IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 1, 2021, accepted June 21, 2021, date of publication July 14, 2021, date of current version July 26, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3097038

The Feasibility of Wearable and Self-Report Stress Detection Measures in a Semi-Controlled Lab Environment

SARA ARISTIZABAL[1,2], (Member, IEEE), KUNJOON BYUN[1,2], NADIA WOOD[1,2], AIDAN F. MULLAN[3], PAIGE M. PORTER[4], CAROLINA CAMPANELLA[1,2], ANJA JAMROZIK[2], IVAN Z. NENADIC[5], AND BRENT A. BAUER[6]

[1]Well Living Lab, Rochester, MN 55902, USA
[2]Delos Living LLC, New York, NY 10014, USA
[3]Department of Biostatistics, Mayo Clinic, Rochester, MN 55905, USA
[4]School for Environment and Sustainability (SEAS), University of Michigan, Ann Arbor, MI 48109, USA
[5]Department of Internal Medicine, University of Michigan Health System, Ann Arbor, MI 48109, USA
[6]General Internal Medicine, Mayo Clinic, Rochester, MN 55905, USA

Corresponding author: Sara Aristizabal (sara.aristizabal@delos.com)

This work was supported by Delos Living, Limited Liability Company (LLC).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Mayo Clinic Institutional Review Board under IRB No. 17-009303.

ABSTRACT Workplace-related stressors, economic strain, and lack of access to educational and basic needs have exacerbated feelings of stress in the United States. Ongoing stress can result in an increased risk of cardiovascular, musculoskeletal, and mental health disorders. Similarly, workplace stress can translate to a decrease in employee productivity and higher costs associated with employee absenteeism in an organization. Detecting stress and the events that correlate with stress during a workday is the first step to addressing its negative effects on health and wellbeing. Although there are a variety of techniques for stress detection using physiological signals, there is still limited research on the ability of behavioral measures to improve the performance of stress detection algorithms. In this study, we evaluated the feasibility of detecting stress using deep learning, a subfield of machine learning, on a small data set consisting of electrodermal activity, skin temperature, and heart rate measurements, in combination with self-reported anxiety and stress. The model was able to detect stress periods with 96% accuracy when using the combined wearable device and survey data, compared to the wearable device dataset alone (88% accuracy). Creating multi-dimensional datasets that include both wearable device data and ratings of perceived stress could help correlate stress-inducing events with feelings of stress at the individual level and help reduce intra-individual variabilities due to the subjective nature of the stress response.

INDEX TERMS Deep learning, perceived anxiety, perceived stress, stress detection model, TSST, wrist-worn wearable.

I. INTRODUCTION

Stress, anxiety, and depression are the most common mental health issues in America [1]. An estimated 55% of individuals in the United States experience daily moderate to high stress, with workplace stress significantly contributing to the mental health crisis [1], [2]. Factors such as longer work hours, high work demands, low salaries, and job insecurity for workers are likely to cause the negative physical, behavioral, and psychological reactions associated with perceived stress in the global workforce [3], [4]. In the U.S., 65% of employees cite work as a significant source of stress [5], and previous research suggests that the physical environment can also contribute to perceived feelings of stress [6]. Factors within the built environment, including light [7], temperature [8], and most notably, sound [7], [9]–[14] have been shown to impact occupants' stress levels and overall well-being.

The associate editor coordinating the review of this manuscript and approving it for publication was Yin Zhang.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Stress encompasses a series of physiological and behavioral reactions in response to new, uncertain, uncontrollable, or unpredictable situations [15]. Stress exists in three different forms: acute, episodic acute, and chronic [16], [17]. Most people experience acute stress occasionally during their everyday life, and because it is short-term, it is not considered to be harmful to one's health. Episodic acute stress refers to acute stress that occurs more frequently [17], [18], while chronic stress refers to stress sustained for a long period of time [19]. Both episodic acute and chronic stress can have detrimental effects on an individual's physical and psychological health [20], [21]. Chronic stress can negatively impact sleep quality, lead to weight gain, and increase the risk of depression, high blood pressure, and cardiovascular disease [22].

Eliminating stress altogether is an unlikely outcome of workplace interventions. However, it may be possible to encourage employees to adopt appropriate stress-coping strategies that alleviate stress and build resiliency to avoid serious health consequences. It is imperative that individuals become aware of the level of stress they are experiencing, as a step toward effectively managing such stress in the future. In recognition of this, there have been numerous attempts to quantitatively measure an individual's stress levels, which are dictated by the autonomic nervous system (ANS). The ANS regulates body functions such as heart rate, respiration, and digestion. In the presence of an event that the body perceives as a threat, the hypothalamus activates the sympathetic branch of the autonomic nervous system (SNS). SNS activation results in the body's arousal and acceleration of the body functions (''fight or flight'' response) [23]. Sustained exposure to a stressor further leads to the release of glucocorticoids into the systemic circulation, which keeps the body on high alert [24], [25]. Following the stressful event, the parasympathetic branch of the autonomic nervous system reduces hormone production, resulting in the slowing down of body functions [23].

Video-cameras and audio recorders have been used to identify stress by extracting semantic, phonetic, and facial features that can be related to arousal or stressful events [26]–[29], whereas behavioral tracking methods consisting of questionnaires have been used to collect and aggregate subjective ratings of varying psychological states, anxiety, and stress [30], [31]. More recently, subtle digital technologies such as features detected from smartphone use and wearable devices have also gradually become an attractive means to detect stress features in a non-invasive manner. These newer technologies serve as an alternative to images and sound collected via video- cameras and audio systems, whose privacy implications can be more serious and evident to users [32]–[35].

Smartphones and wearables allow for the convenient, low-cost and continuous measurement and storage of multiple physiological signals using microelectromechanical systems (MEMS) (*i.e.,* Electrodermal activity (EDA), photoplethysmography (PPG), and acceleration sensors) to measure physiological signs of stress [32], [36]–[40]. New developments in artificial intelligence (AI) and machine learning (ML), such as deep learning methodologies have opened up the possibility of uncovering unique patterns in the data [41], [42], and have provided researchers with the tools to analyze large amounts of data to understand the complex and dynamic stress response and its intra and interindividual variability [43]–[47].

The combination of multidimensional data from these digital technologies, as well as contextual, behavioral, and demographic information, can provide more evidence about the nature, individual signs, and sources of the stress response, thus providing the foundations for better stress identification tools [48].

Previous work in stress detection using wearable sensor data streams has shown that electrodermal activity (EDA) and heart rate (HR) changes are frequently used physiological signals for stress recognition [32].

Villarejo *et al.* [49], were able to correctly classify a stress state from a non-stress state with a 76% success rate using EDA as the main predictor after inducing stress through different tasks that required mental effort. Similarly, Zangroniz *et al.* [50] and Amalan *et al.* [51], showed that supervised learning models using data from wearable devices collecting solely EDA were able to differentiate between a stressful and calm condition with 89% and 93% accuracy, respectively. Zangroniz *et al.* [50], and Amalan *et al.* [51], induced stress using a database of high arousal photographs and the Trier Social Stress Test (TSST), respectively. Setz *et al.* [52] also demonstrated an 82.8% maximum accuracy in discriminating stress from cognitive load using only EDA features, while Kurniawan *et al.* [53], and Anusha *et al.* [54], achieved a 75% precision and 85% accuracy in acute stress classification using only EDA and complex classifiers.

In addition to EDA, the combination of EDA and heart rate (HR) and EDA and heart rate variability (HRV) features have been extensively used to distinguish stressful events [55]–[61]. De Santos Sierra *et al.* previously demonstrated a 99% accuracy in stress discrimination when using a database of EDA and HR signals collected during two stressful tasks: hyperventilation and speech preparation, in combination with conventional machine learning algorithms [62], and fuzzy expert systems [63]. Similarly, Healey and Picard [64] showed that EDA and HR are closely correlated, and accurate predictors of stress levels for those driving a vehicle, and demonstrated a 97% accuracy when differentiating among three different levels of drivers' stress, corresponding to rest, highway and city driving conditions. There have also been numerous attempts at improving the performance of stress detection systems by combining EDA and HRV measurements. Sandulescu *et al*. [65], were able to classify stressful from non-stressful situations using EDA and HRV signals collected with a wrist-worn wearable with 82% accuracy. Similarly, Martinez *et al*. [66], were able to classify different levels of stress (high, medium, low) using EDA and

HRV with a resulting F-measure score of 0.984, 0.970, and 0.943, respectively. Goumopoulos and Menti [67], showed that EDA and HRV features are also valid stress indicators for the senior population, and have the ability to discriminate between calm (baseline) and stressful events. The reliability and reproducibility of EDA and HRV measures for the assessment of autonomic nervous system dysfunction have also been demonstrated before by Posada-Quintero *et al*. [68], and Ghiasi *et al*. [69].

Although most of these stress detection algorithms show a high accuracy for monitoring stress in controlled settings, the classification accuracy drops to 70-80% when assessing stress in real-life environments [70]–[72]. Some of the main reasons for poor detection performance include the unknown situational context of the user, the data quality of non-obtrusive devices, confounding factors such as limb movement and physical activity, improper sensor placement, and a limited battery life [32]. In addition, people experience stress differently. The differential ways in which the body responds to stress poses difficulties when building a stress discrimination algorithm that relies solely on autonomic activity indicators such as heart rate, skin conductance, and heart rate variability, to correctly identify stress events for the broader population [73]. Traditional methods for measuring stress include weighted instruments such as questionnaires and periodic self-reports [67]. Although these methods are considered to be the ''gold standard'' for stress assessment, they can be time-consuming and burdensome for the individual to complete [74]. Nevertheless, periodic self-reports are currently the most appropriate method to ground-truth stressful events [75]. Studies on the potential for combining physiological data from wearable devices with subjective stress data are fairly limited [32], [67], [75], and thus, the impact of self-report behavioral measures (i.e., those collected via questionnaires) on the performance of stress detection algorithms remains largely unexplored.

This paper addresses the limitation in the existing literature by comparing stress detection performance of a deep learning model, a state-of-the-art machine learning approach, using 1) data from a wrist-worn-sensor collecting physiological data in a semi-restricted setting and 2) the physiological data collected with the wrist- worn sensor in combination with behavioral data.

We also explored the potential issues that arise with data collection quality, feature extraction, and classification when wrist-worn-wearables are used to detect stress in a laboratory setting. In this study, we collected data from validated self-assessment measures of perceived stress and anxiety. In order to measure physiological stress response, we collected salivary cortisol samples and raw electrodermal activity, skin temperature, and heart rate data from study participants, using the Empatica E4 wristband (Empatica Inc., *Boston, MA, USA*) and used simple and complex classifiers to distinguish between stress and non-stress periods during the Trier Social Stress Test protocol (TSST), a validated stress induction method. The study was conducted in a simulated waiting area and conference room inside the Well Living Lab. This semi-restricted but naturalistic environment allowed us to capture participants' unbiased reactions while controlling and monitoring some of the environmental conditions in the experimental space. By holding this waiting area and conference room constant across each of our participants' experiences, we tried to eliminate the potential for environmental parameters just prior to, and during the intervention, to differentially impact outcomes seen in our stress measurements, thus strengthening the internal validity of the study, and the ability to attribute changes in the stress response to our intervention.

