Proc. IEEE International Conference on Cognitive Infocommunications, Debrecen, 2017.

Are You Stressed? Detecting High Stress from User Diaries.

Arindam Ghosh, Evgeny A. Stepanov, Morena Danieli, Giuseppe Riccardi
Signals and Interactive Systems Lab,
Department of Information Engineering and Computer Science,
University of Trento, Italy
{arindam.ghosh, evgeny.stepanov}@unitn.it
{morena.danieli, giuseppe.riccardi}@unitn.it

***Abstract*—Knowledge of the complete clinical history, lifestyle, behaviour, medication adherence data, and underlying symptoms, all affect the treatment outcomes. Collecting, analysing and using all these data, while treating a patient can often be very challenging. A doctor can spend only a limited time with a patient. This time is often not enough to learn about all the lifestyle and underlying conditions of a patient's life. Often patients are asked to maintain diaries of their daily activities. Diaries can help to improve adherence by increasing the consciousness of the patients, and can also serve as a way for the doctors to validate this adherence. However, diaries can be cumbersome to parse, and hence increase the task burden of the doctor. In this paper we demonstrate that automatic analysis of diaries can be used to predict the stress level of the diary writers with an F-measure of 0.70.**

I. Introduction

Stress is a complex phenomenon which plays an important role in our daily lives. It has been defined as "the non-specific response of the body to any demand made upon it" [1]. We experience varying levels of stress in our daily lives, and the human body is capable of handling and recovering from a limited amount of it. However, prolonged and chronic exposure to continuous high stress can have negative effects on a person's physical and mental well-being. Prolonged exposure and reaction to stress has become a common explanation for disturbed human behaviour, failure and psychological breakdown. It has been strongly linked to numerous chronic health risks, such as cardiovascular disease [2], [3], diabetes mellitus [4], obesity [5], [6], hypertension [7], and coronary artery disease [8], [9].

Due to the adverse effects of exposure to continuous stress, one of the goals of maintaining a healthy lifestyle is to closely monitor and decrease everyday stress levels. While severe and acute physical or psychological stress is easy to detect due to its immediate visible manifestations, chronic stress in our everyday lives often goes undetected for prolonged periods. By the time people seek out medical assistance, they are already in some advanced form of stress induced exhaustion. For people already suffering from some chronic condition (such as diabetes, depression, etc) this can lead to aggravation of the condition [9]. Since the occurrence of stress is closely related to lifestyle factors, physicians emphasize the importance of making healthy lifestyle choices for lowering stress levels in everyday life. Lifestyle modifications such as regular physical activity, dietary changes, and practicing controlled relaxation and meditation have been shown to be effective for stress reduction [10], [11], [12]. While a physician may advise a patient to walk more, eat healthier, and decrease their stress level; once the patient is out of the clinic, it is often challenging to ensure and track adherence to the prescribed directives. For patients suffering from chronic conditions, researchers have suggested the use of an intermediate care team [13], which can follow the patient and ensure the protocol adherence. However, setting up a care-team is complex and expensive. In most cases, the burden of the tracking and following a patient's lifestyle falls upon family members or the patients themselves.

To simplify the process, physicians often rely upon self-adherence and self-monitoring by patients. To increase self-adherence to healthy lifestyle choices, doctors may advise patients to maintain diaries of their daily activities. Diaries are a popular life-logging tool for storing memories and aiding recollection [14], [15]. Along with tracking activities and logging behaviour, they have also been effective for measuring and monitoring medical adherence [16] and compliance to treatment regimes [17], [18]. Diaries have shown to improve adherence by increasing the consciousness of the patients about their condition. They have proven to be very effective in gaining deep insights into a patient's well-being and can be used by a doctor or a therapist to learn about the patient's behaviour and routines [19]. In psychology research, diaries have been used for promoting psychological recovery for patients suffering from various symptoms such as insomnia, anxiety, depression, and post-traumatic stress disorder [20], [21]. Researchers have demonstrated that maintaining a detailed diary in the form of a reflective journal can not only be used to track stress, but can be therapeutic and can help to even reduce stress levels and improve health management [22].

