Real-time stress detection based on artificial intelligence for people with an intellectual disability

Vries, S. de; Oost, F. van; Smaling, H.; Knegt, N. de; Cluitmans, P.; Smits, R.; Meinders, E.

Assistive Technology
The Official Journal of RESNA

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uaty20

Real-time stress detection based on artificial intelligence for people with an intellectual disability

Stefan de Vries, Fransje van Oost, Hanneke Smaling, Nanda de Knegt, Pierre Cluitmans, Reon Smits & Erwin Meinders

ASSISTIVE TECHNOLOGY
https://doi.org/10.1080/10400435.2023.2261045

Real-time stress detection based on artificial intelligence for people with an intellectual disability

Stefan de Vries, MSc[a], Fransje van Oost, MSc[a], Hanneke Smaling, PhD[b,c], Nanda de Knegt, PhD[d], Pierre Cluitmans, PhD[e], Reon Smits, MSc[a], and Erwin Meinders, PhD[a]

[a]Research and Development, Mentech Eindhoven, Eindhoven, The Netherlands; [b]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands; [c]University Network for the Care sector Zuid-Holland, Leiden University Medical Center, Leiden, The Netherlands; [d]Prinsenstichting, Care center for people with intellectual disabilities, Purmerend, The Netherlands; [e]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT

People with severe intellectual disabilities (ID) could have difficulty expressing their stress which may complicate timely responses from caregivers. The present study proposes an automatic stress detection system that can work in real-time. The system uses wearable sensors that record physiological signals in combination with machine learning to detect physiological changes related to stress. Four experiments were conducted to assess if the system could detect stress in people with and without ID. Three experiments were conducted with people without ID ($n = 14$, $n = 18$, and $n = 48$), and one observational study was done with people with ID ($n = 12$). To analyze if the system could detect stress, the performance of random, general, and personalized models was evaluated. The mixed ANOVA found a significant effect for model type, $F(2, 134) = 116.50$, $p < .001$. Additionally, the post-hoc t-tests found that the personalized model for the group with ID performed better than the random model, $t(11) = 9.05$, $p < .001$. The findings suggest that the personalized model can detect stress in people with and without ID. A larger-scale study is required to validate the system for people with ID.

ARTICLE HISTORY
Accepted 11 September 2023

KEYWORDS
artificial intelligence; challenging behavior; intellectual disability; real-time stress detection; wearables

Introduction

People with an intellectual disability (ID) experience difficulties in conceptual, social, and practical areas of living, according to the DSM-5 (American Psychiatric Association, 2013). They may have difficulty expressing themselves (Smith et al., 2020), which can result in unmet needs that may cause stress (Cappelletti et al., 2015). Stress can be defined as the response to an environmental demand that exceeds the natural regulatory capacity of an organism, especially in uncontrollable and unpredictable situations (Koolhaas et al., 2011). While the stress response to acute stressors can be beneficial, prolonged or chronic exposure to stress affects physical and mental health (Yaribeygi et al., 2017). This may put people with ID that may depend on others to cope with stress at increased risk (Scott & Havercamp, 2014). If the built-up stress goes unnoticed due to communicative deficiencies, stress may exhibit as challenging behavior, for example, aggression and self-injury (Ali et al., 2014; Janssen et al., 2002). Challenging behavior harms the quality of life and may harm fellow residents and caregivers (Bruinsma et al., 2020; Gur, 2016). This further complicates the client's care and may lead to staff dropout (Ryan et al., 2021; ZIN, 2019). To prevent challenging behavior that stems from stress it is important to detect the stress timely. Early notification of stress may help people with ID to inform caregivers about unmet needs, thereby enabling the caregiver to respond adequately and timely. This may reduce the number of behavioral escalations, which can lower the burden on caregivers, reduce costs, improve the quality of life, and limit stress-related health outcomes for people with ID who have difficulties expressing themselves.

In recent years, automatic stress detection has become increasingly effective for the general population. Many different methods have been successful in detecting stress (for a review, see Giannakakis et al., 2019). Studies have demonstrated effective stress detection in both lab and uncontrolled conditions (Can et al., 2019; Sandulescu et al., 2015;, respectively). This was achievable on large windows of data, suitable for detecting stress afterward (Jaques et al., 2017), as well as short windows, suitable for real-time usage (Saeed & Trajanovski, 2017). Many different mathematical methods have been effective, ranging from rule-based algorithms (Salai et al., 2016; Tomczak et al., 2020), to deep learning (Bobade & Vani, 2020). These methodologies have been applied to different data sources like mobile phone usage (Vildjiounaite et al., 2018), facial expressions (Gao et al., 2014), and physiology (Saeed & Trajanovski, 2017).

Not all data sources are practical for a real-time application. Due to advances in wearable sensors, some physiological signals can be recorded in real-time without interfering with daily life. This is true for electrodermal activity (EDA) and heart rate variability (HRV), which have been found highly informative in stress detection (Alberdi et al., 2016). EDA refers to the electric conductivity of the skin, which increases due to sweat secretion (Boucsein et al., 2012), and is considered a biomarker for stress (Cacioppo et al., 2007). The relation between stress

CONTACT Stefan de Vries Stefan.devries@mentechinnovation.eu Mentech Eindhoven, Boschdijk 766, Eindhoven 5624 CL, The Netherlands.

and EDA is especially clear after decomposing the signal into a fast and slow component. The fast component, or skin conductance response (SRC), peaks between 1 to 3 times per minute during rest, and over 20 times per minute during stress (Boucsein, 2012). HRV refers to metrics that express the variation in beat-to-beat time differences (i.e., inter-beat interval; IBI) from the heart (Shaffer & Ginsberg, 2017). Reduced variation is associated with a diminished parasympathetic nervous system and an increased sympathetic nervous system, which is highly associated with stress and physical effort (Appelhans & Luecken, 2006).

Besides EDA and HRV, accelerometers can also be beneficial in stress detection (Wu et al., 2015). Physical activity may elicit similar physiological responses as stress (Mastorakos et al., 2005; Puli & Kushki, 2020). This may lead to false positive stress detections when relying only on physiological signals. Additionally, the intensity of physical movements in people with ID may be informative for the level of stress (Doodeman et al., 2022). Including acceleration may therefore provide context that assists in making more reliable stress predictions.

