Ageing@Work

Smart, Personalized and Adaptive ICT Solutions for Active, Healthy and Productive Ageing with enhanced Workability

Project Acronym:	Ageing@Work
Project Full Name:	Smart, Personalized and Adaptive ICT Solutions for Active, Healthy and Productive Ageing with enhanced Workability
Grant Agreement:	No 826299
Project Duration:	3 years (starting 1 January 2019)

Deliverable 4.2

Worker behaviour & affective traits monitoring methods & infrastructure

Work Package:	WP4
Lead Beneficiary:	KUL
Due Date:	31/10/2021 (M34)
Submission Date:	31/10/2021 (M34)
Deliverable Status:	Final
Deliverable Style:	R
Dissemination Level:	Public
File Name:	D4.2 Worker behaviour & affective traits monitoring methods & infrastructure

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 826299

Authors

Surname	First Name	Beneficiary
McLaughlin	Jack	KUL
Liu	Yunxin	KUL
Evangelia	Zacharaki	UPAT
Pavlou	Michail	UPAT
Ferri	Daniel	MYSPHERA
Lithoxoidou	Evdoxia Irini	CERTH
Segkouli	Sofia	CERTH
Mastoras	Rafail Evangelos	CERTH
Papathomas	Evangelos	CERTH
Hatzikonstantinou	Ioannis	CERTH
Triantafyllidis	Andreas	CERTH

In case you want any additional information or you want to consult with the authors of this document, please send your inquiries to: info-ageingatwork@iti.gr.

Reviewers

Surname	First Name	Beneficiary
Karagiannopoulos	George	HIT
Alocca	Carlo	SAMSUNG

Version History

Version	Date	Modifications made by
v0.1	8/09/2021	Table of Contents created by KUL
v0.2	14/09/2021	Table of Contents modified and subsections defined by UPAT
V0.3	05/10/2021	Summarized description of subsections by UPAT
v0.4	11/10/2021	Input provided by KUL
v0.5	18/10/2021	Input provided by CERTH
v0.6	19/10/2021	Input provided by UPAT
v0.7	21/10/2021	Input provided by MYSPHERA
V0.8	22/10/2021	Merged input and writing the summary by KUL, corrections by UPAT
v1.0	24/10/2021	Submission for internal review by UPAT
v1.1	26/10/2021	Review of version v1.0 provided by HIT
v1.2	29/10/2021	Final version

Disclaimer

Executive Summary

In this deliverable we present the worker behavior and affective traits monitoring methods and infrastructure. The main aim of this is to explore different methods for measuring ongoing levels of negative mental states like stress and depressive disorders in workers and develop rules for identification of stress. We further provide some tools to help manage the different levels of stress identified. We have investigated both advanced methods based on machine learning for inference of worker stress from unobtrusive wearables (such as a smartwatch), and recognition of instantaneous emotions from facial expressions, as well as self-reports for recognition of long-term depressive signs and psychological fatigue.

This deliverable illustrates the need and value in identifying stress for aging workers and presents a summary of the literature on the various behavioural and stress monitoring techniques. While some of these techniques are tried and tested in stress identification, like self-reports, we also explore advanced methods to capture physiological signals, such as skin conductance, heart rate, and skin temperature from wearable technology as well as keystroke analysis. The application of these techniques to stress detection and emotion recognition, as well as their use in online clinical stress management tools like computerised cognitive behavioural therapy (cCBT) is also demonstrated. The infrastructure and algorithms developed for Ageing@Work project, which are used to capture and transform this data for use in stress detection and emotion recognition, are explained in detail.

In addition, we elaborate on the history and effectiveness of cCBT in behavioural interventions as it is an effective method for stress management. We provide some exercises commonly used in such therapy programs to help workers manage ongoing stress. These exercises are administered to workers based upon a set of rules developed from the subjective measures of stress identified in this task, numerous established self-reports. These subjective measures are tested through an online study with the results leading to some alterations to the original rules developed. The outcome is a set of rules which use periodic reporting to identify a workers' stress level and recommend which exercises, if any, would help them manage their stress level. Furthermore, the infrastructure for collecting the self-report data and providing the outcome of the rules' calculation is detailed.

In the future we aim to combine subjective measures with objective measures when detecting stress and evaluate the consistency between the two measurement types.

Table of Contents

List of Figures

List of Tables

List of Terms and definitions

Table 1 Definitions

Abbreviation	Definition
ANN	Artificial Neural Network
cCBT	Computerised Cognitive Behavioural Therapy
CNN	Convolutional Neural Networks
COPSOQ	Copenhagen Psychosocial Questionnaire
ECG	Electrocardiography
EDA	Electrodermal Activity
EEG	Electroencephalography
GAP	Global Average Pooling
GSR	Galvanic Skin Response
HRV	Heart Rate Variability
KNN	K-Nearest Neighbours
MDD	Major Depressive Disorder
ML	Machine Learning
MSC	Mental Health Composite Scale
MQ	Health and Working Conditions Questionnaire from D3.2 (Monthly Questionnaire)
MZ	MoodZoom Questionnaire
NN	Normal to Normal
PHQ9	Patient Health Questionnaire 9 Items
PSS	Perceived Stress Scale
PT	Personality Trait
RBF	Radial Basis Function
RF	Random Forest
RR	Respiration Rate
SF12	Short Form Health Survey 12 items
SVM	Support Vector Machines
VGG	Visual Geometry Group
VUM	Virtual User Model
YLD	Years Lived with Disability

1. Introduction

1.1 Scope of the Deliverable

The deliverable is dedicated to the highly challenging aim of monitoring worker affective traits in realistic workplace environments, be it at the factory or at home with the main emotion of interest in this respect being stress. Measuring stress more accurately and frequently is a necessary step to further propose useful interventions to minimize negative outcomes. The stress level could be measured by both subjective and objective methods using self-reported or technological means, respectively. We have investigated both advanced methods (for the inference of worker stress from unobtrusive wearables such as a smartwatch) and self-reports, while also utilizing contextual cues that could help resolve ambiguities in biosignals-based stress recognition in the wild (such as physical activity and behavioural parameters from the VUM). In addition to acute stress episodes detection, several instruments (questionnaires) ae proposed towards identifying for instance depressive or burn-out signs and user fatigue that may require thorough attention and recommend some intervention (e.g. cognitive behavioral therapy). This latter definition and measurement of stress and depressive signs, indicating long-term effects, was considered as more relevant for the needs of the AW workers than the instantaneous stress levels and was therefore chosen to be integrated to the workers' VUM.

1.2 Relation to Other Activities and Deliverables

This deliverable is related to (i) deliverable D3.2 describing the VUM that includes behavioural parameters that being part of the stress level estimation algorithm (allowing to differentiate stress-induced heart rate changes from heart rate increase due to physical activity), (ii) the middleware and worker smart tracking infrastructure described in deliverable D4.1 for retrieval of parameters of interest, (iii) the Virtual Coach for the delivery of interventions (deliverable D5.3) including the Cognitive Behavioral Therapy, and (iv) deliverable D5.1 presenting the worker Dashboard which visualizes the overall stress level of the worker.

1.3 Structure of the Deliverable

After an introductory section, in Section 2 we outline the rational of monitoring the workers' behaviour and stress, and the importance of delivering Cognitive Behavioural Therapy (CBT) or other behavioural interventions. Following this, we perform in Section 3 a literature review on state-of-the-art methodologies and common practices for monitoring worker behaviour and affective traits. In Section 4 we present the technological infrastructure, algorithms, and methods developed in Ageing@Work to handle data capture and to detect, mainly instantaneous, stress and emotions based on sensors and wearables by exploiting benchmark data found in available databases. Section 5 details the final framework which is based on validated questionnaires and rules developed to flag ongoing stress, especially focusing on long-term effects and depressive signs. This section also elaborates on the solutions used to reduce stress levels, like cCBT. Section 5 ends with information on the future work related to this deliverable, before concluding remarks round out the deliverable in Section 6.

2. Background and Rationale

2.1 Behaviour and Stress Monitoring

Ageing workers not only face many physical challenges but also significant psychological challenges, with two indicated as most common, depression and burnout. Depression can have a great impact on people and leads to impaired functioning in daily life. Depression occurs in about the 7% of the general older population and it accounts for about 5.7% of Years Lived with Disability (YLD) in people over 60 years old. Depression is both undertreated and underdiagnosed in primary care settings, as the symptoms are often overlooked and untreated because they co-exist with other problems encountered by older adults (WHO, 2017).

There is mounting evidence on the link between working stress and burnout, however, there is limited knowledge about the extent to which workers' age is associated with burnout (Marchand and Blanc, 2015). Some findings suggest that age is negatively associated with the occurrence of burnout (Norlund et la., 2015; Marchand et al., 2010). Some studies have also indicated that there is a bimodal relationship between age and burnout (Ahola et al., 2006; Cheng et al., 2013), with burnout being elevated in both younger and older workers. Last, some other studies have shown higher levels of burnout only in older workers (Lindblom et al., 2006; Verdonk et al., 2010). Although, research is not conclusive about the link between age and burnout, working stress is an important issue for the well-being of working population in general and therefore special attention must be paid.

2.2 Computerised Cognitive Behavioural Therapy

Stress has become a major mental and/or health problem in today's workplace. In today's digital age, different technologies (e.g., keystroke features and wearables) have been used to detect stress levels. Prior studies have shown that stress measurement based on new technologies like keystroke dynamics and wearable sensors is consistent with the self-reported stress measurement (Lim et al., 2020; Pakhomov et al., 2020; Sano et al., 2018; Vizer et al., 2009). In the meanwhile, prior research suggests that Cognitive Behavioural Therapy (CBT) or a subset of CBT (i.e., behavioural interventions) is effective in the treatment of depression, anxiety, and stress disorders (Attwood et al., 2012; Bakker et al., 2018; Fuller-Tyszkiewicz et al., 2020; Knowles et al., 2015; Ly et al., 2014; Mewton et al., 2012; Morgan et al., 2017; Newby & McElroy, 2020). However, given that a therapist is necessary, the application of traditional CBT is limited in some situations, particularly in the treatment of stress. Because it is entirely normal for us to experience a range of stress in daily life. Therefore, a simplified CBT is more suitable for reducing stress. Technological advancement has made it possible that computerised, internet and online CBT (cCBT) is effective in treating mental disorders including depression and anxiety (Attwood et al., 2012; Mewton et al., 2012; Morgan et al., 2017; Newby & McElroy, 2020). Moreover, the combination of new technologies (e.g., wearable devices) with online CBT is effective as well in other domains like obsessive-compulsive disorders and insomnia (Aarøen, 2013; Kang et al., 2017; Seung-Gul & Yong-Ku, 2019). However, little is known about the combination of new technologies (e.g., wearables and keystroke features) with CBT in reducing stress and/or depression. We aim to fill this gap by exploring the application of some cCBT exercises in reducing stress among ageing workers. We also examine whether the objective measure of stress (i.e., data obtained from keystroke features and wearable devices) aligns with the subjective measure of stress (i.e., data obtained from self-reported survey).

2.2.1 Stress Detection Using Different Technologies

Measuring stress more accurately and frequently is a necessary step to further propose useful interventions to minimize negative outcomes. The stress level could be measured by both subjective and objective methods (Vizer et al., 2009). A self-reported survey is still a common method when measuring subjective stress. Examples of current measures include the perceived stress/stress perception scale (PSS) and Mental Health Composite Scale (MSC) (Vizer et al., 2009). Recently, researchers have measured objective stress by taking advantage of new technologies like keystroke features (Lim et al., 2020; Vizer et al., 2009) and wearable devices (Pakhomov et al., 2020; Sano et al., 2018). The changes in keystroke features as indicated by the use of the backspace, delete, end, and arrow keys refer to as an indication of experiencing stress (Lim et al., 2020; Vizer et al., 2009). Also, the changes detected by wearable devices as indicated by skin conductance, heart rate, skin temperature refer to as under stress (Pakhomov et al., 2020; Sano et al., 2018). Prior studies on the application of new technologies (i.e., keystroke features and wearable devices) in detecting stress have shown that the objective measures align with subjective measures and show high levels of accuracy. Therefore, we combine both the subjective measures (i.e., data obtained from self-reported survey) with objective measures (i.e., data obtained from keystroke features and wearable devices) when detecting stress. We aim to replicate and validate the prior results about the consistency between the two measures.

2.2.2 The Application of CBT and Behavioural Interventions

Technological advancement has made it possible to design internet (computerised) CBT to treat mental problems (e,g., depression and anxiety). In general, an experimental group receives CBT programs (i.e., a series of several sessions) through online platforms. In contrast, a control group does not receive those CBT programs. cCBT is comparable with face-to-face CBT in terms of treating mental disorders such as depression, anxiety, and distress (Knowles et al., 2015; Seung-Gul & Yong-Ku, 2019). In fact, due to its flexibility, autonomy, low cost, and enhanced privacy, cCBT works pretty well in treating mental problems among patients with mental disorders (Knowles et al., 2015). Also, cCBT is effective in improving mental health for both universal (i.e., without mental disorders) and targeted (i.e., with mental disorders) children and adolescents (Attwood et al., 2012). Moreover, cCBT programs without professional guidance also have the capacity to treat mental problems such as depression and anxiety (Morgan et al., 2017; Seung-Gul & Yong-Ku, 2019). We also design some simplified online CBT programs and aim to explore whether those unguided CBT programs are effective in reducing stress.