Traditionally, patient diaries were pen and paper based. In recent times, with the increase in the usage of smart devices (smart-phones and tablets) capable of recording text, audio, images and video, digital diaries are gaining popularity. These digital diaries are making it easier for patients to maintain regular continuous real-time diaries. Dale et al [23] in 2007, reviewed nine studies to demonstrate that digital diaries or personal digital assistants (PDAs) improved protocol compliance, data accuracy, and acceptance by subjects over pen and paper diaries. Another more recent meta-review by Sharp et al. [24] in 2014, which compared pen and paper based note-taking with digital diaries for monitoring dietary habits, concluded that compliance and patient satisfaction was higher in studies using digital diaries. Walker et al. [25] through a 6-month study compared the use of digital diaries with paper based ones. They showed that the compliance of the patients using digital diaries was about 86.2% compared with 48.3% for patients using pen and paper alternatives. One major disadvantage of paper-based diaries which affects compliance, is that most patients do not carry the diaries with them all the time. Human memory is transient, and patients often forget details of events, feelings, and conditions over time. Often, by the time they write down their experience, some of the details are already lost from memory.

Digital diaries can act as the technological solution capable of monitoring and assisting patients in their daily lives and ensuring that doctors receive a clear and complete picture of the patient's lifestyle. The recent growth in the sensing and processing power of smartphones have turned them into ubiquitous computing devices. The omnipresence of smartphones in people's lives have turned them into effective platforms for life-logging and monitoring users' signals and behaviours in on-the-go scenarios. Smartphones have become a time and cost effective platform for data-collection and research in social [26], health [27], behavioural [28], and clinical [29], [30] sciences. Hence, in our research, instead of paper diaries, we employed a smartphone-based personal agent platform to assist users in proactively collecting daily on-the-spot diary entries about their mental and physical states. Since most users are always in close proximity to their phones, using smartphones as tools to elicit diary entries allows the users to make entries close to when the events occur. A smartphone-based agent provides an added advantage of being able to elicit notes in both spoken, written and image form.

In this paper we explore the use of a smartphone based diary to assist real-time annotations from the patients in on-the-go scenarios. Applying machine learning techniques, we demonstrate that these diary entries can be used to automatically detect the stress level of the patients. Researchers have demonstrated that physiological signals such as galvanic skin response (also known as electrodermal activity), heart rate variability (HRV) and skin temperature can be useful for recognizing stress levels in everyday life. Galvanic Skin Response (GSR) has been proven to be highly successful in predicting stress levels under various settings while driving [31], working in an office [32] and in a call center [33]. In a previous work [34], we demonstrated that by combining different physiological signals (heart rate, skin temperature, galvanic skin response), we can achieve a very accurate prediction of perceived stress in uncontrolled real-life scenarios. However, while there has been a lot of excitement regarding wearable devices, their adoption in everyday life has been limited. Another drawback is that popular wearable devices mostly record a single physiological signal (usually heart rate), at very low sampling frequencies. Most state of the art stress recognition algorithms require signals recorded at high sampling rates to provide acceptable performances. Accounting for the variability in the quality and sampling rate of the various devices is a challenging task. Due to these limitations, we are motivated to explore patient diaries which are less susceptible to noise, for identifying stress levels in on-the-go scenarios.

This paper is organized as follows. In Section II we describe the data collection protocol. In Section III we perform the preliminary data analysis and explain the methodology for aggregating the expert annotations. In Section IV we discuss our machine learning experiments and present the results. Finally in Section V we present our conclusions and future work.

II. Data Collection Protocol

An observational pilot study was conducted with 10 hypertensive and 10 normotensive adults for 10 days each. Adults (male and female) between the ages of 30 and 65 were selected for the study. Patients suffering from essential hypertension and receiving treatment were recruited at the ***Centro Ipertensione Ospedale Molinette*** in Turin, Italy. The healthy control (normotensive) subjects were recruited by a psychologist to rule out hidden hypertension, and any other underlying health problems that might affect the study. The data collection protocol was the same for both groups. The institutional ethics committee of the Azienda Ospedaliera Città della Salute e della Scienza di Torino and the ethics committee of the Università degli Studi di Trento approved the present research study and the data collection protocol. All data was anonymized before analysis. Each participant was provided with an Empatica E3 wearable wristband and an iPhone with an installed agent application capable of recording and transmitting the data to a secure cloud server.

During the first interview, the participants were instructed on how to use the wristband and the application, and were evaluated by a psychologist to identify their perceived stress levels and emotion regulation techniques. During this interview, they also signed an informed consent for participation. Each participant was instructed to maintain a periodic electronic diary to capture daily events, interactions, mood and reflections. To ease the burden of diary writing, they were provided with the option of either writing the diaries as free text, or taking vocal notes which were later transcribed by the system.

The participants were also instructed to record their perceived stress levels at regular intervals – twice a day. The first record of a reported stress-level was done around mid-day for the morning session, and the second record was at the end of the day for the afternoon session.

