An Individual-Oriented Algorithm for Stress Detection in Wearable Sensor Measurements

Martin Karl Moser, Bernd Resch, and Maximilian Ehrhart

Abstract—**Accurately measuring a person's level of stress can have a wide variety of impacts, not only on human health, but also on the perceived feeling of safety when going after daily habits, such as walking, cycling, or driving from one place to another. While there is a vast amount of research done on stress and the related physiological responses of the human body, there is no go-to method when it comes to measuring acute stress in a live setting. This work proposes an advancement of the rule-based stress detection algorithm proposed by Kyriakou et al., to identify moments of stress (MOS) more reliably, through an adaptation, and individualization of the rules proposed in the original paper. The proposed algorithm leverages electrodermal activity (EDA)**

and skin temperature (ST), both recorded by the Empatica E4 wristband, for the assessment of an individual's stress when exposed to an audible stimulus. The algorithm achieves an average recall of 81.31%, with a precision of 46.23%, and an accuracy of 92.74%, measured on 16 test subjects. The tradeoff between precision and recall can be controlled by adjusting the MOS threshold that needs to be reached for an MOS to be detected.

Index Terms—**Electrodermal activity (EDA), individual-specific physiological responses, rule-based algorithm, skin temperature (ST), stress detection.**
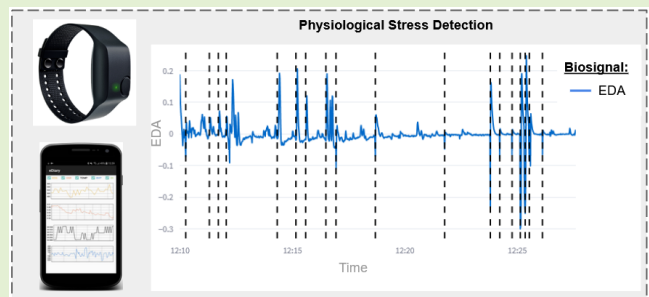
NOMENCLATURE

AC	Alternating current.
ANS	Autonomic nervous system.
DC	Direct current.
DL	Deep learning.
ECG	Electrocardiogram.
EDA	Electrodermal activity.
EDL	Electrodermal level.
EDR	Electrodermal response.
FN	False negative.
FP	False positive.
FPR	False-positive rate.
GNSS	Global navigation satellite system.
GSR	Galvanic skin response.
HR	Heart rate.
HRV	Heart rate variability.
Hz	Hertz.
IBI	Interbeat interval.
LDA	Linear discriminant analysis.
LUCCK	Learning using concave and convex kernels.
ML	Machine learning.
MOS	Moment of stress.
NN	Neural network.
PPG	Photoplethysmography.
SC	Skin conductance.
SCL	Skin conductance level.
SCR	Skin conductance response.
ST	Skin temperature.
SVM	Support vector machine.
TN	True negative.
TP	True positive.
TPR	True positive rate.
VR	Virtual reality.

Martin Karl Moser and Maximilian Ehrhart are with the Department of Geoinformatics, University of Salzburg, 5020 Salzburg, Austria (e-mail: martin.moser@plus.ac.at).

Bernd Resch is with the Department of Geoinformatics, University of Salzburg, 5020 Salzburg, Austria, and also with the Center for Geographic Analysis, Harvard University, Cambridge, MA 02138 USA.

I. INTRODUCTION

STRESS is a natural, but complex reaction of the body in accordance to some environmental threat, also termed

stressor, causing the so-called "fight or flight" response phenomenon [2], which is accompanied by a compound reaction of the physical, physiological, and psychological components of the human body [3]. This change in physical and physiological data, caused by the ANS, is important for human beings to react and adapt to individual situations appropriately. Noninvasive and accurate sensorial monitoring gives researchers the possibility to develop algorithms based on physiological biosignals to identify such situations of high alertness. Giannakakis et al. [4] gave an excellent overview of the individual biosignals that have been related to stress and compared several approaches to detect MOS based on them.

EDA, also known as GSR, and ST are two primary indicators for detecting stress in human beings [1], [2], [3], [5]. EDA is reflected by a person's SCL. Electrodes, which are in constant contact with the human skin, are mounted to a sensor device, where a minimal amount of current is induced, measuring the conductivity of a person's skin. EDA is considered one of the best estimators for the ANS's activation level, due to the connection between the body's sweat-motor system and the parasympathetic component of the ANS [6]. During stress, the ANS reacts through a compound activation of functions, one of them being the opening of sweat glands, located all over the human body [7], which is caused by nerves trying to regain a state of homeostatic equilibrium [8].

ST, another indicative biomarker for stress, regulates the core body temperature, and also maintains a state of homeostasis [7]. After a rise of SC and the production of sweat, ST decreases due to the evaporation of sweat, causing a cooling effect on the human body [9], [10], [11].

Next to EDA and ST, other biometric data sources have been considered to assess an individual's stress level and to enhance an algorithm's robustness, making it applicable to other use cases, especially in a real-world setting [4], [5]. Cardiovascular signals such as HR, IBI, also referred to as the R-R Interval, and HRV are among the most common. The advancement of PPG sensors has made HR, IBI/R-R Interval, and derived HRV, all signals formerly derived from electrocardiogram (ECG) measurements recorded through a chest strap with electrodes located near the heart, accessible via noninvasive, wrist-worn devices such as smartwatches. Due to the temporal peak agreement among PPG and ECG signals, cardiovascular activity-related parameters can be estimated accurately through a PPG sensor [4]. According to [2], cardiac activity, if measured properly, can provide vital information regarding an individual's stress level and further the understanding of the ANS. Based on an extensive literature review conducted by Yu et al. [5], the authors claim that HRV, HR, and EDA are among the most commonly used biosignals to detect and alleviate stress.

While other physiological data sources such as cardiovascular activities can be incorporated into methods for detecting stress, EDA and ST, measured through an unobtrusive, wrist-worn watch, the Empatica E4, are the primary focus of this study. The recorded EDA signal can be decomposed into two components, the SCL, also called EDL, and SCR, also referred to as EDR. Boucsein [8] elaborated on the difference between the two, while suggesting that GSR is an outdated term that is no longer recommended for use, due to its imprecise meaning. He further differentiated electrodermal recordings between endosomatic, that is, recordings that do not make use of an external current measuring potential differences in the skin itself, and exosomatic, that is, recordings that apply either a direct or alternating current, termed dc and ac, respectively, to the skin. Depending on the recording method, the stimulus depicting the physiological reaction of the ANS, reflected by the EDR, can look different [8]. Next to the recording method, the measurement site of the EDA-collecting device matters. Picard et al.'s multiple arousal theory [12] considered how arousals can lead to a multitude of reactions in the human brain at multiple locations, in turn activating a multitude of physiological reactions that differ in terms of prominence, based on the position of the human body they are recorded at. According to their theory, there is a dependence between the signal recording and the location where the noninvasive device is positioned, as shown by the difference in measurements depending on whether a test subject wears the EDA collecting device on the right or left hand. The resulting composite form of the psychophysiological response is then used to relate individual biosignals such as EDA and ST with each other [12].

Understanding how ST is related to stressful events and how it is used in the proposed algorithm requires a basic understanding of sweat, what it constitutes, and how it is produced. In general, there is a differentiation between eccrine sweat glands, located on palms, feet, and forehead, and apocrine sweat glands, which are located in the armpits and groin. The latter gets activated when a body enters the "fight-or-flight" response [7]. During stress, sweat glands located on different parts of the skin are activated, resulting in a change of SC that is measured by EDA [2]. Depending on the humidity level of the environment, sweat evaporates at different paces, causing a cool-down effect to maintain homeostasis [7]. This decrease in ST due to evaporation, after secretion of sweat, should be visible in ST recordings, which is why Kyriakou et al. [1] and Zeile et al. [13] proposed a relationship between an increase in EDA and a decrease in ST during stress. According to [14], ST is proportional to stressor intensity, and a drop in ST could be a sign of experiencing stress. While a drop in ST due to evaporation effects following an increase of EDA makes sense from a logical perspective, several other studies, presented in Section II, suggest the opposite, arguing that when exposed to a stressor, the mean ST gradually increases during the transition from a normal to a stressful state [9]. Compared to the EDA signal, however, changes in ST as a reaction to stress are quite slow [8].

