Clara: Design of a New System for Passive Sensing of Depression, Stress and Anxiety in the Workplace

Juwon Lee[1](✉), Megan Lam[2], and Caleb Chiu[2]

[1] The University of Hong Kong, Pokfulam, Hong Kong
u3534591@connect.hku.hk
[2] Neurum Health, Room 1405, 135 Bonham Strand Trade Centre, 135 Bonham Strand, Sheung Wan, Hong Kong
{meg, caleb}@neurumhealth.com

Abstract. Collective evidence from research on the detriment of mental ill-health in the workplace consistently points to the need for better management of workplace mental health. However, difficulty in making a reliable, unobtrusive measurement of an employee's mental health remains an obstacle in the way of effective interventions at an organizational level. In this paper, a system named Clara is proposed with aims to enable passive measurement, and hence effective management, of workplace mental health. A literature review of different approaches to measure depression, stress, and anxiety is presented, followed by a discussion on the design principles that guided the development of Clara. The overarching system architecture is then outlined, and individual components of the system are explored in finer details. The paper illustrates how Clara, with its passive measurement techniques, has the potential to enable objective assessment of workplace depression, stress and anxiety, allowing for delivery of timely interventions.

Keywords: Workplace mental health · Passive sensing · Machine learning · Depression · Stress · Anxiety

1 Introduction

In recent years, mental health problems in the workplace have been among the top concerns for organizations of varying sizes. In a survey conducted by Mental Health America (MHA) on 17000 workers of different industries, over 35% answered they "always" miss 3–5 days a month because of workplace stress [1]. In the same survey, over 80% answered their personal relationships are being negatively affected by workplace stress [1].

Despite the prevalence of the problem, many suffer in silence. A survey-based study conducted on 44000 employees in America showed that only about half of the people experiencing mental health problems have talked to their employers about their problems [2]. This lack of openness that arises from the stigma associated with mental illnesses makes them altogether a more trying challenge to tackle.

P. Cipresso et al. (Eds.): MindCare 2019, LNICST 288, pp. 12–28, 2019.
https://doi.org/10.1007/978-3-030-25872-6_2

Some attempts have been made to put a price tag to this debilitating phenomenon, and one of such examples is the return on investment analysis report created by PwC on "creating a mentally healthy workplace" [3]. The report, which measured the cost of mental health problems in the workplace as the sum of costs of absenteeism, presenteeism and compensation claims, showed that per year, employers in a single developed country loses on average $11 billion - $4.7 billion in absenteeism, $6.1 billion in presenteeism and $146 million in compensation claims [3]. In [4], the authors measured the cost of poor mental health in the workplace by exploring the association of depressive episodes and work productivity, effectively illustrating that about one-third of the annual $51 billion cost of mental illnesses in America is related to productivity losses.

Incentivized by such tangible implications of mental health problems in the workplace, a growing number of employers have started taking initiatives to improve mental wellbeing of its employees. For example, based on the findings from the American Psychological Association (APA) study [5] that highlighted the importance of support from managerial positions on the mental wellbeing of employees, Unilever began to encourage managers through their global health initiatives to take workshops to recognize signs of mental health distress [6]. Similarly, Delta Air Lines has set up early and active intervention programs to support employees with mental health issues. The programs include the employee assistance program (EAP) that provides employees with unlimited phone consultations with master's-level clinicians, three free face-to-face counseling visits per issue per year and behavioral health leaves [7].

A number of studies have been conducted to evaluate the efficacy of such measures taken by employers at an organizational level. A comprehensive literature review of such studies was provided by [8], which presented comparisons of evaluations on three separate categories of workplace mental health initiatives, namely universal interventions, secondary preventions and tertiary preventions. Universal interventions refer to strategies for improving mental health that are delivered to all employees in a work setting without regard to individual risk factors. On the other hand, secondary preventions and tertiary preventions are targeted to subgroups of employees determined to be at risk [8].

A meta-synthesis of qualitative research suggests that both secondary and tertiary prevention have strong evidence for robust effects and reduced symptoms [8]. However, as these intervention strategies require preliminary identification of target subgroups characterized by high levels of risk factors, a key challenge to delivery of interventions lies in the early detection of such risk factors. The subtlety of mental ailments, which are difficult to detect from an external observation, makes accurate measurement of their presence and severity an extremely challenging task.

The costly harms of workplace mental health problems and the heightened interest around initiatives to address them justify the need for a better employee mental-health management tool with enhanced risk detection capabilities. With aims to deliver more timely, targeted interventions at an organizational level, a new system named Clara is proposed.

In this paper, a literature review on multimodal measurement techniques of mental health is presented, with a focus on measurement of depression, anxiety and stress levels. The literature review is followed by an in-depth description of Clara, covering its design principles, overall system architecture and individual components. The paper concludes with a discussion on Clara's limitations, areas of future research as well as implications on the betterment of workplace mental health management.