II. MATERIALS AND METHODS

A. TRIER SOCIAL STRESS TEST (TSST)

The Trier Social Stress Test (TSST) was used to reliably induce mental stress in study participants throughout the study. The TSST is a widely used laboratory procedure that has been shown to reliably induce physiologically measurable stress responses in individuals [76]. For the TSST, participants are placed in a simulated, socially evaluative situation during which their physiological and psychological stress responses are measured. The test is divided into three sections. The first section is a 45-minute rest period where researchers collect baseline levels of stress. The rest period is immediately followed by a 15-minute stress induction period which is further subdivided into three, 5-minute phases: an anticipatory phase where participants prepare for upcoming tests; a presentation phase where participants are asked to give a recorded speech before a panel of judges who do not emote or react while being recorded with a video camera and microphone, and a mental arithmetic phase during which participants are asked to count backwards in odd-numbered increments. The final section of the test is a 60-minute recovery period, during which participants are debriefed and asked to relax. This recovery period was divided into two 30-min segments (recovery 1 and recovery 2).

B. PHYSIOLOGICAL INDICATORS OF STRESS

1) SALIVARY CORTISOL AS A BIOMARKER FOR STRESS

Cortisol, known as ''the stress hormone'', has been used in stress research for almost two decades, as researchers have shown that cortisol levels increase within a short period after the appearance of a stressor [25], [77]. To verify that participants were stressed during the induction period of the TSST, cortisol was collected at four different time points: just before participants entered the testing room, at the end of the stress induction period, after the first recovery period, and after the second recovery period. Cortisol plays an essential role in the body's stress response and it is secreted by the adrenal medulla [77]. Specifically, the release of cortisol in response to stress leads to an increase in heart rate and blood pressure among other physiological processes [78]. As a result, cortisol is considered a reliable marker of stress and can be measured in blood, saliva, and urine. Salivary cortisol has been primarily used in stress research due to its ease of sampling as well as its non-invasive and inexpensive nature.

2) EMPATICA E4 WEARABLE DEVICE

The continuous monitoring of physiological signals was performed using the Empatica E4 (Empatica Inc., Cambridge, MA). The Empatica E4 is a unobtrusive wearable device for physiological data acquisition [79]. It includes four embedded sensors: a photoplethysmography (PPG) sensor, an electrodermal activity (EDA) sensor, a 3-axis accelerometer, and an optical thermometer. The EDA sensor can measure the conductivity of the skin ranging from 0.01μS to 100μS, and it has a sampling rate of 4Hz. The PPG sensor measures the volume change produced with every heartbeat using light-based technology; pressure increases as the heart pumps the blood into the systemic circulation creating a difference in the amount of light absorbed by oxyhemoglobin. The sensor measures the difference in the light reflected with sampling every 64Hz. The temperature sensor uses an optical infrared thermometer with a resolution of 0.02 °C. Lastly, the three-axis accelerometer, measures the 3 axes of motion X, Y, Z with a sampling rate of 32Hz. The E4 device is small (4cm $\times$ 4cm), lightweight, and comfortable, and wearing it is as easy as wearing a bracelet or a watch. The Empatica E4 allows for data collection using either real-time streaming mode or in- memory recording. For this study, data was collected using the in-memory recording modality that allows for data being stored in the device for up to 60 hours.

To assess markers of sympathetic activity that can help discriminate stress events, electrodermal activity (EDA), heart rate (HR), and skin temperature (ST) were collected throughout the study period.

3) ELECTRODERMAL ACTIVITY

EDA is a measure of the variation in the electrical properties of the skin due to internal or external stimuli, and it is linked to the quantity of sweat secreted by the sweat glands [80]. Therefore, changes in skin conductance have been shown to be a measure of emotional arousal and cognitive load, as the sweat glands in the skin are directly innervated by the sympathetic branch of the ANS [81]. The EDA complex which includes a slow, changing component related to the tonic shifts of electrical conductivity of the skin (SCL), and a phasic component that reflects rapid changes contained in the EDA signal (SCR) [82], has been extensively studied as an indicator of autonomic nervous system activity [83].

4) HEART RATE

In addition to EDA, photoplethysmography (PPG) based metrics have also been used in stress recognition research [84]–[86]. PPG detects blood volume changes in skin capillary beds associated with heart activity [60]. The PPG signal is characterized by peaks and valleys that reflect the systole and diastole phases of the cardiac cycle [87]. The most typical features extracted from the photoplethysmography waveforms are heart rate (HR) and heart rate variability (HRV). HR is defined as the number of times the heart beats per minute, and it increases during stressful events. [88]. HRV is the variation in time between consecutive heartbeats [55]. The activation of the Autonomic Nervous System (ANS) branches translates into a decrease in HRV when there is sympathetic stimulation and an increase during parasympathetic stimulation to restore homeostasis after a stress-related event [55].

5) SKIN TEMPERATURE

Previous research in emotion recognition has shown that under acute stress, the sympathetic nervous system triggers peripheral vasoconstriction which results in short-term changes in skin temperature (ST) from its normal range of 32 to 35°C [89]–[93]. Although ST has the potential to provide more information about the intensity of the stress response, ST can also change in response to physical exertion, the presence of an illness, and environmental conditions such as temperature and humidity. Therefore, the use of skin temperature as an indicator of acute stress is more suitable for laboratory studies than in the field, unless some of the above- mentioned aspects that also affect ST can be reliably measured in external settings.

C. BEHAVIORAL MEASURES

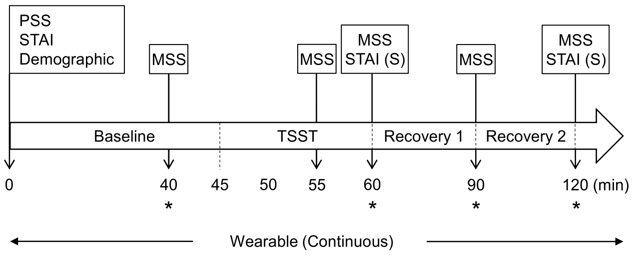
Participants completed surveys regarding their stress and anxiety throughout the study. The State-Trait Anxiety Inventory (STAI) was used to measure two types of anxiety: an individual's general level of anxiety (Trait) and momentary level of anxiety (State) [31]. Each anxiety type was self-reported by participants using 20 items, each rated on a 4-point Likert scale (e.g., from "Almost Never" to "Almost Always"). Both state and trait levels of anxiety were measured at the beginning of the experiment to establish a baseline. Additional state levels of anxiety were measured immediately following the stress induction period and after the full recovery period.

The Perceived Stress Scale (PSS-10) [30] was used to measure individuals' stressful thoughts and feelings for the month prior to the TSST and was asked at the beginning of the experiment (baseline). The Perceived Stress Scale consists of 10 questions assessing the frequency of stressful feelings using a five-point Likert scale from 0 ("Never") to 4 ("Very Often").

To measure subjective "in the moment" feelings of stress, participants were asked to rate their stress using one, seven-point Likert scale item 1 ("Not at all stressed) to 7 ("Extremely stressed") at four different time points: just before entering the testing room, at the end of the stress induction period, after the first recovery period, and after the second recovery period.

D. OVERVIEW OF THE STUDY DESIGN

For this study, approved by the Mayo Clinic Institutional Review Board, eighteen healthy subjects (11 Female, 7 Male; Mean Age = 26.67, SD Age = 4.12), with no history



Note. * - Saliva sample, PSS – Perceived Stress Scale, STAI – State Trait Anxiety Inventory, STAI (S) – State Anxiety only, MSS – "In the Moment" Stress Scale

FIGURE 1. Experiment schedule and data collection details.

of depression, anxiety or mood disorders, no history of drug/alcohol abuse or tobacco use, and women who were not pregnant, were recruited to participate in a single, 2-hour testing session (1:30 pm – 3:30 pm). All subjects read and signed the informed consent form prior to participating in the study.

Two subjects were ultimately excluded from the study due to missing salivary cortisol or wearable data. The testing session took place at the Well Living Lab (WLL), a modular space that can be reconfigured to simulate a variety of indoor environments [94]–[96]. The lab space was reconfigured as a waiting room and a conference room. Prior to the start of the study, subjects were asked to meet with the study coordinator to capture a comprehensive profile of the subject including demographics and health information to understand the participant's general state of physical and mental health prior to enrollment in the study. On the day of the study, at the start of the session, each participant was asked to fill out a baseline survey (PSS and STAI) to establish a reference point for tracking behaviors and physiology (See Fig. 1). Subsequently, participants were outfitted with an Empatica E4 device to be worn throughout the entirety of the study. Participants were also asked to provide salivary cortisol using a Salivette collection kit (Salivette® - Sarstedt, Germany), twice at four separate time points during the experimental session (baseline, at the end of the stress induction period, and at the end of each of the recovery periods), for a total of 8 salivary samples per participant. At each salivary cortisol sample collection, subjects were asked to rate their stress levels using a subjective "in the moment" stress scale. In addition, and prior to the start of the stress procedure, participants were asked to fill out the perceived stress scale (PSS). Building system setpoints (temperature, humidity, and lighting) in the lab space were kept constant during the study.

An overview of the measurement and intervention schedules can be observed in Fig. 1.

E. DATA MANAGEMENT AND MACHINE LEARNING FRAMEWORK

1) DATA PREPARATION AND PRE-PROCESSING

Raw data from the Empatica E4 devices were downloaded after each experimental session and prepared for analysis in MATLAB® and R®. Heart rate and skin conductivity were calculated for the baseline, stress procedure, and two recovery periods. An initial visual inspection of the data was conducted to identify and remove incorrect or unrealistic EDA and HR calculations (over 200 and under 40 beats per minute) from the data set. The data was also interpolated to account for differences in sensor sampling rates.

TABLE 1. Description of data pipelines.

Data Pipeline	Description
Survey and demographics	This data pipeline read the raw data from data lake, enriched the data by standardizing the date formats, and adding the source data file identifier. The resulting processed data was stored in MS SQL table.
EDA, ST, and HR	This data pipeline read the raw data from data lake, enriched the data by adding subject identifiers, standardizing the date formats, and adding the source data file identifier. The data pipeline also labeled the phases based on the TSST timestamps.

The salivary cortisol samples were analyzed using an enzyme immunoassay kit by the Mayo Clinic Medical Laboratories.

Two types of data stores were used as part of the framework: Microsoft® Data Lake and the SQL database. Once the data was pre-processed and imported, enriched data was stored in the Microsoft® SQL database. Automated data pipelines were written in C#.NET programming language to upload the data to Azure Data Lake for further processing using front end tools such as Jupyter notebooks [97] for analysis, and Tableau [98] for data visualization. For the deep learning analysis, the data were flattened and extracted as CSV files before being read by the model. Google Colab was utilized, and the files were available to the model via our secure document storage database. Data pipelines are described in Table. 1.

2) TECHNICAL ARCHITECTURE

In this study, we implemented a full, end-to-end utilization of the Well Living Lab's machine learning framework and ecosystem (Fig. 2). This framework allowed us to automate the data management, processing, analysis, and building of a deep learning model.

Once the data was loaded in the database, various permutations of datasets were created by writing SQL statements and storing the new datasets in new tables (see Fig. 3).