Additionally, adherence to CBT programs is a key determinant of the effectiveness of cCBT for mental disorders like depression and anxiety (Hilvert-Bruce et al., 2012). Smartphones have become an increasingly important part of our daily life. With notifications and reminders sent by smartphones, individuals show higher levels of adherence to different online programs (Kang et al., 2017). Developers and practitioners have designed new applications to reduce stress, depression, and further enhance mental well-being. Researchers also explore the effectiveness of those applications in treating mental problems (Bakker et al., 2018; Fuller-Tyszkiewicz et al., 2020; Ly et al., 2014). Those studies have shown that behavioural interventions delivered with the help of smartphones can reduce perceived stress and further improve mental health (Bakker et al., 2018; Fuller-Tyszkiewicz et al., 2020; Ly et al., 2014). We also deliver online CBT programs through applications installed in smartphones for ageing workers.

2.2.3 The Combination of New Technologies with cCBT

As aforementioned, adherence to CBT programs is a key factor to test the effectiveness of cCBT (Hilvert-Bruce et al., 2012). Recent studies have also focused on the combination of new technologies with cCBT to further increase adherence to cCBT (Aarøen, 2013; Dillon et al., 2016; Kang et al., 2017; Seung-Gul & Yong-Ku, 2019). The combination of cCBT with wearable devices is effective in insomnia treatment (Kang et al., 2017; Seung-Gul & Yong-Ku, 2019). On the one hand, functions such as notifications and reminders increase patients' adherence to digital programs. On the other hand, those notifications and reminders could be delivered more frequently and accurately with the help of wearable devices. Because patients could view their sleeping diaries more regularly (Kang et al., 2017; Seung-Gul & Yong-Ku, 2019). Also, a recent study has shown that behavioural interventions delivered through biofeedback are effective in reducing stress (Dillon et al., 2016). However, it is unclear whether the combination of cCBT with other new technologies like keystroke features is effective in reducing stress. Together, we combine cCBT with both wearable devices and keystroke features to further examine its effectiveness in treating mental disorders.

3. Worker Behaviour & Affective Traits Monitoring Methodologies

Automated stress and affective traits recognition has been examined in various fields of research which consider human emotional reactions, e.g. in marketing, technical equipment, or human–robot interaction. Stress and emotion recognition methods, as stated by (Dzedzickis A et al., 2020), can be classified into two main groups according to the basic monitoring instrument: (i) self-report techniques based on emotions self-assessment by filing various questionnaires and (ii) machine assessment techniques based on measurements of various parameters of human body. Those parameters can include Electroencephalography (EEG), Electrocardiography (ECG), Galvanic Skin Response (GSR), Heart Rate Variability (HRV), Respiration Rate Analysis (RR) and with some more recent approaches exploiting facial features through image based methods.

3.1Common Variables and Metrics in Self-Reported Assessment

Measuring stress accurately and frequently is a necessary step to propose useful interventions to minimize negative outcomes. The stress level could be measured by both subjective and objective methods (Vizer et al., 2009). A self-reported survey is still a common method when measuring subjective stress. Examples of current measures include the perceived stress/stress perception scale (PSS) and Mental Health Composite Scale (MSC) (Vizer et al., 2009).

The behaviour and stress identification of this deliverable leverages the data collected within the project infrastructure. Self-reported surveys are submitted at regular intervals in the app and the answers to these various surveys can assist in identifying levels of stress or depression. Additionally, data from wearable technologies can be utilised on daily and aggregate bases, and keystroke technology is collected and tested for future integration into stress detection.

There are 10 questions (Table 7) which are asked as part of a monthly questionnaire either directly relate to stress – frequency of feeling stressed, negative moods, concentration, and sleep (Åkerstedt et al., 2012) – or relate to variables that effect an individual's ability to manage perceived stress, including locus of control and a support networks (Sur and Ng, 2014). The first of these questionnaires is administered when an individual joins the app, and this initial survey collects additional data including personality traits. There is significant literature on personality traits, with common use of the Big 5 personality traits, and their relationship to perceived stress (Sur and Ng, 2014). Specifically conscientiousness and neuroticism are identified as having a significant correlation to an individual's perceived stress (Sur and Ng, 2014; Saksvik and Hetland, 2011).

In addition to this monthly data, there is a daily collection of emotional data through the MoodZoom questionnaire (Tsanas et al., 2016), as well as the collection of daily data via wearable technology. This information can also be aggregated into a picture of emotional states over a longer horizon, so that if there are any significant shifts these can be acted upon without overreacting to a daily volatility in moods.

Finally, the PHQ-9 (Kroenke and Spitzer, 2001) is a commonly used and effective clinical questionnaire (Levis, Benedetti and Thombs, 2019) related to identification of mental distress or depression, and is answered on a fortnightly basis. It also assists in categorising the severity of an identified issue.

3.2 Physiological and Other Measures of Stress from Sensor Data

Recently, researchers have measured objective stress by taking advantage of new technologies like wearable devices (Pakhomov et al., 2020; Sano et al., 2018) – that measure physiological signals, such as skin conductance, heart rate, skin temperature) – and keystroke patterns (Lim et al., 2020; Vizer et al., 2009). Prior studies on the application of new technologies (e.g., wearable devices and keystroke features) in detecting stress have shown that the objective measures may align with subjective measures and show high levels of accuracy. Therefore, a large research effort has gone towards exploring the use of consumer wearables for measuring physiological data. The emergence of wearable devices and their capability for continuous, energy efficient health monitoring during daily routines has facilitated this process greatly. A summary of devices used for physiological data collection is shown in Table 8 in the Annex.

Sensing devices for vital signs, such as medical or fitness wristbands and clothing or chest strap monitors, are plenty, but the measurements from most of them are affected by the physical activity and motion artifacts. Images and video captured from cameras allow the extraction of facial features and can supplement the physiological measurements. In addition, posture information and behavioral characteristics during job execution can provide valuable information.

An extensive overview on the most common physiological signals and extracted features used for stress detection, along with incorporated data analysis techniques and obtained accuracies, has recently been presented by Smets et al, 2018.

3.2.1 Other Sensor-Based Features for Affective Traits Recognition

Over the past few years, mobile Internet-linked devices have increased in numbers and offered a plethora of possibilities for pervasive monitoring of health and well-being (A. Triantafyllidis et al., 2015). Smartphones, smartwatches and additional mobile devices for the monitoring of health parameters and the environment, with constantly increasing networking and processing capabilities, have been introduced in our daily lives towards enabling better self-monitoring, promoting self-reflection, and improving wellness. Traditionally, data for well-being monitoring is collected through self-reports and answers to given questions. This approach may often be limited due to memory constraints and recall bias (S. D. Gosling, 2020). Nowadays, mobiles emerge as the main means for pervasive health monitoring, because of their wide uptake, and capabilities for sensing, which allow the unobtrusive gathering of rich personalized information in naturalistic contexts. In this light, mobile phones can be utilized for digital phenotyping (J.-P. Onnela and S. L. Rauch, 2016), i.e., 'moment-by-moment quantification of the individual-level human phenotype in situ', facilitating to develop markers to diagnose and treat diseases (I. Barnett et al, 2018).

There have been several research efforts for pervasive and unobtrusive monitoring of health and well-being utilizing mobile devices. Significant correlations between the smartphone-driven features and the disease state of patients with Major Depressive Disorder (MDD), through the usage of location-related data, have been found (Canzian et al., 2020). Also, significant predictors for personality traits, with data derived from mobile application which recorded activity data, conversations (number and duration), smartphone usage and location have been reported (Wang et al., 2018). A different source of data, derived from smart-screen keyboard typing dynamics seems to be able to correlate mental disorder scores for MDD (R. E. Mastoras et al., 2020), Bipolar Disorder (J. Zulueta et al., 2018) and Parkinson's disease (D. Iakovakis et al., 2020), associating psychomotor retardation with keyboard typing. Also, the changes in keystroke features as indicated by the use of the backspace, delete, end, and arrow keys, have been identified as an indication of experiencing stress (Lim et al., 2020; Vizer et al., 2009).

Additionally, the employment of multimodal sensing capabilities, often can increase the accuracy of predictive or diagnostic outcomes (E. Garcia-Ceja et al., 2018). Combining different sources of information from location, activity, typing, surveys and phone usage features has been reported to accurately distinguish the patients with Mild Cognitive Impairment or Alzheimer's Disease from healthy controls (Chen et al., 2019). Additional research studies, grouped by their corresponding field of study, can be found at Table 2 (H. Rashidisabet et al., 2020).

Table 2 State of the art studies using location, activity, keystroke patterns and other features

Research	Key aim	Sensor data	Outcome
Kuster et al. (2018)	Personality traits	Swipe speed, time between touches, touch accuracy, touch duration	Accuracy 0.59-0.67 on big five dimension traits
Wang et al (2018)	Personality traits	Ambient sound, Voice detection, gps, google activity api, phone activity,	Correlation of features and prediction
Kumar et al (2015)	Emotion recognition	Typing dynamics, text-based detection, HRV	Accuracies of 77.67%, 88.7% and 73% with keystrokes, textual and heart rate information respectively
Kołakowska (2015)	Emotion recognition	Typing dynamics on PC	Accuracy ranging from 47%-81% depending on specific emotion.
Ghosh et al (2017)	Emotion recognition	Typing dynamics, Markov chain for emotion,	AUC varying from 66%-84% combining models. Sketchy use of SMOTE and markov.
Wampfler et al (2020)	Emotion recognition	Typing dynamics (heat map of keypreasure, flight-time, down-down interval.	AUC of ~0.80 Valance, Arousal, Dominance prediction

Sağbaş et al (2020)	Stress Detection	Accelerometer, gyroscope, number of deletes while using a keyboard	On average, 87.56% accuracy is achieved.
Exposito et al (2018)	Stress Detection	Key preassure values	found a significant positive correlation (r=0.75, p=0.0081) which indicates that an increase of self-reported stress is usually associated with an increase of typing pressure and vice versa.
Sano et all (2013)	Stress detection	Surveys, GPS, Phone usage statistics, Accelerometer,	75% accuracy of low and high perceived stress recognition using the combination of mobile phone usage and sensor data: either one feature from screen on (SD of % of screen on between 6-9pm), mobility (median or SD of mobility radius), CALL and ACC/SC or the combination of the top 2-3 features from across the modalities.
Wang et al. (2016)	Finding significant predictors for Schizophrenia	Sleep, activity tracking, conversations (number and duration), smartphone usage, location tracking	Accurately predicting aggregated scores of mental health indicators in schizophrenia with a mean error of 7.6% of the score range
Barnett et al. (2018)	Determining how the detected anomalies change from baseline in the two weeks before hospitalization for a schizophrenic episode	Location, screen time, charging times, outgoing messages and calls	Found that the rate of behavioral anomalies detected in the two weeks prior to relapse was 71% higher than the rate of anomalies during other time periods.
Canzian et al. (2015), Mehrotra et al. (2016), Rohani et al. (2018)	Finding significant correlations between the smartphone-driven feature and the disease state of patients with Major Depressive Disorder (MDD) and Bipolar affective disorder (BPAD)	Accelerometer data, GPS, call frequency and duration, location, app usage	(Canzian) Predict PHQ-8 scores with sensitivity and specificity values of 0,74 and 0,78.
Chen et al. (2019)	Demonstrating the feasibility of detecting of cognitive impairment (dementia)	Accelerometer, pace, stride, heart rate, sleep cycle, distance from home, workout sessions, breathe sessions, standing hours, exercise minutes, phone calls, apps, sleep stages, missed calls, new apps, new contacts, energy survey, mood survey, number of steps, stairs climbed, and messages	0.726 (±0.021) Symptomatic vs HC 0.897 (±0.027) Mild AD vs HC
Zulueta et al. (2018)	Characterize the smartphone interactome of BPAD patients and building digital	Keystroke dynamics, accelerometer, average interkey delay, backspace ratio, autocorrect rate	Reported statistical relationship between keystroke meta-data and mood disturbances in subjects with bipolar disorders

	phenotypes to model depression and mania		
Mastoras et al. (2019)	Predicting and detecting depression scales in depressive disorder patients remotely	Keystroke dynamics (e.g., hold time, flight time, speed and press-flight rate) and psychological questionnaires (PHQ-9)	AUC = 0.89 (0.72–1.00; 95% Confidence Interval) and 0.82/0.86 sensitivity/specificity, with the outputted probabilities significantly correlating (>0.60) with the respective PHQ-9 scores.
Hussain et al. (2019)	Predicting mood disorder symptoms	Keystroke dynamics metadata and typing kinematics (e.g., accelerometer)	

3.3 Techniques for Stress and Emotion Recognition

Most of the researchers focusing on stress detection analyze several multi-sensorial physiological measurements aiming to identify which are the mostly related to stressful conditions. One of the earliest identified stress indicators in the human body is the heart rate variability (HRV). Costin et al, 2012 use HRV along with morphological variability of ECG signals to extract time and frequency domain features indicative of stress levels at traffic. Sierra et al., 2010 present a study of stress detection for hyperventilation based on HRV and electrodermal activity (EDA) using Fuzzy Logic. Zubair et al, 2015 propose the use of a wearable band device to measure skin conductance level and acceleration for stress estimation. Overall the principal physiological signals used for stress detection are HRV and EDA or galvanic skin response (GSR). That is because stress is related with the sympathetic nervous system, which controls heart rate, skin conductance and skin temperature.