2 Literature Review on Measurement of Mental Health

The past studies on measurement of affective states including depression, anxiety and stress have often been categorized into three modalities: psychological, physiological and behavioral. Though psychological assessments carried out via questionnaires make up the most widely used method for evaluating such state of mental health [9], their reliability and relevance outside the laboratory setting were often called to question, mainly with regards to the subjectivity introduced from self-report nature. On the other hand, the growing accessibility of digital devices such as mobile phones and smart watches have sparked a noticeable trend in augmenting such self-reported inputs with verifiable data of physiological and behavioral modalities [10].

2.1 Questionnaire-Based Psychological Assessments

Assessment of depression and anxiety have commonly been done in tandem through psychometric instruments including DASS-21 [11], PHQ-9 [12] and GHQ-28 [13]. Likewise, several questionnaire-based assessment methods have been developed for measurement of stress levels. Examples include the Standard Stress Scale (SSS) [14], the Perceived Stress Scale (PSS) [15] and the Stress Response Inventory (SRI) [16]. Together, these questionnaire-based instruments offer clinicians and researchers a standardized means to screen for depression, anxiety and stress. Notably, DASS-21 delivers a set of subscales for each of the three scales it measures, depression, anxiety and stress, allowing for simultaneous yet thorough measurements on the three spectra [11]. For a comprehensive analysis on these measures, their theoretical foci and validation results, we refer to a review by Sakakibara et al. [17].

Despite the widespread usage of such assessment methods, the clinical relevance of their outcomes has often been a source of debate in the psychology community. Concerns frequently raised include reliance on introspective ability of participants, systematic response distortions, and lapses in memory of participants when asked to recall their state from the past [9]. These questionnaire-based assessments have also been criticized for their inability to capture a continuous evolution, and hence subtle changes, of affective states [18].

2.2 Digital Biomarker-Enabled Measurements

With the flux of different sensor-equipped digital devices entering the market, the term "digital biomarker" was introduced to the lexicon of clinical medicine. Its definition is given as "consumer-generated physiological and behavioral measures" collected through connected digital tools [19], and unlike the aforementioned psychological assessments, it offers the benefit of enabling a real-time, continuous and quantitative measurement of an individual's mental state. Many efforts have been made to establish digital biomarkers for mood disorders and affective states, some of which are explored in detail below.

GPS Approach. Studies have shown that GPS data collected from one's mobile phone can act as a powerful predictor for affective disorders including depression and social anxiety [20, 22]. As GPS data captures information about one's range of mobility, home stay and variability in locations visited, signals of mental ill-being were found embedded in the GPS data in multiple occasions. For example, [20] was able to demonstrate the potential of this approach by using fine-grained location data from GPS sensors as a predictor of corresponding, in situ sampling of mental state from 72 study participants. Similarly, in a study conducted with a mobile phone app, Purple Robot [21], higher-level features including circadian movement, normalized entropy and location variance were extracted from raw GPS data of 40 participants for analysis with self-reported depression survey (PHQ-9) [22]. A regression model was trained with the resulting dataset, achieving a cross-validated error of 23.5% in predicting a PHQ-9 score based on the extracted GPS features.

Keystroke Dynamics Approach. Keystroke dynamic refers to the unique characteristics that make up an individual's typing rhythm when using a keyboard or a keypad [23]. Usage of keystroke dynamics in identifying an individual for security applications has been a popular topic for study [24]. Frequently referenced timing features in such studies include time per keystroke, average pause length and pause rate. A variety of keystroke-specific features have been explored as well, including backspace key rate, delete key rate, home key rate and sentence-ending punctuation key rate [23].

Following the promising results in identity authentication applications, keystroke dynamics began to be studied in affective computing. From a field study of participants' keystrokes and their self-reported emotional states, [25] was able to develop classifiers for seven different emotional state, all of which had accuracies above 70%. The study was significant not only in its high achieved accuracy but also in usage of diverse features including those of digraphs (two-letter combinations) and trigraphs (three-letter combinations). Durations between different down keys and up keys of the n-graphs were extracted from participants' keystrokes and were aggregated across participants.

In a separate study exploring the correlations between keystroke dynamics and stress levels, [26] was able to create a KNN-based mode with 75% classification accuracy, further affirming the potential of exploiting continuous monitoring of keyboard interactions for stress detection. On top of keystroke and timing features, the dataset consisted of linguistic features such as language diversity, language complexity and expressivity. Despite the impressive accuracy score, however, it must be noted that introduction of such content-based features may significantly limit the scope of application due to mining of sensitive information.

Mouse Movement Approach. In studies on mouse movements, commonly extracted features are mouse speed, acceleration, direction and number of clicks. Though not as common as studies on keystrokes, a number of studies on mouse movements have demonstrated statistical significance of mouse characteristics in detection of changes in mental states [27, 28].

The authors of [27] designed an experiment that induced emotional states amongst participants by presenting a happy video and a sad counterpart. While the system asked for self-reports on the participants' affective states, it passively gathered mouse movement data without the participants' knowledge. The experiment results showed significant differences in mouse motion characteristics for different levels of arousal. Features that showed particularly high correlation were movement precision, movement smoothness, movement speed and movement acceleration.

