <https://www.mhanational.org/sites/default/files/Mind the Workplace - MHA Workplace Health Survey 2017 FINAL.pdf> (accessed May 07, 2021).

[7] G. Brulé and R. Morgan, "Working with stress: Can we turn distress into eustress," *J. Neuropsychol. Stress Manag.*, vol. 3, pp. 1–3, 2018.

[8] R. Kupriyanov and R. Zhdanov, "The eustress concept: problems and outlooks," *World J. Med. Sci.*, vol. 11, no. 2, pp. 179–185, 2014, doi: <http://dx.doi.org/10.5829/idosi.wjms.2014.11.2.8433>.

[9] R. Faizan and A. ul Haque, "Working Efficiency of Contrasting Genders under Eustress, Distress, Hyper-Stress, and Hypo-Stress," *Prabandhan Indian J. Manag.*, vol. 12, no. 11, p. 32, Nov. 2019, doi: 10.17010/pijom/2019/v12i11/148411.

[10] M. Le Fevre, J. Matheny, and G. S. Kolt, "Eustress, distress, and interpretation in occupational stress," *J. Manag. Psychol.*, vol. 18, no. 7, pp. 726–744, Nov. 2003, doi: 10.1108/02683940310502412.

[11] M. B. Hargrove, W. S. Becker, and D. F. Hargrove, "The HRD eustress model: Generating positive stress with challenging work," *Hum. Resour. Dev. Rev.*, vol. 14, no. 3, pp. 279–298, 2015.

[12] A. Alberdi, A. Azpiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *J. Biomed. Inform.*, vol. 59, pp. 49–75, 2016.

[13] C.-T. Li, J. Cao, and T. M. H. Li, "Eustress or distress," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, Sep. 2016, pp. 1209–1217, doi: 10.1145/2968219.2968309.

[14] S. Shoaib, B. G. Mujtaba, and H. M. Awan, "Overload Stress Perceptions of Public Sector Employees in Pakistan: a Study of Gender, Age, and Education in South Asia," *Public Organ. Rev.*, vol. 19, no. 3, pp. 311–324, Sep. 2019, doi: 10.1007/s11115-018-0405-y.

[15] S. Mittal, S. Mahendra, V. Sanap, and P. Churi, "How can machine learning be used in stress management: A systematic literature review of applications in workplaces and education," *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 2, p. 100110, Nov. 2022, doi: 10.1016/j.ijimedi.2022.100110.

[16] I. Rodríguez, M. W. Kozusznik, and J. M. Peiró, "Development and validation of the Valencia Eustress-Distress Appraisal Scale," *Int. J. Stress Manag.*, vol. 20, no. 4, pp. 279–308, Nov. 2013, doi: 10.1037/a0034330.

[17] A. Di Fabio, J. M. Peiró, I. Rodríguez, and M. W. Kozusznik, "The Valencia Eustress-Distress Appraisal Scale (VEDAS): Validation of the Italian Version," *Sustainability*, vol. 10, no. 11, p. 3903, Oct. 2018, doi: 10.3390/su10113903.

[18] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, "Validation of the Empatica E4 wristband," in *2016 IEEE EMBS International Student Conference (ISC)*, May 2016, pp. 1–4, doi: 10.1109/EMBISC.2016.7508621.

[19] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature," *Psychiatry Investig.*, vol. 15, no. 3, pp. 235–245, Mar. 2018, doi: 10.30773/pi.2017.08.17.

[20] A. M. Shahsavaran, E. Azad Marz Abadi, and M. Hakimi Kalkhoran, "Stress: Facts and theories through literature review," *Int. J. Med. Rev.*, 2015.

[21] W. Liao, W. Zhang, Z. Zhu, and Q. Ji, "A real-time human stress monitoring system using dynamic Bayesian network," in *IEEE computer society conference on computer vision and pattern recognition*, 2005, p. 70.

[22] S. Koldijk, M. A. Neerincx, and W. Kraaij, "Detecting Work Stress in Offices by Combining Unobtrusive Sensors," *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 227–239, Apr. 2018, doi: 10.1109/TAFFC.2016.2610975.

[23] S. Pourmohammadi and A. Maleki, "Stress detection using ECG and EMG signals: A comprehensive study," *Comput. Methods Programs Biomed.*, vol. 193, p. 105482, Sep. 2020, doi: 10.1016/jcmpb.2020.105482.