The participants were also presented with the emotion regulation questionnaire (ERQ), which is a ten item scale [35] designed by Gross and John (2003). Emotion regulation refers to the process that individuals use to feel, express,

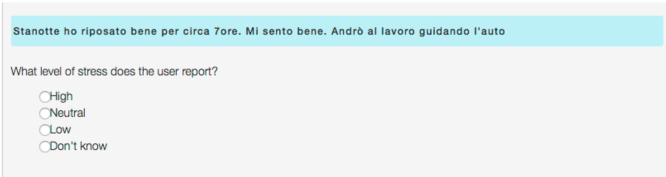

Fig. 1. The screen for the annotation task for the experts to annotate the stress level indicated by the diary entry.

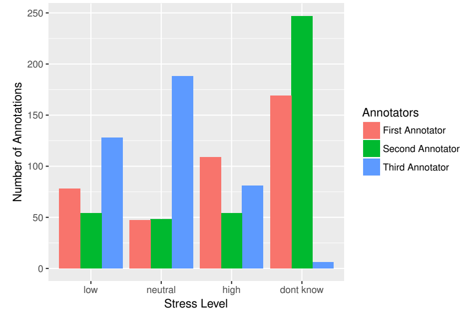

Fig. 2. The distribution of stress level annotations by each of the three expert annotators.

and control the emotions they experience in their daily lives [36]. The questionnaires were used to measure how the respondents use cognitive reappraisal and expressive suppression strategies for emotion regulation. Studies using the emotion regulation questionnaire have demonstrated that increased use of expression suppression strategies is linked to mental problems such as anxiety [37] , stress [38] and depression [39].

Different individuals use the emotion regulation dimensions differently; and this can affect the way they perceive and deal with stress. People who score high on the emotion suppression scale tend to suppress their emotional response to an event, reporting an event as less stressful than they actually feel it to be. Therefore a simple stress level annotation task is more susceptible to be affected by emotion suppression strategies than detailed notes and diaries. Since diaries are more narrative in nature, they do not explicitly require the participants to quantify their stress and emotions; thus, the effect of emotion suppression is less pronounced.

Diaries are rich in contextual information and can provide a deeper insight into the mental health of the users. However, manually parsing and evaluating hundreds of diary entries can be a manual labour intensive process. Therefore, in the following sections we apply machine learning to automatically analyse the patient diaries for identifying the stress level.

III. Data Analysis and Methodology

In this section we analyse the textual data from the diary entries and the stress levels reported by the patients. We collected a total of 245 self-annotations of reported stress from the participants. Of these, 154 sessions were annotated as low stress and 91 sessions were annotated as high stress. We also collected 465 text-based diary entries for these sessions, with 263 diary entries taken during the sessions annotated as low stress and 202 diary entries taken during sessions marked as high stress. Our goal is to use the diary entries to automatically predict the stress levels.

A. Collecting Expert Annotations

The diary entries can be categorized into two types: a) Functional entries, where the users noted down some task they had performed. b) Emotional entries, where the users annotated more detailed events, their reaction to these events, and how it affected their stress level and general well-being. We observed that most short text annotations were functional entries (for example: drank coffee, went for a walk), and did not have any identifiable stress indicators. Therefore, as a preprocessing step, we removed all text entries with shorter than five words.

For the rest of the diary entries, we adopted a supervised machine learning approach to automatically detect the level of stress expressed in these entries. Since the user provided stress annotations were at a session level, and not at the level of each diary entry, the first step was to annotate the stress level of each individual diary entry. Consequently, we needed to obtain ground truth annotations. A group of three experts – psychologists by profession – were recruited for the task. The task involved reading each diary entry and identifying whether the expressed stress level was high, neutral, low. They also had an option of reporting that they were uncertain of the stress level by annotating it with "don't know" (see Fig. 1).

The task of detecting stress levels by reading diary entries, is a *subjective annotation task*. A subjective annotation task is qualitative in nature where the provided annotation or rating not only depends on the content of the task but also on the characteristics and opinions of the annotator. The expected agreement in subjective tasks such as rating movies, or detecting irony or sarcasm from text, is usually low. In our case, even though the annotators were trained psychologists, since the stress annotation task is also subjective in nature, we observed a moderate agreement among the annotators. The figure Fig. 2 shows the distribution of the labels provided by the three annotators.

An important factor for machine learning is the consistency of the annotations. In order to measure the quality and reliability of the collected annotations, we need to calculate the agreement among the annotators. Percent agreement is the simplest measure of agreement – it reports the percentage that the coders agree in their ratings. We obtain a value of 42% agreement for our three expert raters. However, percent agreement does not take into account cases where the agreement might be due to chance. Thus, we compute the Fleiss' Kappa [40], which is a statistical measure of inter-annotator reliability in a multi-annotator setting.