Study presented in [28] shared a similar aim of utilizing mouse movement characteristics for development of an empathetic system. More specifically, the paper presented an experiment that collected mouse movements of 136 participants as they were watching an online course. Based on the timing of mouse movement pauses, the system prompted the participants to record their level of boredom. Mouse movement features were extracted from these records, put together with the labels indicating the level of boredom and fed into a statistical classifier implemented with a tree-based algorithm. Though the specificity of the use case and the highly controlled setting of the experiment make the results of the study not widely applicable, the study has its significance in suggesting a novel methodology for using mouse movements to detect changes in emotional state.

Physiological Data Approach. There has been extensive research conducted on physiological signals that mental loads carry. Wealth of novel smartwatches and fitness trackers being released to the market are enabling better understanding of time-dependent physiology of affective disorders, and [25] provides a general overview. In particular, physiological responses studied in correlation with varying stress levels include skin temperature, heart rate, blood pressure and respiration patterns [25].

3 Clara: A System for Passive Sensing of Depression, Stress, and Anxiety in the Workplace

Hereinafter, a system named Clara is proposed with aims to enable unobtrusive measurement and organizational management of workplace mental wellness. The system uses sensing technology that represents and predicts the internal state of one's mental wellness through digital data, and it provides targeted, personalized tools for management. For employers, the system provides an overall report to enable discovery of insights, which can be utilized to strategize effective interventions when necessary.

3.1 Design Principles

Key design principles that drove the development of Clara are presented. The principles described below outline the framework for decision making used in the process of system development.

Unobtrusive Integration. One key strength of Clara lies in its ability to passively and unobtrusively collect the data that it needs to deliver the service. As one of the ultimate aims of the system lies in work productivity enhancement through better management of workplace mental health, conscious decisions were made throughout the development process to ensure that the system minimizes active inputs from the user as much as possible.

Accurate Sensing. In order to optimize the accuracy of depression, anxiety and stress detection, the system collects behavioral data that have previously been reported to contain statistically meaningful signals of an individual's affective state. Until the accuracy of prediction reaches an acceptable threshold, clinical ground truths of users' mental health are collated every week from their responses to DASS-21, a rigorously validated questionnaire for depression, anxiety and stress screening [11].

Protected Confidentiality. As a large portion of the data collected through the system includes what could be viewed as very personal information, careful thoughts have been put into protecting confidentiality. For instance, historical records of both behavioral data and responses to questionnaires are only viewable by the user himself, and when an organization-wide report is generated for the employer, the aggregated data is presented in a way no specific individual is identifiable.

Protected Information Security. When introducing a new software system, information security is a major concern for many employers. To minimize risk, the system gathers the minimum amount of information needed to extract meaningful features for training the behavioral model. For example, in gathering keystroke data, the exact keystroke is not captured. Instead, binary truth values of whether the keystroke is a special character, a number, a backspace or a delete key are delivered to the server with timestamps. The transfer of any collected data is protected through standard MD5 hashing and encryption algorithms [29].

Engaging User Interface. Lack of engaging user interface is one of the most common reasons behind deleting mental health mobile apps [30]. Therefore, throughout development of Clara, design of user-friendly interface was considered a priority. Examples of user interface on the mobile app are given in Fig. 1.

3.2 System Architecture

The high-level architecture of the backend system is presented in Fig. 2. Descriptions of individual components within the architecture follow.

Data Collection. The current version of Clara collects data from two sources: desktop widget and a mobile phone app. Table 1 displays a list of different data collected through the two mediums. For each type of data collected, the table presents its source, raw data sample, collection trigger event and optionality.

Collection trigger event describes a condition that needs to be met for the system to begin collecting corresponding data points. Whereas collection of keystroke and mouse movement simply begins when employees start using their desktops, more complicated sets of rules are applied to trigger collection of data from the mobile phone. For example, GPS-based location is updated to the system only when the system algorithmically determines that the user has traveled more than a certain distance. This threshold acts as a "distance filter", and its value is dynamically determined by the instantaneous speed captured in the reading. The rationale behind dynamically