[24] S. D. W. Gunawardhane, P. M. De Silva, D. S. B. Kulathunga, and S. M. K. D. Arunatileka, "Non invasive human stress detection using key stroke dynamics and pattern variations," in *2013 International Conference on Advances in ICT for Emerging Regions (ICTer)*, Dec. 2013, pp. 240–247, doi: 10.1109/ICTer.2013.6761185.

[25] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert, "Discriminating Stress From Cognitive Load Using a Wearable EDA Device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010, doi: 10.1109/TITB.2009.2036164.

[26] "Polar Electro." <https://www.polar.com/us-en/sensors/h10-heart-rate-sensor/>.

[27] "Mini Mouse Macro," 2023. <https://sourceforge.net/projects/minimousemacro/reviews/>.

[28] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-aho, and P. A. Karjalainen, "Kubios HRV – Heart rate variability analysis software," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 210–220, Jan. 2014, doi: 10.1016/jcmpb.2013.07.024.

[29] M. Wu, "Trimmed and winsorized estimators," Michigan State University, 2006.

[30] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159–170, Sep. 2017, doi: 10.1016/j.jbi.2017.08.006.

[31] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *J. Neurosci. Methods*, vol. 190, no. 1, pp. 80–91, 2010.

[32] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Mar. 2016, pp. 1–10, doi: 10.1109/WACV.2016.7477553.

[33] Y. M. Lim, A. Ayesh, and M. Stacey, "Detecting cognitive stress from keyboard and mouse dynamics during mental arithmetic," in *2014 Science and Information Conference*, Aug. 2014, pp. 146–152, doi: 10.1109/SAI.2014.6918183.

[34] E. A. Sağbaş, S. Korukoglu, and S. Ballı, "Stress Detection via Keyboard Typing Behaviors by Using Smartphone Sensors and Machine Learning Techniques," *J. Med. Syst.*, vol. 44, no. 1, p. 3903, Oct. 2018, doi: 10.3390/su10113903.

4, p. 68, Apr. 2020, doi: 10.1007/s10916-020-1530-z.

[35] H. Kang, "The prevention and handling of the missing data," *Korean J. Anesthesiol.*, vol. 64, no. 5, p. 402, 2013, doi: 10.4097/kjae.2013.64.5.402.

[36] A. Bernardes *et al.*, "How Reliable Are Ultra-Short-Term HRV Measurements during Cognitively Demanding Tasks?", *Sensors*, vol. 22, no. 17, p. 6528, Aug. 2022, doi: 10.3390/s22176528.

[37] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems*, 2017, vol. 30, [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>.

[38] M. Kožusznik, I. Rodríguez, and J. M. Peiró, "Cross-national outcomes of stress appraisal," *Cross Cult. Manag. An Int. J.*, vol. 19, no. 4, pp. 507–525, Oct. 2012, doi: 10.1108/13527601211269996.

[39] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[40] C.-P. Hsieh, Y.-T. Chen, W.-K. Beh, and A.-Y. A. Wu, "Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection," in *2019 IEEE International Workshop on Signal Processing Systems (SiPS)*, Oct. 2019, pp. 330–335, doi: 10.1109/SiPS47522.2019.9020321.

[41] D. L. Paulhus, "Measurement and Control of Response Bias," in *Measures of Personality and Social Psychological Attitudes*, Elsevier, 1991, pp. 17–59.

[42] N. Tateyama, K. Ueda, and M. Nakao, "Development of an active sensing system for distress detection using skin conductance response," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Sep. 2019, pp. 1–6, doi: 10.1109/ACII.2019.8925442.

[43] M. B. Morshed *et al.*, "Advancing the Understanding and Measurement of Workplace Stress in Remote Information Workers from Passive Sensors and Behavioral Data," 2022, [Online]. Available: [https://www.researchgate.net/profile/Mehrab-Bin-Morshed-2/publication/363885542\\_Advancing\\_the\\_Understanding\\_an\\_d\\_Measurement\\_of\\_Workplace\\_Stress\\_in\\_Remote\\_Informatio\\_n\\_Workers\\_from\\_Passive\\_Sensors\\_and\\_Behavioral\\_Data/links/63336c8213096c2907d43f18/Advancing](https://www.researchgate.net/profile/Mehrab-Bin-Morshed-2/publication/363885542_Advancing_the_Understanding_an_d_Measurement_of_Workplace_Stress_in_Remote_Informatio_n_Workers_from_Passive_Sensors_and_Behavioral_Data/links/63336c8213096c2907d43f18/Advancing).