Besides physiological signals, several studies show that behavioral characteristics can determine stress condition. For example, the work presented by Andren et al, 2005 includes a system designed to compute stress level based on the user's keyboard typing pattern. Giannakakis et al., 2017 used facial images to extract movement characteristics and to determine stress level. One more study made by Mozos et al., 2017 uses wireless sensors to extract physiological signals (EDA, HRV and photoplethysmogram) and sociometric features, such as speech, acceleration and direction of movement, for their stress detection model.

Many works have taken advantage of the multi-modal manifestation of stress and have combined information from different sensors to increase effectiveness. Early works by Liao et al, 2005 and Healey et al., 2005 showed that the accuracy of inferred stress increased with the number of information sources, which is expected as some physical symptoms, such as fast heart rate and rapid breathing, are not unique to stress. In more recent studies, modalities exploiting behavioral patterns, facial expressions and eye movements from camera or Kinect, or complete pose estimation, are combined for psychological stress modelling Carneiro et al., 2017.

In respect to emotion recognition, the methods mainly facial expressions. Generally facial emotion recognition is divided into three major stages, as explained by Azizan, et al.

1. Face Detection,
2. Feature Extraction, and
3. Emotion Classification.

At first stage, which is a pre-processing stage, an image of a face is detected and facial components of the face will be detected from the region. The facial components can be an eyes, brows, nose, and mouth. In the second stage, an informative features will be extracted from different parts of the face. In the last stage, a classifier need to be trained before been used to generate labels for the Emotions using the training data. The most common approaches utilize Convolutional Neural Networks (CNNs) which usually fuse together stages (2) and (3).

Emotion recognition through image based methods has gotten significant attention in the recent years, with emphasis on deep learning solutions. Sokolov & Patkin, 2018 present a CNN-based cross-platform application for emotion recognition running on smartphones and tablets. An architecture similar to ResNet was used by He et al., 2016 to process real-time video input that was implemented in two ways: a desktop version and a light-weight version for mobile settings in which batch normalization layers were removed and convolutional layer weights were modified. Another mobile emotion recognition technique based on video used a Finite State Machine to identify the temporal phases of expression and then applied SVM on every apex state, Suk et al., 2015. With the constant improvement of the processing power presented in modern mobile phones, it is now possible to utilize neural networks on those devices.

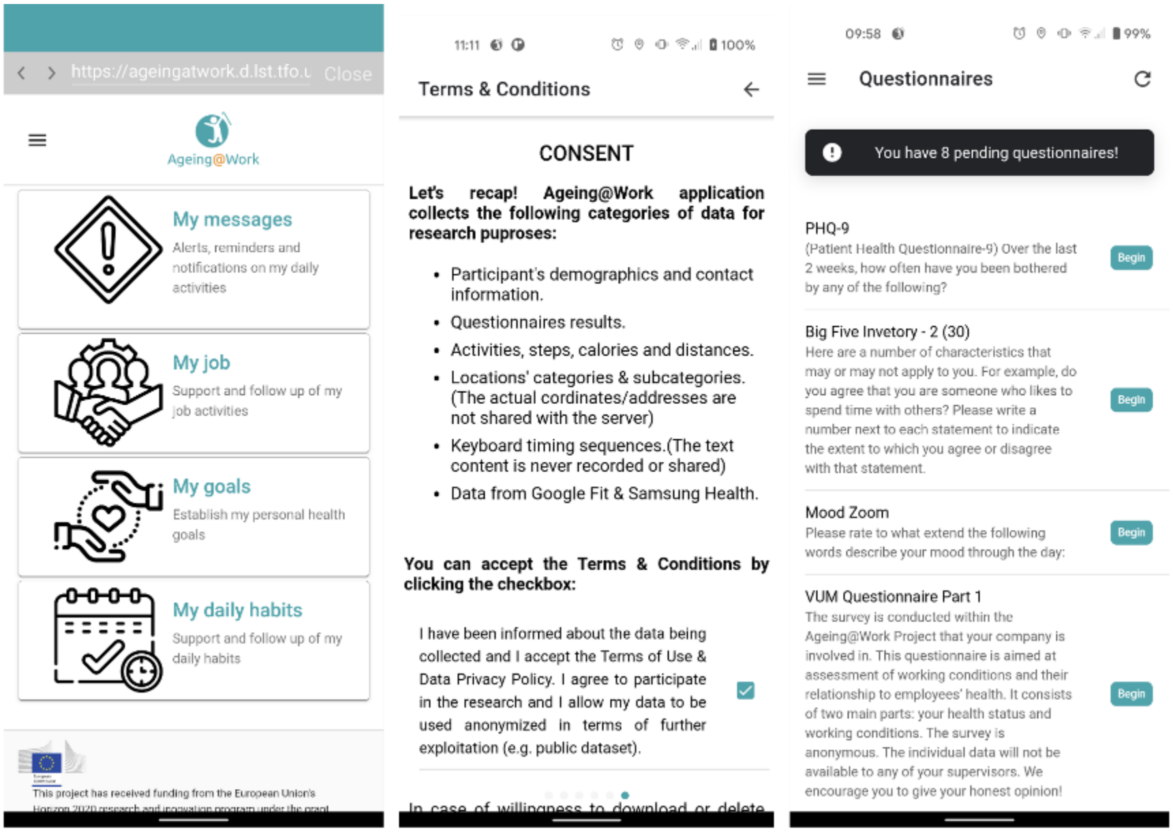
4. AW Infrastructure and Algorithms

4.1 Updates on the AW Application, Middleware Architecture and Data Collection Modalities

In this section, the updates related to the AW application, middleware architecture and data collection modalities will be presented, informing about the AW infrastructure that relates with the worker behaviour & affective traits monitoring methodology.

Since the Deliverable 4.1, the Ageing@Work application has been refactored and been prepared for the prepilots. Keeping the list brief, some of the most important updates are:

- Ageing@Work's data collection has been refactored in order to be in the same structure as Samsung Health and Google fit.
- An overview of avatar's messages for the current day has been added.
- Integrated Avatar's Exercises. Now the users are suggested to do some stretching exercises when they have ignored previous messages for breaks at work.
- The whole application's code has been refactored in order to accommodate translations, both for hardcoded texts as well as for the remotely received questionnaires.
- Integrated the Dashboard within the app as an external browser without the need for additional login.
- Consent form has been included within the app and non-acceptance means no data are being shared with the server.
- Additional questionnaires were included. Currently the following questionnaires are included: PHQ-9, Big Five Inventory (30), MoodZoom (daily), VUM (main questionnaire), SF12 Health Survey, COPSOQ (short version) Work related well-being and finally Workability Assessment.
- Changes have been made after the review process of the pre-pilots. Few bugs were fixed, and the UI was adjusted according to the suggestions to assure user-friendliness.



a) Dashboard Integration b) Consent form c) Questionnaires

Figure 1 Various aspects of the app

Moreover, in order to have a consistent structure between all the different sources (Google Fit, Samsung health and Ageing@Work), the structure of the data being collected by the Ageing@Work was altered, to group each event and provide start and end timestamps. Since all the different data source have the same structure now, their aggregation, based on time, can be done effortless and efficiently. Below an example of the current activities data structure is provided, as send from the application to the server.


```
{
        "distance": 256.6325988769531,
        "duration": 139885,
        "end": 1614708798713,
        "name": "RUNNING",
        "id": 14,
        "intensity": "Vigorous",
        "kcal": 27.871448516845703,
        "mets": 8.539090156555176,
        "start": 1614708658828,
        "steps": 343
}
```


Figure 2 Example of the current activities data structure

Furthermore, Ageing@Work application now also collects phone call events (duration, time of the day and anonymized caller id), charging events, battery level, screen on/off events and connectivity logs (WiFi). These features were added since they have shown promising results, in addition to the previous data being collected, in predicting personality traits (Sadeghian, Alireza, and Marjan Kaedi, 2021). Figure 2presents an overview of the data being collected, while Figure 3 depicts the middleware architecture which includes the Generic Tracking Library.

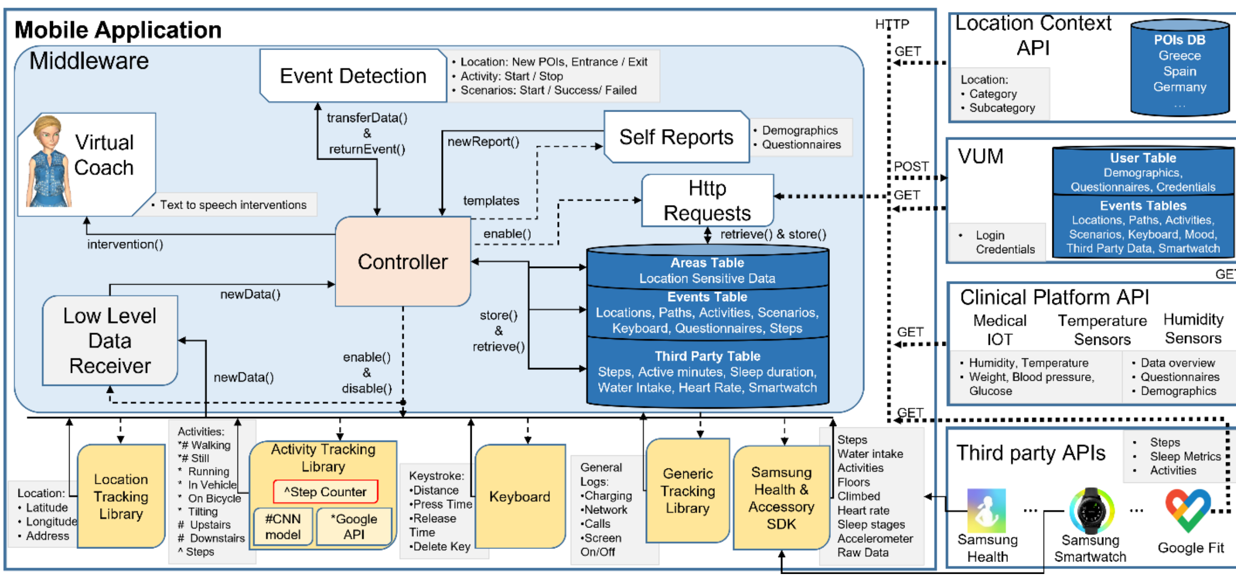


Figure 3 Updated application's middleware architecture

Moreover, the notifications provided by the Virtual Coach were enriched with appropriate messages for each possible detected stress level (low, medium, high). An illustrative example of text communicated to the worker for each level of stress is shown in Figure 4.

Finally, concerning the updates on the data collection, the Ageing@Work application now also accommodates additional questionnaires and in multiple languages for user-friendliness. In detail, the integrated questionnaires are the following:

- Patient Health Questionnaire – 9 (PHQ), 9 items, biweekly
- Big Five Inventory -2 – Short (BFS-2-S), 30 items, once
- Mood Zoom (MZ), 6 item, daily
- Virtual User Model Questionnaire (VUMQ), 103 items, once
- Short Form Health Survey (SF-12), 12 items, once
- Copenhagen Psychosocial Questionnaire (COPSOQ), 41 items, once