Despite the success of many stress detection studies, their generalizability is often questionable (Nkurikiyeyezu et al., 2019). In the first place, generalizability may be harmed by the complex overlap in physical and mental stressors present in daily life. Contextual information, such as movement, may be relevant to distinguish mental stress from physical stress. Another explanation for this may be that a one-size-fits-all solution is not appropriate for stress detection (Alberdi et al., 2016; Taylor et al., 2020). However, training models (i.e., feeding a machine learning algorithm enough data to learn from) for each user in real-world applications is impractical and costly (Matthes et al., 2020). To account for personalization compactly, multi-task learning (MTL) neural networks were introduced (Jaques et al., 2017; Saeed et al., 2018; Taylor et al., 2020). These models capture common characteristics in shared layers and personal characteristics in person-specific layers (see Figure 1). These models have been demonstrated to outperform generalized models while avoiding training a model for each person.

Although many developments have been made in stress detection, most studies have been done with healthy adults. While some studies have investigated this technology in adults and children with autism (Puli & Kushki, 2020; Tomczak et al., 2020), and elderly with dementia (Kikhia et al., 2016), to our knowledge no studies exist in which automatic stress detection was investigated in people with ID. However, on a physiological level stress responses in people with ID don't seem to differ from the general population. For example, compared to the general population, people with ID show no different activity of the hypothalamic-pituitary-adrenal (HPA) axis which involves the central nervous system and the endocrine system during stress (Presland et al., 2013). In addition, it has been found that situations that elicit emotion show physiological changes in people with ID, suggesting that physiology gives information about their emotional state (Frederiks et al., 2019; Vos et al., 2012). Therefore, automatic stress detection based on physiology may be possible for people with ID. The automatic classification of stress based on physiological signals may demonstrate a novel application in care for people with ID. This technology may help people with ID and their caregivers in situations where the continuous presence of a caregiver is required, but who may not be able to continuously observe the client.

Objectives

The first objective of the study was to develop an automatic stress detection system suitable for real-time use. To allow for integration in a real-time system, the study focused on limiting the required computational power, and fully automatic data processing without human interference. The second objective was to assess if the system generalizes to people with ID. For this, the performance of the system was investigated between people with and without ID.

Methods

Participants

To train and validate the system a custom dataset was recorded. Participants were recruited for one of four experiments. The first three experiments were conducted with participants without ID. Participants were recruited through personal connections and online volunteer platforms. Participation was done voluntarily, and no compensation was given. All participants gave written informed consent. Participants had to be at least 18 years old and capable of understanding Dutch or English instructions. Additionally, in the third experiment, people were excluded that could not

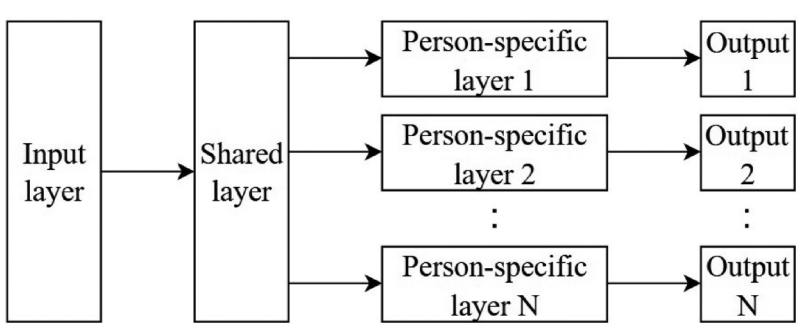

Figure 1. General architecture of multi-task learning neural network with person specific layers.

run on a treadmill. Data for the first three experiments were collected between May 2020 – December 2020.

The fourth group consisted of participants with challenging behavior (i.e., according to psychologists of the care centers) and severe to profound ID, following the DSM-5 criteria they required daily supervision of self-care activities up to 24-hour care. Participants in this group were residents of living facilities. At most eight residents lived together in one living facility. Each resident had their own room, the shared facilities included the bathroom, kitchen, and living room. The type of care differed greatly among clients within a facility, with some clients living relatively independently, while others require very intensive care. The number of clients per caregiver depends on the intensity of the provided care and the time of day. At least one caregiver is present per four clients. The contacted living facilities were known to the researchers through industry events. Participants were selected in collaboration with the caregivers and behavioral experts of these living facilities. The inclusion criteria were ID with challenging behavior, residency at the living facility, willingness to participate from caregivers and legal guardians of the participant, and sensor acceptance from the participant. Data for experiment 4 were collected between June 2019 – October 2020. Informed consent was given by the participant, when possible, otherwise by their legal representatives. The Medical Ethics Review Committee of VU University Medical Center judged that the Medical Research Involving Human Subjects Act did not apply (protocol number 2019.255)

Measurement equipment

A set of wearables was selected to record physiology. The selection was based on improving sensor acceptance and integration in daily care for the participants with ID (Kikhia et al., 2016). Therefore, the number of devices was limited, and only wearable wireless devices were selected. For heart rate, several devices were considered (e.g., Polar H10 (Polar Electro, Kempele, Finland), Zephyr Bioharness (Zephyr Technology Corporation, Annapolis, MD, US)), due to compatibility and affordability reasons Movesense (Movesense, Vantaa, Finland) was chosen. This device recorded the average HR per minute (0.5 – 2 Hz), IBI as an interval in microseconds between consecutive heartbeats (0.5 – 2 Hz), and acceleration on the x, y, and z-axis (26 Hz). Being made for athletes the device can handle movement and is highly accurate even for clinical purposes (Hartikainen et al., 2019; Parak et al., 2021). Empatica E4 wristband (Empatica, Milan, Italy) was used to record EDA in micro Siemens (4 Hz) on the non-dominant wrist, this device has been specifically validated for stress detection (Kyriakou et al., 2019) and challenging behavior (Imbiriba et al., 2020). In addition, it was the only compatible wearable available that recorded EDA in a non-obtrusive fashion. SentiSock (Mentech, Eindhoven, The Netherlands) was used as an additional sensor during the lab experiments to record EDA in micro Siemens (4 Hz) on the dominant foot sole using an EDA sensor integrated into a sock. In separate studies this device has been compared against Empatica for measuring skin conductance and stress detection (De Vries et al., 2022; Leborgne et al., 2023). Data were transmitted using the HUME (Mentech, Eindhoven, The Netherlands), which is a cloud-based platform. The HUME gathers data from wearables sensors using a nearby smartphone and transfers this data to the cloud. In the cloud, this data is processed in real-time using signal-processing and ML methods to come to a binary stress prediction. The stress prediction is visualized in the HUME app of the caregiver, which enables them to see stress in real-time. While HUME is mainly developed to assist people with ID and communicative difficulties, it may apply to other target groups as well.