The deep learning model was designed to predict whether a subject was experiencing stress in a binary classification format (stress or no stress detected). The phases of the TSST protocol were used to manually label the dataset into stress and non-stress periods. For each subject, the baseline and recovery periods were labeled as non-stress, and the segments of the stress procedure (the anticipation, speech, and mental arithmetic task) were merged and labeled as stress regions. Two sets of features were used as *inputs* to the neural network. The first set of features (feature set 1) consisted of

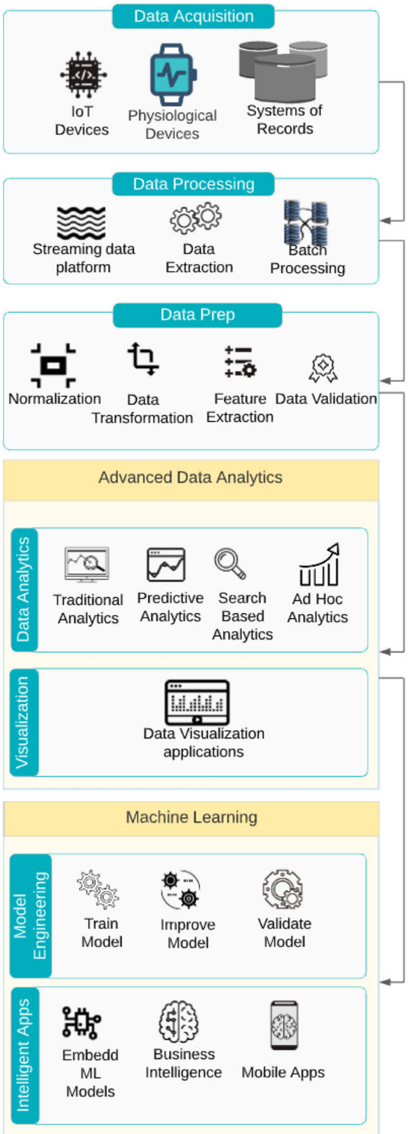


FIGURE 2. Well Living Lab (WLL) machine learning ecosystem.

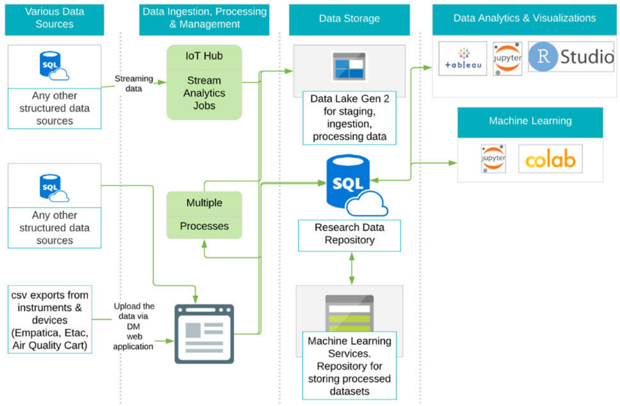


FIGURE 3. Flow diagram of study data through WLL machine learning ecosystem.

signals collected using the Empatica E4 wearable (heart rate, electrodermal activity, and temperature). The second set of features (feature set 2) consisted of data from the wearable combined with subjective information (perceived stress and anxiety ratings).

Statistical analyses were conducted in parallel to the deep learning analyses to assess effect size and variable importance using semi-parametric methods. The aim of this structured analysis was to 1) assess predictors of a stress response independently of one another in a multivariate space, as well as 2) to provide some degree of interpretability for the patterns being detected by the deep learning algorithms.

3) STATISTICAL ANALYSIS

In order to evaluate the underlying relationships between the physiological and behavioral variables of interest and the Trier Social Stress Test, two sets of independent regression models were implemented.

The first set of models reduced the data to the mean level of each HR and EDA predictor of interest for each phase and participant. This reduction resulted in five observations per participant for each of these physiological variables. Survey data were already collected once during each study phase, and therefore were not reduced. Using each physiological variable of interest as the response, independent univariate mixed-effects logistic regression models were fit using the study phase as the predictor. Measurements taken during the baseline period served as the reference level for study phase. A random intercept was included for each participant, and there was no random slope implemented.

The second set of regression models used the full physiological data collected for each participant. Univariate logistic generalized estimating equation models were fit to each of the predictors of interest, using the TSST phase as the binary response. Sequential observations for a given participant were assumed to be correlated following an autoregressive time-series association. Standard errors were computed using the robust Eicker-Huber-White estimator [99].

A follow-up multivariate logistic generalized estimating equation (GEE) model was fit using all predictors of interest. Similar to the univariate models, the within-subject covariance was assumed to be autoregressive, and errors were computed using the Eicker-Huber-White estimator.

4) DEEP LEARNING MODEL FOR STRESS DETECTION

Detection of mental stress has previously been performed using machine learning methods, such as support vector machine and k-means clustering [32]. However, these methods require performing feature extraction on behavioral measurements and physiological signals, which may not be appropriate for the construction of prediction models when working with large amounts of data and, furthermore, unnecessary with the development of deep learning technology [100]. In addition, deep learning methods are more

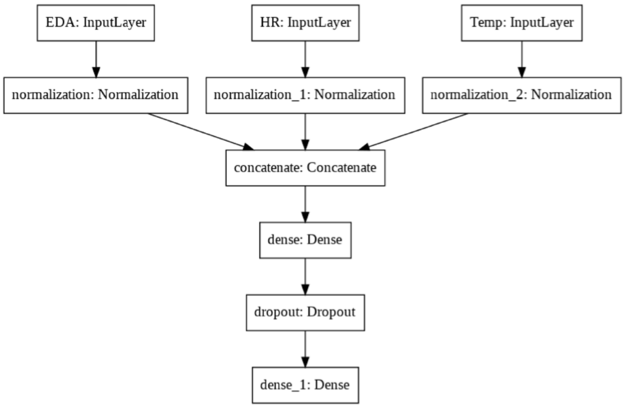


FIGURE 4. Diagram of the proposed neural network with physiological data as inputs. The diagram was generated using keras utility tool.

appropriately capable of handling granular repeated measures data than traditional methods.

In this analysis, we aimed at training a binary classification neural network using Keras 2.0 [101], with Tensorflow implementation (Google Brain, Mountain View, CA, USA). The model was developed using the functional API of Keras, which links all or part of the inputs directly to the output layer, allowing the neural network to determine deep patterns (using the deep path) and simple rules (using the short path) [102]. The functional API can also handle models with various inputs and outputs, shared layers, and non-linear topologies [102]. We developed two neural networks with various input layers. The models were designed to receive data from three different physiological signals (EDA, HR, and Temperature) collected using a wrist-worn wearable as well as self-ratings of stress and anxiety. The physiological data collected through the different segments of the TSST were used as the only inputs for one model, while the physiological signals in combination with the behavioral measurements were used as inputs to the second deep learning model. The neural network was designed to detect stress by discriminating between stress and non-stress periods.

The network was developed using dense and dropout layers as shown in Fig. 4. Dense layers are layers of neurons, in which each neuron is connected to the neurons of the previous layer [103], while dropout layers are regularization layers that help prevent overfitting [104]. The models consisted of a fully connected network structure with two layers. The inputs (physiological and behavioral data) were normalized and concatenated into one vector, which was fed to a hidden dense layer with 32 neurons.

Each neuron used a Rectified Linear Unit (ReLU) activation function (1).

$$ReLU = max\,(0, x) \tag{1}$$

We ran all the inputs for the model specifying to Keras how to connect the layers together (see Fig. 5). The layers in a Keras model are connected pairwise by specifying where the inputs come from when defining each new layer.

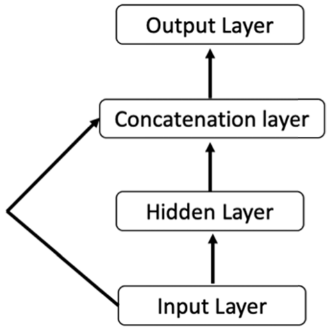


FIGURE 5. Deep learning model layers.

We subsequently created a concatenation layer to concatenate the input and the output of the hidden layer. Our output layer had a single neuron, as we wanted to classify a stressed vs. non-stressed state, using a sigmoid activation function (2).

$$Sigmoid(x) = \frac{1}{(1 + e^{-x})} \tag{2}$$

To configure the model for training, the training and test sets were defined using a 70/30 split. The experimental data was not present in more than one set simultaneously. We used the Adam optimizer [105] with a 0.001 learning rate, and a binary cross-entropy loss function to estimate the error between true and predicted values. Epoch and batch size were 50 and 42, respectively for both models. To evaluate whether subjective ratings of stress and anxiety improve stress recognition above and beyond what could be predicted from data collected from wearables, we evaluated the classification performance of the deep learning model using the following metrics: accuracy (3), sensitivity (4), and specificity (5). The classification was considered to be true positive (TP) if the participant was stressed and was correctly predicted as being stressed. On the contrary, the prediction was considered as false negative (FN) if the participant was incorrectly classified as being non-stressed. False positives (FP) and true negatives (TN) were determined in the same fashion. We had 67,458 instances labeled as stress and 409,078 instances labeled as no-stress in the training set.

$$Accuracy = \frac{(N_{TP} + N_{TN})}{N_{TP} + N_{FN} + N_{TN}} \tag{3}$$

$$Sensitivity = \left(\frac{N_{TP}}{N_{TP} + N_{FN}}\right) \times 100 \tag{4}$$

$$Specificity = \left(\frac{N_{TN}}{N_{TN} + N_{FP}}\right) \times 100 \tag{5}$$

Accuracy provides information about the fraction of instances that were correctly classified by the algorithm. Sensitivity provides information about the number of instances labelled as stress that were scored correctly. Specificity provides information about the fraction of non- stress instances that were classified correctly by the model.

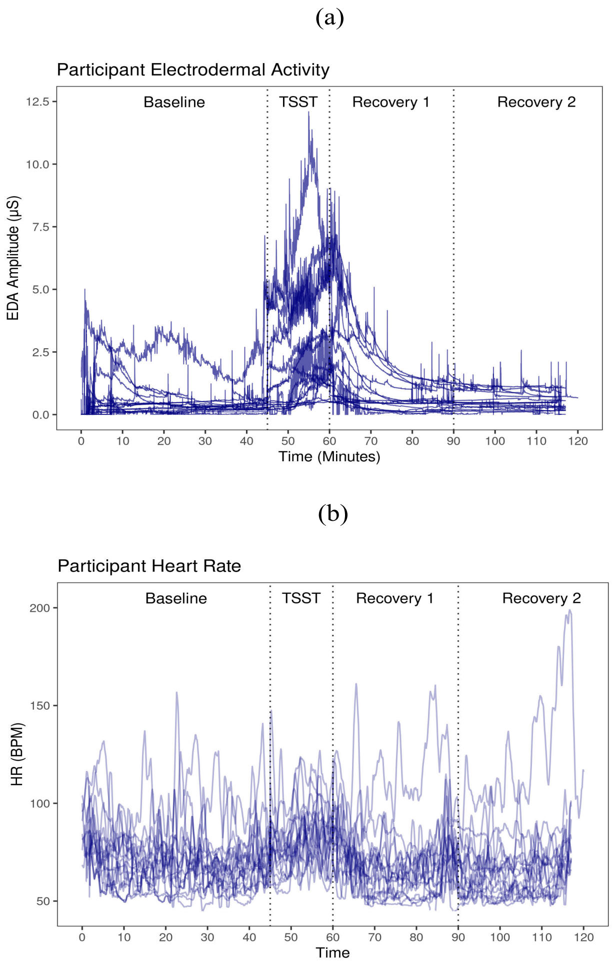


FIGURE 6. Raw (a) heart rate and (b) electrodermal activity values for all participants during the two-hour session. The lines indicate the different phases of the TSST.