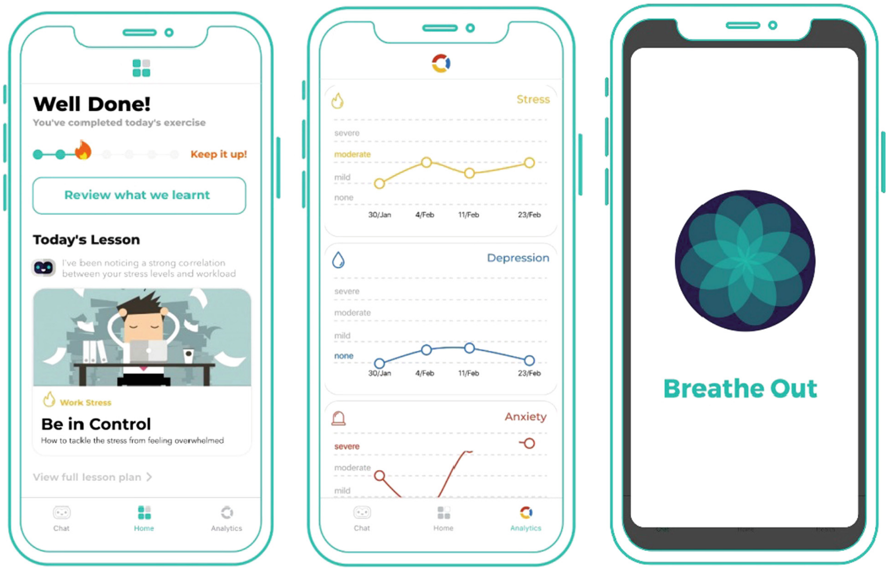

Fig. 1. Clara mobile app user interface

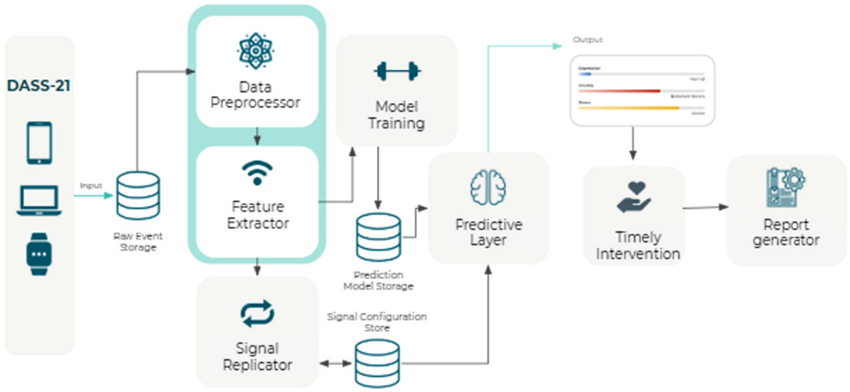

Fig. 2. High-level architecture of the backend system

changing the threshold is to avoid sampling too frequently in the case of high traveling speed. A more detailed description of this implementation can be found in [31], from which the method was taken and adjusted. Likewise, a more detailed description of the trigger methodology for accelerometer implemented in Clara can be found in [32].

Unlike keystroke and mouse movement data that get collected through the desktop widget installed en masse within an organization, the different types of data that get collected on the employee's mobile phone can be selectively opted into the system by each employee. For instance, an employee might choose to provide the system access to his accelerometer data while preventing the system from pulling any data from his

Table 1. Raw data collected by the system

Source	Name	Sample	Trigger event	Optional
Desktop widget	Keystroke dynamic	{timestamp: 1550812094, event_type: KeyUp, is_backspace: True, is_punctuation: False, is_number: False}	Collection triggered when a new key is pressed (event_type = KeyDown) or released (event_type = KeyUp)	No
	Mouse movement	{start_timestamp: 1550812002, end_timestamp: 1550812722, x: 1024, y: 484}	Collection triggered whenever Javascript "onmousemove" event is fired (i.e. when mouse is moved by one or more pixels)	No
Mobile phone app	GPS	{timestamp: 1550989802, latitude: 46.81006, longitude: 92.08174, speed: 18, heading: SW}	Trigger method advanced by [31]. Only collects GPS data when distance travelled exceeds the value of parameter "distance filter", which is dynamically determined by instantaneous speed. Reading is made every 10 min.	Yes
	Accelerometer	{timestamp: 1582091809, a_x: 0.047, a_y: -0.068, a_z: 1.208}	Trigger method advanced in [32]. Accelerometer data collection is initiated when the difference between the magnitude of acceleration in one reading and that of the previous reading is greater than the threshold (T = 1.1 m/s2) for 10 consecutive readings. A reading is made every 1 s.	Yes

mobile phone. While higher level of personalization in both screening and delivery of care incentivizes employees to give access to a variety of data points, the degree of flexibility enabled by allowing employees to selectively provide access makes the system capable of catering to a larger group of people with varying levels of comfort in data sharing.

Collection of aforementioned data occurs concurrently with assessment of the employee's mental state through standardized questionnaires. The current version of Clara focuses on measuring employee's depression, anxiety and stress levels, and employees are prompted to fill out the DASS-21 questionnaires every week. The responses to the questionnaires act as clinical ground truth labels for employees' mental state of the week.

Data Preprocessing. Raw data stored in the central database go through series of preprocessing steps before reaching feature extraction. Common high-level preprocessing steps across different data types consist of handling insufficient and erroneous data and de-noising signals.

As keystroke dynamics and mouse movements are captured rather seamlessly with close-to-zero noise via desktop widgets, no particular cleaning or preprocessing strategy specific to these data types is presented. However, in handling GPS and accelerometer data, thorough considerations were made to ensure that the system captures from the data as much meaningful signals as possible. The details of these preprocessing are given below.

GPS Data. Shielding effects and battery exhaustion are some common reasons behind GPS signal losses. To handle missing GPS data in the system, a matrix imputation method based on matrix factorization (MF) [33] is implemented per interval of 7 days. It is worth noting that such MF-based techniques have traditionally been used to provide collaborative recommendations in location-based services [33].