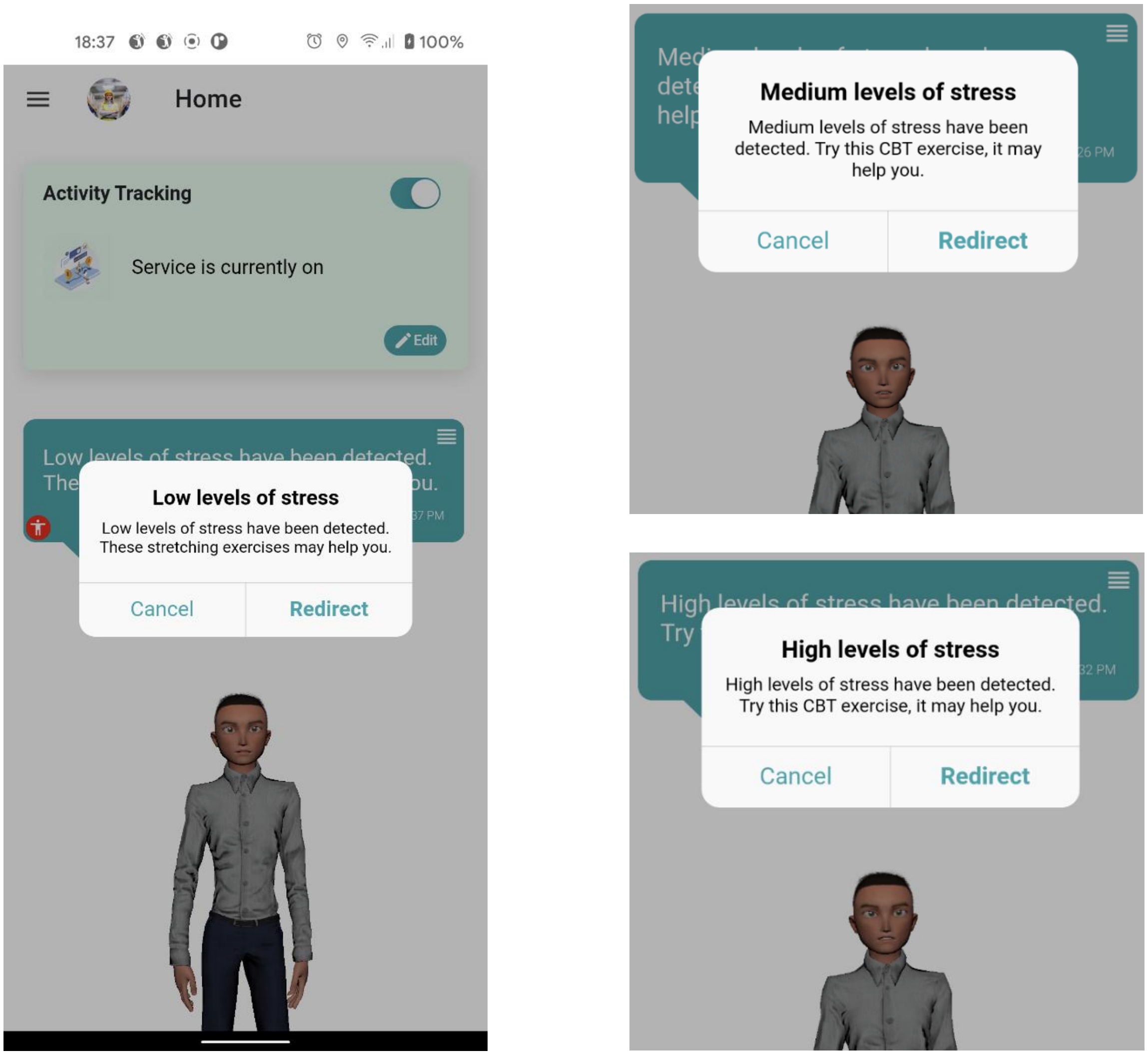


Figure 4 Notifications for stress management by the Virtual Coach

Table 3 Components connected with the application's middleware

Name	Data	Details
Location Tracking Library	Latitude, Longitude, Address	Uses cellular network or GPS. Methods for geocoding and battery optimizations
Activity Tracking Library	Google Activity Recognition: • Walking, Still, Running, In Vehicle, On Bicycle, Tilting CNN model: • Walking, Still, Upstairs, Downstairs Step Counter: • Steps, Activity Intensity, Calories, METs	Googles Activity Recognition: • Orientation/Position independent. Battery optimization implemented. CNN based on UCI HAR dataset: • Orientation/Position depended. Battery restricted. Step Counter: • Provides step counting updates every few seconds.
Location Context API	Locations Category and Subcategory	Custom API that uses OpenStreetMap's files to categorize locations.

Generic Phone Usage Tracking Library	• Charging Events, Battery Levels • Connectivity (Wifi) events • Phone calls (duration, timestamp, anonymized, SHA-256, caller id) • Screen on/off events	Listener service that tracks generic phone usage data.
Keyboard	Press/Release Time, Distance, Deletion	Uses AOSP keyboard files modified to record keystroke Time Sequences.
Self-Reports	Demographics, Questionnaires	Users are prompted to answer self-administered questionnaires periodically and demographic related questions.
Clinical Platform API	Weight, Humidity, Luminescence, Temperature, Blood pressure, Blood Glucose, Questionnaires	Receives IoT derived data and store them under strict protocols. Offers data visualization over aggregated periods and management tools for the administration (e.g. add new Questionnaires).
Samsung Accessory SDK	Raw data : Accelerometer, Gyroscope, Heart rate	Provides methods for raw data retrieval from Samsung Smartwatches.
Samsung Health SDK	Steps, Water intake, Activities, Floors Climbed, Heart rate, Sleep stages	Provides various health / activity related data while using Samsung phone or smartwatch.
Google Fit	Steps, Sleep Metrics, Activities, Heart rate	Provides various health / activity related data while using an Android phone.
Virtual User Model Server (VUM)	Provides: Login credentials Stores: All the non-sensitive data above	All the above data, collected by the middleware, are stored on an SQLite database inside the phone. Periodically, if Wi-Fi is available, these data are sent to VUM. Additionally, methods for a notification broadcast system are provided.

4.2 Implemented ML Algorithms for Stress Detection

Several machine learning based approaches for stress and emotion recognition were developed as part of the Ageing@Work project and evaluated on available benchmark datasets. Those ML-based approaches have been published in (Liakopoulos et al, 2021) and presented also briefly in this and the following Section for completeness. We have tackled the problem of stress detection with two distinct approaches. The first exploits statistical information extracted from biosignal analysis in a feature classification scheme, whereas the second classifies 2D images of spectrograms (from ECG) using a 2D CNN architecture. A schematic diagram of the two approaches is illustrated in the figure below.

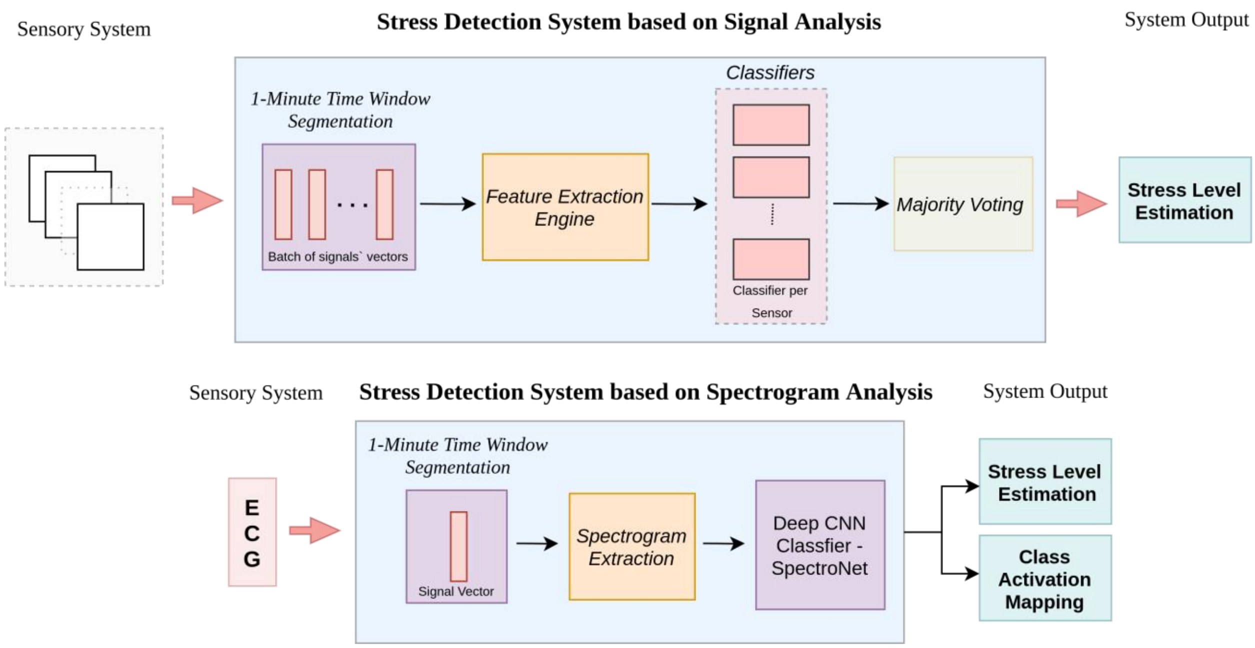


Figure 5 Implemented ML-based stress detection schemes based on feature fusion (top) or spectrogram analysis (bottom)

The preprocessing step is taking as input the raw ECG and EDA and extracts several statistical features for stress detection that have been proposed for physiological signal analysis. In addition to the raw ECG, the heart rate variability signal is calculated using a peak detector on the ECG and is analysed in time and frequency domain. Each signal is split into non-overlapping time windows of 1-minute length used for feature extraction. Data are being scaled using zero-mean and unit variance technique in [0, 1] range.

Typical statistical properties that we include into our feature extraction method for both ECG and produced HRV signals are: mean, standard deviation, median, range of values, min and max values. Moreover, for HRV analysis we incorporate also second order statistics which are based on the difference of two consecutive time distances between heart beats or normal-to-normal (NN) differences described by Camm et al., 1996. More specifically, we detect the NN intervals, i.e. all intervals between adjacent QRS complexes resulting from sinus node depolarizations, and calculate the

1) square root of the mean squared differences of successive NN intervals,
2) number of interval differences of successive NN intervals greater than 50 ms, and
3) standard deviation of NN interval. Moreover, through frequency domain analysis, HRV`s signal energies are measured for specific frequency bands using Power Spectrum Density method. Using these measurements, we obtain the total amount of energy contained in each band and their normalized values.

Moreover, EDA is measured from certain sensors placed on participants' hands. From the obtained signal, we calculate the same statistical properties as used for ECG and store them for further analysis.

Upon feature extraction and fusion, classification of the available data instances is performed to identify the participant's stress condition. For determining the best classification method, several well-known machine learning (ML) algorithms are being deployed and tested on different data subsets. The classification techniques that we select to examine are: support vector machines (SVM) with radial basis function (RBF) kernel, random forest (RF), k-nearest neighbors (KNN) and an artificial neural network (ANN). The specific algorithms are lightweight and operate in near real-time, and are therefore suitable for mobile devices.

In addition to using tubular-like data for classifying stress from non-stress states, we extend our study to 2D data analysis using the spectrograms extracted from the ECG time-windows used in the previous task. With this experiment we aimed to identify and exploit the frequency patterns that relate to the existence of stressful conditions in a participant's recording. For this purpose we developed a custom CNN. The network`s backbone is based on VGG architecture of layer blocking (Simonyan et al., 2014), consisting of a 2D CNN, a Batch Normalization and an Activation Function (Leaky ReLU). The head of the network is a simple global average pooling (GAP) layer followed by a fully-connected layer and a sigmoid and softmax layer. The produced feature maps are classified using the cross-entropy loss function. The GAP layer is also offering valuable information about class activation mapping, i.e. the regions on the image (frequencies localized in time) that are activating the network and determine the final classification decision. An example of a calculated spectrogram and the obtained activation map is shown in the following figures.

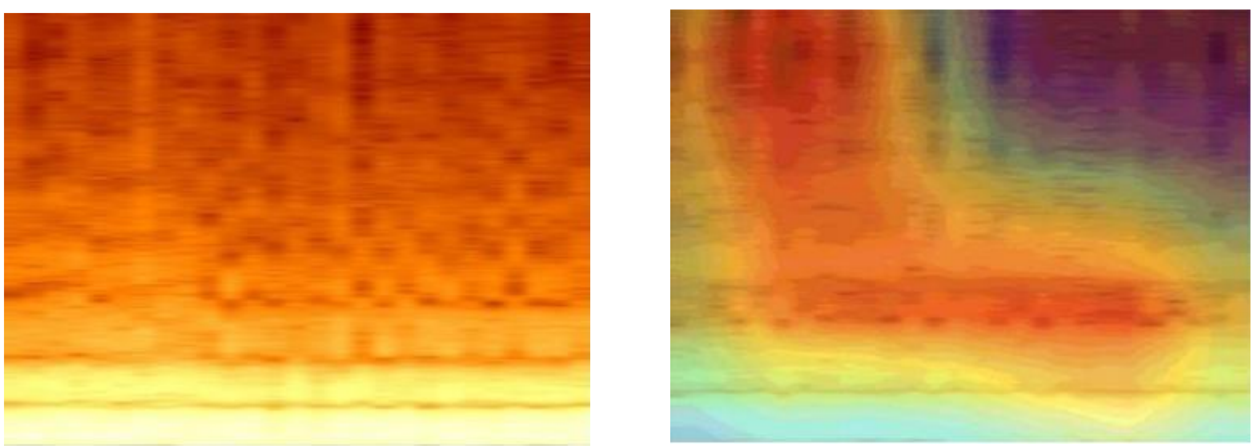


Figure 6 Calculated spectrogram and obtained activation map

During training, a variety of overfit-preventing techniques are being applied, like dropout, structural pruning and early stopping. The network is optimized for mobile platforms (~$700K parameters, 241 MFLOPS for network operations). It has been quantized and pruned using different percentages for each layer.

The method described above follows a feature-level fusion approach where available data from different sensors (Physio, Kinect and Face Camera) are combined, in order to compose a unified vector of attributes for each instance. This data fusion task demands a precise temporal alignment of the recordings of the multiple sensors. In this section we also examine a late integration scheme in which a different classification model is trained with data from each individual sensor and then the final classification is obtained by fusing the various model decisions (Pippa et al., 2018). In this decision-level fusion scheme, we investigate different classification techniques and select the best for each sensor, in order to maximize the final system`s accuracy. In respect to the fusion rule, we apply majority voting (Parhami et al., 1994), i.e. the ensemble chooses a class when it receives the highest number of votes, whether or not the sum of those votes exceeds 50%. Majority voting is the most common fusion technique because it does not require a priori knowledge. Another of its advantages is simplicity of implementation and flexibility in changing or adapting the decision rule with the aim of optimizing performance (Texier et al., 2019).

We first evaluated features from each sensor of the SWELL dataset separately (35 features from heart rate and skin conductance, 41 features from FaceReader and 94 features of body posture) using different classifiers and then fused the multi-modal features in order to examine the benefit of using multiple sensors. The results of the feature-level fusion scheme are shown in the following table. For a single sensor, the best performance is observed when body posture features from Kinect 3D sensor are used. When the multiple multi-modal features are fused, the classification accuracy and F1-score show increased values for all classifiers, with best performing classifier being the ANN (F1-score=95.28% and accuracy=95.07%).