III. RESULTS

A. PHYSIOLOGICAL AND BEHAVIORAL RESPONSES TO INDUCED STRESS

The measurements of the electrodermal activity (EDA) and heart rate (HR) across all subjects on each stage of the TSST (baseline, stress procedure, and the first and second recovery periods) are shown in Fig. 6(a) and (b), respectively. The mean EDA during the baseline, stress procedure, recovery 1, and recovery 2 periods were 0.46, 2.23, 1.02, and 0.53 μS, respectively. Similarly, the mean HR during the baseline, TSST, recovery 1 and recovery 2 periods were 69.9, 81.4, 69.8, 68.7 BPM, respectively.

To measure the relationship between the physiological and self-reported data, the time periods of the averaged self-ratings of stress using an "in the moment" stress scale, and the subjective ratings of state anxiety values are illustrated in Fig. 7(a) and (b). The boxplots feature the data distribution based on the minimum, first quartile, median, third quartile, and maximum.

The variation in cortisol levels of all participants who underwent the Trier Social Stress Test is shown in Fig. 8. The mean values of cortisol across the study conditions were 97.3, 149.6, 167.7, and 112.9 nmol/L, for the baseline, stress tests, and recovery 1 and recovery 2 periods, respectively. Cortisol levels did not temporally mirror the exact stress pattern experienced by the study subjects such that cortisol increased at the exact time a stressor was introduced. This result is expected given the time lag between the transfer of cortisol from the plasma to saliva; with peak cortisol production often being achieved 15-30 min after the stress event [106], [107].

TABLE 2. Model effects predicting average physiological variables by study phase, relative to the baseline phase.

		Stress Tasks	
	β	95% CI	*p-value*
Heart Rate	11.50	(7.51 – 15.45)	**< 0.001**
Skin Temperature	-0.21	(-0.67 – 0.25)	0.374
EDA	1.78	(1.19 – 2.37)	**< 0.001**
		Recovery 1	
	β	95% CI	*p-value*
Heart Rate	-0.14	(-4.10 – 3.83)	0.947
Skin Temperature	-0.25	(-0.71 –0.21)	0.300
EDA	0.56	(-0.03 – 1.15)	0.071
		Recovery 2	
	β	95% CI	*p-value*
Heart Rate	-1.19	(-5.16 – 2.78)	0.564
Skin Temperature	-0.19	(-0.65 – 0.27)	0.439
EDA	0.07	(-0.52 – 0.66)	0.817

B. MEASURES AND STATISTICAL ANALYSIS

Linear mixed-effects models were fit using participant's average physiological measurements in each phase as the response, with the study phase as the predictor of interest. Model effects are given in Table. 2. Compared to the baseline, participants had a higher average heart rate during the stress phase, as well as EDA. Additionally, there was no difference in any of these variables between the baseline and recovery periods.

Similar mixed-effects models were implemented to evaluate participant-reported stress and cortisol between the different TSST phases (Table. 3). Participants gave higher ratings of perceived stress immediately before and after the stress procedure phase, compared to baseline ($p < 0.001$ for both). This difference did not persist in either of the two recovery phases. Participants' reports of state anxiety were also higher following stress tasks compared to baseline ($p = 0.002$) but had no difference between the baseline and rest. Moreover, measured cortisol was higher immediately following the stress procedure ($p = 0.005$) and in the first recovery period ($p = 0.001$) compared to baseline. The cortisol levels measured during the second recovery period did not differ from baseline.

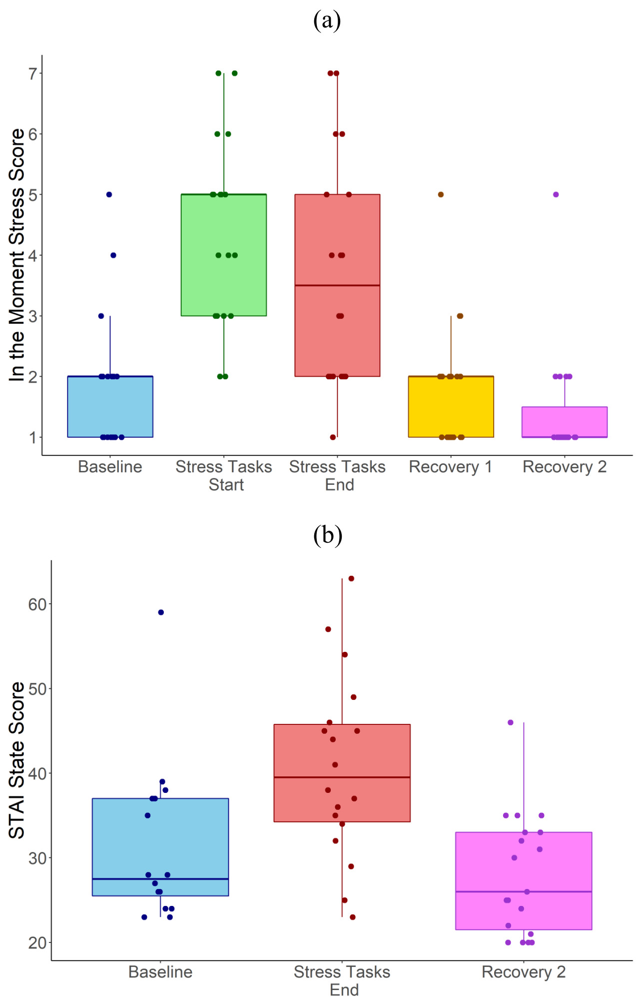


FIGURE 7. Time course of the averaged (a) "in the moment stress" and (b) state anxiety across the different phases of the TSST. The boxplots feature the data distribution based on the minimum, first quartile, median, third quartile, and maximum.

To further evaluate the continuous relationship between stress and behavioral and physiological variables, a series of logistic generalized estimating equations (GEE) were fit using the full dataset of heart rate and electrodermal activity measurements [108]. GEE is a semi-parametric approach that allows to impose some structure to the data. Due to the temporal nature of the data collection, sequential observations for a participant were assumed to be correlated under an autoregressive time series relationship. Table. 4 provides the resulting odds ratios from these models. Effects not given for the univariate models indicate that the model did not converge, and no estimates could be generated; effects missing from the multivariable model indicate that the inclusion of the variable caused the model not to converge, and the variable was excluded.

From the multivariate and univariate analysis, we find that a 1-point increase in the "in the moment" stress scale score was associated with a 444% increase in the odds of being stressed (OR = 5.43, p < 0.001). No other variables showed a significant impact on predicting stress.

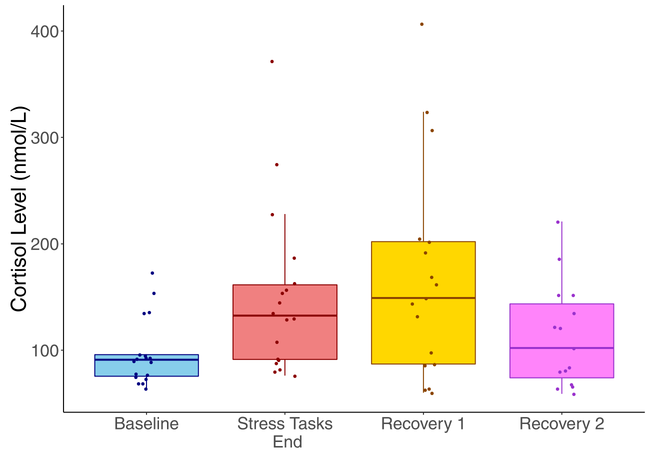


FIGURE 8. Time course of the averaged salivary cortisol across the different phases of the TSST. The boxplots feature the data distribution based on the minimum, first quartile, median, third quartile, and maximum.

TABLE 3. Model effects predicting survey response and cortisol level by study phase, relative to the baseline phase.

		Stress Task Start	
	β	95% CI	*p-value*
MSS	2.58	(1.93 – 3.23)	**< 0.001**
STAI	---	---	---
Cortisol	---	---	---
		Stress Task End	
	β	95% CI	*p-value*
MSS	1.87	(1.20 – 2.53)	**< 0.001**
STAI	9.50	(3.94 – 15.05)	**0.002**
Cortisol	52.3	(17.5 – 87.1)	**0.005**
		Recovery 1	
	β	95% CI	*p-value*
MSS	-0.05	(-0.70 – 0.60)	0.876
STAI	---	---	---
Cortisol	68.6	(33.3 - 104.2)	**< 0.001**
		Recovery 2	
	β	95% CI	*p-value*
MSS	-0.42	(-1.07 – 0.23)	0.215
STAI	-3.15	(-8.63 – 2.31)	0.268
Cortisol	9.0	(-27.8 – 46.2)	0.639

C. DEEP LEARNING ALGORITHM CLASSIFICATION

We compared the performance of the proposed model with various input layers with two different sets of input variables. A confusion matrix and loss functions were used to compare classification ability in addition to the performance metrics described in section II.

TABLE 4. Model effects predicting stress phase.

	Univariable		Univariable	
	OR (95% CI)	*p-value*	OR (95% CI)	*p-value*
Heart Rate	1.002 (0.996, 1.008)	0.466	1.004 (0.992, 1.016)	0.514
Skin Temperature	---	---	---	---
EDA	---	---	---	---
MSS	5.438 (2.260, 13.060)	**<0.001**	5.435 (2.261, 13.068)	**<0.001**
STAI	1.000 (0.999, 1.001)	0.718	1.000 (0.997, 1.003)	0.910

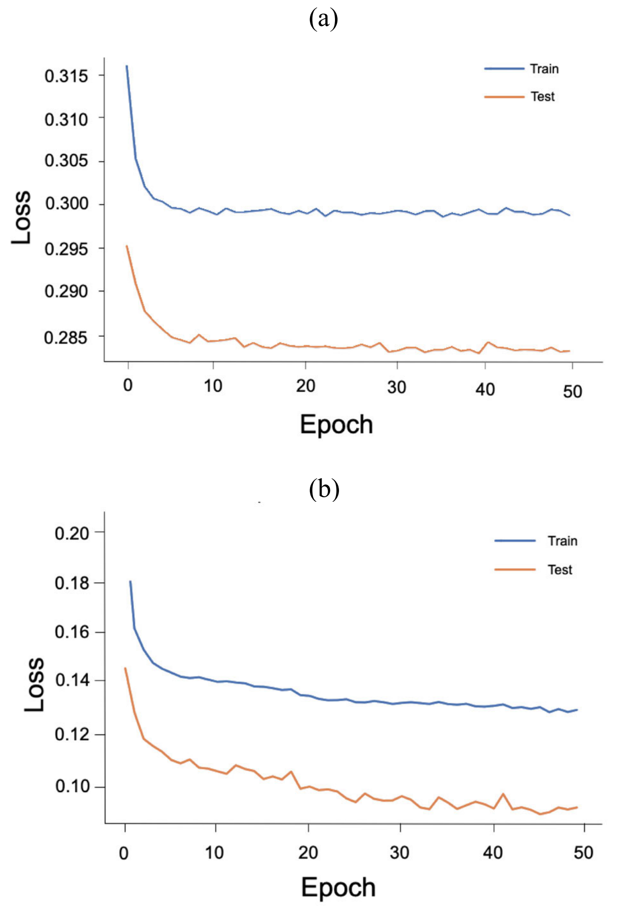


FIGURE 9. Loss function profiles for (a) feature set 1 and (b) feature set 2.