Automatic and accurate stress detection opens up several application areas that can have a positive impact on the health status of our society. Depending on the context applied, use cases of stress detection can range from continuous health tracking to detect early-stage illnesses, controlling long-term health issues elicited through chronic stress, to enhancing traffic safety through threat detection and localization based on georeferenced biosignals. This research aims to drive the development of stress detection algorithms leveraging physiological biosignals recorded through nonobtrusive wearable

sensor technology, that is, the Empatica E4 [15], and addresses the following research questions.

1) How can EDA and ST data, measured by wearable sensor devices, be used to model individual physiological stress responses?
2) Which signal properties are the most indicative and reliable indicators for stress, and how can they be formalized in a rule-based stress detection algorithm?
3) To what degree does the incorporation of individual-based normalization result in more exact and robust stress detection?

Taking into consideration these research objectives, the main contributions of this article are as follows.

1) Stress-indicative rules considering subject-specific biosignal differences lead to more accurate results in a laboratory setting, where acute stress is induced through an audio stimulus.
2) While subject-centric EDA properties such as rise time, amplitude, and response slope are properties that can be used to detect acute stress, no such pattern is found in the ultrashort-term ST variations.
3) The task-specific choice of evaluation metric needs to consider the distribution of class labels available in the data. While accuracy is a commonly used metric in time series classification, it is not the ideal choice when dealing with a highly imbalanced dataset, such as the physiological stress dataset used in this study.

II. RELATED WORK

Existing research, specifically in the area of real-time stress detection, shows there is no agreement on which biosignals one should use to accurately detect stress. Cardiovascular data is one of the most common sources of information leveraged in current algorithms which are integrated in smart wearables. This is due to the affordability, the wearing comfort, and the increasing accuracy of PPG sensors, merely mimicking a signal that is recorded through ECG sensors, which is currently considered the gold standard for cardiovascular biosignal monitoring [16].

As mentioned in Section I, EDA, more specifically the phasic part of the signal termed EDR, representing the arousal of the human body when exposed to a stressor, is one of the leading biosignals used in relation to stress detection. This section describes the current state-of-the-art stress detection algorithms leveraging EDA. However, as stated by Babaei et al. [17], the results presented should be treated with caution, as there is a multitude of methodological issues that can arise when conducting research using EDA, ranging from proper positioning the device on the test subject to ensure valid recordings, preprocessing, and analyzing the recorded data, to reporting results of the analysis.

Kurniawan et al. [18] proposed a classification algorithm to detect stress based on EDA features, that is, central tendency and dispersion measures, rise time, peak height, the total number, and the cumulative amplitude. With a result of 70% accuracy in differentiating between stressful and nonstressful situations, the authors conclude that EDA alone is not sufficient to classify stress accurately. They also claim that

signal recordings vary not only between test candidates, but also within the same test candidate when identical experiments are conducted on different days. Durán Acevedo et al. [19] proposed a methodology for academic stress detection during virtual examinations by leveraging an electronic nose and GSR to determine the stress level of a student. According to the authors, the LDA algorithm used achieved an accuracy of 100% on a standardized raw EDA signal, meaning the proposed methodology can perfectly differentiate between stress and relaxation phases. However, they also mention that further research needs to be done to understand why individuals show different physiological response characteristics when exposed to the same stress situation [19]. Navea et al. [20] propose an Android application to measure stress during mobile communications, that is, exchanging text messages. While the application can measure a person's stress level, the writers also mention that Short Message Service (SMS) composition might not be the best stress trigger to differentiate between low- and high-stress levels based on an EDA signal. Bastiaansen et al. [21] researched the relation between psychophysiological biosignals and experiences, more specifically the relation of SCL with respect to tourism, hospitality, and leisure experiences. They compared the change in SCR, measured through the Empatica E4 wristband, during a real-world roller coaster ride and one simulated based on a VR add-on. A comparison of EDA recordings between VR and non-VR roller coaster experiences shows that the VR-induced experience leads to a weaker, but similar psychophysiological response. This proves the potential of VR devices in creating psychological responses and emotions in human beings, which could ultimately be tied to an improved tourism and leisure experience. Sánchez-Reolid et al. [6] provided an overview of studies that use ML techniques in combination with EDA features to detect stress. They identified that supervised algorithms are used more extensively than unsupervised ones, with SVMs and NNs being the most popular choices as they seem to lead to the best results. Another interesting paper with regard to EDA is proposed by Zhang et al. [22], who compared supervised and unsupervised ML approaches and used them to detect motion artifacts within the electrodermal signal recordings.

Considering the relation between alterations in ST and stress, Yamakoshi et al. [23] found that there is a change in peripheral ST, mainly affected by changes in the blood volume, which can be used to assess a person's stress level while driving a car. However, they also pointed out that peripheral ST is influenced by environmental factors such as ambient temperature and humidity, as well as different clothing that is worn when the experiment is conducted [23].

With regard to ST-derived features, Zhai and Barreto [24] leveraged the slope of a filtered ST signal at specific time segments instead of the mean ST values, due to the unknown temporal delay and the decreased sensitivity of the signal during the exposure to a stressor. This goes hand in hand with Boucsein's claim [8] that an EDA response is much more immediate and prominent when comparing it to the ST response in relation to stress. Zhai and Barreto [24] concluded that the slope of the filtered ST signal shows

a negative trend at the occurrence of an induced stress stimulus.

While previous studies present promising results leveraging individual signals to detect stress, Ahmadi et al. [25] showed that further research on the correlations among physiological biosignals under stress is needed to understand the composite response of the human body and to design more accurate and robust algorithms to identify stress.

A number of studies have used EDA-derived features in combination with ST to come up with algorithms that can be applied to many different use cases, such as stress identification, derivation and explanation of emotions, as well as the assessment of sleep quality [26], [27], [28], [29], [30], [31], [32].

Ali et al. [26] introduced a subject-independent emotion recognition system based on EDA, ST, and ECG data, which classifies four different emotions using a cellular NN, a new architecture proposed by the authors. The model is evaluated on the MAHNOB dataset [33] and achieves an average accuracy of 71.05% on a test set consisting of six subjects.

Zhao et al. [27] proposed an algorithm to detect four different emotions based on recordings of BVP, EDA, and ST from the Empatica E4 wristband. Stressors are induced via video stimuli, and 28 different features are fed into an SVM model that achieves an average accuracy of 75.56% in differentiating between the four emotional states. Kleckner et al. [28] developed an automated rule-based procedure to assess the quality of EDA recordings based on four defined rules relating to EDA and ST signal recordings, while Sabeti et al. [29] proposed a new supervised ML methodology called LUCCK to determine the importance and weighting of the individual biosignals during classification.

Jebelli et al. [30] proposed an ML-based framework for the identification of occupational stress in workers based on physiological signals collected through a noninvasive wearable device. The proposed methodology can achieve a prediction accuracy of 84.48% in differentiating between stress and nonstress situations, and an accuracy of 73.28% with regard to differentiating between low-, medium-, and high-stress levels. Chowdhury et al. [31] used several ML approaches based on multimodal sensor data to predict study participants' perceived exertion, to assess the intensity of specific physical activities. More specifically, HR, ST, and EDA are used as input to the ML models to compare the importance of individual features. According to the authors' results, HR-based features in the time and frequency domain seem to be the most significant in the input space of the individual ML algorithms tested. While HR-based features are the most significant when a single biosignal is used, the extension of the algorithm's input space to include EDA and ST led to an SVM model that achieved an accuracy of 85.2%. Aristizabal et al. [32] evaluated the feasibility of detecting stress using a deep NN taking as input EDA, ST, and HR, all recorded through the Empatica E4 device. The network obtained an accuracy of 88% on the binary classification task of differentiating between stress and nonstress moments, considering the 18 test subjects who volunteered to participate in the study.
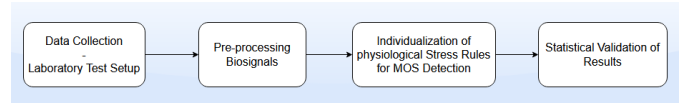

Fig. 1. Methodological workflow.

