ITEA Office
High Tech Campus 69 - 3
5656 AG Eindhoven
The Netherlands
T + 31 88 003 6136
E info@itea3.org
W www.itea3.org

ITEA 3 is a EUREKA strategic ICT cluster programme

Mental Health and Productivity Boosting in the Workplace

D3.1 - Specification of multi-source data collection and analytics

Edited by: Elena Vildjiounaite (VTT), Daesub Yoon (ETRI)
Date: 14 October 2021
Version: V01.0
Contributing partners: Diego Fuentes (Hi-Iberia), Omar Nasir (Helvar), Davor Stjelja (Granlund), Sampsa Puttonen (FIOH), Elena Vildjiounaite (VTT), Matti Vakkuri (Haltian), Kai Kousa (Hintsa), Daesub Yoon (ETRI), Fátima Rodrigues (ISEP).

Table of Contents

Project acronyms

AI	Artificial Intelligence
SVM	Support Vector Machines
LSTM	Long short-term memory network
CNN	Convolutional neural network
HMM	Hidden Markov Model
HR	Heart Rate
HRV	Heart Rate Variability
PIR	Passive Infrared Sensors
IEQ	Indoor Environmental Quality
ECG	Electrocardiogram
XGBoost	Xtreme Gradient Boosting
IAQ	Indoor Air Quality
PANAS	Positive and Negative Affect Schedule
PSS	Perceived Stress Scale
SSRS	Stress Self-Rating Scale
PPI	Peak-to-Peak Interval
GDPR	General Data Protection Regulation

1. Introduction

The main goal of Mad@Work is to facilitate prevention of mental health problems, employee absenteeism and resigning. Therefore focus of data analysis work is more on assessing long-lasting mental conditions than on detection of one-time acute stress. Furthermore, it is very important to be able to assess subjective feelings of the employees: for example, if the individual feels stressed on regular basis, this feeling may result in absenteeism and/ or resigning before it affects health, and deep dissatisfaction with the workplace may develop before objective health data start displaying warning signs. It is because the way people react to work demands and deal with problems is very subjective and unique to each individual.

The main challenge of data analysis work in Mad@Work project is to develop methods to assess mental conditions, suitable for long-term real life use. Another goal is to detect potential stressors, e.g., interruptions, poor environmental conditions etc., in order to reduce stressors automatically or to facilitate self-awareness and job crafting. This challenge results in the following requirements:

- data collection and analysis methods should be privacy-safe
- sensors should be unobtrusive and convenient in long term
- training of data analysis methods should not require notable data labelling efforts from end users
- ability to adapt to personal preferences and needs with respect to sensors (e.g., somebody may be wary of video sensors) and outputs (e.g., somebody may want to get as detailed analysis results as possible during some particularly demanding period, whereas another person may be fine with monthly reports), as well as to specifics of real life conditions, such as that none of the sensors will be able to monitor the target individual all the time.

These challenges are addressed by privacy-by-design approach, guided by Nixu; by choice of sensors and algorithms and by complementary efforts of different partners. VTT focus is on collection and analysis of behavioural data by very unobtrusive sensors (by the virtual computer usage sensor especially, and also by in-office sensors), whereas the main focus of Portuguese, Spanish and Korean partners is on collection and analysis of video data and data from wearables, which are more obtrusive (wearables because the employees need to wear and charge them, and video cameras are traditionally perceived as a Big Brother technology). On the other hand, behaviour-based stress detection is not expected to be as accurate as results of video and wearable data analysis. Hintsa focuses on analysis of application interaction data and chat data, and other Finnish partners focus on indoor environmental quality assessment and prediction because IEQ optimisation is an important trade-off between energy saving and human productivity; poor IEQ can be a stressor.

Since the majority of the stress detection studies to date utilised lab data or rather short-term real life data from only one sensor, publicly available stress datasets are not abundant and do not satisfy Mad@Work requirements. Real life stress can last longer, than stress in lab studies, but be less prominent. In case of behaviour-based stress detection, it was not even studied which data features would allow to detect stress. In case of stress detection via analysis of video or wearables data, it is needed to provide for specifics of uncontrolled data collection (e.g., user motion) and to map short-term acute stress or emotions into long-lasting conditions, e.g., dissatisfaction with one's job and/ or risk of burnout.

In addition, it is not well-known how to collect data in real life so that employees and employers will accept it. Hence data collection in Mad@Work serves two purposes: (1) to provide training and test data for algorithm development and (2) to study user acceptance and specifics of long-term real life data acquisition. The most important novelty is therefore focus on real life data and user convenience.

It is also worth noting that sensor-based stress detection is a relatively new research area generally, and there are not many studies into stress detection via analysis of video data and behavioural data separately, not to speak about their fusion. Use of context in fusion is not well studied either.

At this stage Mad@Work partners focus on unimodal data collection and analysis; work on fusion will start at later stages. Consequently, this deliverable describes data collection and analysis work per partner. This deliverable describes also preliminary fusion plans of partners, but they may change later, depending on unimodal analysis results.

2. Data sources and analytics

2.1. Data Source: HI-Iberia (Spain)

2.1.1. Data description

The main objective for HI-Iberia is to develop and deploy a video-based stress detection system which is able to assess the mental health, concretely stress and emotions, in people working with their PC from **clips of video** recorded through their webcam. Thereby, it is intended that time granularity should be (almost) real-time when the monitored person is recording in front of his/her webcam. Additionally, this stress detection system will be complemented with **physiological data** and **online self-questionnaires** whenever it is possible. Below, different kinds of data are explained in detail.

Clips of video

Clips of video are recorded by the monitored people participating in the pilot through his/her own webcam once a recording session is launched by each of them through the Mad@Work Web App. Such recording session should be launched at least twice per day, and besides the clips of video will have a maximum duration of 1 minute, feature which is managed by the Mad@Work web app by stopping the video clip recording session when it is reached.

Specially for an initial data collection phase, a significant number of clips of video are required from as many different people as possible (currently, 13 people with at least 10 videos each one) in order to perform an accurate training of the deep learning algorithms which are being developed for the video-based stress detection system.

Then, a set of conditions should be taken into account for a correct recording of the video clips:

- the lighting of the workplace is correct, allowing a correct identification of the pilot participant.
- the recording plane should be centred on the pilot participant's face
- no other persons should appear in the same clip of video.

Additionally, when the recording session is finalized, and before storing it, it is necessary to express whether you felt "relaxed" or "stressed" in order to tag the video and facilitate further training of the algorithm, as it is shown below.

All clips of video are stored in the data repository with its corresponding date and tag following the privacy, security and ethical specifications.

Physiological data

Physiological data are measured by wearable devices (bracelets or watches) worn by each of the monitored people participating in the pilot. Concretely, a low-cost commercial wearable device (*Smart Bracelet Watch E66*) is being used and provides the following information:

- Body temperature
- Heart rate
- Respiratory rate
- ECG + PPG ECG monitor
- Blood oxygen monitor

- Blood pressure monitor
- Calories
- Steps
- Sleep duration

In this case, physiological data measured by the wearable device are sent via Bluetooth to the monitored people's mobile for further processing by a mobile application developed by HI-Iberia outside the project scope. So, data are stored in HI-Iberia's servers in order to avoid problems with the protection of user data.

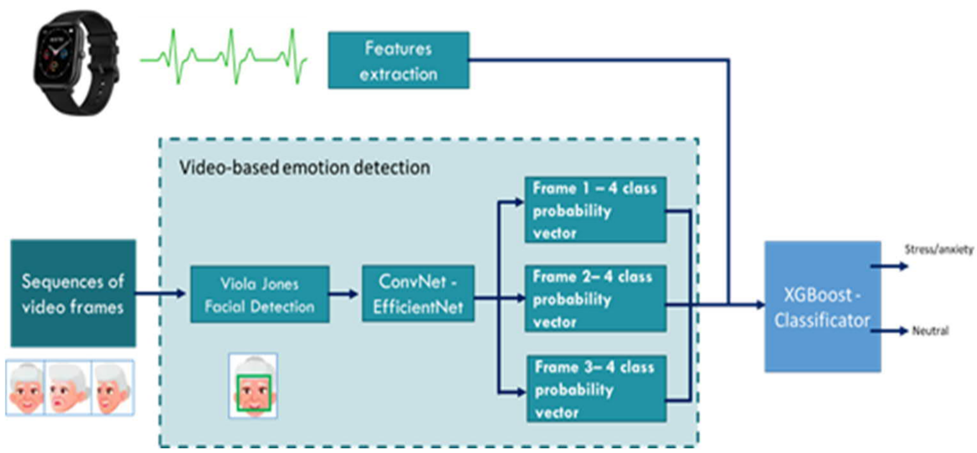
Additionally, it is important to consider for a correct collection of physiological data that the monitored people should wear the device during all working hours, and even more important, the monitored people have to wear and check that the device is working correctly when a recording session is launched.