Experimental protocol

Experiments 1–3 were conducted in a controlled environment (Mentech, Eindhoven, The Netherlands). These experiments were designed to induce stress but differed in experimental tasks and levels of physical activity. All experiments started with applying the sensors, followed by a sitting resting period of 3 minutes. Afterward, multiple stress-inducing tasks of 3–5 minutes were conducted. Each task was followed by 3 minutes of rest. The task order in all experiments was not randomized. After each task, the arousal dimension of the Self-assessment Manikin Scale (SAM) (Bradley & Lang, 1994) was filled in by the participant to indicate their stress levels. This scale presents five pictures ranging from a very drowsy to a highly active person.

The study applied two types of stressors that aimed to affect physical and emotional stress. Physical stress can be induced by mental workload, while emotional stress may be triggered by real or perceived threats (Hong, 2023). To induce physical stress the Stroop task (Stroop, 1992) and an arithmetic task were applied. During the Stroop task participants were asked to read the words for different colors printed in different font colors as fast as possible. The arithmetic task participants were asked to solve questions similar to Figure 2. Both tasks were done for approximately 5 minutes. The protocol was applied in two different experiments. During experiment 1 participants were asked to act out physical movements that were assumed to resemble daily life motions that may introduce signal noise (e.g., waving, cleaning a surface, and rocking back and forth). While in experiment 2 participants walked on a treadmill. The

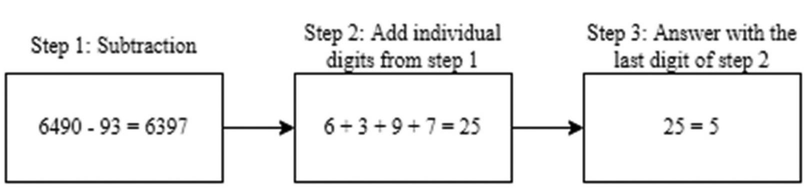

Figure 2. Arithmetic task example.

third experiment aimed to induce emotional stress through VR videos and games. VR has been investigated in the context of stress detection and seems like a promising method (Dammen et al., 2022). Additionally, it was assumed that the ability to immerse the participant in a fear-inducing environment may stimulate emotional stress. Participants were informed they could remove the headset at any point, for example, if they experienced motion sickness. An overview of the content can be found in Table 1. The genre of the chosen content was horror as this induced a strong emotional reaction and was available. All videos were passive experiences in which the participant could look all around. The games required active participation from the participant. The HTC Vive Pro VR headset was used to show the VR content.

In experiment 4, no stress was induced but periods of stress and relaxation were observed during normal daily activities. These recordings were made at the living facility of the participants with ID. Both the Movesense HR+ and Empatica E4 were worn and were complemented with video recordings. A minimum of three recording sessions were planned with the caregivers for each participant. If no stressful episode occurred during the session, a new recording session was planned. The recordings took between 2 to 8 hours. This was dependent on the participants' sensor acceptance and daily schedule. Subsequently, two healthcare professionals (i.e., professional caregiver and psychologist) labeled stress and periods of no stress independently. Both professionals knew the client, one usually provided care, and the other was involved as a behavioral expert for the client. Prior to analyzing the recordings the healthcare professionals agreed on what behavior reflected which arousal state. Based on these definitions they labeled the video recording by listing start and end timestamps and if the behavior was stress. Moments where they did not agree were discussed to come to an agreement.

Data analysis

The data analysis consisted of signal processing, feature extraction, data cleaning, model training, and model validation. With all steps, the limitations of the real-time platform were taken into account. This meant that at most 45-second windows of data could be analyzed at a time.

The signals (i.e., IBI, EDA, acceleration) were filtered in different ways to ensure the best reduction of noise and ensure usability in real-time. The IBI signal was filtered using the real-time IBI correction algorithm by Rand et al. (2007). This filter uses a buffer to correct the signal locally according to a set of rules that detect missed and ectopic heartbeats. The EDA signal was filtered using discrete wavelet transforms (DWT). The EDA signal was then split into a phasic and tonic component using the cvxEDA algorithm (Greco et al., 2016). For the acceleration signal, a Butterworth low-pass filter was used, since motion caused by a person is related to high frequencies (Bayat et al., 2014).

The cleaned signals were used to extract features for each second. An overview of the features can be found in Table 2. The features served as the input array for model training, and the collected labels served as the ground truth. The labels were binarized by excluding the middle SAM value and assigning no-stress to the lower values and stress to the higher values. Next, the features and labels were temporally aligned. Any moments that were not associated with a label were excluded from model training. In total, three models were trained for each experiment. Training was done using the python package Keras (version 2.3.1) with the TensorFlow (version 1.14) backend. The first model was a baseline random chance model, the second a neural network for generic use (referred to as the general model), and the third model was the MTL neural network with personalized layers (referred to as the personalized model). The architecture of the personalized model was based on similar studies that applied MTL to stress detection problems (Jaques et al., 2016; Saeed & Trajanovski,

Table 1. Virtual reality (VR) content used in experiment 3.

Content name	Content type	Source link
The Exorcist	video	https://www.youtube.com/watch?v=Zd-k_jrgDJk&t=13s
The Nun	video	https://www.youtube.com/watch?v=evzsN1BGR6A&t=23s
Bloody Mary	video	https://www.youtube.com/watch?v=gV7u6NYyLpA&t=6s
Nightmare	video	https://www.youtube.com/watch?v=QOgrFZxyQhk&t=46s
Slender Man	video	https://www.youtube.com/watch?v=lYJgTFkCNYU
Lights out	video	https://www.youtube.com/watch?v=7v-dG9Rq_aY&t=18s
The Conjuring 2	video	https://www.youtube.com/watch?v=lYJgTFkCNYU
Five Nights at Fredies	game	https://store.steampowered.com/app/732690/FIVE_NIGHTS_AT_FREDDYS_HELP_WANTED/
Funhouse	game	https://store.steampowered.com/app/468700/NVIDIA_VR_Funhouse/

Table 2. Overview of features extracted on physiological signals, including the signal and window length.