A. Research Gaps

Previous research shows that several different methodologies have been trialed to come up with a go-to method to assess an individual's level of stress. However, as results, the large number of different approaches, and the combinations of biometric signals show, there is no specific set of biosignals and methodology that is used for the detection of stress. A large factor contributing to this is the different time frames, in which stress is tried to be detected. This is accompanied by the nonstandardized data collection methodologies that are applied. Not every physiological signal and the derived biosignal feature is suitable for detecting acute stress in a short-term time domain of seconds, so the focus lies on the identification of meaningful stress-related physiological features, which work across individuals to measure their stress level. An advantage of the rule-based methodology proposed in [1] is the comprehensibility of the algorithm's decision for stress classification, as opposed to black box ML or deep-learning approaches proposed by Ali et al. [26], Sabeti et al. [29], Jebelli et al. [30], Chowdhury et al. [31], and Aristizabal et al. [32], where results of the algorithms are harder to interpret due to the latent nature of the physiological features being involved. This research gap drives the content of this study where the focus is to further enhance the rule-based MOS detection algorithm proposed in [1] through individualizing the rules proposed in the original paper, adapting the hard-coded threshold values toward individual-oriented ones that take into consideration the characteristics of each specific subject to deal with potential biases of sensor measurements.

III. METHODOLOGY

This section introduces the process of collecting data to detect stress in a laboratory testing environment and the preprocessing steps applied to the collected biosignals, followed by the introduction to how stress rules are defined and statistically validated. The focus of the proposed methodology lies on the individualization of stress markers that take into consideration subject-specific physiological reactions that differ among individuals. Fig. 1 illustrates the methodological workflow employed.

A. Data Collection—Laboratory Test Studies

The data considered in this study consists of 16 individuals, eight male and eight female subjects between the age of 18 and 55, who volunteered to participate in a laboratory experiment. Participants were asked to join the study through an email campaign that was sent out by email to employees and students at the University of Salzburg. The individuals who agreed to participate were split into groups of 4–5 people per testing session, where each laboratory experiment lasted a period of at least 12 min. To account for equal test conditions, the same

location and the same time, that is, 10 A.M. to 2 P.M., was used to conduct the experiment. Participants were also asked to not take any stimulants within a 12-h period before the study. Furthermore, to avoid any distractions caused by other participants in the same experiment session, all test candidates were placed on a chair with their backs facing each other. They were instructed not to interact or perform any physical action.

During the experiment, ten acoustic stressors, created through an airhorn sound and amplified by a speaker, were induced to generate reference points. The aim was to induce stress at known points in time, where the acoustic stimulus causes a reaction in the human body, which allows for a clear connection between the stimulus and the physiological footprint that is visible within the biometric signals recorded by the sensors. These reference points also serve as the ground-truth labels, based on which the algorithm's rules are fit, and its performance is evaluated in a later stage. The speaker was placed in the middle of a semicircle so that the distance between each participant and the amplifier was equal. The time span between the airhorn sounds was chosen at random in an interval between 60 and 90 s.

For the laboratory experiments, we used the Empatica E4 wristband [15], a medically certified, nonobtrusive device, measuring various biosignals, including PPG-based cardiovascular data such as HR, HRV, R-R Intervals (the time between two successive R peaks), together with EDA and ST, which are the main biosignals used in the proposed rule-based algorithm. Both signals are recorded with a sampling frequency of 4 Hz. While EDA measurements lie in a range between 0.01 and 100 μS, with a resolution of 900 pS, ST recordings are measured with a resolution of 0.02 °C and with an accuracy of $\pm$0.2 °C, when in the 36 °C–39 °C range. According to the Empatica E4 datasheet [15], the wristband measures STs between -40 °C and 115 °C and ambient temperatures between -40 °C and 85 °C.

To store the data in an anonymized bio-geodatabase, the E4 device connects itself to the University of Salzburg in-house-developed Android-based eDiary app [34] and transmits the physiological data to a local SQLite database. Next to the sensor measurements, relevant data recorded by the smartphone itself, that is, location, altitude, GNSS accuracy, and speed are stored in the bio-geodatabase. Aside from the ability to connect with the Empatica E4 wristband, the eDiary app also has the capability to connect another sensor, the Zephyr BioHarness chest strap. More information about the eDiary app can be found in [34].

B. Preprocessing Biometric Data

Several factors impact the accuracy of the measurements of wrist-worn wearable devices. Next to environmental influences such as temperature and humidity, physical activity and movements of the sensor during measurement are common factors contributing to noise in recordings.

As mentioned previously in Section I, the EDA biosignal consists of a tonic component referred to as SCL, and a phasic component, also known as SCR or EDR, representing fast-changing reactions in the signal. Through the application of appropriate frequency filters, noise can be removed and the
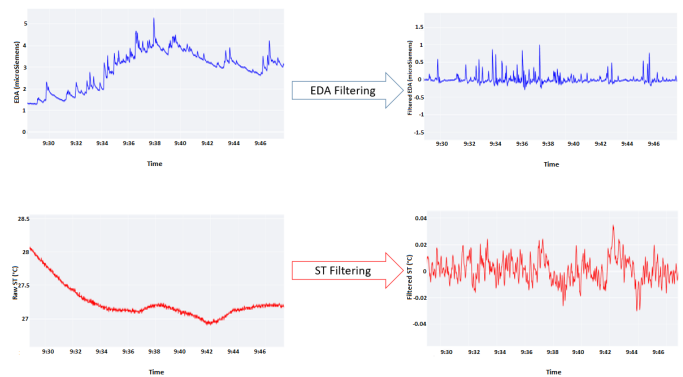

Fig. 2. Raw versus filtered EDA and ST signal.

individual components of the signal can be separated, enabling the extraction of the phasic SCR peaks, which are of interest when it comes to detecting stress through electrodermal recordings [35].

Before filtering the signal for specific frequencies, a general filter to exclude recordings that fall outside the range of valid values mentioned in the sensor's data specifications is applied. This first step of prepreprocessing removes noise and other measurement artifacts, created through environmental influences and other factors, which can be attributed to improper sensor positioning or movement, leading to a loss of contact between the sensor and the human skin. After applying the value filter, missing data points are interpolated linearly, so all measurements fall into the valid range of values as documented in [15]. This is a significant difference compared to the preprocessing steps performed in [1], since the interpolation of missing data points is done on a 4-Hz frequency level, before individual signals are downsampled to 1 Hz, and therefore generating more accurate estimates for the missing values.

Following the interpolation, frequency filtering is applied, where cutoff frequencies were determined based on a frequency spectrum analysis, performed through a fast Fourier transformation (FFT), which converts a signal's time-domain representation to a frequency-domain representation. The identified cutoff frequencies were cross-checked with existing literature to determine their validity. Fig. 2 shows the difference between the raw signal recordings on the left and the preprocessed signals on the right.

Instead of using a low-pass filter cutoff frequency of 3 Hz, as suggested by Chênes et al. [36], or 5 Hz, as used in [1], a first-order, low-pass Butterworth filter with a cutoff frequency of 1 Hz was applied to filter out noise in the EDA recordings. Visual examination of the FFT outputs revealed the noise in the frequency spectrum, and cutoff frequencies were determined accordingly. While this cutoff frequency works well for our stationary laboratory setting, where body movements are unlikely to have an impact on the device's recordings, higher cutoff frequencies for the low-pass filter, that is, 3–5 Hz, should be considered in a nonstationary setting, to filter out noise effectively. After applying the low-pass filter, a first-order high-pass Butterworth filter with a cutoff frequency of 0.05 Hz is applied, as suggested in [37] and [38].