2.1.2. Ground truth: source and description

Although physiological data are used to complement the video-based stress detection system, they also serve as ground truth. Thereby, physiological data, which are collected by a wearable device (bracelet or watch) worn by each of the monitored people participating in the pilot, should be obtained during all working hours continuously. Anyway, it is important to remark that there are numerous studies reporting that objective physiological stress markers do not always well correlate with subjective feelings.

2.1.3. Data analysis methods

Since a video-based stress detection system is being developed to assess the mental health, concretely stress and emotions, in people working with their PC from clips of video, it is intended to follow the below approach:



Based on the paper *Video-Based Stress Detection through Deep Learning*[1] written by H.Zhang et al, an emotion detection model is being developed with 4 basic emotions (sad, happy, angry and neutral) instead of 6. This video-based emotion detection module consists of the following phases:

1. Each video is split into frames, and for each of these frames, a face detection process is carried out using the Viola Jones algorithm. Once having the concerned region (face detector output), each of the frames are passed through a pre-trained EfficientNet convolutional neural network for the emotion detection. For each of the frames, the output

will be a vector of length 4, which will indicate the probability with which this frame belongs to each of the basic emotions. This process is carried out in windows of 3 consecutive frames.

2. Using the physiological data collected by the wearable device, mainly heart rate and electrocardiogram (ECG) data, although we do not rule out the use of other physiological data potentially interesting for the module implementation, a digital transformation process (DFT, Wavelet) is applied to such data in order to get a data characterisation so that such characteristics can be used during the classification phase.
3. The set of 3 vectors of length 4 extracted from the emotion detection model and the extracted features of the ECG and heart rate data serve as input to a binary classifier based on XGBoost (Xtreme Gradient Boosting) which will determine whether the user is suffering from a stress/anxiety episode or not.

It is important to remark that the training of this classifier starts with the dataset created for Mad@Work, which contains videos labelled as stress/anxiety or neutral, and the physiological data collected by the wearable devices at the time the video was recorded. Results of the stress detection will be stored in a database.

Therefore, this approach goes beyond the state-of-the-art by combining clips of video and physiological data in order to detect stress episodes considering emotions previously in a semi-supervised way, since clips of video are labelled but physiological data are unlabelled. Additionally, this approach should be comfortable for long-term real-life use since it is a video-based stress detection system which provides user a private self-diagnostics of stress episodes thus helping him to keep a good mental health.

[1] https://res.mdpi.com/d_attachment/sensors/sensors-20-05552/article_deploy/sensors-20-05552-v2.pdf

2.1.4. Privacy-by-design protection

Apart from signing an Informed Consent with each people participating/using this Mad@Work tool, in which they are informed about performing some video recording sessions voluntarily, a privacy-by-design protection is implemented in the following ways:

– Very restricted access (authorization and authentication) to raw video data stored in the data repository. Only very few project members will have access to them.
– Physiological data are stored in a SQL database, which provides support user management and privilege management to protect user's privacy.
– All data are stored in a pseudonymized way.

2.2. Data Source: Helvar (Finland)

2.2.1. Data description

Helvar's primary focus is to develop algorithms for smart buildings, which incorporate Helvar's existing sensor data and utilise additional sensors to give insights on how to improve environmental conditions. A typical lighting installation contains an array of Helvar's PIR motion sensors which can be used to estimate the amount of activity in a space. Additional sensors will be deployed to measure Indoor Air Quality like CO2 and TVOC. PIR data is irregular in nature, meaning the data is generated when an event happens. Internal PIR measurements are converted into Boolean variables which represent state of occupancy at a given point in time. IAQ data has a pre-defined frequency, meaning it is sampled at regular intervals. The datum typically contains a timestamp and a measurement value. Lastly, the building floorpan is used for determining the grid of PIRs and their proximity to IAQ sensors. Helvar is also experimenting with audio sensors which could be used to augment PIR data to make occupancy readings more precise.

2.2.2. Ground truth: source and description

The ground truth is the IAQ value itself, n steps ahead. The objective is to estimate the impact of occupancy on IAQ.

2.2.3. Data analysis methods

Data analysis involves supervised machine learning techniques that consume a vector consisting of features synthesise from occupancy data and timestamps, where as the target variable is the future IAQ value. For occupancy estimation, PIR data can be combined with spectral features of audio sensors, which then serves as input to the machine learning model. The eventual goal is to use these predictors of IAQ as an input to a BMS system, either by providing recommendations or through direct integration, and then assess the impact on IAQ.

2.2.4. Privacy-by-design protection

PIR data only provides the state of occupancy, so privacy is a non-concern. Similarly, IAQ data is a measure of environmental conditions in a space and does not incorporate any privacy intrusive measurements.

2.3. Data Source: FIOH (Finland) w/ VTT

2.3.1. Data description

The objective is to study whether the following can be detected fairly unobtrusively:

- mental states (negative stress, positive excitement and calm state)
- potential stressors: workload, interruptions, lack of skills etc.

The main VTT goal is to develop organisational barometer. The first focus group study with the line managers inside VTT[1] have demonstrated that time granularity of org. barometer should be approximately a month, so VTT aims at evaluating mental states and stressors on monthly basis. Nevertheless VTT will also study daily and weekly accuracies.
The main goal is to collect data unobtrusively in both office work conditions and remote work conditions and to not require employees to maintain any additional devices, so VTT will collect:

- computer usage data
    - keyboard: timestamp and duration of key press events, timestamp and duration of “delete” and “backspace” presses because use of the latter buttons was reported (by other researchers) to correlate with negative stress

[1] Vildjiounaite E., Kallio J., Kantorovitch J., Kyllönen V., Räsänen P., Ronkainen J. (2020) Towards Social Enterprise with Internet of Office Desks. In: Kreps D., Komukai T., Gopal T.V., Ishii K. (eds) Human-Centric Computing in a Data-Driven Society. HCC 2020. IFIP Advances in Information and Communication Technology, vol 590. Springer, Cham. https://doi.org/10.1007/978-3-030-62803-1_29

  - mouse data: mouse coordinates and timestamps
  - application switch data: application category and timestamp when this application was moved to the foreground.
- mobile phone usage data:
  - battery level
  - screen on/ off
  - application switch data: application category and timestamp when this application was moved to the foreground
  - activity (as detected by Google)
  - location in a form of anonymous numerical location identifier

Collection of mobile usage data is optional; test subjects can volunteer to provide phone data, but it is not the necessary condition for participating in the pilot. Test subjects can also volunteer to provide phone location data or provide only other than location data, if they do not permit the phone app to access location data. Since phone data can be used for stress detection only if the individual uses his/ her phone relatively regularly, and this is not the case for many knowledge workers, it is planned to use phone data mainly for detecting personal context, for example, at home vs. at work. This information would help for example when the test subject uses different computer peripherals in different places, and it affects mouse patterns.

Applications in computer usage data are assigned to one of the following categories:

- web (all browsers);
- communication (Teams, Skype etc.);
- documents (Word, PowerPoint etc.);
- other (everything which does not fall into the above groups, e.g., SW development tools, simulators etc.).

There is also “unknown” category, it is logged on rare occasions when operating system cannot return application information. Applications in mobile phone usage data are assigned to these categories plus couple of other categories, such as “navigation”, because VTT is using mobile phone data collection app, developed in the previous project.

Figure 1, Figure 2, Figure 3 and Figure 4 present examples of collected data.

```
1621372484.9800005,key,0.2497604
1621372490.1702437,key,0.0948393
1621372490.1702437,key,0.2952516
1621372492.5150325,Backspace,0.0997047
1621372493.1660213,key,0.1002393
1621372493.4002209,key,0.0750344
1621372494.4353843,Backspace,0.1101246
1621372495.525358,key,0.1402814
1621372516.5753067,Delete,0.1002636
1621372572.3650208,key,0.1450350
1621372572.4550424,key,0.0900216
1621372572.6352296,key,0.0598316
1621372572.8399994,key,0.0946255
1621372572.9301624,key,0.0674915
1621372667.4501767,Delete,0.1102588
```

Figure 1: Example of keyboard data log: each line contains (1) timestamp; (2) key; (3) duration of keypress

```
1621493877.49493,1120,715,,,
1621493877.5118167,1120,716,,,
1621493877.5149243,1119,720,,,
1621493877.5251348,1117,723,,,
1621493877.5348194,1116,725,,,
1621493877.5424325,1115,727,,,
1621493877.54498,1114,730,,,
1621493877.5551643,1113,735,,,
1621493877.5648258,1110,737,,,
1621493877.57092,1109,741,,,
1621493877.5750098,1108,745,,,
1621493877.5897617,1106,748,,,
1621493877.5949063,1105,752,,,
1621493877.610126,1103,757,,,
1621493877.6147702,1100,769,,,
1621493877.6249485,1097,776,,,
1621493877.6320174,1096,782,,,
1621493877.6402748,1096,788,,,
```