Fig. 9 shows the loss function graphs for the deep learning model with (a) the wearable only data and (b) the wearable measurements in combination with self-reported stress and anxiety as inputs. The loss function in this case a cross-entropy loss function defines the error between the estimated output (predicted labels) and the true output (true labels) and therefore can be considered a measure of the accuracy of the model. Each graph represents the loss function for training and test data sets. The x-axis represents the epoch or the number of times the training data is passed through the neural network, and the y-axis represents the loss. The closer the loss value gets to zero, the more similar the estimated value is to the true value and thus the more accurate the model is. A comparison of our findings shows that the profile for the model with feature set 2 (wearable + self-reports) shows the loss value closest to zero. Table. 5 summarizes the performance metrics across the two sets of features. The combination of raw physiological metrics from the wearable and subjective survey responses resulted in an improvement in stress/non-stress classification, with 96% of the regions identified correctly compared to 88% when only physiological metrics were used as inputs to the model. Temperature, heart rate, and skin conductance data were able to score only 24% of the stress periods correctly while the sensitivity of the model was enhanced when subjective survey responses were included in the feature set (sensitivity = 73%), as can be seen in Table. 5. Specificity measures also showed a slight improvement when using feature set 2.

TABLE 5. Summary of performance metrics across all feature sets.

	Sensitivity	Specificity	Accuracy (%)
Feature Set 1	0.2425	0.9876	88.72
Feature Set 2	0.7338	0.9959	96.05

IV. DISCUSSION

The study aimed to investigate the feasibility of using wrist-worn wearable data, normally available through research-grade and commercial off-the-shelf devices, together with self-ratings of stress and anxiety to identify acute stress using deep learning methods. Previous studies have demonstrated the potential for wearable data streams of measures such as EDA, HR, and HRV to detect acute stress events using machine learning methods [51], [56], [57], [109]. In this study, we compared the results of a deep learning model using only wearable data and wearable in combination with survey data to detect moments of stress. Psychological stress was induced using the TSST protocol and the resulting subjective stress and anxiety ratings, salivary cortisol levels, and wearable- physiological parameters, across the different experimental conditions: baseline, stress tasks, recovery 1, and recovery 2 were used as inputs to a deep learning model for stress recognition. Initial inspection of the data revealed that participants felt more stressed in the anticipation, mental arithmetic, and speech portion of the TSST compared to the baseline and recovery periods. The skin conductivity and heart rate values were on average 2.7 times and 16% higher in the stress task than in any other condition (Fig. 6(a) and (b)), while the skin temperature values did not significantly change between the baseline, stress, and recovery periods. These results are consistent with physiological reactivity to the TSST shown in previous literature [51], [110], [111] and indicate an activation of the hypothalamic-pituitary-adrenal axis and the sympathetic branch of the ANS during the stress tasks [112]. Importantly, subjects' perceived anxiety and ''in the moment'' stress levels were consistent with the wearable results (Fig. 7(a) and (b)). The mean STAI (state) and ''in the moment'' stress scores were statistically significant and differed across the various stages of the experiment. These results are also consistent with previous studies that have shown an association between subjective and objective reactions to stress [113]. However, it is important to note that although previous literature have introduced different validated measures of perceived stress and anxiety as part of the TSST, this study introduced a single question to assess ''in the moment'' feelings of stress. Although this short question has not been validated before, MMS scores were consistent with physiological reactions (EDA and HR) to the TSST.

Cortisol levels were higher following the stress tasks, specifically at the start of the first recovery period compared to the other stages of the experiment (see Fig. 8). These results are consistent and confirm that the salivary testing procedures produced accurate results in this study, insofar as they align with salivary cortisol fluctuations observed in previous studies using the TSST procedure, where salivary cortisol often peaks 15-30 minutes after a stress event and then decreases sharply [114]. Self-ratings of stress were higher during and immediately after the stress procedure, indicating that cortisol lags behind the self-ratings of stress and wearable measurements.

To assess the associations of physiological and behavioral measurements in the dataset with stress, we implemented both a structured analysis, generalized estimating equations (GEE), and unstructured analysis, deep learning. GEEs provide a semi-parametric approach that allows for the determination of effect sizes and importance from our variables of interest while accounting for the within-participant correlation and high frequency of data collection. By applying this structure, we can derive variable-level interpretations of our data that are not available under deep learning. However, this model set failed to converge, which resulted in not having interpretable results. In this case, we could not determine the direction nor degree of the associations between certain variables of interest and the stress event. This would suggest that either our data did not provide sufficient information to the model due to the low participant number, or possibly the high degree of within-participant correlation did not follow the imposed structure of the GEE.

In parallel to this analysis, we evaluated the feasibility of using a deep learning approach for stress/non-stress classification. Deep learning methods have the ability to handle data points that have significant associations with each other and therefore remove some of the constraints that are introduced when using GEE on multilevel correlated data. In addition, these techniques are powerful enough to discover unstructured patterns in the data even when working with relatively small datasets consisting of numerical and categorical inputs. As such, we expected to inform our analysis using this method from the outset.

We evaluated the classification performance on two different sets of features. The binary classifier was able to achieve an accuracy of 88% when only raw wearable data streams were used as the input. However, contextual data such as behavioral responses to stress and anxiety helped the model to better predict stress segments in the data (accuracy = 96%). The precision of the model was higher than the recall, indicating that the deep learning model was able to discriminate 96% of the stress events with 73% of precision. Overall, our findings indicated that the data from the wearable could further be enhanced by including short survey responses regarding stress and anxiety. Previous studies have shown that a combination of different physiological measurements improves the ability of algorithms to correctly identify stress events [63], [64], [115], [116]. However, few studies have also attempted to detect stress using a combination of physiological and behavioral responses. Kyriakou *et al*. [117] were able to demonstrate an 84% accuracy in stress detection when using a rule-based approach based on skin temperature and electrodermal activity in addition to perceived stress measures collected through e-diaries and subject interviews [117]. Similarly, Gjoreski *et al.* [47] developed a method to monitor stress with a wrist-worn device using ecological momentary assessment and a stress log together with a machine learning approach. The method was able to detect 70% of stress events with a 95% precision [47]. Smets *et al.* [48], have previously shown that there is an association between wearable-measured physiological indicators and self-reported everyday stress and have highlighted the importance of these multi-dimensional data sets for stress detection given the person-to-person variability. These results further build upon this body of literature by demonstrating that self-report, alongside measures of physiological indicators of stress, further inform the use of deep-learning models when detecting stress between individuals. This study, however, is not without limitations. The data collected here come from a relatively small sample (18 participants) with a greater number of non-stress than stress instances. This may have resulted in a class imbalance problem which could have impacted the model performance accuracy level [118]. In addition, the data collected in the study has a high degree of variability which makes a standard set of models inappropriate for handling this type of data and highlights the importance of exploring unconstrained methods such as deep learning for stress discrimination. As such, it is advisable that future research be conducted using these measures and performing the TSST with a greater number of participants, potentially building a more variable dataset. Furthermore, the physiological data used in this study was obtained from a research-grade wearable. Future work should evaluate the performance of these types of models with data from consumer wearable devices given their widespread availability and their implications for use by the general public as opposed to the research community. Additionally, the dataset used in this study was the result of an induced stress procedure, which allowed for a more careful analysis of the stress response, but which simplified the stress detection task in comparison to what would be observed under field conditions [47]. Future studies should perform analyses under various stress conditions that more closely resemble day-to-day stress experiences. The study presented in this manuscript is a first step towards investigating the feasibility of using data streams from wearable devices that reflect autonomic nervous activity, specifically sympathetic activation, in combination with subjective information of psychological states (i.e., stress and anxiety ratings), to identify different levels of mental stress in a non-invasive manner. Acute stress triggers cardiovascular changes such as increased heart rate, skin conductivity, and acceleration of other body functions. Additionally, experiencing stressful situations results in highly subjective feelings (e.g., perceived stress and positive and negative affect), dependent upon people's personality and mindset [114]. Building large multi-dimensional datasets that include not only wearable data but also ratings of perceived stress and anxiety, could help reduce the intra-individual variabilities that come with the subjective and dual nature of the stress response, as well as help correlate real-life situations with perceptions of stress, and therefore aid in the development of a more robust stress detection method.

Nowadays, commercial wearables such as the Apple Watch (Apple Inc, Cupertino, CA), Fitbit (Alphabet Inc, San Francisco, CA), and the Halo band (Amazon.com Inc, Seattle, WA) include PPG, and in some cases, EDA and ECG sensors. Fitbit in particular uses primarily EDA for stress tracking. The use of these types of sensors for emotion detection in real-world environments is challenging, and the reliability of the measurements could be improved by prompting the user to indicate a true vs. a false stress event when the device picks up on signs of stress. Implementing this type of strategy for a limited period of time could help companies differentiate stress moments from electronic or motion artifacts and therefore help build algorithms that better understand and discriminate particular individual daily stressors. Future research will evaluate the reproducibility of this framework in real-world scenarios using consumer-facing wearables.

V. CONCLUSION

The goal of this research was to assess the feasibility of detecting stress events in a semi-restricted environment using physiological and self-reported stress, anxiety data, and deep learning methods. The measurements in the dataset came from 18 subjects who underwent the Trier Social Stress Test at the Well Living Lab. Our results are similar to those found in the literature, and they suggest that mental stress increases the levels of salivary cortisol, self-ratings of stress, heart rate, and skin conductance. More importantly, these findings show that deep learning methods allow for physiological and behavioral measures, collected at vastly different frequencies, to be combined into a single model that outperforms either data type individually. However, this was a small and highly correlated dataset and therefore a larger amount of data is needed to confirm the obtained results and improve the classification performance. Nevertheless, this framework shows potential in terms of being able to rapidly collect stress and anxiety ratings through different day-to-day activities that can help correctly label physiological and other sensor data as stress and non-stress and thus, contribute to the personalization of this type of algorithm.

ACKNOWLEDGMENT

The authors thank the study participants for their time and commitment and the staff of the Well Living Lab for their support. We acknowledge Jo Bernau, Carole Wolfe, Kevin Hovde, Christian Ramos, Syed Shabih Hassan, and Jeyakumar Raman for their contributions to the study. The Well Living Lab was launched in 2016 as a collaboration between Delos Living and Mayo Clinic; neither party provided any oversight nor editing of the manuscript. The funding did not influence any aspect of this study design and reporting of results.

REFERENCES

[1] Centers for Disease Control and Prevention. (2018). *Mental Health in the Workplace*. [Online]. Available: https://www.cdc.gov/workplacehealthpromotion/tools-resources/workplace-health/mental-health/index.html

[2] Gallup. (2019). *Gallup Global Emotions 2019*. [Online]. Available: https://www.gallup.com/analytics/248909/gallup-2019-global-emotions-report-pdf.aspx

[3] K. Bhui, S. Dinos, M. Galant-Miecznikowska, B. de Jongh, and S. Stansfeld, "Perceptions of work stress causes and effective interventions in employees working in public, private and non-governmental organisations: A qualitative study," *BJPsych Bull.*, vol. 40, no. 6, pp. 318–325, Dec. 2016, doi: 10.1192/pb.bp.115.050823.

[4] P. M. Bongers, C. R. de Winter, M. A. Kompier, and V. H. Hildebrandt, "Psychosocial factors at work and musculoskeletal disease," *Scandin. J. Work, Environ. Health*, vol. 19, no. 5, pp. 297–312, Oct. 1993, doi: 10.5271/sjweh.1470.