A similar logic applies to the ST recordings, where a lower cutoff frequency in the low-pass filter can be used in the

Fig. 4.　EDA feature distribution comparison [amplitude (left) and slope (right)].
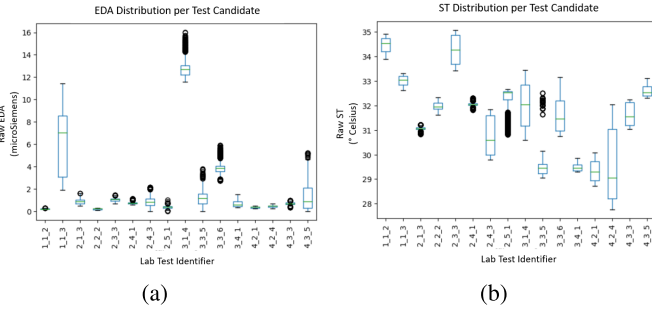
Fig. 3.　(a) Distribution of EDA measurements. (b) Distribution of ST measurements.

stationary lab test setting. Instead of applying a second-order Butterworth filter with a low-pass cutoff frequency of 1 Hz, as in [1], a low-pass frequency filter of 0.1 Hz is used for our lab test signals. To extract the relevant features of the ST signal, a second-order high-pass Butterworth filter with a cutoff frequency of 0.01 Hz is utilized [39]. In combination, these two filters form a bandpass filter that contains the small fluctuations connected to stress-related events [39].

After applying these changes to the preprocessing steps, EDA and ST signals are stored in a clean version of the bio-geodatabase, filtered, interpolated, and then downsampled from 4- to 1-Hz frequency. These steps are applied to smooth the individual signals and to add georeferences by joining information with the smartphone's Global Positioning System (GPS) data based on common timestamps.

C. Individualization of Stress Detection

Physiological biosignal measurements as reactions of the human body, recorded through wearable sensor technology, have several biases inherent in the device's recordings. These biases can be attributed to a number of different factors. According to [8], [18], and [40], demographics such as age, gender, and ethnicity significantly impact the quality of measurements recorded through wearable sensor technology that is placed on the human skin [8], [18], [40]. Aside from these demographic influences, an individual's diet [2] and the general acceptance level of biometric systems can influence the stress level an individual experiences when wearing a wearable device, as emotions might be suppressed or manipulated to hinder or trigger voluntary-caused physiological reactions of the body. Kitsiou et al. [41] conducted a study among Greek students, where they showed that demographic and social factors have an impact on the acceptance level of biometric systems. These factors constrain the algorithm design process and show the need for subject-oriented formulations of biosignal rules. Fig. 3(a) and (b) illustrates the aforementioned differences by displaying a comparison of raw EDA and ST recordings, respectively, taken from the 16 test candidates.

As can be seen in Fig. 3(b), ST recordings among the test subjects differ significantly and suggest that hard-coded threshold values, as proposed in [1], might not be applicable to every individual. To address these biases, an individualization of stress rules based on physiological signals is proposed.
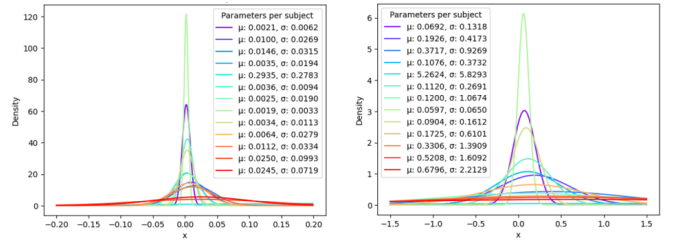
These individual-oriented stress rules, where threshold values are set through subject-dependent baseline calculations, should consider these differences and provide more reliable indicators for determining an individual's stress level.

An expressive example of this is the fourth rule of the MOS algorithm proposed by Kyriakou et al. [1], where hard-coded threshold values of 8° and 10° with regard to the EDA slope between a local minimum and maximum in the signal are used as an indicator for stress. While these values work in some cases, they fail to take into consideration differences among individuals and assume that the magnitude of the signal response is the same for every test subject.

Fig. 4 shows a subset of test candidates, comparing their individual distributions of EDA features, that is, the amplitude increases the value and the slope, both calculated between a local EDA minimum and a local EDA maximum. This comparison of slope and amplitude distributions shows a significant difference among individuals and justifies the individualization of rules, as there exists a general pattern in terms of value distributions, which can be distinguished based on magnitude. The uniform behavior of distributions and the differences in magnitude is leveraged to set the individual threshold values in the algorithm. Resulting means and standard deviations for each EDA-based feature are used as baseline values, from which the proposed EDA stress detection rules described in Section III-D are derived. A thorough discussion on baseline calculations for physiological parameters, however, is outside the scope of this article.

As existing literature shows [9], [14], [42], there is no general agreement on whether ST increases or decreases after experiencing an MOS. More specifically, the time delay of ST cooling and subsequent warming after experiencing stress remains an open question in peripheral ST research [42]. Boucsein [8] also mentioned that the reaction in the ST, visible in the recorded biosignal, is delayed; however, it is not stated by how much, and it also is likely to be subject-specific. The ST rule proposed in [1] considers a 3–6-s decrease, exactly 3 s after a GSR/EDA onset as a sign for an MOS. After evaluating this rule on our lab test dataset, results show that this is the least significant rule formulated in this article, as it is barely ever met due to its strictly set threshold values. Throughout the algorithm design phase, insights from Fig. 3(b) are taken into consideration, and an exhaustive grid search is applied to determine a set of rules representing ST characteristics that occur as a reaction to the audio stress stimulus. The final paragraph in Section III-D further elaborates on the results of this grid search.
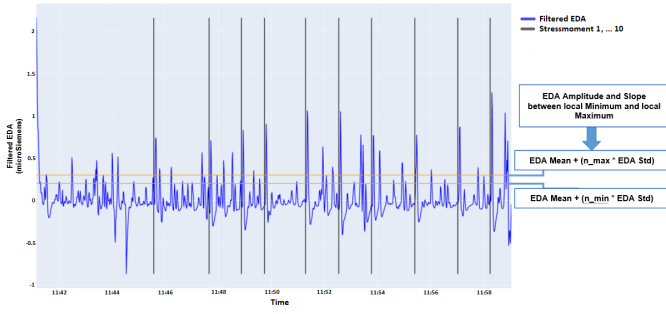
Fig. 5. Principle of individualized EDA stress detection rules.

D. Stress Detection Algorithm Based on EDA and ST

1) Electrodermal Activity: After applying the preprocessing steps outlined in Section III-B, the resulting signal reflects the phasic component of a person's EDA, the SCR, reflective of the response of the ANS to an induced stress stimulus. Taking the insights gained from Section III-C into consideration, individual-oriented rules are formulated, where thresholds are dependent on the distribution of an individual's baseline values, that is, the mean and standard deviation of the EDA amplitude and the EDA slope. Through the incorporation of the different response magnitudes, individualized stress rules are formulated.

The principle of the individual-oriented EDA rules is illustrated in Fig. 5, where the mean of an EDA feature $\pm n$ times the standard deviation of the feature is used as the threshold values for individual rules.

To fine-tune the rules that serve as stress indicators, variations in EDA and ST signals are considered and parameters are determined based on multiple grid search runs, testing several thousand parameter combinations to identify the best blending. The ternary scoring system used in [1] is adopted and adjusted to a quaternary design, where each rule can contribute up to 1.0 points (classes: 0.0, 0.25, 0.5, 1.0).

2) Rule 1—Candidate Generation Based on EDA Rise Time: The first rule [see (1)] computes the rise time between a local EDA onset (starting from a local minimum) and a local EDA peak and serves as a candidate generation step. It identifies stress point candidates $\text{MOS}_{\text{Candidates}}$, where the rise time between local minima and maxima of EDA is used as a precondition to check for potential stress situations. According to [8], stress situations lead to a rise in EDA, which varies between 1 and 6 s, where reactions after 6 s can be considered spontaneous and can potentially be attributed to some other event [35]. Since the EDA response can be seen as a linear time-invariant process [38], we can assume a linear mapping between the stressor and the SCR amplitude. Thus, the rise time of the EDA can be directly connected to the stressor. Taking this into consideration, a minimum duration of 1–6 s, following an electrodermal onset, needs to be met for a specific EDA rise onset to be checked further. The upper limit for the time range is set to 6 s to make sure no potential MOS is missed by the algorithm

$$\text{MOS}_{\text{Candidate}} = \begin{cases} 1, & \text{if } 1 \leq t_{\text{EDApeak}} - t_{\text{EDAonset}} \leq 6 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where t_{EDAonset} and t_{EDApeak} refer to the indices of the local EDA onsets and local EDA peaks, respectively.