Figure 2: Example of mouse data log: each line contains (1) timestamp; (2) x, y coordinates


```
1591775306.9533823,other
1591775360.4263725,other
1591775362.9249978,communications
1591775372.199845,other
1591775395.722469,other
1591775398.0076675,other
1591775481.6862295,communications
1591775493.3002622,other
1591775731.6465738,web
1591775743.3422346,web
1591776140.3978071,other
1591776175.2045474,other
1591776177.4481328,other
1591776185.684313,unknown
1591776197.3226645,other
1591776206.564086,other
1591776212.0984373,documents
```


Figure 3: Example of application switch log: each line contains (1) timestamp; (2) application category


```
{"timeStamp":"2020-08-17T13:29:54.265Z",
"dataType":"PHONE ACTIVITY","value":"STILL",
"geoLocation":{"name":"0"}

{"timeStamp":"2020-08-17T13:37:32.718Z",
"dataType":"PHONE APP IN FOREGROUND","value":"0,44033#8,2593#",
"geoLocation":{"name":"0"}}

{"timeStamp":"2020-08-17T13:51:24.583Z",
"dataType":"PHONE_BATTERY","value":"97#true",
"geoLocation":{"name":"0"}}

{"timeStamp":"2020-08-17T14:00:57.858Z",
"dataType":"PHONE SCREEN","value":"SCREEN_ON",
"geoLocation":{"name":"0"}}

{"timeStamp":"2020-08-17T14:00:58.266Z",
"dataType":"PHONE_SCREEN","value":"SCREEN_OFF",
"geoLocation":{"name":"0"}}

{"timeStamp":"2020-08-17T14:01:22.317Z",
"dataType":"PHONE_APP_IN_FOREGROUND","value":"8,35007#",
"geoLocation":{"name":"0"}}
```


Figure 4: example of mobile phone data log: each line contains timestamp, data type, data value and anonymous location identifier, which is 0 since location data collection is not enabled

Data are collected in event-driven way: keyboard presses, computer and phone applications, phone activity and locations are recorded each time when the respective event occurs. Mouse coordinates are recorded depending on mouse type and its built-it settings. Data are collected continuously.

In addition, VTT is going to offer test subjects, on voluntary basis, to test other data sources in smaller-scale pilots:

- video analysis SW from Portugal or Spain
- physiological data (from BEIA or from Finnish company HealthZilla): resting heart rate and heart rate variability
- in-office environmental sensors (if employees return to offices): motion (depth cameras, logging head trajectories, and/ or PIR sensors, logging motion events) and indoor environmental quality data, such as temperature, CO2, noise etc.

Offering additional data sources to the test subjects will allow to study what do they really want to use in real work. Collected data will provide additional input to fusion methods, eventually allowing to get more information about subjects' conditions and the stressors.

2.3.2. Ground truth: source and description

The main focus of VTT is organisational barometer, and its main goal is to improve positive mental health of knowledge workers and this way to prevent mental health problems, such as burnout. Often, negative subjective feelings come first, then accumulate and start to manifest themselves in physiological conditions. Hence in order to detect risk of employee dissatisfaction and to ensure work engagement and positive motivation of the employees, it is important to recognise their subjective feelings. Therefore self-reports of test subjects will serve as ground truth.

In addition, FIOH will collect objective physiological data, such as heart rate and saliva samples. FIOH will also conduct cognitive tests. These data will also serve as ground truth in sensor data analysis, but it is worth noting that numerous studies reported that objective physiological stress markers do not always well correlate with subjective feelings.

Self-reports are collected on every workday for the first 2 weeks and then 3 times per week (Monday, Wednesday, Friday) for the rest of the pilot, but test subjects have an option to submit self-reports voluntarily also on other workdays. Self-reports are obtained as follows: the application reminds test subjects to submit self-report at time moment, specified by test subject: (default is 4pm). Then the test subject can fill the form, presented in the Figure 5 below. The form was designed so that submitting a self-report would take very little time from the subjects. In order to maintain interest of test subjects in submitting the reports, progress bar is gamified: it is in the form of progressing pictorial story, the bottom left corner of Figure 5 shows one image of the progress bar and points obtained by submitting self-reports to date.

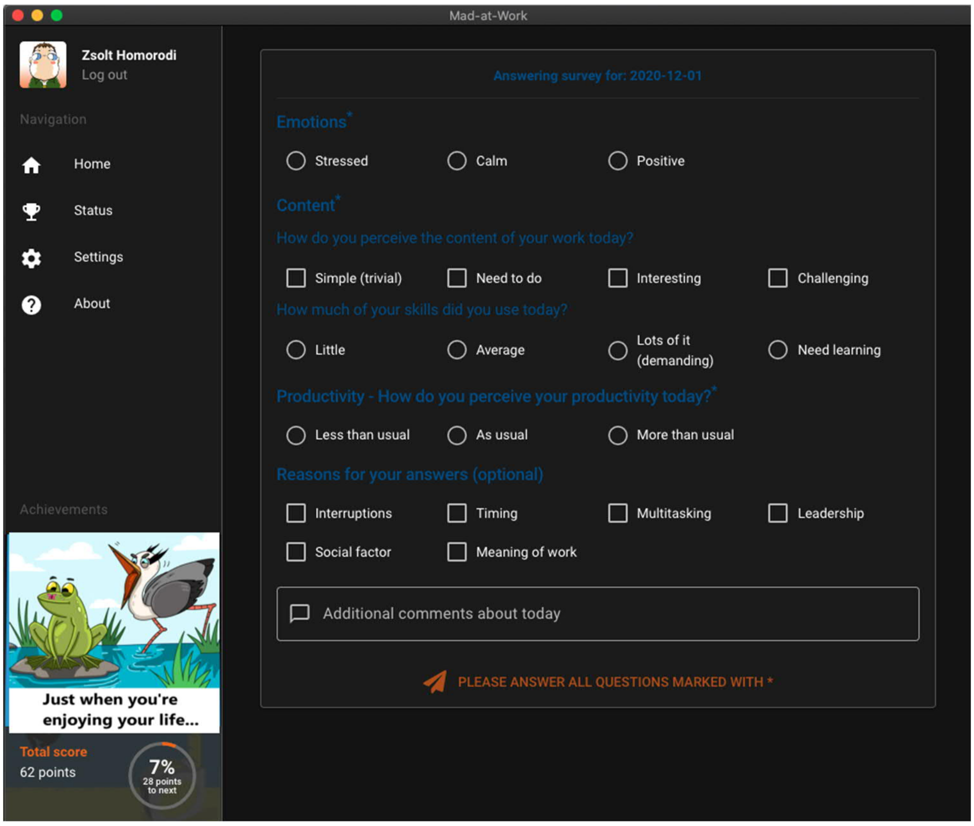


Figure 5: Self-reporting UI.

2.3.3. Data analysis methods

VTT collects behavioural data, and ways, how stress or work dissatisfaction manifest themselves in human behaviours, highly depend on a person. Therefore stress detectors on the basis of behavioural data need to be personalised. VTT will aim at comparing purely person-specific models (i.e., models, trained only on the data of the target person) with the models, obtained via transfer learning.

To the best of our knowledge, to date the majority of studies into stress detection on the basis of computer usage data utilised lab data or short-term data from certain events, such as user exams, so analysis of real life data of knowledge workers is a novel problem. The main difference is that user tasks are known during lab studies and exams, and knowledge about the tasks is typically used in data analysis, whereas in real life user tasks are not known.

To date, person-specific stress detectors on the basis of behavioural data were reported to be up to 20% more accurate than “one-model-fits-all” stress detectors, but these accuracy gains were obtained at the cost of notable self-reporting efforts: every test subject was asked to provide about 100 self-reports, sometimes more. This self-reporting efforts is unrealistically high for a real life org. barometer.

Thus VTT will aim at developing partially supervised data analysis methods, unlike the majority of the current behaviour-based stress detectors, which utilise fully supervised methods. VTT will use autoencoders, semi-supervised SVM and HMM, semi-supervised classifier ensembles, active learning and transfer learning. And VTT will compare accuracies of these methods with that of fully supervised learning methods. VTT will use autoencoders, self-labelling and semi-SVM SW, implemented in open source Python libraries (e.g., scikit-learn and Keras), and also existing implementations of transfer learning, but generally, current state of the art does provides only a few partially supervised learning methods. Hence VTT will use also own SW, developed in the previous and the current project.