Feature	Signal	Window length in seconds	Reference
Slope	EDA signal	20	Cho et al. (2017)
Mean	SCR	20	Cho et al. (2017)
Max	SCR	20	Cho et al. (2017)
Number of peaks	SCR	20	Cho et al. (2017)
Slope	IBI	20	Shaffer and Ginsberg (2017)
Standard deviation	IBI	20	Shaffer and Ginsberg (2017)
Root-Mean-Square of Successive Differences	IBI	20	Shaffer and Ginsberg (2017)
Magnitude	Acceleration on chest	20 & 40 seconds	(Garcia-Ceja et al. (2015)

EDA = Electro Dermal Activity; SCR = Skin Conductance Response; IBI = Inter-Beat Interval.

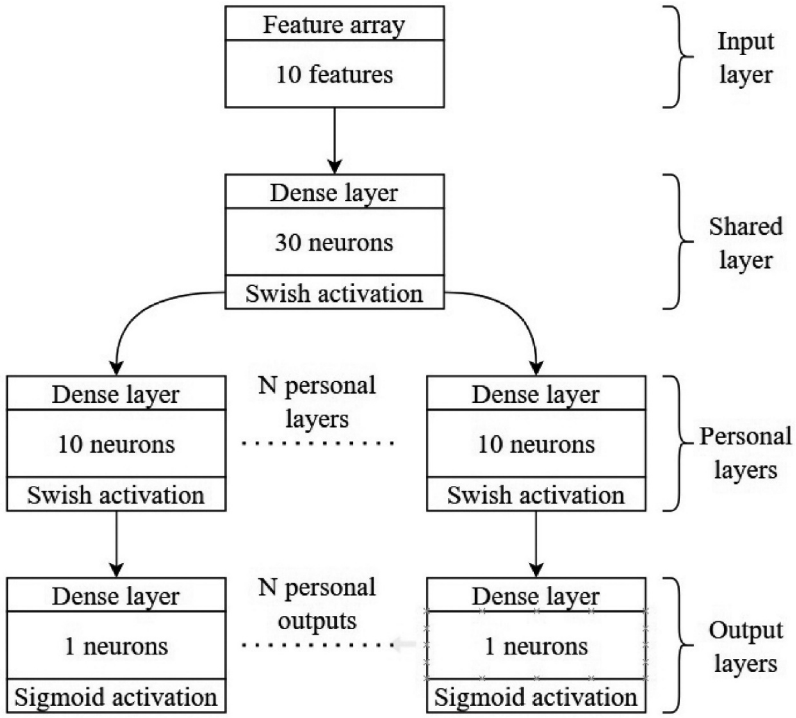

Figure 3. Personalized model architecture using the multi-task learning neural network framework.

2017; Taylor et al., 2020) and is shown in Figure 3. The general model used the same architecture as the personalized model, instead of the personalized layers, it contained only one output layer. The models were compiled using the binary cross-entropy loss function with the optimizer Adam and a learning rate of 0.0001. To assess the performance of the models, a leave-one-out cross-validation (LOOCV) was used. Predictions were made after each iteration on the left-out data, which were used to calculate the balanced accuracy.

For experiment 4, a trained model from experiments 1–3 was applied to validate the performance in the ID group. Through this approach, it could be investigated if models trained on people without ID generalized to people with ID.

Statistical analysis

To compare the balanced accuracy of the different models in the groups, a mixed ANOVA was done using Pingouin (version 0.5.1). Assumptions of normality, homogeneity of variances, and sphericity were checked using the python packages SciPy (version 1.7.3) and Pingouin. The four different experiments were used as between-subject differences and the random chance, general, and personalized models were used as within-subject differences. The dependent variable was balanced accuracy. Afterward, post-hoc T-tests were done to assess the differences between experimental groups, models, and interactions (e.g., the model difference between experimental groups). The p-values in the post-hoc tests were corrected using the Bonferroni method. Statistical power was calculated post-hoc with G-power 3, using the F-scores from the mixed ANOVA procedure and the sample size of the smallest group.

Results

Descriptives of the sample are given in Table 3. In total 21 participants were excluded, 7 from experiment 1, and 14 from experiment 3. The reason for exclusion was either corrupt sensor data, or not experiencing any stress during the experiment based on their self-reported SAM scores. In experiment 3 no reports on motion sickness were made.

All models were trained using the LOOCV procedure. This resulted in a balanced accuracy for each person on each model, shown in Figure 4. Mixed ANOVA was used to assess the differences between the models and experimental groups.

Prior to the analysis, the assumptions for mixed ANOVA were tested. Normality was assessed per within-group, the p-values were 0.62, 0.98, and 0.07 for the groups random, general, and personal, respectively. Sphericity was tested using Mauchly's test and was met, $\chi^2(2) = 1.93$, $p = .38$. Finally, the homogeneity of variances was assessed with Levene's test and was not violated, $F = 1.03$, $p = .42$. The results

Table 3. Demographical information of each experimental group.