[44] J. Yang, K. Wang, X. Peng, and Y. Qiao, "Deep Recurrent Multi-instance Learning with Spatio-temporal Features for Engagement Intensity Prediction," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, Oct. 2018, pp. 594–598, doi: 10.1145/3242969.3264981.

[45] R. Holder, R. K. Sah, M. Cleveland, and H. Ghasemzadeh, "Comparing the Predictability of Sensor Modalities to Detect Stress from Wearable Sensor Data," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, Jan. 2022, pp. 557–562, doi: 10.1109/CCNC49033.2022.9700682.

[46] L. Potter, J. Scallon, D. Swegle, T. Gould, and G. E. Okudan Kremer, "Establishing a Link between Electrodermal Activity and Classroom Engagement," in *IIE Annual Conference. Proceedings*, 2019, pp. 988–993.

[47] C. M. Stoney, M. C. Davis, and K. A. Matthews, "Sex Differences in Physiological Responses to Stress and in Coronary Heart Disease: A Causal Link?", *Psychophysiology*, vol. 24, no. 2, pp. 127–131, Mar. 1987, doi: 10.1111/j.1469-8986.1987.tb00264.x.

[48] M. T. Matteson and J. Mulyanevich, *Controlling work stress: Effective human resource and management strategies*, Jossey-Bass, 1987.

[49] M. A. Rosales, A. A. Bandala, R. R. Vicerra, and E. P. Dadios, "Physiological-Based Smart Stress Detector using Machine Learning Algorithms," in *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Nov. 2019, pp. 1–6, doi: 10.1109/HNICEM48295.2019.9073355.



**Mohamad Awada** received his B.Sc. degree in civil engineering, and M.Sc. degree in construction management from the American University of Beirut, Lebanon, in 2016 and 2018, respectively. He worked as a research assistant for the year of 2019 at the American University of Beirut in the construction management department. He is currently pursuing his Ph.D. in civil and environmental engineering at the University of Southern California where he earned his M.Sc. degree in computer science in 2022. The presented work is part of his Ph.D. research about healthy office spaces.



**Burçin Becerik-Gerber** received the Doctor of Design (D.Des.) degree from Harvard University, Cambridge, MA, USA, in 2006. She is a Professor of Civil and Environmental Engineering with the University of Southern California, Los Angeles, CA, USA. To date, she has authored over 180 peer-reviewed journal and conference papers. Her work has received support worth over \$12M individual and collaborative grants from a variety of agencies. Her research focuses on the design, automation, control, and visualization of user centered responsive and adaptive built environments. Prof. Becerik Gerber received the NSF CAREER Award in 2014, the Viterbi Junior Research Award in 2016, the Mellon Mentoring Award in 2017 and CETI (Celebration of Engineering & Technology) Award in 2018. In 2012, MIT's Technology Review has named her as one of the world's top young innovators under the age of 35. She was one of the ten Rutherford Visiting Fellows of Alan Turing Institute-UK's national institute for data science and artificial intelligence. She has been serving as an Associate Editor for ASCE's Journal of Computing in Civil Engineering since 2011 and Nature's Scientific Reports since 2021.



agents and social robots. Dr. Lucas is a member of Association for Computing Machinery



**Shawn C. Roll** received the Ph.D. degree in health and rehabilitation science from Ohio State University, Columbus, OH, USA, in 2011. He is an Associate Professor and the Director of the Ph.D. program in the Mrs. T.H. Chan Division of Occupational Science and Occupational Therapy with the University of Southern California, Los Angeles, CA, USA. He is a licensed occupational therapist and registered musculoskeletal sonographer. To date, he has received nearly \$6M in funding on individual and collaborative research

prevention, and rehabilitation for work-related injuries, as well as understanding the relationships among the person, the environment, and participation in activities (e.g., daily occupations) to promote health and well-being within the adult working population. Dr. Roll is a member of numerous professional organizations, including the American Occupational Therapy Association (AOTA), American Institute of Ultrasound in Medicine (AIUM), and the Society for the Study of Occupation. He has been named a Fellow of the AOTA and the AIUM and inducted into the AOTF Academy of Research. He is an Associate Editor for the Journal of Diagnostic Medical Sonography.



**Ruying Liu** received her B.Sc. degree in Civil Engineering in 2017 from the University of Edinburgh, UK. She received her M.Sc. degree in Civil and Environmental Engineering with an area of emphasis in Structural Engineering, Mechanics and Materials in 2018 from the University of California, Berkeley. She worked as a structural designer from 2018 to 2020 in Bay Area. She is currently pursuing her Ph.D. in Civil and Environmental Engineering at the University of Southern California.