Another novelty is analysis of behavioural data, obtained in real life: to date, the majority of works studied stress detection on the basis of computer usage data and motion data, obtained sensor data and ground truth in short-term lab sessions. Lab studies, however, do not provide data on prolonged stress, and previous studies demonstrated that stress does not manifest itself in real life in a same way as in lab studies. In addition, in lab studies often task knowledge is used for feature extraction, whereas in real life task knowledge is not available.

Accuracies of the methods will be evaluated against self-reports, and when possible, against objective physiological data, according to the traditional accuracy metrics, e.g., false positive and false negative rate, F1 score etc. VTT will evaluate accuracies of accessing stress on daily, weekly and monthly basis, as well as accuracies of detecting periods of long-lasting stress.

The overall approach should be comfortable for long-term real life use because of the two factors: (1) VTT uses sensors, not requiring any maintenance efforts from the end users; (2) VTT aims at requiring rather limited self-reporting efforts of the end users: we assume that employees may want to provide self-reports explicitly from time to time, but they do not want to do it regularly for a long time.

2.3.4. Privacy-by-design protection

Privacy protection is integrated in data collection and storage in the following ways:

- we do not collect any detailed computer and phone usage data, only application categories and events of keypresses (i.e., no content of keypresses, no addresses of visited web pages, no location coordinates etc.).
- all data are pseudonomysed, and only a few project members (from FIOH) know, how to link real identities of test subjects to their pseudonyms; data analysts cannot know, whose data they analyse
- access to data is protected by passwords and log auditing system: i.e., we log, who accessed the data.
- user can check and modify collected computer usage data as it is stored on her computer
- user can disable data collection on computer and phone any time
- user can disable sending of computer and phone data any time

Privacy and security measures were designed in collaboration with Nixu, the main security and privacy expert in this project.

2.4. Data Source: Granlund (Finland)

2.4.1. Data description

Granlund is developing a Digital Twin for buildings, which will be used mostly to give the insight in how the building is operated from the perspective of occupancy, IAQ and energy consumption. It will use the data gathered by IAQ sensors and energy meters and processed with AI technologies to provide insights. For example, it will compare the common first/last occupancy of the day in rooms with HVAC schedule and set-points to see how much is there unnecessary energy spent.

At the same time, it will give an insight to asset managers how people use the working space and could it be more optimised. It will also give an insight on how good building and spaces are regarding the IAQ.

Pilot building is one of the buildings of Vaasa hospital in Finland. The data collected in the building comes from three sources. First is building management system (BMS) data, which consists of datapoints collected by sensors, meters and other devices in building's systems. Second is installed IAQ IoT sensors in three rooms, which measure temperature, humidity, CO2, TVOC, PM values, O3, HCHO, CO and noise in decibels. Third is the occupancy ground truth collection from one room. The data collected from these sources ranges from minute level to hourly level.
Additionally, static information from building drawings and plans is used in establishing Digital Twin.

2.4.2. Ground truth: source and description

Ground truth data collected from the pilot building is occupancy. Data is collected from an infrared camera sensor for counting people installed in one of the meeting rooms. Ground truth collected in this way is used for training the machine learning algorithm for occupancy estimation.

2.4.3. Data analysis methods

Data analysis is done using various range of data-driven methods, such as simpler unsupervised data mining techniques and more complicated supervised machine and transfer deep learning. Data collected with BMS and IoT sensors on IAQ (temperature, CO2, TVOC, humidity) and energy sub-meters is used for understanding the IAQ and Energy patterns using data analysis methods. And as previously described this data is further processed with ML to get the insight on the occupancy.
Novelty is to scale previously developed ML models for this purpose with minimal need for re-training. Generally, the state-of-the-art models have been developed using data from one room or building. Since every building is different it is not sure how easy it is to implement these models on a wider scale. This will be investigated in Granlund's Mad@Work work.

2.4.4. Privacy-by-design protection

The infrared camera sensor for counting people has been installed in one of the meeting rooms in pilot building. This type of camera sensor does counting using infrared image without the use of RGB camera, so the privacy is not an issue. Furthermore, the sensor processes the image on sight and it stores only the number of entries/exits. The data can only be accessed from sight or using a VPN. Similarly, BMS data can only be access from sight, while IoT data is in the secure cloud, with random IDs which do not match sensor with building and room in any way.

2.5. Data Source: Haltian (Finland)

2.5.1. Data description

Halitan is further developing its Haltian Empathic Building digtal twin concept for smart office solution focuses on improving employee well-being and happiness. A complete and end-to-end smart office solution that combines technology, culture and physical space into one.

Empathic Building for smart office is one of the only solutions focused on employee experience. Instead of being focused on creating reports for optimisation of facility management costs or providing information on the building structure, electrical part or HVAC systems our solution focuses on the end-user – the tenant.

The aim is that it improves the end-user – employee – interactions with the spaces, environment and technology to find the right space, colleagues and voice their feelings

Data description can be found from the address https://docs.empathicbuilding.com/master/

2.5.2. Ground truth: source and description

Existing public datasets (e.g., SWELL-KW, WESAD) for human stress detection will be used where stress, inducing methods, physiological and behavioural modalities and data annotations available.[2]

2.5.3. Data analysis methods

In time series analysis AR, MA, ARMA, and ARIMA models, intervention models, outlier detection, transfer function models, time series regression, GARCH model, vector time series models, cointegrated processes are used.[3]

2.5.4. Privacy-by-design protection

> *"The design and implementation of privacy requirements in systems is a difficult problem and requires the translation of complex social, legal and ethical concerns into systems requirements"*[4]

The aim is to execute the work by following the guidelines of Privacy by design and an engineering practice:

1. Proactive not reactive, Preventative, not Remedial
2. Privacy as the default
3. Privacy Embedded into Design
4. Full functionality - Positive Sum not Zero Sum
5. End-to-end security - Lifecycle Protection
6. Visibility and Transparency
7. Respect for User Privacy

---

[2] B. Mahesh, E. Prassler, T. Hassan, and J.-U. Garbas, "Requirements for a Reference Dataset for Multimodal Human Stress Detection," in 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 492–498, IEEE, 2019.
[3] Little, Todd D. The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2, Statistical Analysis. Oxford University Press, 2013.
[4] *S. Gürses, C. Troncoso, and C. Diaz, Engineering privacy by design reloaded, in Proc. Amsterdam Privacy Conf., 2015, pp. 1–21.*

2.6. Data Source: Hintsa (Finland)

2.6.1. Data description

Data of the clients that are acquired from the answered surveys and from the user activity in our mobile app. Data includes wellbeing data, such as exercises, steps, sleep, and data on the work culture and wellbeing like burnout risk or turnover intentions. It also includes the progression of the studied elements, although the data time series is currently incomplete.

The most present data analysis investigates chat activity in our mobile app and how the increased engagement can positively support the wellbeing journey and lead to better results in our clients.

2.6.2. Ground truth: source and description

Opinions of the clients and of their coaches.

2.6.3. Data analysis methods

The main goal of our data collection and analysis is to find out the success of our coaching programmes and also the success of the specific micro actions that we recommend to our clients. Moreover, we want to create a system of tracking the progress of our programmes, which will allow us to intervene in the case when the programme is not as successful as was expected. With the progress tracking we also want to ensure that we collect all the available data along the client's journey.

Our aim is to create insightful visualisations that will support different departments in sales and delivery of our programmes. We want to be able to see the data on the progress and the success of our programmes on the individual level, as it provides important insights to our coaches and empowers them to create tailored interventions to match each client's needs. Correspondingly, we aim to create visualisation dashboards on aggregate levels, to be able to see the progress and success of not only programmes of different companies, but overall success of each of our programme types. The goal is also to build the visualisations that we are able to drill the results down to a specific subgroups based on a geography, type of work, gender, age group, etc. That way we will be able to observe what are the needs of each subgroup of clients and ultimately what interventions have the highest success rate for them. Additionally, these visualisations will support our sales team, as they will be able to see the success of our programmes regarding each studied element.

Regarding methods, we intent on using descriptive analysis, classifications and statistical testing of the significance of the improvement within the programme. We will also test the significance of the differences between each subgroups.

2.6.4. Privacy-by-design protection

Various measures to collect and store the clients data securely and according to GDPR. Analysis of aggregated data of many clients will utilise anonymous or pseudonomysed data, see https://www.hintsa.com/privacy/

2.7. Data Source: Ageas w/ Portuguese partners

2.7.1. Data description

The main objective of the Portuguese consortium is to detect if a collaborator is under stress during his working period, either at the office or at home.

In order to act as quickly as possible and minimize employees' stress episodes, the stress detection will be in real-time, during the working period. If the worker presents stress signals that are beyond a security limit, he will be immediately notified, otherwise he will receive a weekly report, by email, summarily describing his stress levels.