Group	Average age (years)	SD age	N participated	N included in analysis	N male
Experiment 1	35.5	16.5	18	11	7
Experiment 2	24.2	5.8	14	14	10
Experiment 3	34.7	14.8	48	34	18
Experiment 4	45.0	13.2	12	12	6

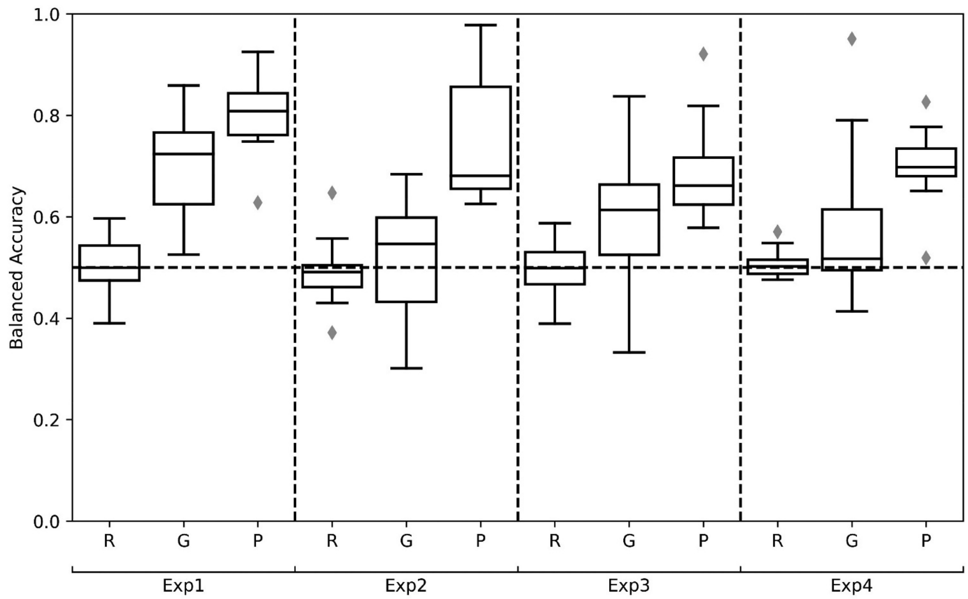

Figure 4. Distribution of balanced accuracies for each model on each experiment. Where R is the random chance model, G is the personalized model, P is the personalized model and Exp# indicates the experiment.

of the mixed ANOVA showed a significant effect for experimental group, $F(3, 67) = 7.23$, $p < .001$, $\eta_p^2 = 0.24$. Additionally, a significant effect was found for model type, $F(2, 134) = 116.50$, $p < .001$, $\eta_p^2 = 0.64$. Based on these findings the statistical power was assessed using the smallest η_p^2 and the sample size of the smallest group ($n = 11$). The statistical power was found to be 0.81. Post-hoc tests were conducted to assess the study's objectives.

Validation of stress detection system

The random, general, and personalized models differed significantly from each other. In each experimental group, the personalized model performed better than the random and general model, $T(70) = 15.03$, $p < .001$, $\eta_p^2 = 0.66$, $T(70) = -8.32$, $p < .001$, $\eta_p^2 = 0.24$, respectively. In turn, the general model performed significantly better than the random model, $T(70) = 5.50$, $p < .001$, $\eta_p^2 = 0.19$. However, the within-comparison found that the general model did not significantly differ from the random model in experiments 2 and 4. Contrary, the personalized model did outperform the random model significantly in all experimental groups. For an overview of the within-comparison results see Table 4.

Generalizability to people with ID

A comparison between the different experimental groups was conducted to assess if the system generalized to people with ID. Experimental group 4 did not differ from groups 2 and 3, $T(23.98) = -0.14$, $p = 1.00$, $\eta_p^2 = 0.001$; $T(19.34) = 0.50$, $p = 1.00$, $\eta_p^2 = 0.007$, respectively. However, group 4 did differ from group 1, $T(18.62) = 3.49$, $p = .015$, $\eta_p^2 = 0.35$. Although, groups 2 and 3 also differed significantly from group 1, $T(19.94) = 3.54$, $p = .012$, $\eta_p^2 = 0.35$; $T(13.93) = 3.57$, $p = 0.019$, $\eta_p^2 = 0.34$, respectively.

Discussion

This study had two main objectives. Firstly, to develop a stress detection system given real-time data processing constraints. Secondly, to assess if the system would generalize to people with ID. For the first objective it could not be concluded that the general model was capable of detecting stress. Guided by

Table 4. Post-hoc analysis of model performance differences per group.

Group	Model 1	Model 2	T-value	DoF	*P*-value	η_p^2
Exp 1	General	Personal	−4.67	10	.01	0.23
Exp 1	General	Random	5.75	10	.002	0.56
Exp 1	Personal	Random	10.76	10	< .001	0.82
Exp 2	General	Personal	−5.38	13	.002	0.49
Exp 2	General	Random	0.64	13	1.00	0.01
Exp 2	Personal	Random	5.77	13	.001	0.63
Exp 3	General	Personal	−4.75	33	< .001	0.15
Exp 3	General	Random	4.44	33	.001	0.28
Exp 3	Personal	Random	9.69	33	< .001	0.65
Exp 4	General	Personal	−4.17	11	.02	0.40
Exp 4	General	Random	1.00	11	1.00	0.04
Exp 4	Personal	Random	9.05	11	< .001	0.74

DoF = Degrees of freedom; P-value is corrected using Bonferroni; η_p^2 = *Partial Eta squared*.

recommendations from previous research, the personalized model was investigated (Saeed & Trajanovski, 2017). In contrast to the general model, the personalized model did perform significantly better than the random chance and general model in all groups. It was therefore concluded that the proposed system can detect stress in lab conditions, but it does require a personalized stress detection model. This further underlines that personalized models result in better-performing systems than one-size-fits-all methods (Alberdi et al., 2016; Taylor et al., 2020).

For the second objective, it was analyzed if the models generalized to people with ID. While the ANOVA found a significant effect on model performance for the different groups, it was found in the post-hoc analysis that control group 1 significantly differed from all other groups. The group with ID did not differ from control groups 2 and 3. This suggests that the proposed system performed similarly for people with and without ID. In addition, the personalized model performed significantly better for the group with ID than the random chance and general model, while the latter two did not differ from each other at this point. Therefore, model personalization was necessary for people with ID as well, and resulted in an accuracy of 70 percent on average.

Overall, the system performs similarly for people with and without ID. This underlines the findings from the literature that physiology provides information about the emotional state of people with ID (Vos et al., 2012). Similar to people with autism (Puli & Kushki, 2020; Tomczak et al., 2020), it is possible to predict stress in people with ID based on physiology. People with ID may significantly benefit from this technology as it may notify caregivers about needs they may not be able to express themselves (Smith et al., 2020). In practice, this could mean that stress is detected earlier, which may prevent escalation into challenging behavior and limit negative physical and mental health related to prolonged stress (Penley et al., 2002; Scott & Havercamp, 2014), such as the reduced risk of cardiovascular disease (Steptoe & Kivimäki, 2012) and depression (Hartley et al., 2009).