The value of Fleiss' Kappa we obtain is 0.338. This value indicates a fair agreement among the three annotators. As a

TABLE I
PAIRWISE COHEN'S κ FOR EACH OF THE ANNOTATOR PAIRS.

Annotator Pair	κ
one and two	0.46
one and three	0.32
two and three	0.23

next step, we identify if the agreement is the same between all the three annotators; and calculate pairwise agreements using Cohen's Kappa [41].

From the pairwise Cohen's Kappa among the three annotator's reported in Table I, we can observe that annotators *one* and *two* have the highest agreement, and the annotator *three* has a comparatively low agreement with the other two annotators.

In order to arrive at a single label for the annotation, we heuristically obtain aggregate ground-truth stress labels for the diary entries as follows:

- An entry is considered as a high stress, if **any** of the annotators labeled it as such.
- If none of the annotators labeled the entry as high stress, majority voting of the expert labels is used as a final label.
- In case majority cannot be decided, the entry is marked as low stress.

The labels generated after applying the heuristics are used as a supervision signal to train and evaluate supervised classification models.

IV. EXPERIMENTS AND RESULTS

A. Text Pre-processing and Feature Extraction

Before the application of machine learning to classify the stress level from a diary entry, we perform automatic text pre-processing. We apply standard text pre-processing steps widely used in Natural Language Processing (NLP): tokenization, lemmatization, lowercasing, and removal of the stop-words (using Italian stop-word list).

We make use of three kinds of features – user, linguistic, and psycholinguistic. While linguistic and psycholinguistic features are extracted from the pre-processed individual diary entries, the user features characterise the authors of the entries.

a) User Features: Stress perception is dependent on the individual coping strategies towards stress. The hypothesis is grounded on the observation that people respond to different stressful events in different ways in accordance with the regulatory styles they adopt. Two underlying dimensions have been proposed as primary factors of the emotion regulation process: (a) reappraisal and (b) suppression. Reappraisal is a strategy deployed before the activation of an emotional response, while suppression is a strategy that is deployed when an emotional response has already been triggered. It has been suggested that the suppression strategy may require some effort to manage the emotional response, thus reducing the cognitive and affective resources of an individual. Cognitive reappraisal and expressive suppression scores are used as user features; and they are extracted from the emotion regulation questionnaires that the users had been given during the selection process. The feature values are normalized between 0 and 1.

b) Linguistic Features: We have experimented with three different representations of the text: bag-of-words, bag-of-pos-tags and word embedding vectors.

In the *bag-of-words* (BoW) representation, each diary entry is represented as a tf-idf weighted vector of the training lexicon. The bag-of-words representation is one of the simplest to generate. However, despite its simplicity it has been shown to be highly accurate in document classification.

To generate the *bag-of-pos-tags* (POS) representation, the diary entries were automatically annotated for Part-of-Speech (POS) tags using The treetagger toolkit. Similar to the bag-of-words representation, each diary entry is represented by a tf-idf weighted vector of all possible part of speech tags (38).

For the *word embedding representation* (WE) [42], we use cbow vectors (size: 400, window: 5) pre-trained on Italian Wikipedia dump. Each diary entry is represented as term-frequency weighted average of per-word vectors (400 dimensional vector).

c) Psycholinguistic Features: For extracting the psychilinguistic features we use the Linguistic Inquiry and Word Count (LIWC) [43]. LIWC is a hand coded lexicon of words categorized into several linguistic and psychological categories. LIWC has been widely used for psychological and sociological research in the fields of health, personality, deception, dominance, etc. Using the word counts in the text, the LIWC program produces a vector of per-category scores. The LIWC categories describe positive or negative emotions (happiness, sadness, kindness, anger, etc.), different function words (pronouns, articles, quantifiers, etc.), and paralinguistic properties (accents, fillers, and disfluencies), among others. The categories correlate with various psychological traits, and often provide indications about social skills, personality, etc.

In this study, we have used the Italian lexicon of LIWC [44], which contains 85 word categories. This 85 dimensional vector is used for the classification task.

B. Machine Learning

Using the document representations and the features described in the previous section we detect **high stress** diary entries. The task is cast as binary classification into **high stress** and **other**. As mentioned earlier, the supervision signal is the consensus of the expert annotations.

To ensure user-independence of the models, we use Leave-One-Subject-Out (LOSO) cross-validation setting (20 folds – 1 fold per user). We experiment with different feature combinations through the fusion of their vectors. Even though, we have experimented with different algorithms such as Naive Bayes, Random Forest, AdaBoost, and Support Vector Machine (SVM) with linear kernel. The best results were obtained using SVMs, and for the rest of the paper we report performances of these SVM models.