Two applications were specifically developed to be installed in the employees' laptop:

– a video-based application that collects, in a non-intrusive way, physiological features of the worker, such as pupil diameter, eye gazing, eye blinking, heart rate variability and perceived emotions through facial expressions. It should be noted that, although this is a video application, no image of the workers' faces will be saved, but only vital signs collected through the video.

– an application that alerts the worker to fill a questionnaire whose main goal is to collect the worker perception of how stressful the work period was and possible reasons for this; and another monthly questionnaire that includes self-assessment scales.

The pilot will be done with 20-40 Ageas' employees (knowledge workers in the insurance industry) from different business units. The employees have not been selected yet, but those who will participate in the pilot will be volunteers.

Data will be collected for three months, but any worker is free to turn off the video-based application at his own will, whenever he needs, so the data collected may not correspond to all working hours of three months in full.

2.7.2. Ground truth: source and description

The way people react to everyday situations and deal with stress is very subjective and unique to each individual. So, it is essential in a stress detection system to collect the psychological state and subjective feelings of workers as grand truth. To accomplish this, the application pops up a reminder for the worker to fill a self-report questionnaire twice a day, at the end of morning and afternoon working periods, at the hours previously defined by the worker. The main goal is to collect the worker's perception of how stressful the work period was and why this was the case. Besides this daily report, the worker will be also reminded to fill a monthly questionnaire that includes self-assessment scales. These scales will assess his mental health state, such as his level of stress, anxiety, depression, as well as his workload level during the month period.

The daily form was designed so that submitting a self-report takes little time from the workers and, more importantly, serves as ground truth and complements the data collected by the video-based tool. Below are the questions that compose the daily quiz.

What level of stress are you feeling right now? 1 2 3 4 5 6 7 8 9 10 No stress ○○○○○○○○○○ Extreme Stress	How do you perceive the content of your work this morning/afternoon? ○ Simple (trivial) ○ Obligation ○ Interesting ○ Challenge
How much did you need to use your skills this morning/afternoon? ○ Little (not demanding at all) ○ Moderately (not too demanding) ○ Very (very demanding) ○ I needed to learn new skills	Productivity - How do you perceive your productivity this morning/afternoon? ○ Less than usual ○ As usual ○ More than usual
Reasons for your answers: ☐ Interruptions ☐ Multi-tasking ☐ Leadership ☐ Relationship with colleagues ☐ Importance of work to you ☐ Time available ☐ Personal factors ☐ Workload	Indicate your alert level throughout this period: ○ Extremely alert ○ Very alert ○ Alert ○ Reasonably alert ○ Neither alert nor sleepy ○ With some signs of sleepiness ○ Sleepy but effortless to stay awake ○ Sleepy, but with some effort to stay awake ○ Very sleepy, struggling to stay awake, fighting sleep
What emotions are you feeling right now? ○ Riled up ○ Angry ○ Frustrated ○ Cheered up ○ Happy ○ Accomplished ○ Interested	Without manually confirming (i.e. without putting your hand on your wrist or neck), silently count the heartbeats you feel in your body. Keep your hands on the desk or on your legs, and in a sitting position. Click here to get started and when you hear a sound you should enter your heart rate count: ____

2.7.3. Data analysis methods

Several models will be developed based on the physiological data collected by the video-based application. This data will be partially labelled as stress/non_stress, so unsupervised learning will be used to find structure in the data and label it with minimal ground-truth information.

After that, different models will be trained separately, with each kind of physiological data using several classification algorithms, such as forest trees, SVM, XGBoost. These models will be evaluated with the grand truth and the best ones will be combined to obtain a better predictive performance that could be obtained from any of the learning models alone. In order for the ensemble model to be more accurate than any of its individual models, the base learners have to be as accurate and diverse as possible. So, different models considering different sources of data and learning methods will be advantageous.

The main novelty of our approach is the development of a stress detection system with behavioural data collected in a non-obtrusive way, at the office, contrary to the majority of studies that use data collected in controlled lab sessions.

The overall approach will be straightforwardly translated to real-life applications because the final model to be deployed only needs the physiological data collected by the video-application.

2.7.4. Privacy-by-design protection

Concerning privacy protection:

- the data is stored in a Postgres database, which manages database access permissions
- no video-images are collected
- all data are pseudonymised, and only the database manager has access to the corresponding individuals' information

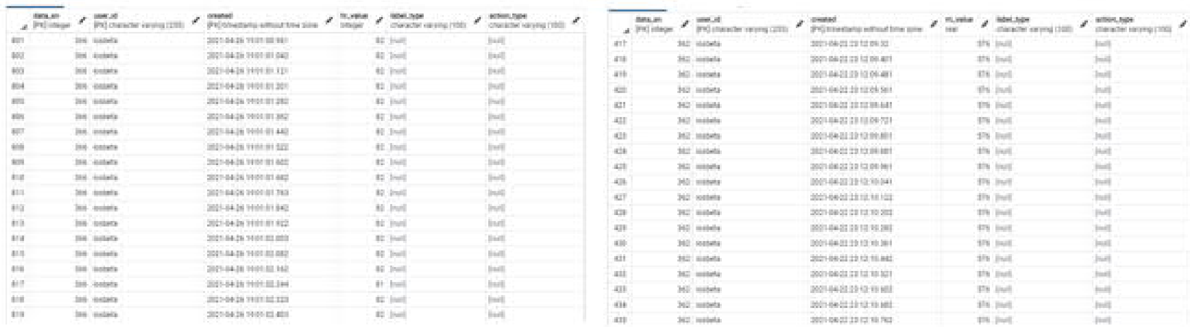
– data analysts only have access to pseudonymised data

Besides the guidelines state above, the whole pilot plan and details are dependent on the privacy and data protection analysis that is being performed by Ageas/Médis' DPO, Compliance and Legal teams, in collaboration with other partners, Nixu in particular.

2.8. Data Source: ETRI w/ Korean partners

2.8.1. Data description

Korean consortium collects data in 4 categories for stress analysis: biological data, environmental data, questionnaire data, and task information. Physiological data is collected by Samsung Galaxy Watch. Samsung Galaxy Watch produces HR and Peak-to-Peak Interval (PPI) data. Figure 6 shows examples of physiological data we collected. The collected data is stored in PostgreSQL.



(a) HR data (b) PPI data

Figure 6: Example of Physiological data stored in PostgreSQL

In addition to biological data, Korea collects environmental data, including light intensity, sound, fine dust, temperature, humidity, distance from the sensor, and air quality. We built environmental devices and collected environmental data as well as physiological data.

	user_id [PK] character varying (255)	humidity_value real	created [PK] timestamp without time zone
33	aduinotest	19.8	2020-11-23 13:36:45.437
34	aduinotest	19.7	2020-11-23 13:36:47.538
35	aduinotest	19.6	2020-11-23 13:36:55.942
36	aduinotest	19.5	2020-11-23 13:37:00.144
37	aduinotest	19.6	2020-11-23 13:37:08.608
38	aduinotest	19.5	2020-11-23 13:37:16.953
39	aduinotest	19.4	2020-11-23 13:37:19.055
40	aduinotest	19.3	2020-11-23 13:37:37.961
41	aduinotest	19.4	2020-11-23 13:37:40.066
42	aduinotest	19.6	2020-11-23 13:37:44.269
43	aduinotest	19.7	2020-11-23 13:37:46.369
44	aduinotest	19.8	2020-11-23 13:37:48.47
45	aduinotest	19.7	2020-11-23 13:37:54.774
46	aduinotest	19.6	2020-11-23 13:37:58.977
47	aduinotest	19.7	2020-11-23 13:38:05.28
48	aduinotest	19.6	2020-11-23 13:38:07.381
49	aduinotest	19.7	2020-11-23 13:38:09.482
50	aduinotest	19.8	2020-11-23 13:38:19.988

(a) Humidity sensor data

	user_id [PK] character varying (255)	illuminance_value real	created [PK] timestamp without time zone
44	aduinotest	1018	2021-01-11 13:48:41.358
45	aduinotest	1028	2021-01-11 13:48:41.371
46	aduinotest	1035	2021-01-11 13:48:41.384
47	aduinotest	1005	2021-01-11 13:48:41.395
48	aduinotest	1007	2021-01-11 13:48:43.197
49	aduinotest	1044	2021-01-11 13:48:46.202
50	aduinotest	1031	2021-01-11 13:48:49.2
51	aduinotest	1049	2021-01-11 13:48:52.194
52	aduinotest	1040	2021-01-11 13:48:55.202
53	aduinotest	1047	2021-01-11 13:48:58.203
54	aduinotest	1008	2021-01-11 13:49:01.198
55	aduinotest	1015	2021-01-11 13:49:04.199
56	aduinotest	1029	2021-01-11 13:49:07.196
57	aduinotest	1033	2021-01-11 13:49:10.195
58	aduinotest	1030	2021-01-11 13:49:13.195
59	aduinotest	1006	2021-01-11 13:49:16.194
60	aduinotest	1011	2021-01-11 13:49:19.191
61	aduinotest	1004	2021-01-11 13:49:22.192

(b) Light sensor data

Figure 7: Example of environmental sensor data stored in PostgreSQL

To analyze stress, Korea also uses task information by accessing the Google Calendar. We periodically collect task data from the Google Calendar and extract useful task information by interpreting task data.