For the purpose of imputing missing spatial values in our system, two separate matrices are constructed using respectively longitudes and latitudes of collected GPS data points. Each element in the matrix represents one GPS data point, and the elements in one row collectively represent data points collected in one day. In each matrix, inferences of missing points are made by factorization, which aims to minimize the errors of the observed points. Using this factorization, both the longitudinal matrix and the latitudinal matrix are reconstructed with missing values imputed.

Accelerometer Data. A particular instance of accelerometer signal is considered insignificant if the motion it captures is not of physical movement of the user. In other words, signals carried by accelerometer when the user simply picks up the phone should not be considered as a valid acceleration input for the model. To correctly identify and remove intervals of such signals, the system calculates the rate of change of acceleration, also known as the jerk. The underlying assumption is that high jerk is associated with physical movement of the individual, whereas low jerk is associated with usage of the phone without the movement. Based on this understanding, a threshold for jerk is defined, and the number of instances of the signal whose jerk is below this threshold is counted in each interval. The ratio of this number to the total number of entries in the interval is used to determine significance of the movement captured in that interval. When this ratio falls under 0.5, the interval is discarded as containing no movement. This approach was previously implemented in [32].

Feature Extraction. Avariety of features are extracted from streams of preprocessed data. A comprehensive summary is found in Table 2. A few features require further elaborations, which are provided below.

Table 2. Features extracted from different data streams

Data stream	Feature extracted	Description and comments	Unit
Keystroke	Average pause length	Total pause time/total number of pauses. (Note. threshold of pause T = 500 ms of no keyboard input)	s
	Pause rate	Total number of pauses/total number of keystrokes	–
	Time per keystroke	Total input time/total number of keystrokes	s
	Adjusted time per keystroke	(Total input time - total pause time)/total number of keystrokes	s
	Digraph timing features	Timing between different sets of key down and key up events in digraphs	s
	Trigraph timing features	Timing between different sets of key down and key up events in trigraphs	s
Mouse movement	Average movement speed	Total distance traveled by the cursor during activity/active time during which distance was measured	pixel/s
	Latest average speed before inactivity	Average of movement speed during 120 s before inactivity	pixel/s
	Mouse inactivity occurrences	Number of inactivity occurrences (Note. threshold of inactivity T = 500 ms of 0 change in coordinate)	–
	Average duration of mouse inactivity	Total duration of inactive sessions/mouse inactivity occurrences	s
	Horizontal-to-total movements ratio	Number of horizontal mouse movements/total number of mouse movements	–
	Vertical-to-total movements ratio	Number of vertical mouse movements/total number of mouse movements	–
GPS	Location variance	Logarithm of sum of squared longitudinal variance and squared latitudinal variance	deg^2
	Time spent in moving	(Note. threshold of stationary state T = 1 km/hour for speed)	h
	Total distance traveled	–	km
	Average moving speed	Total distance traveled/time spent in moving	km/h
	Number of unique locations	Number of unique clusters formed by the DBSCAN aclustering	–
	Entropy	$-\sum p_i log\, p_i$ where p_i = percentage of time that a participant spends in location cluster i	–
	Homestay	Time spent at home. Calculated by identifying the location cluster at which most data points are recorded at 0-6am window	h

(*continued*)

Table 2. (*continued*)

Data stream	Feature extracted	Description and comments	Unit
Accelerometer	Average step duration	Total time for movement recordings/number of steps	s
	Average acceleration magnitude	Root-mean-square sum of accelerations in x, y, and z directions	m/s^2
	Standard deviation of acceleration magnitude	$\sqrt{\frac{1}{M}(\sum_{i=0}^{M-1}(a_i - \bar{a})^2}$ where $\bar{a}$ = average acceleration magnitude	m/s^2
	Average peak acceleration	Average of peak acceleration magnitude in each step	m/s^2

Digraph and Trigraph Features from Keystroke Data. Digraph and trigraph features are extracted by forming two-letter combinations and three-letter combinations in the keystroke data. Digraph features calculated are as follows: difference between timestamps of the first key down and second key down, first key up and second key down, and first key down and last key up. Similarly, trigraph features calculated are as follows: difference between first key down and third key down and first key down and third key up. Also known as duration and latency features, these features are extracted to better capture the rhythm of users' typing on the keyboard.

Directional Features from Mouse Movement. Previous studies have discovered that users frequently make horizontal mouse movements as they gaze through information in paragraphs [34]. In the same studies, vertical mouse movements were more frequently observed when users gazed through dropdowns and menus [34]. These findings suggest coordination between an individual's mouse movements and his reading patterns, which could potentially have indicative signals of his mental state.

For application in our system, mouse movement directions have been categorized into vertical movement and horizontal movement depending on the angle the trajectory of the cursor movement forms with the horizontal axis parallel to the bottom of the laptop screen. Figure 3 illustrates this mapping.