There were several limitations in the study. In the first place, the sample size per group was relatively small. Given the effect sizes, it was found to be sufficient, but a larger scale study would be recommended. The study design could have been improved by randomizing the task order within the experiments. Additionally, baseline levels of stress were not recorded. Participants who reported stress within the relaxing conditions were removed, but the relative stress increase could have been a meaningful addition. By recording baseline stress levels it could also have been assessed if the experimental settings itself induced stress. Next, moments of stress were labeled either through self-reports or observations by behavioral experts. This may have introduced some subjectivity into the process (Alekhine et al., 2020). In addition, binarizing the labels into stress and no-stress removed information about the intensity of stress. There is evidence that regression-based models can perform better for stress detection (Siirtola & Röning, 2020), suggesting that a continuous scale may fit better. All moments that were not labeled, such as moments in between tasks, were excluded, which may have introduced selection bias. It could also have removed meaningful information, such as anticipation for the next stressful task. Requesting labels from these moments could have improved the design. Additionally, the study aimed to demonstrate that stress detection in people with ID is feasible. The model performance was therefore not highly optimized. The performance may be improved by implementing active learning to prompt caregivers for expert-feedback (El-Hasnony et al., 2022). In addition, semi-supervised learning can be used to include unlabeled data (Chebli et al., 2018), which reduces the labeling burden on caregivers. Both methods integrate real-life data, which improves the generalizability to the healthcare use-case. Lastly, the model was developed to be integrated into a real-time system. The model performance may therefore have been limited by the restrictions of the real-time system. Nevertheless, the findings suggest that the proposed stress detection system could detect stress in both people with and without ID. Future work may focus on notifying caregivers about physiological changes preceding stress (Simons et al., 2021), and how automatic stress detection affects physical and mental health, and behavioral outcomes in people with ID.

Conclusions

The study aimed to assess if the proposed automatic stress detection could be used to detect stress. The personalized model was found to be capable of detecting stress in all groups. On average the personalized model had a balanced accuracy of 0.73, while the general model performed at 0.60. These findings further support the notion that a one-size-fits-all approach may not be suitable for stress detection. Furthermore, it was found that the developed system performed largely similarly for people with and without ID. This suggests that the system generalized to people with ID. The proposed technology could, therefore, assist people with ID to express their needs during times of stress. Notifying caregivers during times of stress enables them to help people with ID cope, which may reduce the negative effects of stress, such as challenging behavior. Although the findings are promising, a large-scale validation study is required. The present study, therefore, serves as the first step in this direction.

Disclosure statement

Employees S. de Vries and F. van Oost and owners E. Meinders and R. Smits of Mentech Innovation b.v. have a commercial interest in the findings from the study. The remaining authors have no conflicts of interest to declare.

Funding

The work was supported by the Health~Holland [LSHM21001].

Data availability statement

Participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

References

Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. *Journal of Biomedical Informatics*, *59*, 49–75. https://doi.org/10.1016/j.jbi.2015.11.007

Alekhine, P. F., Rouillac, L., & Granry, J. (2020). Subjective versus objective assessment of short term occupational stress: Bias and analysis of self-assessment of high stress levels. *Journal of Advances in Medicine and Medical Research*, 50–64. https://doi.org/10.9734/jammr/2020/v32i1030519

Ali, A., Blickwedel, J., & Hassiotis, A. (2014). Interventions for challenging behaviour in intellectual disability. *Advances in Psychiatric Treatment*, *20*(3), 184–192. https://doi.org/10.1192/apt.bp.113.011577

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Appelhans, B., & Luecken, L. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology*, *10*(3), 229–240. https://doi.org/10.1037/1089-2680.10.3.229

Bayat, A., Pomplun, M., & Tran, D. A. (2014). A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, *34*, 450–457. https://doi.org/10.1016/j.procs.2014.07.009

Bobade, P., & Vani, M. (2020). Stress detection with machine learning and deep learning using multimodal physiological data. *2020 Second international conference on inventive research in computing applications (ICIRCA)*. https://doi.org/10.1109/ICIRCA48905.2020.9183244

Boucsein, W. (2012). Methods of electrodermal recording. In W. Boucsein (Ed.), *Electrodermal activity* (pp. 87–258). Springer US. https://doi.org/10.1007/978-1-4614-1126-0_2

Boucsein, W., Fowles, D., Grimnes, S., Ben-Shakhar, G., Roth, W., Dawson, M., & Filion, D. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bradley, M., & Lang, P. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Bruinsma, E., van den Hoofdakker, B. J., Groenman, A. P., Hoekstra, P. J., de Kuijper, G. M., Klaver, M., & de Bildt, A. A. (2020). Non-pharmacological interventions for challenging behaviours of adults with intellectual disabilities: A meta-analysis. *Journal of Intellectual Disability Research*, *64*(8), 561–578. https://doi.org/10.1111/jir.12736

Cacioppo, J., Tassinary, L., & Berntson, G. (2007). *Handbook of psychophysiology*. https://doi.org/10.1017/cbo9780511546396

Can, Y. S., Arnrich, B., & Ersoy, C. (2019). Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of Biomedical Informatics*, *92*, 103139. https://doi.org/10.1016/j.jbi.2019.103139

Cappelletti, E. R., Kreuter, M. W., Boyum, S., & Thompson, T. (2015). Basic needs, stress and the effects of tailored health communication in vulnerable populations. *Health Education Research*, *30*(4), 591–598. https://doi.org/10.1093/her/cyv033

Chebli, A., Djebbar, A., & Marouani, H. F. (2018). Semi-supervised learning for medical application: A survey. *2018 international conference on applied smart systems (ICASS)*, 1–9. https://doi.org/10.1109/ICASS.2018.8651980

Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N.-K., & Lee, B. (2017). Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. *Sensors*, *17*(10), 2435. https://doi.org/10.3390/s17102435

Dammen, L., van Finseth, T. T., McCurdy, B. H., Barnett, N. P., Conrady, R. A., Leach, A. G., Deick, A. F., Van Steenis, A. L., Gardner, R., Smith, B. L., Kay, A., & Shirtcliff, E. A. (2022). Evoking stress reactivity in virtual reality: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *138*, 104709. https://doi.org/10.1016/j.neubiorev.2022.104709