There are several factors affecting stress. Company size, working hours, ergonomic risks, biological risks, and job requirements are reported to be highly related to work-related stress of Korean workers. Therefore, various questionnaire collected from service users can be usefully used for stress analysis. The questionnaire data collected in this study are as follows.

- General information
  - User age
  - Gender
  - Education
  - Stressful day
  - Smoking
  - Health
  - Exercise
  - Preferring stress solution
- Personality trait
- Job related Questionnaires
- required job demand
- job conflict
- interpersonal conflict
- organization system

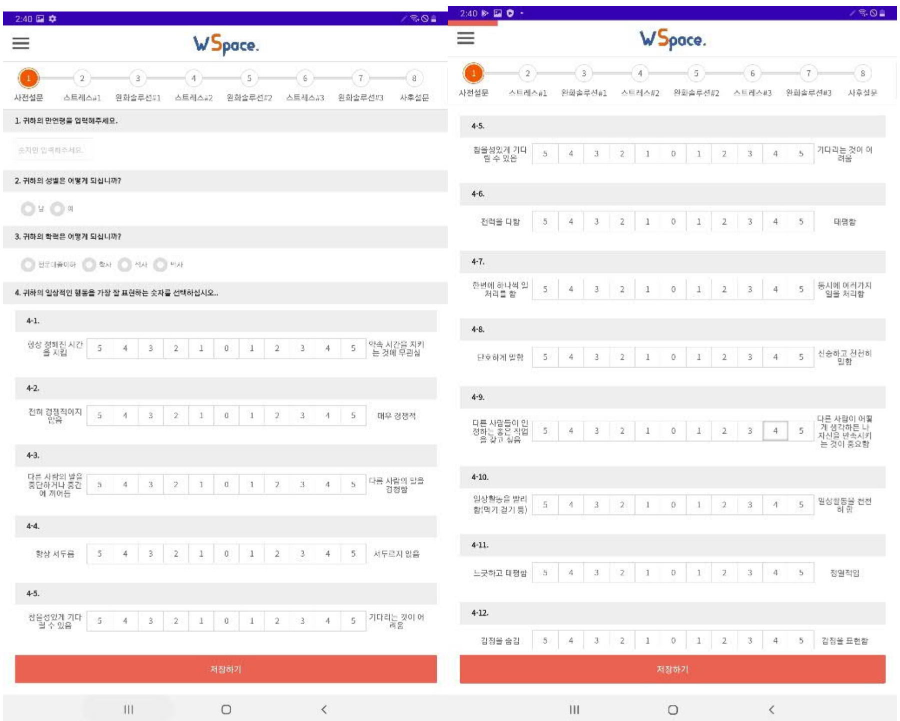


Figure 8: Questionnaires UI

2.8.2. Ground truth: source and description

Korean consortium defines stress research protocol as ground truth and collects questionnaires to validate stress research protocol.

Each part of the stress research protocol is detailed as follows:

- **Reference condition**

The participant has to avoid caffeine and smoking before the experiment begins. Before the experiment, the participant reads and signs a consent form. And then, the participant is equipped with the wearable devices

After the subject is equipped with the wearable devices, the baseline time is recorded as a reference. During the baseline, the subject sits at the chair and takes a rest. The reference condition aims at inducing a neutral affective state.

- **Stress condition**

The subject is exposed to the well-studied Paced Auditory Serial Addition Task (PASAT). The PASAT is a neuropsychological test used to assess capacity and rate of information processing and sustained and divided attention. This test has also been widely used in stress conditions. The subject is given a number every 3 seconds and is asked to add the number they just heard to the number they heard before. This is a challenging task that involves working memory, attention and arithmetic capabilities. The subject gets stressed through the PASAT and the physiological data are recorded as a stress state.

- **Relief condition**

A guided relief condition follows the stress condition. The relief condition aims to decrease the subject's stress and bring the subject back closely to the neutral affective state. The subject is given one of the relief solution candidates: taking a rest with closed eyes, listening to music, drinking tea, and stretching. Relief condition time is recorded as an unstressed state.

In addition to laboratory data applying the stress research protocol, Korea collects daily data using Wfriend. Wfriend is an app application that collects physiological and environmental data from wearable devices and environmental devices. Wfriend also collects daily and weekly questionnaire data. Korea has a plan to collect daily data from October using Wfriend. We use questionnaire data as ground truth for daily data.

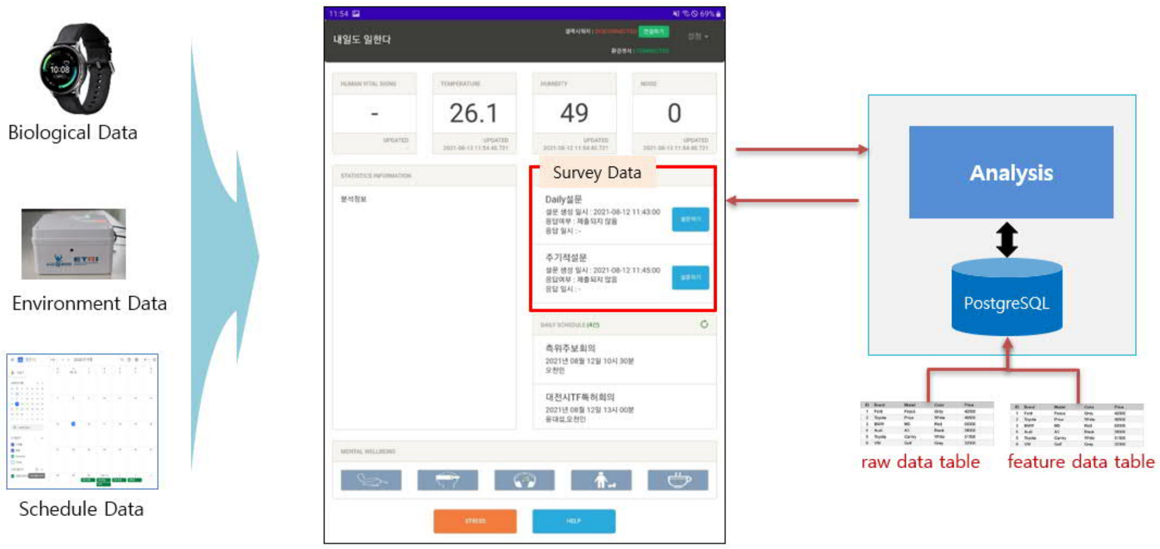


Figure 9: Daily Data Collection using Wfriend App

Stress exceeding human information processing capacities will directly lead to errors in human performance. In such cases, stress can make the required task accomplishment impossible in principle. Furthermore, continuous high levels of stress will cause cardiovascular disease. Thus, the purpose of this study is to assess stress of knowledge workers and provide an appropriate solution when the stress level is high enough to be dangerous to workers. The stress management will help to maintain worker's mental and physical health, and improve work productivity. In order to improve work productivity and job satisfaction of knowledge workers, it is necessary to evaluate the subjective stress level of knowledge workers. Therefore, the subjective stress level can be set as ground truth.

2.8.3. Data analysis methods

The data analysis and evaluation follow the classical data processing chain, consisting of the following steps: pre-processing, segmentation, feature extraction, and classification. The first step is to pre-process data. In this step, we deal with missing data and apply interpolation. After pre-processing, the dataset is segmented using a sliding window. And then, we extract various features from the data segments.

The extracted features, detailed above, serve as input for the classification step. Five machine learning algorithms are applied and compared within the benchmark: Decision Tree (DT), Random Forest (RF), AdaBoost (AB), and Support Vector Machine (SVM). As the entire data processing chain is implemented in Python, we used the scikit-learn implementation of the classifiers mentioned above. For the AdaBoost ensemble learner, a decision tree is used as a base estimator. For the decision-tree-based classification algorithms, information gain is used to measure the quality of splitting decision nodes.

Finally, accuracy and F1-score are used as evaluation metrics. Accuracy represents the number of correctly classified instances out of all samples. The F1-score is defined as the harmonic mean of precision indicating the reliability of the results in a certain class and recall representing a measure of completeness. All models are evaluated using the leave-one-subject-out (LOSO) cross-validation (CV) procedure. Hence, the results indicate how a model would generalize and perform on data of a previously unseen subject.

2.8.4. Privacy-by-design protection

Physiological data, collected from wearable devices, are stored in PostgreSQL. Since relational databases are designed to take into account the privacy of users, they fundamentally support user management and privilege management to protect user's privacy. All we need is to train data for mental health model generation. Therefore, we do not consider more privacy functionalities.