3) Rule 2—EDA Response Amplitude: After having identified potential stress candidates, amplitudes between EDA onsets and EDA peaks are calculated and compared to a threshold amplitude value, which is computed per individual over all EDA onsets and peaks within a specified time interval. As there is an initial transient phase in the signal through the specific configuration of our lab test, the first 5 min should be excluded from these baseline calculations.

The threshold values that need to be reached are given by the mean amplitude value of all EDA increases and n number of standard deviations, $\text{Amp}_{\text{Avg}} + n*\text{Amp}_{\text{SD}}$, which are added to the amplitude average. Since our data showed that the EDA increases of the subjects tend to follow a normal distribution, mean and standard deviation are used to model the extreme events. Leveraging the quaternary scoring system, rule 2, can add a maximum of 1.0 points to the stress score. Depending on the degree of threshold fulfillment and if the EDA onset is part of the stress candidates, the rule adds 0.0, 0.25, 0.5, or 1.0 points to the MOS score, reflecting how indicative an amplitude increase is for stress. Interval ranges, the number of standard deviations above the mean, and the individual contributions of each interval were determined through a comprehensive grid search

$$\text{Score}_{\text{AmpEDA}}$$
$$= \begin{cases} 0.00, & \text{if } \text{Amp}_{\text{EDA}} < \text{Amp}_{\text{Avg}} \\ 0.25, & \text{if } \text{Amp}_{\text{EDA}} \geq \text{Amp}_{\text{Avg}} \\ 0.50, & \text{if } \text{Amp}_{\text{EDA}} \geq (\text{Amp}_{\text{Avg}} + 0.1*\text{Amp}_{\text{SD}}) \\ 1.00, & \text{if } \text{Amp}_{\text{EDA}} \geq (\text{Amp}_{\text{Avg}} + 0.4*\text{Amp}_{\text{SD}}). \end{cases}$$
$$(2)$$

4) Rule 3—EDA Response Slope: The same methodology is applied to determine the individualized threshold values for the EDA response slope. Next to the rise time and the amplitude of an EDA response, the slope between a local EDA onset and a local EDA peak is another indicative EDA feature relating the ANS activity to stress [1], [2], [8]. The mean and the standard deviation are calculated over all EDA onsets and peaks, excluding the first X min that constitutes the transient phase of the signal.

All candidate EDA onsets are checked and compared to the slope threshold values. Depending on the interval an EDA slope falls into, a score of 0.0, 0.25, 0.5, or 1.0 points is added to the MOS score

$$\text{Score}_{\text{SlopeEDA}}$$
$$= \begin{cases} 0.00, & \text{if } \text{Slope}_{\text{EDA}} < \text{Slope}_{\text{Avg}} \\ 0.25, & \text{if } \text{Slope}_{\text{EDA}} \geq \text{Slope}_{\text{Avg}} \\ 0.50, & \text{if } \text{Slope}_{\text{EDA}} \geq (\text{Slope}_{\text{Avg}} + 0.1*\text{Slope}_{\text{SD}}) \\ 1.00, & \text{if } \text{Slope}_{\text{EDA}} \geq (\text{Slope}_{\text{Avg}} + 0.4*\text{Slope}_{\text{SD}}). \end{cases}$$
$$(3)$$

5) Rule 4—MOS Score: After checking rules 2 and 3 at each index of an EDA onset candidate ($\text{MOS}_{\text{Candidate}}$) for each participant, the sum of all rules' point scores is computed

to generate an overall MOS_{Score} for every second in the measurement signal. This score can be interpreted as an overall indicator of stress.

Since each rule can contribute up to 1.0 point, a maximum score of 2.0 points can be reached. The MOS_{Score} of an $MOS_{Candidate}$ EDA onset is calculated as the sum over the point contributions of rules 2 and 3

$$MOS_{Score}(MOS_{Candidate}) = \sum Score_{AmpEDA} + Score_{SlopeEDA}. \tag{4}$$

6) Rule 5—MOS Threshold: The previously calculated MOS_{Score} is compared against a threshold value to determine whether a candidate EDA onset is considered an MOS. If the MOS_{Score} of a specific candidate is greater than or equal to the threshold value, the point is considered an MOS. Depending on the use case applied, the threshold value can be adjusted to fit the problem context

$$MOS = \mathbb{1}_{MOS_{Score}}(MOS_{Candidate}) \tag{5}$$

$$= \begin{cases} 1, & \text{if } MOS_{Score} \geq MOS_{Threshold} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\mathbb{1}$ refers to an indicator function determining the return value for the binary stress classification. If the specified $MOS_{Threshold}$ is reached, the function returns a value of 1, indicating that a particular $MOS_{Candidate}$ coincides with an MOS.

7) Skin Temperature: With regard to ST measurements, several rules considering the increase and decrease time as well as the first derivative of the signal before and after an EDA onset were trialed and grid searched. Even though the parameter space for the grid search was defined based on the findings from other researchers, no clear pattern, indicative of the physiological response representing stress, is visible, leading to the conclusion that ST measurements in the time domain might not be an informative feature to detect stress. This aligns with the diverging opinions in the literature, where there is no agreement on whether ST rises or falls after experiencing stress. However, further research should be conducted to validate this finding, as the test sample of 16 test subjects is limited, and because the time frame in which stress should be detected plays an important role. While the ST signal does not provide a reoccurring indicative response in our acute stress detection setting, it might be a relevant physiological biomarker for longer time frames.

Overall, the best parameter combination for the detection of stress in our stationary test setting uses the rise time, amplitude, and slope of a particular subject's SCR. Subject-specific cutoff values concerning the EDA response amplitude and response slope are determined based on the mean and the standard deviation between local EDA onsets and peaks. A comprehensive parameter grid search, constrained by a parameter search space considering the domain knowledge of experts and related research, revealed that local EDA amplitudes and slopes of 0.1 and 0.4 standard deviations above the mean form indicative thresholds to determine a person's level of stress, assuming that the EDA response complex lasts between 1 and 6 s. However, since cutoff values are quite sensitive parameters and highly dependent on the use case, we recommend grid searching the ideal parameter values and tweaking the $MOS_{Threshold}$ that needs to be met, to fit the study context.

IV. RESULTS

The previously described methodology is evaluated on a test dataset consisting of 16 individuals—eight male and eight female study participants—between the age of 18 and 55. An MOS is considered to be identified correctly if the detected MOS lies within a 4-s time interval after a reference stress moment, induced by the audio stimulus. Since we are dealing with a highly unbalanced dataset, which contains only a small number of induced stress stimuli, compared to the high number of nonstress situations, the metric to evaluate the algorithm's performance has to be chosen with care, to avoid misleading interpretations.

While most papers report accuracy as a measure of evaluation, it is important to keep in mind that accuracy fails to reflect the aforementioned class imbalance and might lead to incorrect conclusions with respect to the performance of the algorithm. An algorithm could achieve a reasonably high accuracy score by simply identifying nonstress moments correctly (TNs), without considering the number of actual stress moments that were classified as such (TPs), because the number of TNs is higher by magnitudes compared to the TPs. Furthermore, the choice of metric an algorithm is evaluated on heavily depends on the problem at hand and the context it is applied to.

Thus, in this study, we refrain from using the misleading accuracy metric and use recall and precision, which take into account the number of TPs, FPs, and FNs.

1) *TP (Correctly Detected MOS):* Detected MOS corresponds to a reference MOS (stress induced in the lab experiment).
2) *FP (Falsely Detected MOS):* Detected MOS does not correspond to a reference MOS.
3) *TN (Correctly Detected Non-MOS):* No MOS detected where there is no reference MOS.
4) *FN [Falsely Detected Non-MOS (missed MOS)]:* No MOS detected where there is a reference MOS.