Cluster-Based GPS Features. Cluster-based GPS features refer to features derived from location clusters, created internally in the system by an unsupervised algorithm called DBSCAN [35]. Our implementation of DBSCAN algorithm requires two parameters, epsilon and min_points. These two parameters are used to separate the GPS coordinates into three different categories as one of the following: i) core point if the number of points around the coordinate is greater than min_points within epsilon, ii) border point if the number of points around the coordinate is less than min_points within epsilon and if there's at least one core point within epsilon, and iii) noise point if the coordinate falls under neither of the two aforementioned categories. Upon removal of noise points, the remaining points are iteratively categorized into a cluster until no neighbor, as determined by the reach of epsilon, is found. When this happens, one of the remaining, uncategorized core points is added to a new cluster and its neighbors are again added iteratively. At the end of all iterations, all points end up being labelled with a cluster id.

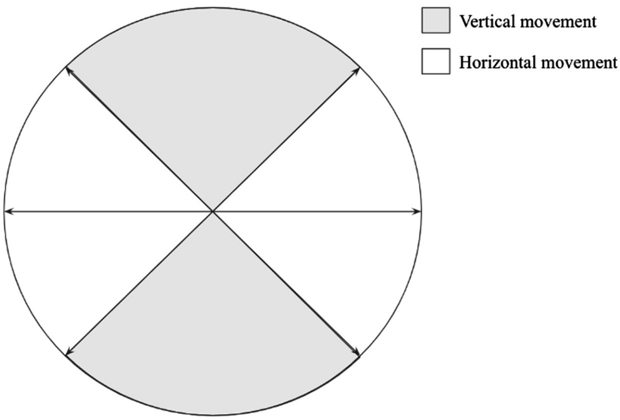

Fig. 3. Directional mapping of mouse movement

The clusters created through this algorithm represent the locations that the user visits, based on the GPS data. The number of unique clusters is therefore indicative of the number of unique locations the user visits, and it is used as a feature in our model. Likewise, for each cluster, a measure of entropy can be calculated using the equation provided in the description box of Table 2. It is worth noting that this implementation has previously been used in [31] to measure the variability of time that an individual spends at different locations.

Step Segmentation from Accelerometer Data. In our system, we consider a single step as the motion that takes place between two consecutive occurrences of the same foot making contact with the ground. A segmentation algorithm advanced in [36] is used to calculate step-based features from the raw accelerometer data. Based on the assumption that the local maxima of the acceleration vector are associated with footfalls, it uses the derivative of the smoothed acceleration, jerk, to identify occurrences of a foot hitting the ground. A low-pass filter, digital Butterworth filter with a critical frequency of 0.05, is used for smoothing purpose before the calculation of jerk.

Once steps are identified within the accelerometer data stream, features such as average step duration and average peak acceleration are calculated.

Model Training and Prediction. Features extracted from the raw data get aggregated into vectors labelled with corresponding ground truth values derived from DASS-21 responses. In such feature aggregation, we consider an interval of 7 days; each interval ends with the day an employee fills in a DASS-21 questionnaire, and the data points collected within the interval are put together with the DASS-21 score calculated from the responses. Per single set of DASS-21 response, three of such feature vectors are constructed, each one labelled with a severity rating for either depression, anxiety or stress. Following the standard DASS-21 scoring scheme, a severity rating that acts as a label for supervised learning can take on one of the following values: normal, mild, moderate, severe and extremely severe.

Upon construction, the labelled feature vectors are fed into a supervised learning algorithm, a variant of Adaboost. The decision to implement an Adaboost variant for model training was motivated by the sparsity of datasets constructed in the system. To better illustrate this concept, consider users who choose to opt-out of accelerometer data collection for concerns ranging from privacy issues to battery drainage. The user's dataset will then have missing values in columns that are properly populated in datasets of other users who did not opt-out of accelerometer data collection. In the latter group of users, accelerometer features may be highly discriminative in predicting affective states, and thus, simply disregarding accelerometer columns would be a poor decision in modeling.

To overcome the problem regarding missing values, a variant of AdaBoost is used. First practical boosting technique introduced, AdaBoost uses weighted ensemble of "weak classifiers" or "base classifiers", which individually has a slightly higher performance than a random classifier, to produce a strong classifier of better performance [37]. The algorithm is also described as a weighted majority-voting scheme because each vote (i.e. classification) from a weak learner is weighted based on its error. For the purpose of binary classification, the traditional AdaBoost algorithm denotes a negative class with -1 and a positive class with +1 [37]. In the variant of Adaboost implemented in Clara, another class, denoted by 0, is added to represent a case in which a base classifier abstains from voting. Advanced in [38], the "Adaboost that abstains" allows a base classifier to "abstain" or output 0 when the particular feature with which it performs learning is missing value. Following this strategy, the model becomes resilient to missing data and is able to handle datasets with large portion of features missing.

With datasets constructed with three labels (depression severity, anxiety severity and stress severity), three models are trained using the aforementioned algorithm. The weights of the learned models are stored in the backend of the system to make real time predictions of corresponding severity levels.