[5] American Psychological Association, United States. (2013). *APA's Work and Well-Being Survey*. [Online]. Available: https://www.apa.org/news/press/releases/2013/03/employee-needs

[6] G. W. Evans and J. M. McCoy, "When buildings don't work: The role of architecture in human health," *J. Environ. Psychol.*, vol. 18, no. 1, pp. 85–94, Mar. 1998, doi: 10.1006/jevp.1998.0089.

[7] A. Liebl, J. Haller, B. Jödicke, H. Baumgartner, S. Schlittmeier, and J. Hellbrück, "Combined effects of acoustic and visual distraction on cognitive performance and well-being," *Appl. Ergonom.*, vol. 43, no. 2, pp. 424–434, Mar. 2012, doi: 10.1016/j.apergo.2011.06.017.

[8] L. Lan, Z. Lian, and L. Pan, "The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings," *Appl. Ergonom.*, vol. 42, no. 1, pp. 29–36, Dec. 2010, doi: 10.1016/j.apergo.2010.04.003.

[9] E. De Croon, J. Sluiter, P. P. Kuijer, and M. Frings-Dresen, "The effect of office concepts on worker health and performance: A systematic review of the literature," *Ergonomics*, vol. 48, no. 2, pp. 119–134, Feb. 2005, doi: 10.1080/00140130512331319409.

[10] M. Haka, A. Haapakangas, J. Keränen, J. Hakala, E. Keskinen, and V. Hongisto, "Performance effects and subjective disturbance of speech in acoustically different office types—A laboratory experiment," *Indoor Air*, vol. 19, no. 6, pp. 454–467, Dec. 2009, doi: 10.1111/j.1600-0668.2009.00608.x.

[11] R. Rylander, "Physiological aspects of noise-induced stress and annoyance," *J. Sound Vib.*, vol. 277, no. 3, pp. 471–478, Oct. 2004, doi: 10.1016/j.jsv.2004.03.008.

[12] S. A. Stansfeld and M. P. Matheson, "Noise pollution: Non-auditory effects on health," *Brit. Med. Bull*, vol. 68, no. 1, pp. 243–257, Dec. 2003, doi: 10.1093/bmb/ldg033.

[13] J. C. Westman and J. R. Walters, "Noise and stress: A comprehensive approach," *Environ. Health Perspect.*, vol. 41, pp. 291–309, Oct. 1981.

[14] G. W. Evans, M. Bullinger, and S. Hygge, "Chronic noise exposure and physiological response: A prospective study of children living under environmental stress," *Psychol. Sci.*, vol. 9, no. 1, pp. 75–77, Jan. 1998.

[15] A. L. Dougall and A. Baum, "Stress, coping, and immune function," in *Handbook of Psychology, American Cancer Society*. Hoboken, NJ, USA: Wiley, 2003, pp. 441–455, doi: 10.1002/0471264385.wei0316.

[16] C. Hammen, E. Y. Kim, N. K. Eberhart, and P. A. Brennan, "Chronic and acute stress and the prediction of major depression in women," *Depression Anxiety*, vol. 26, no. 8, pp. 718–723, Aug. 2009, doi: 10.1002/da.20571.

[17] E. M. Jackson, "STRESS RELIEF: The role of exercise in stress management," *ACSM'S Health Fitness J.*, vol. 17, no. 3, pp. 14–19, May 2013, doi: 10.1249/FIT.0b013e31828cb1c9.

[18] J. Bakker, M. Pechenizkiy, and N. Sidorova, "What's your current stress level? Detection of stress patterns from GSR sensor data," in *Proc. IEEE 11th Int. Conf. Data Mining Workshops*, Dec. 2011, pp. 573–580, doi: 10.1109/ICDMW.2011.178.

[19] N. Schneiderman, G. Ironson, and S. D. Siegel, "Stress and health: Psychological, behavioral, and biological determinants," *Annu. Rev. Clin. Psychol.*, vol. 1, no. 1, pp. 607–628, Apr. 2005, doi: 10.1146/annurev.clinpsy.1.102803.144141.

[20] H. Yaribeygi, Y. Panahi, H. Sahraei, T. P. Johnston, and A. Sahebkar, "The impact of stress on body function: A review," *EXCLI J*, vol. 16, pp. 1057–1072, Jul. 2017, doi: 10.17179/excli2017-480.

[21] S. Vrshek-Schallhorn, C. B. Stroud, S. Mineka, C. Hammen, R. E. Zinbarg, K. Wolitzky-Taylor, and M. G. Craske, "Chronic and episodic interpersonal stress as statistically unique predictors of depression in two samples of emerging adults," *J. Abnormal Psychol.*, vol. 124, no. 4, pp. 918–932, Nov. 2015, doi: 10.1037/abn0000088.

[22] M. M. Larzelere and G. N. Jones, "Stress and health," *Primary Care Clinics Office Pract.*, vol. 35, no. 4, pp. 839–856, Dec. 2008, doi: 10.1016/j.pop.2008.07.011.

[23] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology*, 13th ed. Saunders, 2016.

[24] B. L. Seaward, *Managing Stress: Principles and Strategies for Health and Well-Being*, 8th ed. Burlington, MA, USA: Jones & Bartlett Learning, 2014.

[25] K. Dedovic, A. Duchesne, J. Andrews, V. Engert, and J. C. Pruessner, "The brain and the stress axis: The neural correlates of cortisol regulation in response to stress," *NeuroImage*, vol. 47, no. 3, pp. 864–871, Sep. 2009, doi: 10.1016/j.neuroimage.2009.05.074.

[26] S. Sondhi, M. Khan, R. Vijay, and A. K. Salhan, "Vocal indicators of emotional stress," *Int. J. Comput. Appl.*, vol. 122, no. 15, pp. 38–43, Jul. 2015, doi: 10.5120/21780-5056.

[27] M. Gavrilescu and N. Vizireanu, "Predicting depression, anxiety, and stress levels from videos using the facial action coding system," *Sensors*, vol. 19, no. 17, p. 3693, Aug. 2019, doi: 10.3390/s19173693.

[28] H. Gao, A. Yuce, and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," in *Proc. ICIP*, 2014, pp. 5961–5965, doi: 10.1109/ICIP.2014.7026203.

[29] J. Zhang, X. Mei, H. Liu, S. Yuan, and T. Qian, "Detecting negative emotional stress based on facial expression in real time," in *Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2019, pp. 430–434, doi: 10.1109/SIPROCESS.2019.8868735.

[30] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *J. Health Social Behav.*, vol. 24, no. 4, pp. 385–396, 1983, doi: 10.2307/2136404.

[31] C. D. Spielberger, R. L. Gorsuch, R. Lushene, P. R. Vagg, and G. A. Jacobs, *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA, USA: Consulting Psychologists Press, 1983.

[32] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *J. Biomed. Informat.*, vol. 92, Apr. 2019, Art. no. 103139, doi: 10.1016/j.jbi.2019.103139.

[33] G. M. Slavich, S. Taylor, and R. W. Picard, "Stress measurement using speech: Recent advancements, validation issues, and ethical and privacy considerations," *Stress*, vol. 22, no. 4, pp. 408–413, Jul. 2019, doi: 10.1080/10253890.2019.1584180.

[34] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic stress detection in working environments from Smartphones' accelerometer data: A first step," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 4, pp. 1053–1060, Jul. 2016, doi: 10.1109/JBHI.2015.2446195.

[35] V. Motti and K. Caine, "Users' privacy concerns about wearables: Impact of form factor, sensors and type of data collected," vol. 8976, Jan. 2015, doi: 10.1007/978-3-662-48051-9_16.

[36] A. Sano, S. Taylor, A. W. McHill, A. J. Phillips, L. K. Barger, E. Klerman, and R. Picard, "Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study," *J. Med. Internet Res.*, vol. 20, no. 6, p. e210, Jun. 2018, doi: 10.2196/jmir.9410.

[37] S. Saganowski, P. Kazienko, M. Dziezyc, P. Jakimów, J. Komoszynska, W. Michalska, A. Dutkowiak, A. Polak, A. Dziadek, and M. Ujma, "Consumer wearables and affective computing for wellbeing support," 2020. *arXiv:2005.00093*. [Online]. Available: https://arxiv.org/abs/2005.00093

[38] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 671–676, doi: 10.1109/ACII.2013.117.

[39] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. T. X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," 2014, doi: 10.1145/2632048.2632054.

[40] G. Bauer and P. Lukowicz, "Can smartphones detect stress-related changes in the behaviour of individuals?" in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2012, pp. 423–426, doi: 10.1109/PerComW.2012.6197525.

[41] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, pp. 1–21, Dec. 2015, doi: 10.1186/s40537-014-0007-7.

[42] D. Chen, S. Liu, P. Kingsbury, S. Sohn, C. B. Storlie, E. B. Habermann, J. M. Naessens, D. W. Larson, and H. Liu, "Deep learning and alternative learning strategies for retrospective real-world clinical data," *NPJ Digit. Med.*, vol. 2, no. 1, p. 43, Dec. 2019, doi: 10.1038/s41746-019-0122-0.

[43] R. Ahuja and A. Banga, "Mental stress detection in university students using machine learning algorithms," *Procedia Comput. Sci.*, vol. 152, pp. 349–353, Jan. 2019, doi: 10.1016/j.procs.2019.05.007.

[44] N. Keshan, P. V. Parimi, and I. Bichindaritz, "Machine learning for stress detection from ECG signals in automobile drivers," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 2661–2669, doi: 10.1109/BigData.2015.7364066.

[45] S. Elzeiny and M. Qaraqe, "Machine learning approaches to automatic stress detection: A review," in *Proc. IEEE/ACS 15th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Oct. 2018, pp. 1–6, doi: 10.1109/AICCSA.2018.8612825.

[46] S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," *Biocybernetics Biomed. Eng.*, vol. 39, no. 2, pp. 444–469, Apr. 2019, doi: 10.1016/j.bbe.2019.01.004.

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

[48] E. Smets, E. R. Velazquez, G. Schiavone, I. Chakroun, E. D'Hondt, W. De Raedt, J. Cornelis, O. Janssens, S. Van Hoecke, S. Claes, I. Van Diest, and C. Van Hoof, "Large-scale wearable data reveal digital phenotypes for daily-life stress detection," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–10, Dec. 2018, doi: 10.1038/s41746-018-0074-9.

[49] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla, "A stress sensor based on galvanic skin response (GSR) controlled by ZigBee," *Sensors*, vol. 12, no. 5, pp. 6075–6101, May 2012, doi: 10.3390/s120506075.

[50] R. Zangróniz, A. Martínez-Rodrigo, J. Pastor, M. López, and A. Fernández-Caballero, "Electrodermal activity sensor for classification of calm/distress condition," *Sensors*, vol. 17, no. 10, p. 2324, Oct. 2017, doi: 10.3390/s17102324.

[51] S. Amalan, A. Shyam, A. Anusha, S. Preejith, A. Tony, J. Jayaraj, and S. Mohanasankar, "Electrodermal activity based classification of induced stress in a controlled setting," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2018, pp. 1–6, doi: 10.1109/MeMeA.2018.8438703.

[52] 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.

[53] H. Kurniawan, A. V. Maslov, and M. Pechenizkiy, "Stress detection from speech and galvanic skin response signals," in *Proc. 26th IEEE Int. Symp. Comput.-Based Med. Syst.*, Jun. 2013, pp. 209–214, doi: 10.1109/CBMS.2013.6627790.