Fig. 6 displays an example output of the algorithm for one test candidate. It shows the filtered EDA signal in blue, including the induced stress moments (reference MOS), represented as vertical green lines, and the detected stress moments based on the rules formulated in Section III-D, represented by the dashed gray lines. For this specific individual, the algorithm can perfectly classify all the MOS, resulting in an ideal recall, with considerably high precision.

Depending on the study context and the desired balance between recall and precision, the $MOS_{Threshold}$, the score that needs to be reached for an $MOS_{Candidate}$ to be classified as an MOS, can be adjusted. Hence, the number of TPs, FPs, and FNs is controlled via the $MOS_{Threshold}$, which is suggested to be set to 1.25, to balance the tradeoff between recall and precision, resulting in 58.95% F1-Score. A summary of the results evaluated on the 16 test samples is displayed in Fig. 7(a)–(d), which highlight the tradeoffs that come with individual threshold values.
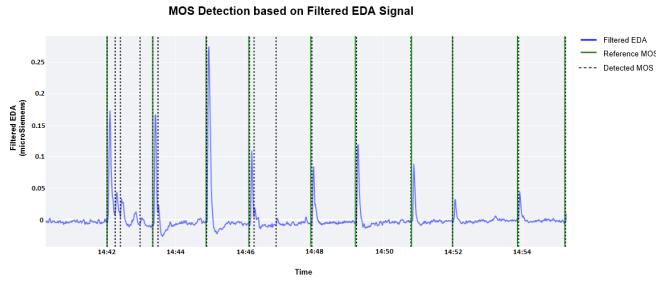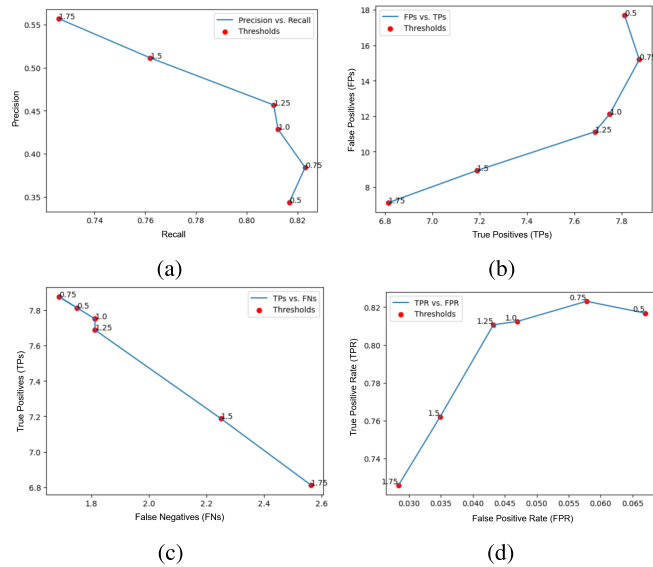
Fig. 6. MOS detection example.

Fig. 7. (a) Avg. precision versus recall. (b) Avg. FPs versus TPs. (c) Avg. TPs versus FNs. (d) Avg. TPR versus FPR).

Setting the $MOS_{Threshold}$ to a value of 1.25, the algorithm achieves an average recall of 81.31%, an average precision of 46.23%, an average F1-Score of 58.95%, and an average accuracy of 92.74%, when evaluated on the 16 test candidates. Changing the $MOS_{Threshold}$ parameter to a lower value will increase the algorithm's recall, however, at the cost of reducing precision, producing a higher number of FPs.

To determine the TPR and the FPR, displayed in Fig. 7(d), the sample size is calculated, which amounts to the total number of EDA onsets, as these points have the potential of receiving an MOS score. The TPR, also known as sensitivity, indicates the probability that an induced stress moment is detected correctly, while the FPR represents the probability that a nonstress moment will be classified as an MOS. The low FPR is also indicative of the highly unbalanced data. Since there is significant variability among individuals in terms of the number of EDA onsets, there is a considerable amount of TNs in the sample, leading to a misleading accuracy value and a low FPR. The high TPR demonstrates the capabilities of the rule-based methodology and can be interpreted as the model's ability to detect the induced MOS, formulated as a probability.

V. DISCUSSION
A. Discussion of the Methodology

The proposed methodology aims to address the biases of wearable sensor measurements, mentioned in Section III-C,

and take into consideration the distinctiveness of stress responses among individuals. However, several factors have an influence on the proposed algorithm, which are worth discussing.

The preprocessing procedure applied to the biosignals, which should take into consideration the study context, plays a critical role in the study outcomes. For a stationary laboratory setting, lower cutoff frequencies for the low-pass frequency filter can be used, as there is a low chance of having noise in the recordings which can be attributed to movement. In a non-stationary setting, however, where individuals are continuously in motion, these frequencies need to be tweaked to account for the motion artifacts in the signal. Using a higher frequency value for the low-pass filter is suggested in such contexts. In general, we suggest adjusting the cutoff frequencies based on the use case the stress detection algorithm is applied to, ideally determined and fine-tuned by a frequency spectrum analysis.

Once proper frequency filters are applied, the next step is to identify suitable stress indicator thresholds for the rule-based stress detection algorithm. As threshold values for EDA-based stress rules are set based on individual-oriented baseline parameters, prior physiological recordings of the individual, at least for a number of timesteps, are assumed. To make the stress detection algorithm applicable to a live setting, the average baseline values of previous study participants, taking into consideration the study context, can be used as initial threshold values, before getting adjusted by the individual-based calculations. A larger and more diverse sample from different study contexts would increase the algorithm's robustness through more accurate initial threshold values and at the same time improve the validity of the conclusions that can be drawn from the results.

While the algorithm uses only a single physiological parameter (EDA) to identify stress, it fails to incorporate other vital health information and causes a potential single point of failure, in case the signal recordings are poor. Additionally, the methodology assumes that the Empatica E4 wristband is used for signal measurements recorded in a strictly controlled, stationary laboratory setting. The integration of ST features was tried, however, the irregularities of the ST response in accordance with a stressor made it impossible to formulate stress-indicative rules that could be incorporated into the algorithm. As stated in [23], this can be attributed to the multitude of influences on the peripheral ST recordings, that is, environmental factors such as temperature and humidity, as well as clothing that is worn during the study.

The integration of other physiological biosignals, that is, cardiovascular features, as proposed in other studies [16], [21], [25], [27], [31], [32], will be the next step in our research, aiming to increase the robustness of our algorithm and making it applicable to other use cases, where EDA measurements might be affected by motion artifacts.

B. Discussion of the Results

The adapted, individual-oriented stress detection algorithm proposed in this work results in an 81.31% recall, 46.23%

precision, and 92.74% accuracy in the binary classification setting for stress.

Depending on the desired recall and precision value, the $MOS_{Threshold}$ parameter can be adjusted. Lower MOS threshold values are likely to result in a higher recall value, however, at the cost of lower precision, meaning the algorithm falsely classifies non-MOS points as MOS. Higher MOS thresholds lead to more robust results, where there is relatively high certainty that a detected MOS reflects some stress situation that was experienced.

While a threshold score of 1.75 achieves relatively high precision, as displayed in 7(a), it also results in a lower recall, a lower number of correctly classified MOS, indicated by the number of TPs in Fig. 7(b), and a higher number of FNs, as can be seen in Fig. 7(c). Since the number of FN reflects the stress moments that were missed by the algorithm, setting the threshold too strictly comes with the cost of not identifying situations where a person was in fact experiencing stress. On the contrary, when the threshold score is defined too loosely, that is, 0.5, the precision drops quite drastically, and the number of falsely classified MOS, reflected by the number of FPs, rises, as can be seen in Fig. 7(b). For threshold values set to 1.25 and below, the probability that an MOS detected by the algorithm coincides with an induced stress moment is above 80%. The FPR can be neglected in this case as it is heavily affected by the imbalanced dataset, caused by the large number of TNs, the non-MOS sequences of the time series that were correctly classified as such.