In terms of user interfaces, the stress dashboard provides users with their stress information. Since the dashboard is public to all users, it does not provide personal information such as name and age for privacy. Therefore, users can not check if which member gets stressed or not in the group.

3. Data fusion analytics

3.1. Fusion plan: HI-Iberia (Spain)

3.1.1. Data modalities

The main objective for HI-Iberia is to develop and deploy a long-term mental health assessment system which is able to extract some long-term trends of mental health in people working with their PC. As it is shown below, it is intended to fuse **historical results (stress and emotions)** extracted from the video-based stress detection system with a time granularity that it is (almost) real time, with **online self-questionnaires** which are launched through the Mad@Work Web App once per week.

Therefore, it is intended that time granularity should be weekly since both data sources will be combined to have the same granularity, that is, historical results will be contemplated along each week.

Concretely for the self-questionnaires, they are launched through the Mad@Work Web App once per week and of course, all monitored people participating in the pilot have to answered it. These self-questionnaires are based on questionnaires commonly used in stress experiments in the laboratory and in everyday life, such as:

- Patient Health Questionnaire-4 (PHQ-4),
- Perceived Stress Scale (PSS), 10 items (once per month)
- Stress Self-Rating Scale (SSRS),
- NASA-TLX,
- Self-Assessment Manikin and Positive and Negative Affect Schedule (PANAS)

However, HI-Iberia is going to use the Perceived Stress Scale (PSS) questionnaire applied weekly instead of monthly.

PERCEIVED STRESS SCALE

The questions in this scale ask you about your feelings and thoughts during the last month. **In each case, you will be asked to indicate by circling *how often* you felt or thought a certain way.**

Name ______________________ Date ________

Age _____ Gender (Circle): **M F** Other ______________________

0 = Never 1 = Almost Never 2 = Sometimes 3 = Fairly Often 4 = Very Often

1. In the last month, how often have you been upset because of something that happened unexpectedly?	0	1	2	3	4
2. In the last month, how often have you felt that you were unable to control the important things in your life?	0	1	2	3	4
3. In the last month, how often have you felt nervous and "stressed"?	0	1	2	3	4
4. In the last month, how often have you felt confident about your ability to handle your personal problems?	0	1	2	3	4
5. In the last month, how often have you felt that things were going your way?	0	1	2	3	4
6. In the last month, how often have you found that you could not cope with all the things that you had to do?	0	1	2	3	4
7. In the last month, how often have you been able to control irritations in your life?	0	1	2	3	4
8. In the last month, how often have you felt that you were on top of things?	0	1	2	3	4
9. In the last month, how often have you been angered because of things that were outside of your control?	0	1	2	3	4
10. In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?	0	1	2	3	4

Additionally, it is intended to include other kind of data like computer usage data collected by VTT in future stages of the technical development.

3.1.2. Data fusion algorithms

The envisioned long-term mental health assessment system goes beyond the state-of-the-art by advancing multimodal fusion methods that combine data from several asynchronous and heterogeneous data sources with different temporal granularity, which are not always available. For such multimodal fusion we will use event-based fusion methods, classifier ensembles and methods for analysing sequential data, e.g. HMM and LSTM, and adapt these methods to the needs of uncontrolled environments.

The key of this assessment system is to perform a decision-level fusion with the above-mentioned data in order to extract relevant statistics like number of stress episodes per time, time of day when stress episodes usually occur, etc., which could be interesting for monitored people, as well as combining such data with the self-questionnaires in order to define a user pattern and extract relevant characteristics to provide more accurate recommendations which reduce long-term stress episodes.

Since the data to fuse have different time granularities, it is intended to group and fuse historical results (stress and emotions) along each week in order to align such time granularity with the self-questionnaires' time granularity.

3.2. Fusion plan: Helvar (Finland)

3.2.1. Data modalities

We plan to fuse together PIR occupancy with IAQ data collected from indoor spaces. We are also developing fusion of audio sensor data with PIR to have improved occupancy sensing.

3.2.2. Data fusion algorithms

For audio data, we are using spectral analysis to extract estimated energy signatures and run machine learning to determine if the space is occupied or not. For IAQ analysis, the occupancy estimation is fed to a supervised learning algorithm which predicts the amount of short-term increase or decrease in CO2 values. Then these two data sources will be combined. Future plans include to add other sensors, e.g., CO2 or humidity.

3.3. Fusion plan: FIOH (Finland) w/ VTT

3.3.1. Data modalities

We plan to experiment with fusion of different data sources, indicative of stress: computer keyboard data, mouse data and application data, and also (if test subjects volunteer for it) physiological data and video analysis results of other partners. Depending on return to offices, we will use also motion data. If there will be enough phone data (its collection is optional), we will use also it at least for context detection. These data types have different time granularities, because e.g. keyboard presses typically occur much more frequently than change of application.

On the other hand, collection of computer usage data is continuous, whereas video and physiological data are collected occasionally. Hence the overall plan is to perform fusion of data sources of different availability, different time granularity and different accuracy: for example, facial expression data is expected to provide more accurate estimate of the subjects' current condition than his/ her behaviour data.

We also plan to use in fusion data on potential stressors, such as e.g., noise data or phone call data.

3.3.2. Data fusion algorithms

For fusion of similar data types we plan to experiment with feature-level fusion, when features are extracted from reasonably long time windows (e.g., a few hours). Therefore AI methods for fusion will be similar to the methods for unimodal data analysis.

We also plan to experiment with decision-level fusion for similar and dissimilar data types, especially with probabilistic decision-level fusion methods.

The main goal of data fusion research is to find an appropriate trade-off between increasing accuracy of data analysis and decreasing data collection efforts, where the latter include costs of sensors (i.e., we aim at finding the most appropriate sets of sensors) and data labelling efforts, required from the end users. Therefore we aim at developing a practical system, satisfying the requirements of the org. barometer.

The main novelty of our approach is the aim at developing a practical system and thus to provide for fusion of asynchronous data sources and for different availability of data sources, where availability depends on user preferences and user actions. For example, if the test subject is allowing use of the camera only in certain situations, or is not typing currently, these data types will be unavailable for some time.

Another novelty is the aim to develop methods, that provide sufficiently accurate results according to the requirements of org. barometer (e.g., on monthly basis), because the majority of the current methods were developed on the lab data and aim at accurate classification of short time periods (e.g., a few minutes/ an hour).

3.4. Fusion plan: Granlund (Finland)

3.4.1. Data modalities

We plan to fuse together the previously mentioned collected data from the buildings with the processed data. For example, using the IAQ and ML we would estimate the occupancy in the room, then this occupancy would be compared with the HVAC system schedule, set-points, energy consumption and raw IAQ data. Using this building managers would have better insight in how good performance of their building is regarding the space usage, IAQ and energy consumption and where could it be improved.

3.4.2. Data fusion algorithms

Data fusion algorithms will be as described in 2.5.3. We will use different data-driven methods. During the pre-processing data with smaller sampling time will be resampled to match the largest sampling time in the dataset. Some data will be fused in post-processing phase, where for example more static data such as HVAC system schedules will be compared with the occupancy estimated by ML on a daily level.

3.5. Fusion plan: Haltian (Finland)

3.5.1. Data modalities

Existing public datasets (e.g., SWELL-KW, WESAD) for human stress detection will be used where physiological and behavioural modalities are available.

3.5.2. Data fusion algorithms

Will be same as described in Section 2.5.

3.6. Fusion plan: Hintsa (Finland)

3.6.1. Data modalities

We are going to fuse clients' data that are acquired from the answered surveys and from the user activity in our mobile app. Data includes wellbeing data, such as exercises, steps, sleep, and data on the work culture. It also includes the chat activity in our mobile app. These data are asynchronous and irregular, as users chat and use the mobile app when it suits them. Therefore data will be combined over time, and also importance of each data type will be studied.
Then we will also combine data of different employees to get insights (1) into overall company situation and (2) into success of each coaching programme for different cases.

3.6.2. Data fusion algorithms

Data fusion methods will be same as described in section 2.6

3.7. Fusion plan: Ageas w/ Portuguese partners

3.7.1. Data modalities

The several physiological features, pupil diameter, eye gazing, eye blinking, heart rate variability and facial expressions that will be used to develop stress detection models are collected by the same video-based application, so data fusion will be simple since they have the same date/time key for each user. Our first goal is to study in separate the contribution of each of the physiological signals to stress detection, and only then study which combination of physiological signals works best. This combination will be done not only at data level but also at model level through ensemble learning. In addition, models may also be tailored to a specific worker or business units, or instead be generic to all workers.