Table 4 Multi-modal feature fusion scheme

Modality (No features)	Dataset	Classifier	Val. Setting	F1-score	Accuracy
Physiology (35 features)	SWELL	SVM	10CV	73.08 %	75.00 %
		RF		78.00 %	80.71 %
		KNN		72.09 %	73.04 %
		ANN		76.61 %	77.55 %
FaceReader (41 features)	SWELL	SVM	10CV	81.31 %	79.53 %
		RF		80.25 %	77.72 %
		KNN		79.71 %	76.61 %
		ANN		82.78%	86.60 %
Body posture (94 features)	SWELL	SVM	10CV	86.09 %	87.79 %
		RF		87.80 %	87.33 %
		KNN		87.94 %	87.89 %
		ANN		90.52 %	90.38 %
Physiology, FaceReader & Body Posture (171 features)	SWELL	SVM	10CV	94.57 %	94.65 %
		RF		92.65 %	92.65 %
		KNN		94.95 %	94.45 %
		ANN		95.28 %	95.07 %
2D Spectrogram	WESAD	CNN	Random Split	96.73 %	96.79 %
		CNN	LOSO	79.43 %	82.35 %

Moreover, we evaluated a CNN on 2D spectogram images using the WESAD dataset. Although the results are not directly comparable due to differences in the incorporated datasets, they indicate that the proposed CNN classifier based only on ECG can achieve better or similar performance with the multi-sensor features classified by standard ML techniques (evaluated by 10CV). This might be attributed to the fact that the spectograms obtained from raw ECG retain the whole frequency content and its variation over time.

The next set of experiments targeted the decision-level fusion on the SWELL dataset. In this case the predictions of the classifiers performing best for each individual sensor were identified (from the experiments shown table) and combined. The selected sets of sensors and classifiers for decision-level fusion were the following:

- Physio Sensor: Random Forest (80.71% accuracy)
- Face Camera: ANN (86.6% accuracy)
- Kinect Sensor: ANN (90.38% accuracy)

Fusion of prediction is performed by majority voting and the results are illustrated in the .

Table 5 Stress detection performance for decision-level fusion

Modality (#features)	Dataset	Fusion Technique	F1-score	Accuracy
Physiology, FaceReader & BodyPosture	SWELL	Majority Voting	97.69%	97.64%

Moreover, in Table 6, we are comparing our method with the original SWELL study by Koldijk et al., 2018. Using our data preprocessing method and the model hyperparameters' optimization, we achieve higher accuracy for all modalities.

Table 6 Accuracy in comparison with the benchmark method on SWELL dataset

Modality	Method	Koldijk et al. [29]	Ours
Physiology	SVM	64.1%	80.71% (RF)
FaceReader	SVM	75.4%	86.6% (ANN)
BodyPosture	SVM	83.4%	90.38% (ANN)
Physiology, FaceReader & BodyPosture	SVM Majority	89.3% No score	94.65% (SVM) 97.69%

4.3 Implemented ML Algorithms for Emotion Recognition

For emotion recognition we implemented a CNN architecture that exploits facial expressions in 2D images captured from a mobile phone's camera. The architecture is a modified version of an architecture used in the FER2013 challenge. The architecture of the network is shown in Figure 7. We used ReLU as activation function and binary cross entropy as loss function.

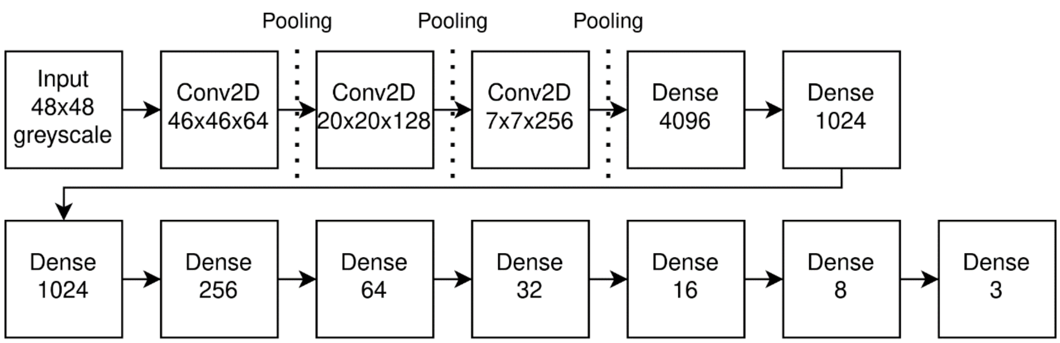


Figure 7 Implemented CNN architecture for facial emotion recognition

Changes in lighting which can happen from frame to frame, along with image blurriness from movement and inaccurate face detection can unintentionally influence the network's predictions. In order to circumvent this and to increase the overall robustness of the classifier, we use the network in batches of images, evaluate every image in the batch and finally select the most dominant emotion displayed within this time window. The most dominant emotion is selected by simple voting of each frame, weighted by the confidence of the face detection algorithm.

All steps of the emotion recognition pipeline including the integration of the network into the Android platform are illustrated in the following figure and detailed next.

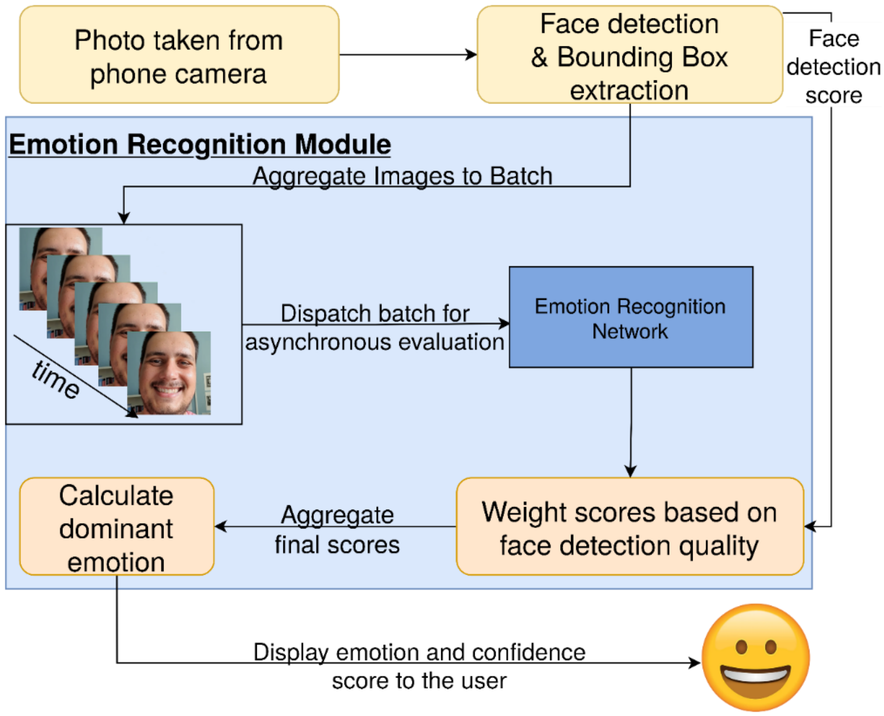


Figure 8 Face detection and emotion recognition pipeline

1. The existing repository (https://github.com/amineHorseman/facial-expression-recognition-using-cnn} was ported to Tensorflow 2.0 and Keras framework in order to work on latest standards and export compatible TensorFlow Lite (an optimized FlatBuffer format) models for the Android platform.

2. An Android application was developed to detect the face of the user and to extract the surrounding bounding box utilizing the face detection of Camera2 API from Google, which is compatible with most Android smartphones. The application also assigns a confidence score **W** to each detected face. This confidence score assumes a floating point value between 0 and 1.

3. The cropped image that contains only the face is then converted to grayscale and it is reshaped to 48*48 in order to be compatible with the CNN input size.

4. Inference is performed using batches of 10 images (as explained in the previous paragraph) which are dispatched asychronously to the neural network. The CNN detects the emotion captured per frame and yields a decision score p_c for each emotion class **C**. The emotion with the highest score from the network is later multiplied by the face detection confidence score to receive the final weighted value of the frame's most dominant emotion.

5. After all images in the batch have been evaluated, scores of the same emotion class are added together across frames and the class with the highest sum is used to determine and display the final emotion to the user.

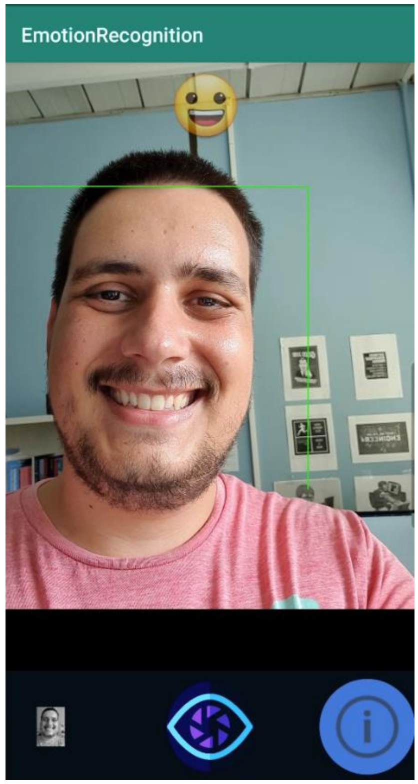


Figure 9 User interface of Android application for face detection and emotion recognition

An example of the execution of our application is visualized in the following figure. The detected face is marked by a green bounding box and the corresponding cropped grayscale image patch (extracted from the bounding box) is illustrated on the bottom left part of the figure. This image patch is subsequently scaled and introduced as input to the CNN.

The classification results of the three emotional states are illustrated in the confusion matrix in the following confusion matrix. The results indicate that the network recognizes the *positive* and *negative* emotional states better than the *neutral* state. This was not surprising for us, as the facial features of those emotions have generally more discriminative power compared to the *neutral* class.

5. AW Framework for Behaviour and Stress Identification

5.1 Online Study based on Self-reports

Preliminary rules were formulated to identify stress in individuals. These rules were developed leveraging existing literature, however, as they utilized data from numerous self-reported sources we sought to provide some empirical support to show they are indeed correlated; the final rules are outlined in the following subsection. This short study was conducted to test the relationship between the various questionnaire variables collected by the project, as well as potential extra questionnaires to use for stress and mental health detection.

Subjects (n = 100) were sampled from Prolific (www.prolific.co), with the constraints that they were from UK or central European countries, spoke English, and were between 40 and 65 in order to reflect the demographics of the Aging at Work population.

Variables in the online study are summarized below:

- Monthly Questionnaire (MQ) contains ten relevant questions: 4 identify stress (Q_Stress), 2 identify locus of control (Q_Control), 3 identify support network (Q_Support), and 1 is quality of life (QoL).
- MoodZoom (Tsanas et al., 2016) is asked daily in the app: 6 questions about subjects' current mood.
- Big 5 Personality Trait (John and Srivastava, 1999) questions are answered at onset of app usage. We test all 5 traits however only Conscientiousness and Neuroticism are significantly historically associated with stress.
- PHQ-9 (Kroenke and Spitzer, 2001) is asked biweekly as a measure of depression and mental stress: a summed score of the values which is associated with severity of depression.
- K10 (Andrews and Slade, 2001; Kessler and Mroczek, 1994; Kessler et al., 2000) is a proposed additional questionnaire to be used when high ongoing stress or depression levels are detected, and it identifies if someone has a mental distress disorder. We test if it is necessary or if PHQ-9 is sufficient.

All correlation tests mentioned below refer to Pearson correlations with a 95% confidence interval unless otherwise stated.

5.1.1 Monthly Questionnaire and Stress Identification

Table 7 summarizes the variables (selected from the questionnaires) that are used for stress assessment. After the online study, the questions which lowered the internal consistency of measure or lowered the correlation to the stress measure, were dropped from the rules, as explained next.

Table 7 - Questions from Monthly Health Survey relating to stress

27. How often do you have negative feelings such as blue mood, despair, anxiety, depression?
28. How well are you able to concentrate?
29. How well are you able to handle multiple things at the same time?
30. How satisfied are you with your sleep?
31. How often have you been stressed during the last 4 weeks?
32. I am confident in my ability to solve problems that I might face in life
33. How satisfied are you with your personal relationships?
34. How satisfied are you with the support you get from your friends?
35. How often do you socialize with friends / neighbours / relatives?
36. How would you rate your quality of life?

First, we tested for the internal consistency of the questions from MQ (Table 7) used to calculate stress: questions 27, 28, 30, and 31. The four questions had a Cronbach's alpha of 0.769 which is acceptable to combine their values into a summed score called Q_Stress. Each of these questions were positively correlated to one another, ranging from a rho of 0.37 up to 0.67 (all *p value's* < 0.001).

Next, we tested for the internal consistency of the questions from MQ used to calculate control, as this can have a relationship to how stress is managed (Q_Control). The two items had a Cronbach's alpha of 0.685 which borders on acceptable; they were significantly correlated to one another (rho = 0.52, $p < 0.001$). We proceeded to compare the correlation of Q-Stress and Q_Control, and as expected they were negatively correlated (rho = -0.5, $p < 0.001$). We also compared each question's correlation to Q_Stress individually, question 29 had a rho of -0.37 ($p = 0.0001$) and the other, question 32, had a rho of -0.5 ($p < 0.0001$). Given that the internal consistency (alpha = 0.685) was questionable and one of the questions alone had a stronger and more significant effect (in the expected direction), we proposed keeping that question alone as the Q_Control variable and dropping the other (keep question 32 and drop 29).