Comparing these results to state-of-the-art approaches, which employ models of higher complexity, demonstrates that simpler, more interpretable models can achieve equal or better performance when it comes to the identification of stress. While direct comparison of the results is constrained by the study context, the time frame in which an MOS should be detected, the sensor equipment used, and the reported performance metric, the proposed algorithm achieves an accuracy value that is at least 4%–5% higher than the best results reported in [30], [31], [32], and [33], which leverage more complex models such as SVMs, NNs, and other ML approaches to identify stress. Compared to the rule-based methodology proposed in [1], our stress rules take into consideration the differences in individuals' physiological responses, resulting in a more robust and precise algorithm for detecting stress.

A significant advantage of the rule-based approach is that detected stress moments can be justified and explained through the individual rules that are met, and the individual stress scores that contribute to the overall MOS_{Score}, which needs to meet the user-defined $MOS_{Threshold}$.

The previously stated results should be interpreted with caution as there are several factors that can impact an individual's stress level. In our laboratory testing environment, an audio stimulus was used to induce stress, which is different from other studies. Karthikeyan et al. [9] used text and colors, that is, the Stroop color test, whereas Bastiaansen et al. [21] used VR experiences to induce stress. Depending on the stress-inducing stimulus, the physiological response of the ANS might be different. Furthermore, the same stress stimulus was

induced up to ten times, so individuals are likely to get used to the stressor, causing a less intense physiological reaction of the human body due to familiarity [19]. Future work will consider different stimuli to see whether there is a difference in the psychophysiological stress reaction in accordance with a particular type of stressor.

Aside from the experimental setup, the context-dependent preprocessing of the data and the demographic as well as environmental factors that can impact sensor measurements, the relatively small sample size of 16 test candidates, arising from the exclusion of individuals with poor signal recordings caused by improper wearing, motion artifacts and general inoperability of the sensor device, should be taken into account. While the sample represents female and male test subjects equally, social and demographic characteristics of individuals have an impact on the response to a stress-causing stimulus [2], [41] and should be researched further. Even though the subject selection was performed based on a comparative exploratory data analysis, revealing highly homogeneous data with respect to the patterns of physiological responses to stress, the limited sample size of 16 test subjects should be considered when interpreting the results of this work. On the positive side, the labeled dataset is of high quality, collected in a strictly controlled laboratory setting, and can be employed in future studies by researchers around the world as the dataset will be made openly available.

VI. CONCLUSION

This work proposes a novel, individual-oriented algorithm for detecting acute stress in a laboratory experiment setting, leveraging EDA measurements obtained through a nonobtrusive wearable, the Empatica E4 wristband. EDA features with threshold values considering individual-based physiological signal characteristics are used to form indicative rules for measuring a person's stress level. An initial rule considering the rise time between a local EDA onset and a local EDA peak is used to determine potential stress moments, termed $MOS_{Candidates}$. For each of these $MOS_{Candidates}$, rules considering the EDA amplitude value and the EDA response slope are checked, where threshold values, which determine whether the rule is met, are calculated independently for each test subject. Each rule can add a score of up to 1.0 point (classes: 0.0, 0.25, 0.5, 1.0), following a quaternary scoring system, reflecting the certainty of the algorithm that a specific $MOS_{Candidate}$ coincides with an induced MOS. If the sum of all rules' contributions meets a user-defined threshold, an $MOS_{Candidate}$ is classified as an MOS. The algorithm achieves an average recall of 81.31%, an average precision of 46.23%, an average F1-Score of 58.95%, and an average accuracy of 92.74%, evaluated on a test consisting of 16 subjects. The results indicate that stress can be detected accurately based on a single physiological parameter, namely EDA. More specifically, it is the phasic component of the EDA signal, the EDR, which allows the identification of stress.

While a drop in ST after an increase of SC, that is, EDA, should be visible in the ST signal due to the evaporation of sweat, the analysis of ST reactions analogously to stress showed that there are no regular patterns that can be extracted

from the signal to form stress-indicative rules. This suggests that other physiological parameters should be considered to detect stress more accurately and robustly in an acute setting of just a few seconds. Future research will incorporate cardiovascular biosignals into the rule-based algorithm and trial more complex models to see whether they outperform the individualized rule-based approach proposed in this work.

The main findings of this article are that EDA is an indicative physiological parameter for assessing a subject's stress level. Due to the differences in physiological responses to stress stimuli, it is important to use individual-oriented threshold values, when using a rule-based framework for detecting stress. ST changes corresponding to a stress reaction are not indicative in an acute stress detection setting, as there is no recurring pattern in the signal. This finding is in line with existing research, where there is no agreement on whether an increase or decrease of the ST signal occurs after being exposed to a stress-causing stimulus.

Finally, threshold values for filtering frequencies of individual biosignals should be chosen considering the study context, and whether motion artifacts due to movement can lead to the inclusion of noise in measurements. Similarly, the choice of evaluation metric is dependent on the use case and should be selected with care, as accuracy might not be the most indicative performance measure when dealing with a highly imbalanced classification problem such as stress detection.

REFERENCES

[1] K. Kyriakou et al., "Detecting moments of stress from measurements of wearable physiological sensors," *Sensors*, vol. 19, no. 17, p. 3805, Sep. 2019.

[2] I. Ladakis and I. Chouvarda, "Overview of biosignal analysis methods for the assessment of stress," *Emerg. Sci. J.*, vol. 5, no. 2, pp. 233–244, Apr. 2021.

[3] B. Zhang, Y. Morère, L. Sieler, C. Langlet, B. Bolmont, and G. Bourhis, "Stress recognition from heterogeneous data," *J. Image Graph.*, vol. 4, no. 2, pp. 116–121, 2016.

[4] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 440–460, Jan. 2022.

[5] B. Yu, M. Funk, J. Hu, Q. Wang, and L. Feijs, "Biofeedback for everyday stress management: A systematic review," *Frontiers ICT*, vol. 5, p. 23, Sep. 2018.

[6] R. Sánchez-Reolid, M. T. López, and A. Fernández-Caballero, "Machine learning for stress detection from electrodermal activity: A scoping review," *MDPI Appl. Sci.*, pp. 1–29, Nov. 2020. [Online]. Available: https://www.researchgate.net/publication/346617804_Machine_Learning_for_Stress_Detection_from_Electrodermal_Activity_A_Scoping_Review

[7] L. B. Baker, "Physiology of sweat gland function: The roles of sweating and sweat composition in human health," *Temperature*, vol. 6, no. 3, pp. 211–259, Jul. 2019.

[8] W. Boucsein, *Electrodermal Activity*. Boston, MA, USA: Springer, 2012.

[9] P. Karthikeyan, M. Murugappan, and S. Yaacob, "Descriptive analysis of skin temperature variability of sympathetic nervous system activity in stress," *J. Phys. Therapy Sci.*, vol. 24, no. 12, pp. 1341–1344, 2012.

[10] C. H. Vinkers et al., "The effect of stress on core and peripheral body temperature in humans," *Stress*, vol. 16, no. 5, pp. 520–530, Sep. 2013, doi: 10.3109/10253890.2013.807243.

[11] T. Lobstein and J. Cort, "The relationship between skin temperature and skin conductance activity: Indications of genetic and fitness determinants," *Biol. Psychol.*, vol. 7, nos. 1–2, pp. 139–143, Sep. 1978.

[12] R. W. Picard, S. Fedor, and Y. Ayzenberg, "Multiple arousal theory and daily-life electrodermal activity asymmetry," *Emotion Rev.*, vol. 8, no. 1, pp. 62–75, Jan. 2016.

[13] P. Zeile, B. Resch, M. Loidl, A. Petutschnig, and L. Dörrzapf, "Urban emotions and cycling experience—Enriching traffic planning for cyclists with human sensor data," *J. Geographic Inf. Sci.*, vol. 4, no. 1, pp. 204–216, 2016.

[14] K. A. Herborn et al., "Skin temperature reveals the intensity of acute stress," *Physiol. Behav.*, vol. 152, pp. 225–230, Dec. 2015.

[15] *E4 Wristband | Real-Time Physiological Signals | Wearable PPG, EDA, Temperature, Motion sensors*, Empatica, Boston, MA, USA, 2015.