3.7.2. Data fusion algorithms

AI algorithms for data fusion will be similar to the algorithms used for unimodal data analysis. Our goal is to study which physiological signal or combination of signals will be more accurate to stress detection since in the literature there is not a consensual choice of the best ones.

3.8. Fusion plan: ETRI w/ Korean partners

3.8.1. Data modalities

In order to build a reliable stress detection system, it is important to understand that stress is primarily a physiological response to a stimulus triggered by the sympathetic nervous system (SNS). Therefore, we need to build up a stress model using physiological data from multiple sources. Collecting physiological data using embedded devices is quite intrusive. Hence, these modalities are only available on specific occasions. Wearable devices are only minimally intrusive. Devices like smartphones and watches are already popular among users. Therefore, Korean consortium collects various physiological data using wearable devices. We also collect questionnaires, environmental information, and task information for stress analysis. We collect questionnaires and task information using web interfaces.

We access the Google Calendar for task information. To get task information, we create a unified id and add it to all google calendars as a member. The mental management system can periodically collect and interpret task data since we already defined the task template for google calendar.

3.8.2. Data fusion algorithms

Korean consortium detects the stress state by the mental health index, which is based on multiple source data fusion and calculated by adding static mental health index, dynamic mental health index, and task load index.

The static mental health index is based on questionnaires and computed using self-reports. The stress information is extracted from questionnaires. The dynamic mental health index is decided by biological signal information and environmental information. The classification algorithms are the same as the methods for analyzing biological signals in the unimodal user interface. The environmental information consists of light intensity, sound, fine dust, temperature, humidity, distance, and air quality. The task load index is based on task schedules and current tasks. It depends on the type of task schedules and tasks such as deadline, evaluation, presentation, and paper writing.

The difference of our approach from conventional approaches is to calculate multiple mental health indexes, including static mental health index, dynamic mental health index, and task load index. We calculate multiple indexes and assign weights to each index. As a result, we get a unified stress index and determine mental health.

4. Conclusions

The Mad@Work project focuses on assessing long-lasting mental states and assessing employees' subjective emotions rather than the detection of one-time acute stressors. To achieve this purpose well, the Mad@Work project collects a variety of data to analyse the stress and emotional state of knowledge workers. Project partners collect facial video information using webcams, collect physiological data (ECG, heart rate, respiration rate, blood pressure) using wearable devices (bracelets, watches), and collect computer usage data and mobile phone usage data of workers. In addition, partners will use a variety of IoT sensors to collect information on air quality, lighting, noise, and activity in the workplace, and collect their own online questionnaires. Each partner of the Mad@Work consortium uses the collected data in the unimodal form or fuses the multimodal data, and then applies artificial intelligence-based machine learning techniques to increase the accuracy of stress state detection for individuals. Project partners' detail activities are as follows:

Hi-Iberia collects facial video clip information by webcam, physiological data (ECG, heart rate, respiratory rate, blood pressure) by wearable devices (bracelets or watches), and online self-questionnaires. After segmenting the video into frames, Hi-iberia performs the face detection process using the Viola Jones algorithm, and emotion detection using the pre-trained EfficientNet convolutional neural network. This video-based emotion detection module is being developed to recognise 4 basic emotions (sad, happy, angry and neutral). Although physiological data are used to complement the video-based stress detection system, they also serve as ground truth. Hi-Iberia fuses historical results (stress and emotions) extracted from the video-based stress detection system with a time granularity that is (almost) real time, with online self-questionnaires which are launched through the Mad@Work Web App once per week. Because the data to be fused has different time granularities, historical results (stress and emotion) are grouped and consolidated weekly to match those time units with the self-questionnaires' time granularity. For such multimodal fusion we will use event-based fusion methods, classifier ensembles and methods for analysing sequential data, for example HMM and LSTM, and adapt these methods to the needs of uncontrolled environments.

Helvar uses a PIR sensor, CO2 and TVOC to collect the amount of activity in a space and indoor air quality information. Helvar develops algorithms for smart buildings to provide insights on how to use information on lighting and air quality in workplaces to improve environmental conditions. Data analytics uses supervised machine learning techniques to predict future IAQ values. They plan to fuse together PIR occupancy with IAQ data collected from indoor spaces and to fuse audio sensor data with PIR to improve occupancy sensing. For audio data, we are using spectral analysis to extract estimated energy signatures and run machine learning to determine if the space is occupied or not. For IAQ analysis, the occupancy estimation is fed to a supervised learning algorithm which predicts the amount of short-term increase or decrease in CO2 values. Then these two data sources will be combined.

FIOH/VTT studies whether mental states (negative stress, positive excitement, and calm states), potential stressors (workload, distractions, lack of skills, and etc.) can be detected unnoticed. They collect computer usage data (keyboard, mouse and application switch data). They optionally collect mobile phone usage data (battery level, screen on/off, application switch data, activity, etc.) according to the consent of the participant. Collected data will provide additional input to fusion methods, eventually allowing to get more information about subjects' conditions and the stressors. They aim to develop a partially supervised data analysis method to create a purely individual model (i.e. a model trained only on the data of the target person). They use autoencoders, semi-supervised SVMs and HMMs, semi-supervised classifier ensembles, active learning and transfer learning. They plan to experiment with fusion of different data sources, indicative of stress: computer keyboard data, mouse data and application data, and also (if test subjects volunteer for it) physiological data and video analysis results of other partners. For fusion of similar data types they plan to experiment with feature-level fusion, when features are extracted from reasonably long time windows (e.g., a few hours). They also plan to experiment with decision-level fusion for similar and dissimilar data types, especially with probabilistic decision-level fusion methods.

Granlund is developing a Digital Twin for buildings, which will be used mostly to give the insight in how the building is operated from the perspective of occupancy, IAQ and energy consumption. IAQ IoT sensors are installed in rooms, which measure temperature, humidity, CO2, TVOC, PM values, O3, HCHO, CO, noise in decibels, and the occupancy. Data analysis is done using simpler unsupervised data mining techniques and more complicated supervised machine and transfer deep learning. They plan to fuse together the previously mentioned collected data from the buildings with the processed data. For example, using the IAQ and ML they would estimate the occupancy in the room, then this occupancy would be compared with the HVAC system schedule, set-points, energy consumption and raw IAQ data. Some data will be fused in post-processing phase, where for example more static data such as HVAC system schedules will be compared with the occupancy estimated by ML on a daily level.

Halitan is developing its Haltian Empathic Building digital twin concept for smart office solution which focuses on improving employee well-being and happiness. Halitan uses previously published data sets (eg SWELL-KW, WESAD) for human stress detection and analysis. They use AR, MA, ARMA and ARIMA models, intervention models, outlier detection, transfer function models, time series regression, GARCH models, vector time series models, and cointegration processes for time series data analysis.

Portuguese consortium detects if a collaborator is under stress during his working hours, either at the office or at home. They install a video-based application that collects pupil diameter, eye gazing, eye blinking, heart rate variability, and facial expressions on employees' laptops, as well as a survey application that collects stress size and stress reasons during working hours. Portuguese consortium data will be partially labelled as stress/non-stress, so unsupervised learning will be used to find structure in the data and label it with minimal ground-truth information. After that, different models will be trained separately, with each kind of physiological data using several classification algorithms, such as forest trees, SVM and XGBoost. The several physiological features, pupil diameter, eye gazing, eye blinking, heart rate variability and facial expressions that will be used to develop stress detection models are collected by the same video-based application, so data fusion will be simple since they have the same date/time key for each user. Developed models may also be tailored to a specific worker or business units, or instead be generic to all workers.

Korean consortium collects data in 4 categories for stress analysis. An application is installed for pilot testing on the android-based Galaxy Tab. This app can collect biological data, environmental data, schedule data and questionnaire information. When workers wear smart watches on their wrists, HR and HRV physiological data are collected. The app collects schedule information within the scope permitted by the user in connection with Google Calendar. When the app is connected to an environmental device, it collects environmental data such as light intensity, sound, fine dust, temperature, humidity, distance from the sensor, and air quality. In addition, a questionnaire request is periodically popped up on the app screen to collect subjective stress states and stressors when workers respond to the questionnaire. To predict stress or steady state, the research team will use machine learning algorithms such as decision trees, random forests, AdaBoost and SVM. In addition, physiological data, questionnaires, environmental information, and schedule information for stress analysis will be fused and analyzed. The approach of the study is based on the fusion of multiple source data, and seeks to detect the stress state calculated through the fusion of static mental health state, dynamic mental health state and workload state.

To protect the privacy of the Mad@Work project, all data is stored in a pseudonymous fashion, providing very limited access (authorization and authentication) to collected and analysed data. Physiological data and survey data are stored in an SQL database to support user management and permission management to protect users' personal information. The Mad@Work project will provide novel and state-of-the-art research results as we develop a system to assess mental state by collecting real-life data without disrupting knowledge workers.