Proceeding onto the support network questions (33, 34, and 35) of MQ, we tested their internal consistency and found a Cronbach's alpha of 0.721 which is acceptable. By further testing pairwise correlations we found a strong correlation between two of the variables – question 33 and 34 (rho = 0.65, $p < 0.001$). However, when testing question 35, we found a weaker correlation with question 33 (rho = 0.32, $p = 0.001$) and question 34 (rho = 0.43, $p < 0.001$). Question's 33 and 34 relate to satisfaction of relationships with family and friends, whereas question 35 relates to frequencies of family/friend interactions. Given that people may have different frequencies that leave them satisfied with a feeling of support, we proposed dropping question 35 from this combined variable (Q_Support). Doing so increased the Cronbach's alpha of Q_Support to 0.784, and this variable was negatively correlated (as expected) to Q_Stress (rho = -0.298, $p = 0.003$).

We also tested the single QoL question from MQ, which was negatively correlated to Q_Stress as was expected (rho = -0.5, $p < 0.001$).

5.1.2 Daily Mood Assessment

Moving on to the MZ questions we test each one's correlation to Q_Stress. The questions around elation and anger are not significantly correlated to stress. Anxiousness is correlated (rho = 0.44, $p < 0.001$) , as is sadness (rho = 0.45, $p < 0.001$) and irritability (rho = 0.25, $p = 0.01$) to a lesser extent. Additionally, energetic is significantly negatively correlated (rho = -0.36, $p < 0.001$). As such, anxiousness and sadness may be useful indicators to include after the pilot phase, however, consideration is to be made regarding the frequency of these measures, as they are daily measures and an overreaction to a bad day should be avoided.

5.1.3 Personality Traits

As per previous literature we find a significant correlation between numerous personality traits and stress. Conscientiousness and Q_Stress have a correlation with rho = 0.47 ($p < 0.001$) as does Neuroticism and Q_Stress (rho = -0.61, $p < 0.001$). Significant correlations here were expected, however, the direction of the latter relationship is unexpected. Openness and Agreeableness do not have any significant correlation, yet Extraversion is weakly correlated (rho = 0.33, $p < 0.001$). Given the unexpected direction of Neuroticism we suggest removing it as an influence on the rules used for stress detection until enough significant real life data can be gathered to determine its relationship to stress in this system.

5.1.4 Depression and Mental Stress

Continuing on to the PHQ-9 questions, we first check for internal consistency even though it is an established questionnaire. As expected we find a strong Cronbach's alpha of 0.884. We repeat the same process for the K10 questionnaire and find an similar level of internal consistency between the questions with a Cronbach's alpha of 0.94.

Testing the correlation between the MQ measure of stress identification (Q_Stress) and the PHQ-9 depression disorder measure we find a strong significant correlation (rho = 0.77, $p < 0.001$). We also find a strong correlation (rho = 0.72, $p < 0.001$) between Q_Stress and the K10 which identifies mental stress disorders. Furthermore, we evidence a significant correlation between the PHQ-9 value and K10 value (rho = 0.86, $p < 0.001$). This implies the inclusion of the PHQ-9 into the stress rule identification and severity is a beneficial action, and negates the necessity of adding the K10 questionnaire to the process.

5.2 Final Rule-Based Decision Making for Stress Assessment

Classification of behavior and emotion and affective traits is difficult because of the subject-dependent nature of the physiological stress response and lack of a stress reference, the influence of physical activity and the lower signal quality due to motion artifacts. We therefore adopted a rule-based decision approach that is more robust to the previous challenges and requires less data to optimize its internal parameters. The initial rules for stress detection leveraged literature on the topic and have since been slightly adjusted based upon the information from the study above.

Currently the primary source of stress detection is the monthly questionnaire, with the individuals QoL and the four stress related questions. In addition to the four questions relating to stress, the question regarding locus of control[1] and the two questions on support network[1] are considered as an influences on the individuals capacity to handle high levels of stress. The personality trait information is also integrated into the rules, not in regard to the identification of stress but in combination with the other information for the potential of an individual to effectively handle stress. There is an existing dataset of over 14000 observations which summaries the average scores on each personality trait (median and range also) of individuals over 40 and under 60 (John and Srivastava, 1999). Using this we create bands of average, high and low levels of each trait. Originally, we used categorisation of personality for conscientiousness and neuroticism, however, the study above displayed contradictory results to the literature regarding neuroticism so we have dropped it for now as an influence on stress. Whilst this was surprising it is not unexpected, as different situational stressors can affect the relationship (Sur and Ng, 2014). After the pilot phase it is also possible to include the MoodZoom answers in a similar capacity, to see if self-reported emotional states and those identified through wearable technology align daily and in an aggregated monthly manner.

The aforementioned data form rules for identification of stress levels. If a medium level of stress is identified for that month, then the individual is sent a small cCBT exercise, specified in the following section, to assist in reducing their stress. If a higher level of stress is identified, then the latest PHQ-9 data is used to assess the possible severity of mental distress or depression. Based up the results the individual would be sent either the usual cCBT exercises plus an additional behavioural experiment exercise (see section below) or recommended to contact a mental health professional. If an individual was recommended to the cCBT exercises in multiple consecutive months (i.e. the exercises are not effective in lowering their stress) then they would be recommended to seek out a mental health professional.

In more detail, the indicators that are extracted to calculate the final score and recommend actions are:

- Conscientiousness (Extracted from the initial questionnaire “the Big Five Inventory”)
- Support (Extracted from the monthly questionnaire “VUM”)
- Control (Extracted from the monthly questionnaire “VUM”)
- Stress (Extracted from the monthly questionnaire “VUM”)
- Quality of Life (Extracted from the monthly questionnaire “VUM”)

The final score corresponds to the level of action to be taken, with low score corresponding to no action (everything is fine), mid scores corresponding to some of the CBT exercises, and high scores flagging the need for extra evaluation.

All the indicators have a direct or inverse relation with the final score:

- A higher *Conscientiousness* corresponds to higher final scores (direct)
- A higher *Support* corresponds to lower final scores (inverse)
- A higher *Control* corresponds to lower final scores (inverse)
- A higher *Stress* corresponds to higher final scores (direct)

[1] One question each was dropped from locus of control and support networks after the online study.

- A higher *Quality of Life* corresponds to lower final scores (inverse).

In the case the user is flagged for extra evaluation, through the *PHQ-9 questionnaire*, depending on the answers either no extra actions are taken, extra CBT exercises are recommended, or the user is referred for professional help.
Finally, if the user is flagged for extra evaluation repeatedly, but there is no improvement, the action from the *PHQ-9 questionnaire* is increased even if the previous threshold was not reached.

5.3 Integration within the AW Middleware and Workers' Dashboard

The stress identification pipeline can be summarized in the following steps:

1. The worker answers the within-app questionnaires.
2. The questionnaires answers are sent to the server where they are saved.
3. An external service written in node, using the existing REST API, requests periodically the aforementioned questionnaires, which consist of "the Big Five Inventory", "VUM" and "PHQ-9", including the answers and scores for each user.
4. A score, estimating the stress levels of the user, is being calculated using a combination of questions from all the different questionnaires, as described in Section 5.2. The rules are applied by a lookup table depending on the input indicators.
5. The computed score is returned into the Ageing@Work server through the same REST API and is stored into the database. Additionally, depending on the score, an upcoming notification event will be stored, with the appropriate action, as described in Section 5.4.
6. A service, within the Ageing@Work application, periodically requests from the server the upcoming action events. Whenever a pending action is received, a notification is displayed which, if clicked, redirects the user to the home page, where the avatar informs the user about the system's suggested action to reduce her/his stress level.

5.4 CCBT Content for Ageing@Work

The following cCBT exercises are a frequent and introductory part of larger online courses in cognitive behavioural therapy. They can be used to help reduce stress and provide tools for managing stress. However, it is not part of this deliverable to create or supply a clinical mental health program, which is why if high levels of stress are detected and not easily manageable then we recommend seeking professional help. The following exercises are delivered to the user via the virtual agent in the form a link to an online document which they can use or download.

5.4.1 Exercises for Medium Stress Level

Solution-focused CBT:

We highly recommend you take the CBT exercise in the absence of noise and/or other external distractions, in other words, you should choose a time and place where you are less likely to be interrupted. Also, please refrain from other activities such as listening to music, browsing the internet, chatting with family members and/or friends.

Instructions:

In the following exercise(s), you will be asked to close your eyes and vividly imagine yourself thinking, feeling, and behaving in ways that would demonstrate that you were resolving your problems.

Cognitive restructuring:

Think about a recent recurring problem that you felt some kind of emotional upset, such as anxiety, sadness, or anger.

Thought diary

Please describe the situation which gave rise to your identified automatic negative thought.	Please describe how this situation made you feel emotional.	Please write down your identified automatic thought.
To what extent do you believe this thought?		%

For and against

Please use this column to list all the evidence which supports this thought	Please use this column to list all the evidence which goes against this thought

Alternative thought

Is there an alternative thought or explanation which fits better with the evidence presented above? If so, write it in the box below	
Thought:	To what extent do you believe this new thought? %

Strength-focused solution

Description of time when the problem was not happing or less severe. What was different?	What did you do to make this happen? What kind of actions did you take (or not take).	What strength were you possibly using here?

5.4.2 Additional Behavioral Exercise for High Stress Level

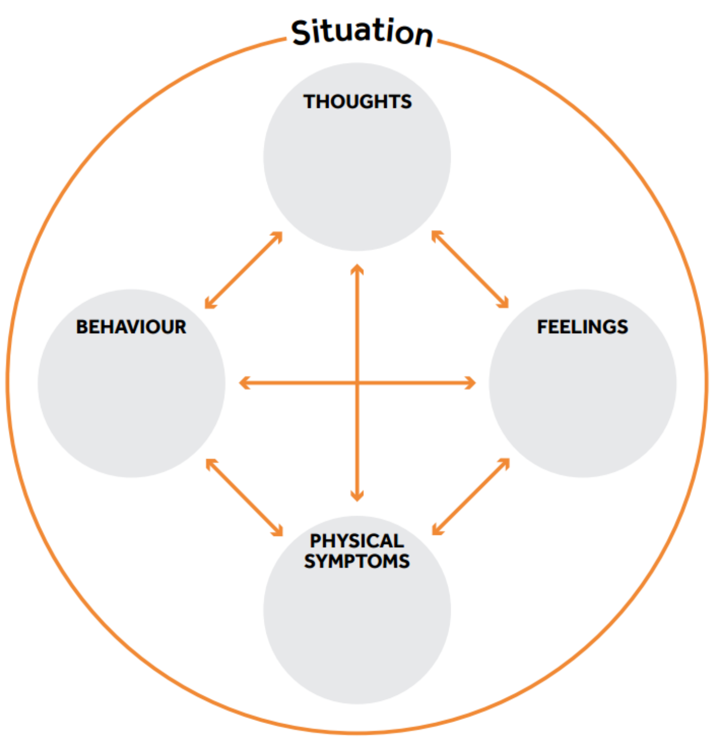
The same text and exercises as section 5.5.1 would be used, plus the addition of the below paragraph – awareness text – between the ‘solutions based CBT’ and ‘introduction’ paragraphs. The behavioral exercise below would then follow the exercise from section 5.5.1.

Awareness text:

Feeling happy is not a ‘default’ human emotion, it is perfectly normal for us to experience a range of emotions/stress and for our mood to fluctuate over time. Sometimes, low levels of stress may help us to improve performance. For instance, feeling stressed about a presentation to a more senior level of staff may lead to more thorough preparation. There may have been a lot of occasions (e.g., making an important phone call and visiting the dentist) that made you feel stressed. Under some sort of stress is entirely normal—we are not supposed to always feel relaxed and happy.

Behavioural experiment:

Please think about the problem again and complete the following table.