De Vries, S., Smits, R., Tataj, M., Ronckers, M., Van Der Pol, M., Van Oost, F., Adam, E., Smaling, H., & Meinders, E. (2022). *Accurate stress detection from novel real-time electrodermal activity signals and multi-task learning models*. https://doi.org/10.54941/ahfe1001845

Doodeman, T. W. M., Schuengel, C., & Sterkenburg, P. S. (2022). Expressions of stress of people with severe intellectual disabilities and sensitive caregiving to regulate stress: A qualitative study. *Journal of Intellectual & Developmental Disability*, *47*(4), 308–317. https://doi.org/10.3109/13668250.2021.2011171

El-Hasnony, I. M., Elzeki, O. M., Alshehri, A., & Salem, H. (2022). Multi-label active learning-based machine learning model for heart disease prediction. *Sensors*, *22*(3), 1184. https://doi.org/10.3390/s22031184

Frederiks, K., Sterkenburg, P., Barakova, E., & Feijs, L. (2019). The effects of a bioresponse system on the joint attention behaviour of adults with visual and severe or profound intellectual disabilities and their affective mutuality with their caregivers. *Journal of Applied Research in Intellectual Disabilities: JARID*, *32*(4), 890–900. https://doi.org/10.1111/jar.12581

Gao, H., Yüce, A., & Thiran, J.-P. (2014). Detecting emotional stress from facial expressions for driving safety. *2014 IEEE International conference on image processing (ICIP)*, 5961–5965. https://doi.org/10.1109/ICIP.2014.7026203

Garcia-Ceja, E., Osmani, V., & Mayora-Ibarra, O. (2015). Automatic stress detection in working environments from smartphones' accelerometer data: A first step. *IEEE Journal of Biomedical and Health Informatics*, *20*, 1053–1060. https://doi.org/10.1109/JBHI.2015.2446195

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2019). *Review on psychological stress detection using biosignals*. https://doi.org/10.1109/TAFFC.2019.2927337

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2016). cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering*, *63*(4), 797–804. https://doi.org/10.1109/TBME.2015.2474131

Gur, A. (2016). Challenging behavior, functioning difficulties, and quality of life of adults with intellectual disabilities. *International Journal of Developmental Disabilities*, *64*(1), 45–52. https://doi.org/10.1080/20473869.2016.1221233

Hartikainen, S., Lipponen, J. A., Hiltunen, P., Rissanen, T. T., Kolk, I., Tarvainen, M. P., Martikainen, T. J., Castren, M., Väliaho, E.-S., & Jäntti, H. (2019). Effectiveness of the chest strap electrocardiogram to detect atrial fibrillation. *The American Journal of Cardiology*, *123*(10), 1643–1648. https://doi.org/10.1016/j.amjcard.2019.02.028

Hartley, S. L., MacLean, W. E., & Abbeduto, L. (2009). Depression in adults with mild intellectual disability: Role of stress, attributions, and coping. *American Journal on Intellectual and Developmental Disabilities*, *114*(3), 147–160. https://doi.org/10.1352/1944-7588-114.3.147

Hong, K. (2023). Classification of emotional stress and physical stress using a multispectral based deep feature extraction model. *Scientific Reports*, *13*(1), 2693. https://doi.org/10.1038/s41598-023-29903-3

Imbiriba, T., Cumpanasoiu, D. C., Heathers, J., Ioannidis, S., Erdoğmuş, D., & Goodwin, M. S. (2020). Biosensor prediction of aggression in youth with autism using kernel-based methods. *Proceedings of the 13th ACM international conference on PErvasive technologies related to assistive environments*, 1–6. https://doi.org/10.1145/3389189.3389199

Janssen, C., Schuengel, C., & Stolk, J. (2002). Understanding challenging behaviour in people with severe and profound intellectual disability: A stress-attachment model. *Journal of Intellectual Disability Research: JIDR*, *46*(6), 445–453. https://doi.org/10.1046/J.1365-2788.2002.00430.X

Jaques, N., Rudovic, O., Taylor, S., Sano, A., & Picard, R. W. (2017). Predicting tomorrow's mood, health, and stress level using personalized multitask learning and domain adaptation. *AffComp@IJCAI*.

Jaques, N., Taylor, S., Nosakhare, E., Sano, A., & Picard, R. W. (2016). *Multi-task learning for predicting health, stress, and happiness*.

Kikhia, B., Stavropoulos, T. G., Andreadis, S., Karvonen, N., Kompatsiaris, Y., Sävenstedt, S., Pijl, M., & Melander, C. (2016). Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors*, *16*(12), 1989. https://doi.org/10.3390/s16121989

Koolhaas, J. M., Bartolomucci, A., Buwalda, B., De Boer, S. F., Flügge, G., Korte, S. M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., Van Dijk, G., Wöhr, M., & Fuchs, E. (2011). Stress revisited: A critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews*, *35*(5), 1291–1301. https://doi.org/10.1016/j.neubiorev.2011.02.003

Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., Liedlgruber, M., Wilhelm, F., Osborne, T., & Pykett, J. (2019). Detecting moments of stress from measurements of wearable physiological sensors. *Sensors (Basel, Switzerland)*, *19*(17), 3805. https://doi.org/10.3390/s19173805

Leborgne, F., Smits, R., Gencheva, M., De Vries, S., Meinders, E., Cluitmans, P., Lee, C., & Kuhlmann, J. (2023). *The development of a washable and durable smart textile to measure electrodermal activity for early stress recognition*. https://doi.org/10.54941/ahfe1002872

Mastorakos, G., Pavlatou, M., Diamanti-Kandarakis, E., & Chrousos, G. P. (2005). Exercise and the stress system. *Hormones (Athens, Greece)*, *4* (2), 73–89. https://doi.org/10.2217/17455057.2.4.561

Matthes, S., Han, Z., Qiu, T., Michel, B., Klinger, S., Shen, H., Liu, Y., & Altakrouri, B. (2020). *Personalized stress detection with self-supervised learned features*. https://www.semanticscholar.org/paper/Personalized-Stress-Detection-with-Self-supervised-Matthes-Han/95f19123d9fc0351e2311e93390604c81141f74b