[16] M. A. Hasnul, N. A. A. Aziz, S. Alelyani, M. Mohana, and A. A. Aziz, "Electrocardiogram-based emotion recognition systems and their applications in healthcare—A review," *Sensors*, vol. 21, no. 15, p. 5015, Jul. 2021.

[17] E. Babaei, B. Tag, T. Dingler, and E. Velloso, "A critique of electrodermal activity practices at CHI," in *Proc. CHI Conf. Human Factors Comput. Syst.*, New York, NY, USA, May 2021, pp. 1–14.

[18] 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.

[19] C. M. Durán Acevedo, J. K. Carrillo Gómez, and C. A. Albarracín Rojas, "Academic stress detection on university students during COVID-19 outbreak by using an electronic nose and the Galvanic Skin Response," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102756.

[20] R. F. Navea, P. J. Buenvenida, and C. D. Cruz, "Stress detection using Galvanic Skin Response: An Android application," *J. Phys., Conf. Ser.*, vol. 1372, no. 1, Nov. 2019, Art. no. 012001.

[21] M. Bastiaansen, M. Oosterholt, O. Mitas, D. Han, and X. Lub, "An emotional roller coaster: Electrophysiological evidence of emotional engagement during a roller-coaster ride with virtual reality add-on," *J. Hospitality Tourism Res.*, vol. 46, no. 1, pp. 29–54, Jan. 2022.

[22] Y. Zhang, M. Haghdan, and K. S. Xu, "Unsupervised motion artifact detection in wrist-measured electrodermal activity data," 2017, arXiv:1707.08287.

[23] T. Yamakoshi et al., "Feasibility study on driver's stress detection from differential skin temperature measurement," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 1076–1079.

[24] J. Zhai and A. Barreto, "Stress detection in computer users based on digital signal processing of noninvasive physiological variables," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 1355–1358.

[25] N. Ahmadi et al., "Quantifying occupational stress in intensive care unit nurses: An applied naturalistic study of correlations among stress, heart rate, electrodermal activity, and skin temperature," *Human Factors*, vol. 64, no. 1, pp. 159–172, Feb. 2022.

[26] M. Ali, F. Machot, A. Mosa, M. Jdeed, E. Machot, and K. Kyamakya, "A globally generalized emotion recognition system involving different physiological signals," *Sensors*, vol. 18, no. 6, p. 1905, Jun. 2018.

[27] B. Zhao, Z. Wang, Z. Yu, and B. Guo, "EmotionSense: Emotion recognition based on wearable wristband," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct. 2018, pp. 346–355.

[28] I. R. Kleckner et al., "Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1460–1467, Jul. 2018.

[29] E. Sabeti et al., "Learning using concave and convex kernels: Applications in predicting quality of sleep and level of fatigue in fibromyalgia," *Entropy*, vol. 21, no. 5, p. 442, Apr. 2019.

[30] H. Jebelli, B. Choi, and S. Lee, "Application of wearable biosensors to construction sites. I: Assessing workers' stress," *J. Construct. Eng. Manage.*, vol. 145, no. 12, Dec. 2019, Art. no. 04019079.

[31] A. K. Chowdhury, D. Tjondronegoro, V. Chandran, J. Zhang, and S. G. Trost, "Prediction of relative physical activity intensity using multimodal sensing of physiological data," *Sensors*, vol. 19, no. 20, p. 4509, Oct. 2019.

[32] S. Aristizabal et al., "The feasibility of wearable and self-report stress detection measures in a semi-controlled lab environment," *IEEE Access*, vol. 9, pp. 102053–102068, 2021.

[33] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.

[34] A. Petutschnig, S. Reichel, K. Měchurová, and B. Resch, "An eDiary app approach for collecting physiological sensor data from wearables together with subjective observations and emotions," *Sensors*, vol. 22, no. 16, p. 6120, Aug. 2022.

[35] J. Schumm, M. Bächlin, C. Setz, B. Arnrich, D. Roggen, and G. Tröster, "Effect of movements on the electrodermal response after a startle event," *Methods Inf. Med.*, vol. 47, pp. 186–191, Feb. 2008.

[36] C. Chênes, G. Chanel, M. Soleymani, and T. Pun, "Highlight detection in movie scenes through inter-users, physiological linkage," in *Social Media Retrieval*. London, U.K.: Springer, Oct. 2013, pp. 217–237.

[37] D. R. Bach, K. J. Friston, and R. J. Dolan, "An improved algorithm for model-based analysis of evoked skin conductance responses," *Biol. Psychol.*, vol. 94, no. 3, pp. 490–497, Dec. 2013.

[38] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Apr. 2016.

[39] V. Shusterman, K. P. Anderson, and O. Barnea, "Spontaneous skin temperature oscillations in normal human subjects," *Amer. J. Physiol.-Regulatory, Integrative Comparative Physiol.*, vol. 273, no. 3, pp. R1173–R1181, Sep. 1997.

[40] E. A. M. Alrahawe, V. T. Humbe, and G. N. Shinde, "An analysis on biometric traits recognition," *Int. J. Innov. Technol. Exploring Eng. (IJITEE)*, vol. 8, no. 7S, pp. 307–312, May 2019.

[41] A. Kitsiou, C. Despotidi, C. Kalloniatis, and S. Gritzalis, "The role of users' demographic and social attributes for accepting biometric systems: A Greek case study," *Future Internet*, vol. 14, no. 11, p. 328, Nov. 2022.

[42] T. Hui and R. Sherratt, "Coverage of emotion recognition for common wearable biosensors," *Biosensors*, vol. 8, no. 2, p. 30, Mar. 2018.

Martin Karl Moser received the bachelor's degree in business statistics from the University of Mount Olive, Mount Olive, NC, USA, in May 2019, and the master's degree in data science from the University of Salzburg, Salzburg, Austria, in June 2022, where he is currently pursuing the Ph.D. degree with the Department of Geoinformatics (Z_GIS).

He is also a Researcher with Z_GIS, University of Salzburg. Within the department, he works at the Geo-social Analytics Laboratory, with a focus on topics related to emotion sensing based on wearable sensors, specifically for urban planning and related applications. He is a U.S. College Sports Consultant for Smarthlete GmbH, Vienna, Austria. His main research interests include the processing and development of algorithms leveraging psychophysiological sensor data and the different geo-related use cases the findings can be applied to.

Mr. Moser is a recipient of the NCAA Postgraduate Scholarship.

Bernd Resch received the Ph.D. degree in "Live Geography" (real-time monitoring of environmental geo-processes) from the University of Salzburg, Salzburg, Austria, in 2009, and the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

He is an Associate Professor with the Department of Geoinformatics (Z_GIS), University of Salzburg, and a Visiting Scholar with Harvard University, Cambridge. He heads the Geo-social Analytics Laboratory and the iDEAS:lab, Z_GIS. He is an Executive Board Member of Spatial Services GmbH, Salzburg. His research interests include understanding cities as complex systems through analyzing a variety of digital data sources and focusing on developing machine-learning algorithms to analyze human-generated data like social media posts and physiological measurements from wearable sensors. The findings are relevant to a number of fields, including urban research, disaster management, epidemiology, and others.

Dr. Resch received the Theodor Körner Award for his work on "Urban Emotions." Among a variety of other functions, he is an Editorial Board Member of the *International Journal of Health Geographics* (IJHG), *International Journal of Geo-Information* (IJGI), and *PLOS ONE*, a scientific committee member for various international conferences (having chaired several conferences), and a Speaker of the Faculty of Digital and Analytical Sciences at PLUS.

Maximilian Ehrhart was born in Schönwies, Province Tyrol, Austria, on March 4, 1999. He received the bachelor's degree in digitalization, innovation and society from the University of Salzburg, Salzburg, Austria, in 2022. He is currently pursuing the master's degree in artificial intelligence with Johannes Kepler University Linz, Linz, Austria.

He is working as a Research Assistant at the University of Salzburg. During his bachelor and master studies, he has been working as a Research Assistant, focused on the development of methods for stress detection based on physiological measurement data using machine-learning and deep-learning techniques. His research interests include deep-learning methods for time-series analysis and prediction and generative models for time series.