[54] A. S. Anusha, P. Sukumaran, V. Sarveswaran, S. S. Kumar, A. Shyam, T. J. Akl, S. P. Preejith, and M. Sivaprakasam, "Electrodermal activity based pre-surgery stress detection using a wrist wearable," *IEEE J. Biomed. Health Inf.*, vol. 24, no. 1, pp. 92–100, Jan. 2020, doi: 10.1109/JBHI.2019.2893222.

[55] 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 Invest.*, vol. 15, no. 3, pp. 235–245, Mar. 2018, doi: 10.30773/pi.2017.08.17.

[56] S. Boonnithi and S. Phongsuphap, "Comparison of heart rate variability measures for mental stress detection," in *Proc. Comput. Cardiol.*, Sep. 2011, pp. 85–88.

[57] P. Melillo, M. Bracale, and L. Pecchia, "Nonlinear heart rate variability features for real-life stress detection. Case study: Students under stress due to university examination," *Biomed. Eng. OnLine*, vol. 10, no. 1, p. 96, 2011, doi: 10.1186/1475-925X-10-96.

[58] I. Mohino-Herranz, R. Gil-Pita, M. Rosa-Zurera, and F. Seoane, "Activity recognition using wearable physiological measurements: Selection of features from a comprehensive literature study," *Sensors*, vol. 19, no. 24, p. 5524, Dec. 2019, doi: 10.3390/s19245524.

[59] A.-F. The, I. Reijmerink, M. van der Laan, and F. Cnossen, "Heart rate variability as a measure of mental stress in surgery: A systematic review," *Int. Arch. Occupational Environ. Health*, vol. 93, no. 7, pp. 805–821, Oct. 2020, doi: 10.1007/s00420-020-01525-6.

[60] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *J. Biomed. Informat.*, vol. 59, pp. 49–75, Feb. 2016, doi: 10.1016/j.jbi.2015.11.007.

[61] S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy, "Stress detection in working people," *Procedia Comput. Sci.*, vol. 115, pp. 359–366, Jan. 2017, doi: 10.1016/j.procs.2017.09.090.

[62] A. S. de Sierra, C. S. Ávila, G. B. del Pozo, and J. G. Casanova, "Stress detection by means of stress physiological template," in *Proc. 3rd World Congr. Nature Biologically Inspired Comput.*, Oct. 2011, pp. 131–136, doi: 10.1109/NaBIC.2011.6089448.

[63] A. de Santos Sierra, C. S. Avila, J. G. Casanova, and G. B. del Pozo, "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4857–4865, Oct. 2011, doi: 10.1109/TIE.2010.2103538.

[64] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005, doi: 10.1109/TITS.2005.848368.

[65] V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. Mozos, "Stress detection using wearable physiological sensors," in *Artificial Computation in Biology and Medicine*. Jun. 2015, pp. 526–532, doi: 10.1007/978-3-319-18914-7_55.

[66] R. Martinez, E. Irigoyen, A. Arruti, J. I. Martin, and J. Muguerza, "A real-time stress classification system based on arousal analysis of the nervous system by an F-state machine," *Comput. Methods Programs Biomed.*, vol. 148, pp. 81–90, Sep. 2017, doi: 10.1016/j.cmpb.2017.06.010.

[67] C. Goumopoulos and E. Menti, "Stress detection in seniors using biosensors and psychometric tests," *Procedia Comput. Sci.*, vol. 152, pp. 18–27, Jan. 2019, doi: 10.1016/j.procs.2019.05.022.

[68] H. F. Posada-Quintero, T. Dimitrov, A. Moutran, S. Park, and K. H. Chon, "Analysis of reproducibility of noninvasive measures of sympathetic autonomic control based on electrodermal activity and heart rate variability," *IEEE Access*, vol. 7, pp. 22523–22531, Feb. 2019, doi: 10.1109/ACCESS.2019.2899485.

[69] S. Ghiasi, A. Greco, R. Barbieri, E. P. Scilingo, and G. Valenza, "Assessing autonomic function from electrodermal activity and heart rate variability during cold-pressor test and emotional challenge," *Sci. Rep.*, vol. 10, no. 1, pp. 1–3, Mar. 2020, doi: 10.1038/s41598-020-62225-2.

[70] M. Ciman and K. Wac, "Individuals' stress assessment using human-smartphone interaction analysis," *IEEE Trans. Affect. Comput.*, vol. 9, no. 1, pp. 51–65, Jan. 2018, doi: 10.1109/TAFFC.2016.2592504.

[71] M. Sysoev, A. Kos, and M. Pogačnik, "Noninvasive stress recognition considering the current activity," *Pers. Ubiquitous Comput.*, vol. 19, no. 7, pp. 1045–1052, Aug. 2015, doi: 10.1007/s00779-015-0885-5.

[72] E. Vildjiounaite, J. Kallio, V. Kyllönen, M. Nieminen, I. Määttänen, M. Lindholm, J. Mäntyjärvi, and G. Gimel'farb, "Unobtrusive stress detection on the basis of smartphone usage data," *Pers. Ubiquitous Comput.*, vol. 22, no. 4, pp. 671–688, Aug. 2018, doi: 10.1007/s00779-017-1108-z.

[73] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, M. Al'absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. E. Wittmers, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proc. 10th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, Apr. 2011, pp. 97–108.

[74] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "CStress: Towards a gold standard for continuous stress assessment in the mobile environment," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. (UbiComp)*, 2015, pp. 493–504, doi: 10.1145/2750858.2807526.

[75] Y. S. Can, D. Gokay, D. R. Kılıç, D. Ekiz, N. Chalabianloo, and C. Ersoy, "How laboratory experiments can be exploited for monitoring stress in the wild: A bridge between laboratory and daily life," *Sensors*, vol. 20, no. 3, p. 838, Feb. 2020, doi: 10.3390/s20030838.

[76] M. A. Birkett, "The trier social stress test protocol for inducing psychological stress," *J. Visualized Exp.*, vol. 56, p. e3238, Oct. 2011, doi: 10.3791/3238.

[77] K. E. Hannibal and M. D. Bishop, "Chronic stress, cortisol dysfunction, and pain: A psychoneuroendocrine rationale for stress management in pain rehabilitation," *Phys. Therapy*, vol. 94, no. 12, pp. 1816–1825, Dec. 2014, doi: 10.2522/ptj.20130597.

[78] C.-J. Huang, H. E. Webb, M. C. Zourdos, and E. O. Acevedo, "Cardiovascular reactivity, stress, and physical activity," *Frontiers Physiol.*, vol. 4, p. 314, 2013, doi: 10.3389/fphys.2013.00314.

[79] M. Garbarino, M. Lai, S. Tognetti, R. Picard, and D. Bender, "Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition," in *Proc. 4th Int. Conf. Wireless Mobile Commun. Healthcare-'Transforming Healthcare Through Innov. Mobile Wireless Technol.*, 2014, pp. 39–42, doi: 10.4108/icst.mobihealth.2014.257418.

[80] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, and K. H. Chon, "Electrodermal activity is sensitive to cognitive stress under water," *Frontiers Physiol.*, vol. 8, p. 1128, Jan. 2018, doi: 10.3389/fphys.2017.01128.

[81] H. D. Critchley, "Review: Electrodermal responses: What happens in the brain," *Neuroscientist*, vol. 8, no. 2, pp. 132–142, Apr. 2002, doi: 10.1177/107385840200800209.

[82] W. Boucsein, M. J. Christie, R. Edelberg, W. W. Grings, D. T. Lykken, and P. H. Venables, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, no. 8, pp. 1017–1034, Aug. 2012, doi: 10.1111/j.1469-8986.2012.01384.x.

[83] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, "A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments," *Psychophysiology*, vol. 49, no. 1, pp. 1017–1034, 2013.

[84] R. Zangróniz, A. Martínez-Rodrigo, M. López, J. Pastor, and A. Fernández-Caballero, "Estimation of mental distress from photoplethysmography," *Appl. Sci.*, vol. 8, no. 1, p. 69, Jan. 2018, doi: 10.3390/app8010069.

[85] M. Abbod, Y.-R. Chiou, S.-H. Yang, S.-Z. Fan, and J.-S. Shieh, "Developing a monitoring psychological stress index system via photoplethysmography," *Artif. Life Robot.*, vol. 16, no. 3, pp. 430–433, Dec. 2011, doi: 10.1007/s10015-011-0976-y.

[86] Y. Zheng, T. C. H. Wong, B. H. K. Leung, and C. C. Y. Poon, "Unobtrusive and multimodal wearable sensing to quantify anxiety," *IEEE Sensors J.*, vol. 16, no. 10, pp. 3689–3696, May 2016, doi: 10.1109/JSEN.2016.2539383.

[87] M. Ghamari, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *Int. J. Biosensors Bioelectron.*, vol. 4, no. 4, pp. 195–202, 2018, doi: 10.15406/ijbsbe.2018.04.00125.

[88] C. Schubert, M. Lambertz, R. A. Nelesen, W. Bardwell, J.-B. Choi, and J. E. Dimsdale, "Effects of stress on heart rate complexity—A comparison between short-term and chronic stress," *Biol. Psychol.*, vol. 80, no. 3, pp. 325–332, Mar. 2009, doi: 10.1016/j.biopsycho.2008.11.005.

[89] K. A. Herborn, J. L. Graves, P. Jerem, N. P. Evans, R. Nager, D. J. McCafferty, and D. E. F. McKeegan, "Skin temperature reveals the intensity of acute stress," *Physiol. Behav.*, vol. 152, pp. 225–230, Dec. 2015, doi: 10.1016/j.physbeh.2015.09.032.

[90] T. Oka, K. Oka, and T. Hori, "Mechanisms and mediators of psychological stress-induced rise in core temperature," *Psychosomatic Med.*, vol. 63, no. 3, pp. 476–486, May 2001, doi: 10.1097/00006842-200105000-00018.