Intervention Delivery and Report Generation. Trigger rules for intervention delivery are as follows: (i) when the predicted severity of at least one affective states out of depression, anxiety and stress reaches "extremely severe" and (ii) when the predicted severity of at least two affective states out of depression, anxiety and stress reach "severe".

A set of interventions are available for users to choose from. One example of an effective intervention built into the system is respiratory biofeedback [39]. In this intervention, the user is given a set of instructions for breathing and relaxation exercises, and the effects of these exercises are captured and made visually available to the user real time. A screenshot of breathing exercise delivered to user as part of this intervention is provided in Fig. 1.

To enable discovery of key insights around the state of workplace mental wellbeing at an organizational level, reports are generated and delivered to employers monthly.

4 Limitations and Future Work

As it currently stands, the system has a few limitations that need modifications through future works. More appraisals around buy-in and trust from different stakeholders are needed to further validate the system.

4.1 Naive Handling of Missing GPS Data

The current method of handling missing GPS data involves usage of MF-based imputation. A brief study on this methodology using a small sample of GPS dataset shows that the approach lacks the robustness needed for a more meaningful usage of the GPS data. Measuring the accuracy of the method involves taking out values from a latitudinal/longitudinal matrix, reconstructing the matrix using imputation and comparing the output with the original values of the matrix. In the preliminary study with a dataset whose missing values were limited to less than 10% of the size of the whole dataset, the MF-based imputation yielded data points, more than 15% of which had greater difference from the ground truth values than the accepted threshold (T = 500 m).

In order to improve the significance of GPS signals captured, a more effective strategy in handling missing GPS data should be implemented. Some suggestions can be found in [40], where a principled statistical approach is explored based on weighted resampling of the observed data.

4.2 Exclusion of Physiological Signals

As explored in Sect. 2.2.4, physiological signals such as heart rate and skin temperature have been reported as powerful indicators of an individual's affective state. While the current version of Clara only collects behavioral data, integration of physiological signals in the future versions is expected to improve predictive capability of the system. For capturing such signals, however, sensor hardware is required. The upcoming system update will therefore include addition of a pipeline that gathers data from such sensors-equipped wearables and aggregates them to the central storage system before corresponding preprocessing and feature extractions.

5 Conclusion

Clara is an unobtrusive, data-driven monitoring and management solution for workplace mental health. In this paper, we have addressed the limitations of questionnaire-based mental health assessments that Clara overcomes and presented a review of passive sensing technologies that power Clara. We have discussed the overall system architecture of Clara before delving into the details of implementing the smaller components within the system. Last, based on the limitations identified in the first version, we have proposed areas of future work.

With their detrimental effects on productivity, affective disorders at the workplace have an extremely high price tag. Clara, marked by its timely interventions based on passive monitoring technology, has the potential to provide employers of the workplace a solution to effectively manage the mental wellbeing of their employees.

6 Conflict of Interest Disclosure

The co-authors Megan Lam and Caleb Chiu reported holding Board Director positions at Neurum Ltd. No other authors have disclosures to declare.

References

1. Mind the Workplace Report. http://www.mentalhealthamerica.net/tags/mind-workplace-report. Accessed 12 Jan 2019
2. Higginbottom, K.: Poor Mental Health Is Widespread in the Workplace. https://www.forbes.com/sites/karenhigginbottom/2018/09/14/poor-mental-health-widespread-in-the-workplace/#5a67868e72f8. Accessed 12 Jan 2019
3. Creating a mentally healthy workplace - headsup.org.au. https://www.headsup.org.au/docs/default-source/resources/beyondblue_workplaceroi_finalreport_may-2014.pdf?sfvrsn=6. Accessed 12 Jan 2019
4. Dewa, C.S., Thompson, A.H., Jacobs, P.: The association of treatment of depressive episodes and work productivity. Can. J. Psychiatry **56**, 743–750 (2011)
5. Work and Well-Being Survey - apaexcellence.org. https://www.apaexcellence.org/assets/general/2016-work-and-wellbeing-survey-results.pdf
6. Headspace Homepage, https://www.headspace.com. Accessed 15 Jan 2019
7. Partnership for Workplace Mental Health. http://www.workplacementalhealth.org/Case-Studies/Delta-Air-Lines. Accessed 16 Jan 2019
8. Graham, A.: Review: brief interventions reduce drinking in patients not seeking treatment. Evid.-Based Ment. Health **5**, 116 (2002)
9. Razavi, T.: Self-report measures: an overview of concerns and limitations of questionnaire use in occupational stress research
10. Jones, G., Wright, J., Regele, O., Kourtis, L., Pszenny, S., Sirkar, R., Kovalchick, C.: Evolution of the digital biomarker ecosystem. Digit. Med. **3**, 154 (2017)
11. Lovibond, P., Lovibond, S.: The structure of negative emotional states: comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories. Behav. Res. Ther. **33**, 335–343 (1995)
12. Kroenke, K., Spitzer, R.L.: The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Ann. **32**, 509–515 (2002)
13. Griffiths, T.C., Myers, D.H., Talbot, A.W.: A study of the validity of the scaled version of the general health questionnaire in paralysed spinally injured out-patients. Psychol. Med. **23**, 497 (1993)
14. Pruebas Herramienta de Documentalista. https://www.neps-data.de/Portals/0/Working%20Papers/WP_XLV.pdf. Accessed 20 Jan 2019
15. Cohen, S., Kamarck, T., Mermelstein, R.: A global measure of perceived stress. J. Health Soc. Behav. **24**, 385 (1983)
16. Koh, K.B., Park, J.K., Kim, C.H., Cho, S.: Development of the stress response inventory and its application in clinical practice. Psychosom. Med. **63**, 668–678 (2001)