Nkurikiyeyezu, K., Yokokubo, A., & Lopez, G. (2019). Importance of Individual differences in physiological-based stress recognition models. *2019 15th international conference on intelligent environments (IE)*. https://doi.org/10.1109/IE.2019.00006

Parak, J., Salonen, M., Myllymäki, T., & Korhonen, I. (2021). Comparison of heart rate monitoring accuracy between chest strap and vest during physical training and implications on training decisions. *Sensors (Basel, Switzerland)*, *21*(24), 8411. https://doi.org/10.3390/s21248411

Penley, J. A., Tomaka, J., & Wiebe, J. S. (2002). The association of coping to physical and psychological health outcomes: A meta-analytic review. *Journal of Behavioral Medicine*, *25*(6), 551–603. https://doi.org/10.1023/A:1020641400589

Presland, A. D., Clare, I. C. H., Broughton, S., Luke, L., Wheeler, E., Fairchild, G., Watson, P. C., Chan, W. Y. S., Kearns, A., & Ring, H. A. (2013). Hypothalamic-pituitary-adrenal (HPA) axis activity in adults with intellectual disabilities: A preliminary investigation: HPA axis activity in adults with intellectual disabilities. *Journal of Intellectual Disability Research*, *57*(6), 539–551. https://doi.org/10.1111/j.1365-2788.2012.01557.x

Puli, A., & Kushki, A. (2020). Toward automatic anxiety detection in autism: A real-time algorithm for detecting physiological arousal in the presence of motion. *IEEE Transactions on Biomedical Engineering*, *67* (3), 646–657. https://doi.org/10.1109/TBME.2019.2919273

Rand, J., Hoover, A., Fishel, S., Moss, J., Pappas, J., & Muth, E. (2007). Real-time correction of heart interbeat intervals. *IEEE Transactions on Biomedical Engineering*, *54*(5), 946–950. https://doi.org/10.1109/TBME.2007.893491

Ryan, C., Bergin, M., & Wells, J. S. G. (2021). Work-related stress and well-being of direct care workers in intellectual disability services: A scoping review of the literature. *International Journal of Developmental Disabilities*, *67*(1), 1–22. https://doi.org/10.1080/20473869.2019.1582907

Saeed, A., Ozcelebi, T., Lukkien, J., Erp, J. V., & Trajanovski, S. (2018). Model adaptation and personalization for physiological stress detection. *2018 IEEE 5th international conference on data science and advanced analytics (DSAA)*, (pp. 209–216). IEEE.

Saeed, A., & Trajanovski, S. (2017). Personalized driver stress detection with multi-task neural networks using physiological signals. *ArXiv*.

Salai, M., Vassányi, I., & Kósa, I. (2016). Stress detection using low cost heart rate sensors. *Journal of Healthcare Engineering*, *2016*, 1–13. https://doi.org/10.1155/2016/5136705

Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., & Mozos, Ó. M. (2015). Stress detection using wearable physiological sensors. *International Work-Conference on the Interplay Between Natural and Artificial Computation*. https://doi.org/10.1007/978-3-319-18914-7_55

Scott, H., & Havercamp, S. (2014). Mental health for people with intellectual disability: The impact of stress and social support. *American Journal on Intellectual and Developmental Disabilities*, *119*(6), 552–564. https://doi.org/10.1352/1944-7558-119.6.552

Shaffer, F., & Ginsberg, J. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*, 5. https://doi.org/10.3389/fpubh.2017.00258

Siirtola, P., & Röning, J. (2020). Comparison of regression and classification models for user-independent and Personal stress detection. *Sensors*, *20*(16), 4402. https://doi.org/10.3390/s20164402

Simons, R., Koordeman, R., De Looff, P., & Otten, R. (2021). Physiological measurements of stress preceding incidents of challenging behavior in people with severe to profound intellectual disabilities: Longitudinal study protocol of single-case studies. *JMIR Research Protocols*, *10*(7), e24911. https://doi.org/10.2196/24911

Smith, M., Manduchi, B., Burke, É., Carroll, R., McCallion, P., & McCarron, M. (2020). Communication difficulties in adults with intellectual disability: Results from a national cross-sectional study. *Research in Developmental Disabilities*, *97*, 103557. https://doi.org/10.1016/j.ridd.2019.103557

Steptoe, A., & Kivimäki, M. (2012). Stress and cardiovascular disease. *Nature Reviews Cardiology*, *9*(6), 360–370. https://doi.org/10.1038/nrcardio.2012.45

Stroop, J. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Taylor, S., Jaques, N., Nosakhare, E., Sano, A., & Picard, R. W. (2020). Personalized multitask learning for predicting tomorrow's mood, stress, and health. *IEEE transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2017.2784832

Tomczak, M. T., Wojcikowski, M., Pankiewicz, B., Lubinski, J., Majchrowicz, J., Majchrowicz, D., Walasiewicz, A., Kilinski, T., & Szczerska, M. (2020). Stress monitoring system for individuals with autism spectrum disorders. *IEEE Access*, *8*, 228236–228244. https://doi.org/10.1109/ACCESS.2020.3045633

Vildjiounaite, E., Kallio, J., Kyllönen, V., Nieminen, M., Määttänen, I., Lindholm, M., Mäntyjärvi, J., & Gimel'farb, G. (2018). Unobtrusive stress detection on the basis of smartphone usage data. *Personal and Ubiquitous Computing*, *22*(4), 671–688. https://doi.org/10.1007/s00779-017-1108-z

Vos, P., De Cock, P., Munde, V., Petry, K., Van Den Noortgate, W., & Maes, B. (2012). The tell-tale: What do heart rate; skin temperature and skin conductance reveal about emotions of people with severe and profound intellectual disabilities? *Research in Developmental Disabilities*, *33*(4), 1117–1127. https://doi.org/10.1016/j.ridd.2012.02.006

Wu, M., Cao, H., Nguyen, H.-L., Surmacz, K., & Hargrove, C. (2015). Modeling perceived stress via HRV and accelerometer sensor streams. *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. https://doi.org/10.1109/EMBC.2015.7318686

Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *Excli Journal* *16*, 1057–1072. *Doc1057; ISSN 1611-2156*. https://doi.org/10.17179/EXCLI2017-480

ZIN. (2019). *Screeningsrapport Gehandicaptenzorg Zinnige Zorg*.