Four steps: planning the experiment; carrying out the experiment; observing the results, and reflecting on what the results mean

Plan and carry out the experiment

<table>
<tr><td colspan="2" rowspan="3">Specific cognition/belief to be targeted:</td><td colspan="2">What are you going to do to test out the validity of this cognition/belief?</td></tr>
<tr><td>Where are you going to do it?</td><td></td></tr>
<tr><td>When are you going to do it?</td><td></td></tr>
<tr><td>How much do you believe this thought? (0-100%)</td><td></td><td>Who will be involved?</td><td></td></tr>
<tr><td colspan="2">What do you think will actually happen?</td><td colspan="2">What else might happen?</td></tr>
<tr><td>How likely do you think this is to happen? (0-100%)</td><td></td><td>How likely do you think this is to happen? (0-100%)</td><td></td></tr>
<tr><td colspan="4">What may stop you from doing this behavioural experiment? What things might get in your way?</td></tr>
<tr><td colspan="4">How will you overcome these potential difficulties?</td></tr>
</table>

Reflect on the behavioural experiment

<table>
<tr><td colspan="2">Write your original prediction here:</td><td colspan="2">What actually happened?</td></tr>
<tr><td>How likely did you originally feel this was to happen? (0-100%)</td><td></td><td colspan="2"></td></tr>
<tr><td colspan="4">What was the original belief/cognition that you were testing with this Behavioural Experiment?</td></tr>
<tr><td>How much did you believe this thought before the experiment? (0-100%)</td><td></td><td>How much do you believe this thought now? (0-100)</td><td></td></tr>
<tr><td colspan="4">What have you learnt as a result of this Behavioural Experiment?</td></tr>
<tr><td colspan="3">How might you change your original thought to account for this new learning?</td><td>How much do you believe this thought now? (0-100)</td></tr>
<tr><td colspan="4">Based on this new thought, what may you do differently in the future?</td></tr>
<tr><td colspan="4">Are there any further behavioural experiments that may be useful?</td></tr>
</table>

5.5 Future Work

The current implementation of the Ageing@Work application supports data collection of activity information from wearable devices, smartphones, and third-party apps, keystroke typing data, location information, data from medical and environmental IoT devices and mood self-reports. All of these data have been reported, each individually but also in combinations, to be suitable for monitoring and classification purposes among healthy controls and patients with mental disorders. Bearing in mind that depression is considered the most common mental illness with an estimated share of 25–30% (D. Vigo et al., 2016), we aim to study if the multimodal data that are being collected can yield correlations with the stress level estimated by expert-based rules (Section 5.2) that use self-reported standards. Furthermore, since studies on personality traits do not need medical examination and are suitable for the personalization of the Virtual Coach, which is already using the extroversion score as a mean for appearance modifications, we may consider evaluating our future dataset using them. It should be noted though, that since the outcomes from personality traits test yield many different scales˙ of extroversion, agreeableness, conscientiousness, neuroticism and openness to experience, a larger participant pool would be required for any relative results, in contrast with depression scales that can be examined as a binary classification problem or multi class problem if enough participants are recruited. Another plausible study could be performed by exploiting the self-reports from the Mood Zoom Questionnaire (A Tsanas et al, 2016), which has demonstrated promising adherence results and strong correlation with established clinical questionnaires.

In this direction, when enough data will have been collected from the pilot studies, we aim to perform data analysis between all the different modalities – activity, location, typing and generic behaviour logs (calls, connectivity, charging, breaks@work, stairs, leisure walks, etc.) – and the results obtained from the self-reported questionnaires. In addition, predictive ML models could be developed to estimate worker's personality traits and depressive tendency. More specifically, from the collected data, features will be extracted, and correlation analysis will be performed with the answers/scores of the self-reported questionnaires. Additionally, the scores of the questionnaires, if sufficient in number, will be used as ground truth to train ML models for each different modality. The outcomes of the different ML models could be combined to increase our system's predictive performance. The Figures below depict few of the possible features for each tracking modality.

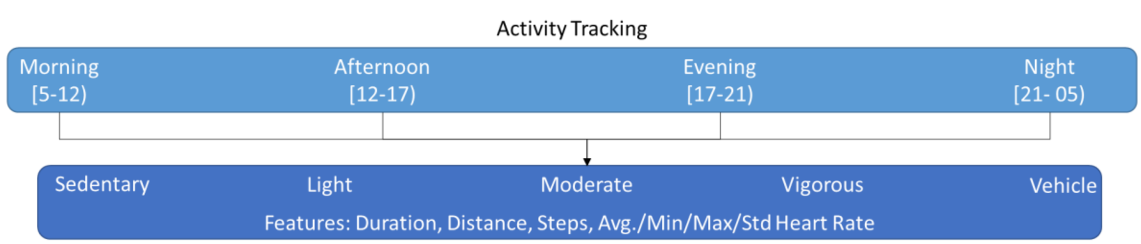


Figure 10 Activity tracking features

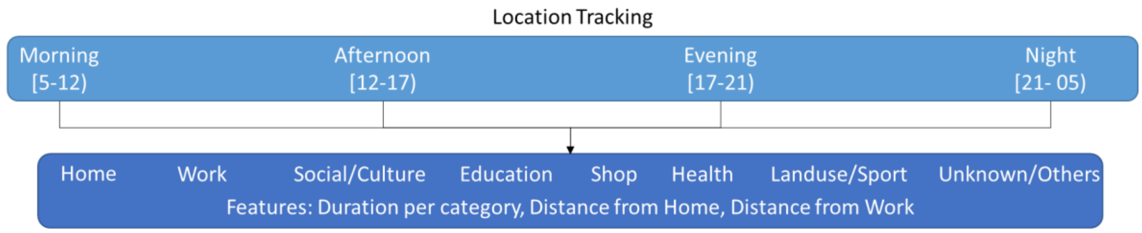


Figure 11 Location tracking features

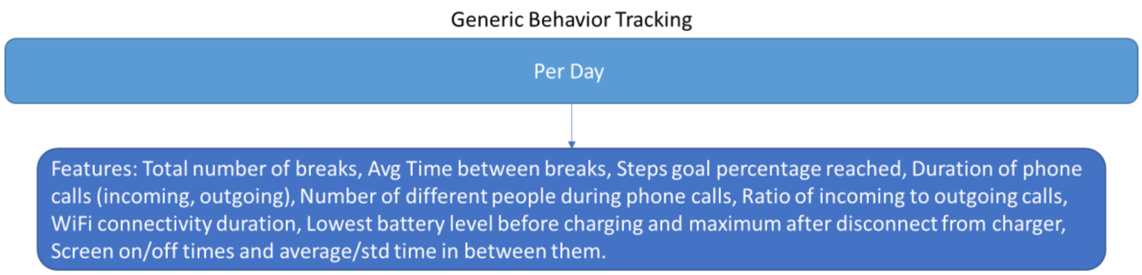


Figure 12 Generic behavior tracking features

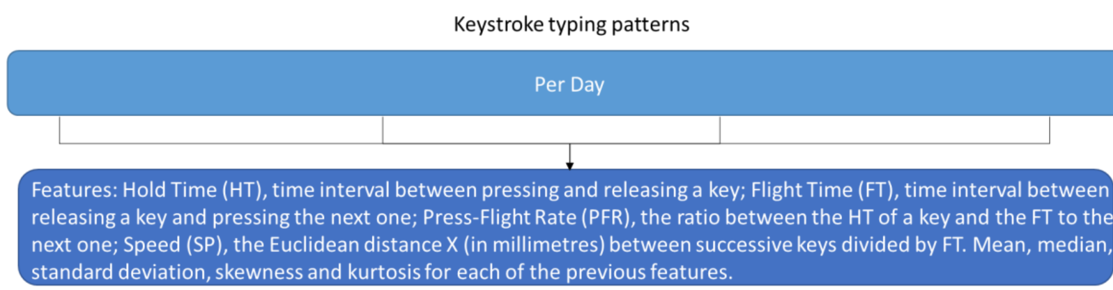


Figure 13 Keystroke typing pattern features

6. Conclusions

Identifying all factors which collectively contribute to stress and negative emotions is of ultimate importance for accurate recognition of the psychological state of the individual. However, when faced with the task on identifying stress and other behaviours using technological means, there are multiple challenges that should be handled, in respect to both data collection and analysis, namely:

- Requirement for unobtrusiveness and continuous, ambulatory monitoring
- Equipment should not interfere with the job activities, i.e. should not extend beyond the clothes
- Daily working settings, i.e. context, is unknown
- Technical limitations of wireless wearable devices when moving out of the lab and into ambulatory settings, especially into locations with no wireless coverage at all

Moreover, the recognition of affective states is especially challenging due to subject-dependent nature of facial patterns, emotional expression, and physiological stress response, while the classification task is also challenged by the lack of a stress and emotion reference.

Our more advanced data-driven approach for recognition of stressful conditions is mainly applicable for office environments and uses features from ECG and EDA biosignals analyzes facial expression patterns when the emotional state is sought. The results on previous (open source) data showed that our classification models outperform previous work of Koldijk et al., 2018 on the same benchmark dataset (Koldijk et al., 2014). Moreover, the proposed CNN-based spectogram analysis revealed that the temporal variation of the spectrum of frequencies seems to have high discriminative power for stress identification.

Such ML-based approaches, although seem like a promising approach, they have some limitations:

1) They need a large amount of annotated data for model training, which is under the current settings not available
2) Continuous tracking is required that is computationally intensive and data demanding
3) Most importantly, they provide useful insights only on instantaneous stress and emotional levels, while offer only a small value on the long-term stress and worker's fatigue. As discussed by Powell and Enright (1990), anxiety and stress are a normal part of daily life and can be managed adequately by an individual, whereas prolonged or more serious stress is what requires a form of intervention or help. Consequently, daily fluctuations in mood or stress are not an adequate measure of serious prolonged stress, and the potential volatility in these measures caused by an isolated bad day or a pressure situation at work could lead to unnecessary flagging of stress intervention.

For the aforementioned reasons and limitations of instantaneous stress and emotions monitoring with the use of sensors, IoT and machine learning, we have finally integrated a rule-based approach using variables from self-reports for stress assessment. The incorporated methodology does not rely on daily questionnaires to identify stress, but rather utilises validated questionnaires over a longer period. Potentially in the future we may use changes in aggregated emotional levels over monthly periods.

The performed user study showed that the questions use to identify stress are internally consistent and the additional interaction variables – such as locus of control, support networks, personality traits, and the PHQ9 – are sufficiently correlated to our measure of stress to help form a measure of the severity of the stress level. In the future, if more data become available, we aim to combine subjective measures (i.e., data obtained from self-reported surveys) with objective measures (i.e., data obtained from wearable devices or keystroke features) when detecting stress and evaluate the consistency between the two measurement types.

7. References

Aarøen, J. A. (2013). Using Wearable Technology in the Treatment of OCD, 1–10.

A. Kolakowska, “Recognizing emotions on the basis of keystroke dynamics,” Proc. - 2015 8th Int. Conf. Hum. Syst. Interact. HSI 2015, pp. 291–297, 2015.

A. Mehrotra, R. Hendley, and M. Musolesi, “Towards multi-modal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction,” UbiComp 2016 Adjun. - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., pp. 1132–1138, 2016.

A. Sano and R. W. Picard, “Stress recognition using wearable sensors and mobile phones,” Proc. - 2013 Hum. Assoc. Conf. Affect. Comput. Intell. Interact. ACII 2013, pp. 671–676, 2013.

A. Triantafyllidis, C. Velardo, D. Salvi, S. A. Shah, V. Koutkias, and L. Tarassenko, “A Survey of Mobile Phone Sensing, Self-reporting and Social Sharing for Pervasive Healthcare,” IEEE J. Biomed. Heal. Informatics, pp. 1–1, 2015.

A. Tsanas et al., “Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder,” J. Affect. Disord., vol. 205, pp. 225–233, Nov. 2016.

Ahola K, Honkonen T, Isometsa E, et al. Burnout in the general population. Results from the Finnish Health 2000 Study. Soc Psychiatry Psychiatr Epidemiol 2006;41:11–17.

Ainsworth, Barbara E., et al. "Guide to the assessment of physical activity: Clinical and research applications." Circulation 128 (2013): 2259-2279.

Åkerstedt, T., Orsini, N., Petersen, H., Axelsson, J., Lekander, M., & Kecklund, G. (2012). Predicting sleep quality from stress and prior sleep – A study of day-to-day covariation across sixweeks. Sleep Medicine, 13(6), 674–679. https://doi.org/10.1016/j.sleep.2011.12.013

Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler psychological distress scale (K10). Australian and New Zealand journal of public health, 25(6), 494-497.

Attwood, M., Meadows, S., Stallard, P., & Richardson, T. (2012). Universal and targeted computerised cognitive behavioural therapy (Think, Feel, Do) for emotional health in schools: Results from two exploratory studies. Child and Adolescent Mental Health, 17(3), 173–178. https://doi.org/10.1111/j.1475-3588.2011.00627.x

Azizan, Illiana & Fatimah, K.. (2020). Facial Emotion Recognition: A Brief Review.

Bakker, D., Kazantzis, N., Rickwood, D., & Rickard, N. (2018). A randomized controlled trial of three smartphone apps for enhancing public mental health. Behaviour Research and Therapy, 109(February), 75–83. https://doi.org/10.1016/j.brat.2018.08.003

Bushman, Barbara A. "Wouldn't you like to know: how can I use METs to quantify the amount of aerobic exercise?." ACSM's Health & Fitness Journal 16.2 (2012): 5-7.

C. J. Soto and O. P. John, "Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS," J. Res. Pers., vol. 68, pp. 69–81, Jun. 2017.

Cheng Y, Chen IS, Chen CJ, Burr H, Hasselhorn HM. The influence of age on the distribution of self-rated health, burnout and their associations with psychosocial work conditions. J Psychosom Res 2013;74:213–220.

D. A. Rohani, M. Faurholt-Jepsen, L. V. Kessing, and J. E. Bardram, "Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: Systematic review," JMIR mHealth uHealth, vol. 6, no. 8, 2018.

D. Iakovakis, S. Hadjidimitriou, V. C.-S. reports, and undefined 2018, "Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage Parkinson's disease," nature.com, Accessed: Nov. 18, 2020. [Online]. Available: https://www.nature.com/articles/s41598-018-25999-0/.

D. Vigo, G. Thornicroft, and R. Atun, "Estimating the true global burden of mental illness," The Lancet Psychiatry, vol. 3, no. 2. Elsevier Ltd, pp. 171–178, Feb. 01, 2016.