[91] T. Hui and R. Sherratt, "Coverage of emotion recognition for common wearable biosensors," *Biosensors*, vol. 8, no. 2, p. 30, Mar. 2018, doi: 10.3390/bios8020030.
[92] E. K. Zavadskas, A. Kaklauskas, M. Seniut, G. Dzemyda, S. Ivanikovas, V. Stankevic, C. Simkevicius, and A. Jarusevicius, "Web-based biometric mouse intelligent system for analysis of emotional state and labour productivity," in *Proc. 25th Int. Symp. Autom. Robot. Construct. (ISARC)*, Jun. 2008, doi: 10.3846/isarc.20080626.429.
[93] M. T. Quazi, S. C. Mukhopadhyay, N. K. Suryadevara, and Y. M. Huang, "Towards the smart sensors based human emotion recognition," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2012, pp. 2365–2370, doi: 10.1109/I2MTC.2012.6229646.
[94] S. Aristizabal, P. Porter, N. Clements, C. Campanella, R. Zhang, K. Hovde, and C. Lam, "Conducting human-centered building science at the well living lab," *Technol.|Archit. Des.*, vol. 3, no. 2, pp. 161–173, Jul. 2019, doi: 10.1080/24751448.2019.1640535.
[95] N. Clements, R. Zhang, A. Jamrozik, C. Campanella, and B. Bauer, "The spatial and temporal variability of the indoor environmental quality during three simulated office studies at a living lab," *Buildings*, vol. 9, no. 3, p. 62, Mar. 2019, doi: 10.3390/buildings9030062.
[96] A. Jamrozik, C. Ramos, J. Zhao, J. Bernau, N. Clements, T. V. Wolf, and B. Bauer, "A novel methodology to realistically monitor office occupant reactions and environmental conditions using a living lab," *Building Environ.*, vol. 130, pp. 190–199, Feb. 2018, doi: 10.1016/j.buildenv.2017.12.024.
[97] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. B. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and Jupyter Development Team, *Jupyter Notebooks—A Publishing Format for Reproducible Computational Workflows*. 2016, doi: 10.3233/978-1-61499-649-1-87.
[98] A. Deadorf, "Tableau (version 9.1)," *J. Med. Library Assoc.*, vol. 104, no. 2, pp. 182–184, Apr. 2016.
[99] H. White, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, vol. 48, no. 4, pp. 817–838, 1980, doi: 10.2307/1912934.
[100] S.-H. Song and D. K. Kim, "Development of a stress classification model using deep belief networks for stress monitoring," *Healthcare Inf. Res.*, vol. 23, no. 4, pp. 285–292, Oct. 2017, doi: 10.4258/hir.2017.23.4.285.
[101] F. Chollet. (2015). *Keras*. [Online]. Available: https://keras.io
[102] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, and H. Shah, "Wide & deep learning for recommender systems," 2016. Accessed: Oct. 18, 2020, *arXiv:1606.07792*. [Online]. Available: http://arxiv.org/abs/1606.07792
[103] A. Géron, *Neural Networks and Deep Learning*. Newton, MA,USA: O'Reilly, 2018.
[104] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
[105] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Jan. 2017, Accessed: May 19, 2021. [Online]. Available: http://arxiv.org/abs/1412.6980
[106] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'—A tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, nos. 1–2, pp. 76–81, 1993, doi: 10.1159/000119004.
[107] M. Qi, H. Gao, L. Guan, G. Liu, and J. Yang, "Subjective stress, salivary cortisol, and electrophysiological responses to psychological stress," *Frontiers Psychol.*, vol. 7, Feb. 2016, doi: 10.3389/fpsyg.2016.00229.
[108] J. R. Wilson and K. A. Lorenz, "Generalized estimating equations logistic regression," in *Modeling Binary Correlated Responses using SAS, SPSS and R*, J. R. Wilson and K. A. Lorenz, Eds. Cham, Switzerland: Springer, 2015, pp. 103–130, doi: 10.1007/978-3-319-23805-0_6.
[109] R. Castaldo, L. Montesinos, P. Melillo, S. Massaro, and L. Pecchia, "To what extent can we shorten HRV analysis in wearable sensing? A case study on mental stress detection," in *Proc. Nordic-Baltic Conf. Biomed. Eng. Med. Phys.*, Jun. 2017, pp. 643–646, doi: 10.1007/978-981-10-5122-7_161.
[110] J. Hellhammer and M. Schubert, "The physiological response to trier social stress test relates to subjective measures of stress during but not before or after the test," *Psychoneuroendocrinology*, vol. 37, no. 1, pp. 119–124, Jan. 2012, doi: 10.1016/j.psyneuen.2011.05.012.
[111] Y. Shiban, J. Diemer, S. Brandl, R. Zack, A. Mühlberger, and S. Wüst, "Trier social stress test *in vivo* and in virtual reality: Dissociation of response domains," *Int. J. Psychophysiol.*, vol. 110, pp. 47–55, Dec. 2016, doi: 10.1016/j.ijpsycho.2016.10.008.
[112] M. Ziegler, "Psychological stress and the autonomic nervous system," in *Primer on the Autonomic Nervous System*. 2004, pp. 189–190, doi: 10.1016/B978-012589762-4/50051-7.
[113] A. Oldehinkel, J. Ormel, N. M. Bosch, E. M. C. Bouma, A. M. Van Roon, J. G. M. Rosmalen, and H. Riese, "Stressed out? Associations between perceived and physiological stress responses in adolescents: The TRAILS study," *Psychophysiology*, vol. 48, pp. 52–441, Apr. 2011, doi: 10.1111/j.1469-8986.2010.01118.x.
[114] Y. Xin, J. Wu, Z. Yao, Q. Guan, A. Aleman, and Y. Luo, "The relationship between personality and the response to acute psychological stress," *Sci. Rep.*, vol. 7, no. 1, p. 16906, Dec. 2017, doi: 10.1038/s41598-017-17053-2.
[115] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders, "Towards mental stress detection using wearable physiological sensors," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 1798–1801, doi: 10.1109/IEMBS.2011.6090512.
[116] J. R. Rojas, J.-H. Hong, and A. Dey, "Stress recognition: A step outside the lab," in *Proc. PhyCS*, 2014, pp. 107–118, doi: 10.5220/0004725701070118.
[117] K. Kyriakou, B. Resch, G. Sagl, A. Petutschnig, C. Werner, D. Niederseer, M. Liedlgruber, F. Wilhelm, T. Osborne, and J. Pykett, "Detecting moments of stress from measurements of wearable physiological sensors," *Sensors*, vol. 19, no. 17, p. 3805, Sep. 2019, doi: 10.3390/s19173805.
[118] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, p. 27, Mar. 2019, doi: 10.1186/s40537-019-0192-5.

SARA ARISTIZABAL (Member, IEEE) was born in Medellin, Colombia, in January 1989. She received the B.S. degree in biomedical engineering from Escuela de Ingenieria de Antioquia, Medellin, in 2011, and the Ph.D. degree in biomedical engineering and physiology from the Mayo Clinic Graduate School of Biomedical Sciences, Rochester, MN, USA, in 2017, with an emphasis in medical imaging.

She is currently a Research Scientist with the Well Living Lab, a research collaboration between Delos and the Mayo Clinic, Rochester, and with Delos Labs, NYC, USA. She is also an Assistant Professor with the Department of Medicine, Mayo Clinic College of Medicine, Rochester. Her research interests include the development of biometric indicators to assess health and wellness using wearable and non-invasive physiological sensors, the impact of biophilic design on human behavior, and the effect of the indoor built environment on human health.

KUNJOON BYUN was born in Seoul, South Korea. He received the B.A. degree in psychology from the University of Wisconsin-Madison, USA, in 2011, and the M.A. degree in psychology from College of William and Mary, USA, in 2016.

From 2017 to 2018, he worked as a Project Manager at Carnegie Mellon University, PA, USA. He is currently a Senior Research Analyst at the Well Living Lab, Rochester, MN, USA. His research interests include the areas of human cognition, emotion, and judgment and decision making.

Mr. Byun's previous work has been presented at American Psychological Association, Society for Judgment and Decision Making, and International Society for Environmental Epidemiology.

NADIA WOOD was born in Lahore, Pakistan, in 1978. She received the B.S. degree in computer science from Winona State University, MN, USA, in 2004. She is currently pursuing the M.C.S. degree in data science from the University of Illinois at Urbana-Champaign, IL, USA.

From 1999 to 2002, she was a Research Assistant with the Virginia Modeling and Simulation Center, Suffolk, VA, USA. From 2005 to 2019, she worked as a Software Engineer and Architect at Mayo Clinic, Rochester, MN, USA. She is currently a Technology Architect at the Well Living Lab, Rochester. Her research interests include ways to uncover relationships between indoor environment and human health using machine learning, data management strategies for advanced analytics, and software development.

AIDAN F. MULLAN was born in Iowa City, IA, USA, in 1996. He received the B.A. degree in mathematics and psychology from Carleton College, in 2018, and the M.A. degree in statistics from the University of California, Berkeley, in 2019.

From 2018 to 2019, he was a Consultant for biomedical research with the Department of Radiology, University of Iowa. Since 2019, he has been working as a Statistician with Mayo Clinic, where he become the Lead Statistician for emergency medicine, in 2020. His research interests include improving interpretability of machine learning and deep learning techniques, development of real-time clinical prediction tools for the diagnosis of sepsis and time-critical conditions, and the characterization and progression of early-onset Parkinson's Disease.

PAIGE M. PORTER was born in Chicago, IL, USA. She received the B.A. degree in psychology and the M.S. degree in environmental science with a specialization in behavior, education, and communication from the University of Michigan, Ann Arbor, in 2017 and 2021, respectively.

From 2017 to 2019, she was a Behavioral Research Analyst with the Well Living Lab, a collaboration between Delos and the Mayo Clinic, Rochester, MN, USA. Her research interests include community-wide behavior adaptations to emerging climate change and energy challenges, and how to instill adaptation and leadership competences through formal education pathways. She is a member of the American Society of Adaptation Professionals. She was a recipient of the 2019 Academic Merit Fellowship with School for Environment and Sustainability, University of Michigan.

CAROLINA CAMPANELLA received the B.S. degree in biology from Purdue University, West Lafayette, IN, USA, in 2005, and the Ph.D. degree in cognitive and developmental psychology from Emory University, Atlanta, GA, USA, in 2014.

From 2014 to 2016, she was a Postdoctoral Research Fellow in cognitive neuroscience with the University of Massachusetts, Amherst, MA, USA, where she studied the impact of napping on cognitive performance in early childhood. She is currently a Behavioral Scientist and the Vice President with Delos Living LLC, where she works as a part of the Research and Development Team to develop technology-based solutions to improve health and well-being in the built environment and supports marketing and product development by conducting research on the customer experience.

ANJA JAMROZIK received the B.S. degree in psychology and cognitive science from McGill University, Montreal, Canada, in 2007, and the Ph.D. degree in cognitive psychology from Northwestern University, Evanston, IL, USA, in 2014.

From 2015 to 2019, she was a Research Consultant with the Well Living Lab, and has been with Delos Living LLC, since 2020. In 2021, she joined Composer as a Founding Researcher and the Strategists. She also consults on user and customer experience research and product strategy for technology companies and has a long-standing interest in people's experience in the built environment.

IVAN Z. NENADIC received the B.A. degree in mathematics and physics from the Saint Olaf College, Northfield, MN, USA, in 2006, the Ph.D. degree in physiology and biomedical engineering from the Mayo Graduate School, Rochester, MN, USA, in 2011, and the M.D. degree from the Mayo Clinic School of Medicine, Rochester, in 2020.

From 2011 to 2013, he was a Postdoctoral Fellow with the Mayo Clinic Basic Ultrasound Research Laboratory, where he was a Research Associate, from 2014 to 2015. He was an Assistant Professor with the Mayo Clinic College of Medicine, from 2015 to 2020. In 2020, he joined the Department of Internal Medicine, University of Michigan Hospital, Ann Arbor, MI, USA. He has authored more than 50 articles. His research interests include ultrasound and magnetic resonance imaging, shear wave elastography, physiological signal processing, physiological systems modeling, wearable technology, and machine learning.

Dr. Nenadic is currently the Editor-in-Chief of one textbook.

BRENT A. BAUER received the M.D. degree from the Mayo Clinic Medical School, in 1988.

He has been a Professor of medicine with the Mayo Clinic, MN, USA, since 1992. His main research interest includes the scientific evaluation of integrative medicine therapies. He has authored several book chapters and over 100 articles on this topic and the Medical Editor of the *Mayo Clinic Guide to Integrative Medicine* and *The Science of Integrative Medicine*, and the Great Courses video series. He is currently the Founder of the Integrative Medicine Program, Mayo Clinic. He is also the former Medical Director of the Well Living Lab, a collaboration between Delos and Mayo Clinic Center for Innovation. He was a recipient of numerous national awards, including the New Leadership Development Award, the Glaxo Welcome, in 1999, and the Edison Award for the Well Living Lab, in 2018.