17. Sakakibara, B.M., Miller, W.C., Orenczuk, S.G., Wolfe, D.L.: A systematic review of depression and anxiety measures used with individuals with spinal cord injury. Spinal Cord **47**, 841–851 (2009)
18. Gross, C.: The standard stress scale (SSS): measuring stress in the life course. In: Blossfeld, H.-P., von Maurice, J., Bayer, M., Skopek, J. (eds.) Methodological Issues of Longitudinal Surveys, pp. 233–249. Springer, Wiesbaden (2016). https://doi.org/10.1007/978-3-658-11994-2_14
19. Rock Health Homepage, https://www.rockhealth.com/reports/. Accessed 02 Feb 2019
20. Chow, P.I., Fua, K., Huang, Y., Bonelli, W., Xiong, H., Barnes, L.E., Teachman, B.A.: Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J. Med. Internet Res. **19**, e62 (2017)
21. CBITs TECH Website, https://tech.cbits.northwestern.edu/purple-robot/. Accessed 03 Feb 2019
22. Hellhammer, D., Meinlschmidt, G., Kumsta, R.: Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. F1000 - Post-publication peer review of the biomedical literature (2016)
23. Epp, C., Lippold, M., Mandryk, R.L.: Identifying emotional states using keystroke dynamics. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI 2011 (2011)
24. Maheshwary, S., Pudi, V.: Mining keystroke timing pattern for user authentication. In: Appice, A., Ceci, M., Loglisci, C., Masciari, E., Raś, Zbigniew W. (eds.) NFMCP 2016. LNCS (LNAI), vol. 10312, pp. 213–227. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61461-8_14
25. Malhi, G.S., Hamilton, A., Morris, G., Mannie, Z., Das, P., Outhred, T.: The promise of digital mood tracking technologies: are we heading on the right track? Evid. Based Ment. Health **20**, 102–107 (2017)
26. Solanki, R., Shukla, P.: Estimation of the user's emotional state by keystroke dynamics. Int. J. Comput. Appl. **94**, 21–23 (2014)
27. Tsoulouhas, G., Georgiou, D., Karakos, A.: Detection of learner's affective state based on mouse movements. J. Comput. **3**, 9–18 (2011)
28. Hibbeln, M., Jenkins, J.L., Schneider, C., Valacich, J.S., Weinmann, M.: How is your user feeling? Inferring emotion through human-computer interaction devices. MIS Q. **41**, 1–21 (2017)
29. Heron, S.: Advanced encryption standard (AES). Netw. Secur. **2009**, 8–12 (2009)
30. Torous, J., Nicholas, J., Larsen, M.E., Firth, J., Christensen, H.: Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid. Based Ment. Health **21**, 116–119 (2018)
31. Farhan, Y., Morillo, W., Lu, B., Kamath, R., Bamis, W.: Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. In: 2016 IEEE Wireless Health (WH) (2016)
32. A simple method for reliable footstep detection on embedded sensor platforms. http://ozeo.org/lib/exe/fetch.php?media=bodytrack:libby_peak_detection.pdf
33. Zheng, V.W., Zheng, Y., Xie, X., Yang, Q.: Towards mobile intelligence: Learning from GPS history data for collaborative recommendation. Artif. Intell. **184–185**, 17–37 (2012)
34. Rodden, K., Fu, X., Aula, A., Spiro, I.: Eye-mouse coordination patterns on web search results pages. In: Proceeding of the Twenty-Sixth Annual CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI 2008 (2008)
35. Louhichi, S., Gzara, M., Abdallah, H.B.: A density based algorithm for discovering clusters with varied density. In: 2014 World Congress on Computer Applications and Information Systems (WCCAIS) (2014)

36. Pakhomov, A., Sicignano, A., Sandy, M., Goldburt, E.T.: A novel method for footstep detection with extremely low false alarm rate. In: Unattended Ground Sensor Technologies and Applications V. (2003)
37. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**, 119–139 (1997)
38. Conroy, B., Eshelman, L., Potes, C., Xu-Wilson, M.: A dynamic ensemble approach to robust classification in the presence of missing data. Mach. Learn. **102**, 443–463 (2015)
39. Meuret, A.E., Wilhelm, F.H., Roth, W.T.: Respiratory biofeedback-assisted therapy in panic disorder. Behav. Modif. **25**, 584–605 (2001)
40. Barnett, I., Onnela, J.-P.: Inferring mobility measures from GPS traces with missing data. Biostatistics (2018)