Dillon, A., Kelly, M., Robertson, I. H., & Robertson, D. A. (2016). Smartphone Applications Utilizing Biofeedback Can Aid Stress Reduction. Frontiers in Psychology, 7(JUN). https://doi.org/10.3389/FPSYG.2016.00832

Dzedzickis A, Kaklauskas A, Bucinskas V. Human Emotion Recognition: Review of Sensors and Methods. Sensors (Basel). 2020;20(3):592. Published 2020 Jan 21. doi:10.3390/s20030592

E. A. Sağbaş, S. Korukoglu, and S. Balli, "Stress Detection via Keyboard Typing Behaviors by Using Smartphone Sensors and Machine Learning Techniques," J. Med. Syst., vol. 44, no. 4, 2020.

E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. J. Oedegaard, and J. Tørresen, "Mental health monitoring with multimodal sensing and machine learning: A survey," Pervasive and Mobile Computing, vol. 51. Elsevier B.V., pp. 1–26, Dec. 01, 2018.

F. Hussain et al., "Passive Sensing of Affective and Cognitive Functioning in Mood Disorders by Analyzing Keystroke Kinematics and Speech Dynamics," Springer, Cham, 2019, pp. 161–183.

Fuller-Tyszkiewicz, M., Richardson, B., Little, K., Teague, S., Hartley-Clark, L., Capic, T., … Hutchinson, D. (2020). Efficacy of a smartphone app intervention for reducing caregiver stress: Randomized controlled trial. JMIR Mental Health, 7(7). https://doi.org/10.2196/17541

Hilvert-Bruce, Z., Rossouw, P. J., Wong, N., Sunderland, M., & Andrews, G. (2012). Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behaviour Research and Therapy, 50(7–8), 463–468. https://doi.org/10.1016/j.brat.2012.04.001

I. Barnett, J. Torous, P. Staples, L. Sandoval, M. Keshavan, and J.-P. Onnela, “Relapse prediction in schizophrenia through digital phenotyping: a pilot study,” Neuropsychopharmacology, vol. 43, no. 8, pp. 1660–1666, 2018, [Online]. Available: https://doi.org/10.1038/s41386-018-0030-z.

J. Zulueta et al., “Predicting mood disturbance severity with mobile phone keystroke metadata: A biaffect digital phenotyping study,” J. Med. Internet Res., vol. 20, no. 7, pp. 1–10, Jul. 2018, Accessed: Oct. 16, 2020. [Online]. Available: https://www.jmir.org/2018/7/e241/.

J.-P. Onnela and S. L. Rauch, “Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health,” Neuropsychopharmacology, vol. 41, no. 7, pp. 1691–1696, 2016.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research (Vol. 2, pp. 102–138). New York: Guilford Press.

K. Kroenke, R. L. Spitzer, and J. B. W. Williams, “The PHQ-9: Validity of a brief depression severity measure,” J. Gen. Intern. Med., vol. 16, no. 9, pp. 606–613, 2001, Accessed: Dec. 10, 2020. [Online]. Available: /pmc/articles/PMC1495268/?report=abstract.

K. Kroenke, R. L. Spitzer, J. B. W. Williams, and B. Löwe, “The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: A systematic review,” Gen. Hosp. Psychiatry, vol. 32, no. 4, pp. 345–359, Jul. 2010.

K. M. Anil Kumar, B. R. Kiran, B. R. Shreyas, and S. J. Victor, “A Multimodal Approach to Detect User’s Emotion,” Procedia Comput. Sci., vol. 70, pp. 296–303, Jan. 2015.

Kang, S. G., Kang, J. M., Cho, S. J., Ko, K. P., Lee, Y. J., Lee, H. J., … Winkelman, J. W. (2017). Cognitive behavioral therapy using a mobile application synchronizable with wearable devices for insomnia treatment: A pilot study. Journal of Clinical Sleep Medicine, 13(4), 633–640. https://doi.org/10.5664/jcsm.6564

Kessler R, Mroczek D. (1994). Final Versions of our Non-Specific Psychological Distress Scale. Ann Arbor(MI): Survey Research Center of the Institute for Social Research, University of Michigan.

Kessler RC, Andrews G, Colpe L, et al. (2000). Short Screening Scales to Monitor Population Prevalences and Trends in Nonspecific Psychological Distress. Cambridge(MA): Department of Health Care Policy, Harvard Medical School.

Knowles, S. E., Lovell, K., Bower, P., Gilbody, S., Littlewood, E., & Lester, H. (2015). Patient experience of computerised therapy for depression in primary care. BMJ Open, 5(11). https://doi.org/10.1136/bmjopen-2015-008581

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ‐9: validity of a brief depression severity measure. Journal of general internal medicine, 16(9), 606-613.

L. Canzian and M. Musolesi, “Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis,” UbiComp 2015 - Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., pp. 1293–1304, Sep. 2015, Accessed: Oct. 16, 2020. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2750858.2805845.

L. Kuster, C. Trahms, and J. N. Voigt-Antons, “Predicting personality traits from touchscreen based interactions,” 2018 10th Int. Conf. Qual. Multimed. Exp. QoMEX 2018, 2018.

Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ, l1476. https://doi.org/10.1136/bmj.l1476

Liakopoulos, L., Stagakis, N., Zacharaki, E. I., & Moustakas, K. (2021, July). CNN-based stress and emotion recognition in ambulatory settings. In 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA) (pp. 1-8). IEEE.

Lim, Y. M., Ayesh, A., & Stacey, M. (2020). Continuous Stress Monitoring under Varied Demands Using Unobtrusive Devices. International Journal of Human-Computer Interaction, 36(4), 326–340. https://doi.org/10.1080/10447318.2019.1642617

Lindblom KM, Linton SJ, Fedeli C, Bryngelsson IL. Burnout in the working population: relations to psychosocial work factors. Int J Behav Med 2006;13:51–59.

Ly, K. H., Asplund, K., & Andersson, G. (2014). Stress management for middle managers via an acceptance and commitment-based smartphone application: A randomized controlled trial. Internet Interventions, 1(3), 95–101. https://doi.org/10.1016/j.invent.2014.06.003

M. Exposito, R. W. Picard, and J. Hernandez, “Affective keys: Towards unobtrusive stress sensing of smartphone users,” MobileHCI 2018 - Beyond Mob. Next 20 Years - 20th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. Conf. Proc. Adjun., pp. 139–145, 2018.

Marchand A, Durand P, Haines V, III, Harvey S. The multilevel determinants of workers’ mental health: results from the SALVEO study. Soc Psychiatry Psychiatr Epidemiol 2015;50:445–459.

Marchand, A.; Blanc, M.E.; Beauregard, N. Do age and gender contribute to workers’ burnout symptoms? Occup. Med. 2018, 68, 405–411.

Mewton, L., Wong, N., & Andrews, G. (2012). The effectiveness of internet cognitive behavioural therapy for generalized anxiety disorder in clinical practice. Depression and Anxiety, 29(10), 843–849. https://doi.org/10.1002/da.21995

Morgan, C., Mason, E., Newby, J. M., Mahoney, A. E. J., Hobbs, M. J., McAloon, J., & Andrews, G. (2017). The effectiveness of unguided internet cognitive behavioural therapy for mixed anxiety and depression. Internet Interventions, 10(October), 47–53. https://doi.org/10.1016/j.invent.2017.10.003

Newby, J. M., & McElroy, E. (2020). The impact of internet-delivered cognitive behavioural therapy for health anxiety on cyberchondria. Journal of Anxiety Disorders, 69(October 2019), 102150. https://doi.org/10.1016/j.janxdis.2019.102150

Norlund S, Reuterwall C, Höög J, Lindahl B, Janlert U, Birgander LS. Burnout, working conditions and gender–results from the northern Sweden MONICA Study. BMC Public Health 2010;10:326.

O'Brien, Myles W., et al. "Step rate thresholds associated with moderate and vigorous physical activity in adults." International journal of environmental research and public health 15.11 (2018): 2454.

Pakhomov, S. V. S., Thuras, P. D., Finzel, R., Eppel, J., & Kotlyar, M. (2020). Using consumer-wearable technology for remote assessment of physiological response to stress in the naturalistic environment. PLoS ONE, 15(3), 1–14. https://doi.org/10.1371/journal.pone.0229942

Powell, T., & Enright, S. (1990). Anxiety and Stress Management (1st ed.). Routledge. https://doi.org/10.4324/9781315683867

R. Chen et al., "Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams," Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 2145–2155, 2019.

R. E. Mastoras et al., "Touchscreen typing pattern analysis for remote detection of the depressive tendency," Sci. Rep., vol. 9, no. 1, pp. 1–12, Dec. 2019, Accessed: Oct. 16, 2020. [Online]. Available: https://doi.org/10.1038/s41598-019-50002-9.

R. Wampfler, S. Klingler, B. Solenthaler, V. R. Schinazi, and M. Gross, "Affective State Prediction Based on Semi-Supervised Learning from Smartphone Touch Data," Conf. Hum. Factors Comput. Syst. - Proc., vol. 20, pp. 1–13, Apr. 2020, Accessed: Oct. 16, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3313831.3376504.

R. Wang et al., "CrossCheck: Toward passive sensing and detection of mental health changes in people with schizophrenia," UbiComp 2016 - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., pp. 886–897, Sep. 2016, Accessed: Oct. 16, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/2971648.2971740.

S. D. Gosling, O. P. John, K. H. Craik, and R. W. Robins, "Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers.," J. Pers. Soc. Psychol., vol. 74, no. 5, pp. 1337–1349, 1998, Accessed: Oct. 16, 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/9599447/.

S. Ghosh, N. Ganguly, B. Mitra, and P. De, "TapSense: Combining self-report patterns and typing characteristics for smartphone based emotion detection," Proc. 19th Int. Conf. Human-Computer Interact. with Mob. Devices Serv. MobileHCI 2017, pp. 1–12, Sep. 2017, Accessed: Oct. 16, 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3098279.3098564.

Sadeghian, Alireza, and Marjan Kaedi. "Happiness recognition from smartphone usage data considering users' estimated personality traits." Pervasive and Mobile Computing 73 (2021): 101389

Saksvik, I. B., & Hetland, H. (2011). the role of personality in stress perception across different vocational types. Journal of Employment Counseling, 48(1), 3–16. https://doi.org/10.1002/j.2161-1920.2011.tb00106.x

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J. K., Barger, L. K., Klerman, E., & Picard, R. (2018). Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study. Journal of Medical Internet Research, 20(6). https://doi.org/10.2196/jmir.9410

Seung-Gul, K., & Yong-Ku, K. (2019). Cognitive Behavioral Therapy for Insomnia in the Digital Age. Advances in Experimental Medicine and Biology, 1192, 629–641. https://doi.org/10.1007/978-981-32-9721-0_30

Sur, S., & Ng, E. S. (2014). Extending Theory on Job Stress: The Interaction Between the "Other 3" and "Big 5" Personality Traits on Job Stress. Human Resource Development Review, 13(1), 79–101. https://doi.org/10.1177/1534484313492332

Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Palmius, N., Osipov, M., Clifford, G. D., Goodwin, G. M., & De Vos, M. (2016). Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. Journal of Affective Disorders, 205, 225–233. https://doi.org/10.1016/j.jad.2016.06.065

Verdonk P, Hooftman WE, van Veldhoven MJ, Boelens LR, Koppes LL. Work-related fatigue: the specific case of highly educated women in the Netherlands. Int Arch Occup Environ Health 2010;83:309–321.

Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. International Journal of Human Computer Studies, 67(10), 870–886. https://doi.org/10.1016/j.ijhcs.2009.07.005

W. Wang et al., "Sensing Behavioral Change over Time: Using Within-Person Variability Features from Mobile Sensing to Predict Personality Traits," Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol., vol. 2, no. 3, pp. 1–21, 2018.

World Health Organization (2017). Mental health of older adults. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults.

8. ANNEX I

*Table 8 Summary of devices and physiological parameters used for used for stress detection**

Name of device	Physiological parameters collected
Emotiv [13]	EEG (Electroencephalography)
Waveguard EEG cap [15]	EEG
Biopac MP150 and Empatica E4	Heart Rate
BrainLink Headset [16]	EEG
Shimmer [17]	GSR and PPG (Galvanic Skin Response, Photoplethysmogram)
Shimmer EMG [18]	EMG (Electromyogram)
BioHarness kit, SC Amplifier	Electro CardioGram, Skin Conductance, Electromyogram, Respiration
Bioamplifier Temperature sensor, Piezoelectric Arterial Pulse Transducer	Rate (RR), Arterial Pulse Rate, Skin Temperature
FlexComp Infiniti [19]	Skin Conductance, Blood Volume Pressure, Electrooculogram, Interbeat Intervals
Equivital Life Monitor	Heart Rate Variability
Wearable electrocardiograph with tri-axis Accelerometer	Heart Rate Variability
BioSemi ActiveTwo system	EEG
g.GSRsensor2	GSR
uEye camera	HR
B-Alert X10	EEG, ECG
Truscan32 System	EEG
Tobii T60 Tracker	Points Of Regard (POR)
g.US Bamp	EEG
ECGZ2	Thoracic Electrical Bio-Impedance (TEB)
SleepSense breath belt, g.PULSEsensor, g.GSRsensor.	RR, pulse, GSR
Thought Technology FlexComp Infiniti, BodyMedia Sensewear, Biopac GSR100C Affectiva Q Sensor	Skin Conductivity
Thought Technology FlexComp Infiniti, Biopac ECG100C	HRV
Thought Technology FlexComp Infiniti	EEG
Finapres monitor	BP
LM34 Temperature sensors	Skin Temperature
Biopac	EEG

**reproduced from (Panicker et al, 2019)*