TABLE II
MICRO-F_1-MEASURES FOR THE HIGH-STRESS DETECTION USING SUPPORT VECTOR MACHINES (LINEAR KERNEL).

Model		F_1
Majority Baseline		0.56
Bag-of-words	(BoW)	0.64
Bag-of-pos-tags	(POS)	0.58
Word Embeddings	(WE)	0.63
Psycholinguistic	(LIWC)	0.67
LIWC + User		0.69
LIWC + User + POS		**0.70**

Since in our setup both categories are of interest, we report micro-F_1 score, which is equivalent to accuracy. The baseline performance is the majority (assign the most frequent label in the training data to every diary entry), and has $F_1 = 0.56$.

Table II reports performances of individual feature models and the best feature combinations. We observe that while bag-of-pos-tags representation yields the lowest performance ($F_1 = 0.58$); the individual performances of bag-of-words, word embeddings, and LIWC yield comparable performances. LIWC yields the best individual performance of $F_1 = 0.67$.

Since expression of stress depends upon certain characteristics of the user making the diary entry, we observe that appending user-level features such as their cognitive re-appraisal and emotion suppression scores leads to an improvement in the performances. Combining LIWC and user features yields F_1-measure of 0.69. Further, the fusion of bag-of-pos-tags and LIWC vectors together with user features yields the best performance of $F_1 = 0.70$.

V. CONCLUSION

The ubiquity of smartphones has made it easy for people to maintain digital diaries. In this paper we have described the tasks of data collection, expert annotation of diary entries, as well as automatic detection of the entries referring to high stress. We demonstrate that it is possible to detect high stress events with acceptable accuracy.

One of the limitations of our approach is the subjectivity of expert annotations. We have observed a low to fair inter-annotator agreement; and used a set of simple heuristics to produce a single ground-truth label. A future extension of the current approach is to improve the expert annotation aggregation methodology using more advanced techniques. The proposed methodology and automatic stress detection models are a step-up from traditional diaries both in terms of ease and accuracy of logging and their analysis, which decreases the burden of both doctors and patients keeping them.

REFERENCES

[1] H. Selye, "The evolution of the stress concept: The originator of the concept traces its development from the discovery in 1936 of the alarm reaction to modern therapeutic applications of syntoxic and catatoxic hormones," *American scientist*, vol. 61, no. 6, pp. 692–699, 1973.

[2] H. Iso, C. Date, A. Yamamoto, H. Toyoshima, N. Tanabe, S. Kikuchi, T. Kondo, Y. Watanabe, Y. Wada, T. Ishibashi *et al.*, "Perceived mental stress and mortality from cardiovascular disease among japanese men and women," *Circulation*, vol. 106, no. 10, pp. 1229–1236, 2002.

[3] T. G. Pickering, "Mental stress as a causal factor in the development of hypertension and cardiovascular disease," *Current hypertension reports*, vol. 3, no. 3, pp. 249–254, 2001.

[4] R. S. Surwit, M. S. Schneider, and M. N. Feinglos, "Stress and diabetes mellitus," *Diabetes care*, vol. 15, no. 10, pp. 1413–1422, 1992.

[5] F.-S. Koch, A. Sepa, and J. Ludvigsson, "Psychological stress and obesity," *The Journal of pediatrics*, vol. 153, no. 6, pp. 839–844, 2008.

[6] P. Björntorp, "Do stress reactions cause abdominal obesity and comorbidities?" *Obesity reviews*, vol. 2, no. 2, pp. 73–86, 2001.

[7] M. Esler, N. Eikelis, M. Schlaich, G. Lambert, M. Alvarenga, T. Dawood, D. Kaye, D. Barton, C. Pier, L. Guo *et al.*, "Chronic mental stress is a cause of essential hypertension: presence of biological markers of stress," *Clinical and Experimental Pharmacology and Physiology*, vol. 35, no. 4, pp. 498–502, 2008.

[8] R. von Känel, P. J. Mills, C. Fainman, and J. E. Dimsdale, "Effects of psychological stress and psychiatric disorders on blood coagulation and fibrinolysis: a biobehavioral pathway to coronary artery disease?" *Psychosomatic Medicine*, vol. 63, no. 4, pp. 531–544, 2001.

[9] S. Cohen, D. Janicki-Deverts, and G. E. Miller, "Psychological stress and disease," *Jama*, vol. 298, no. 14, pp. 1685–1687, 2007.

[10] J. D. Spence, P. A. Barnett, W. Linden, V. Ramsden, and P. Taenzer, "Lifestyle modifications to prevent and control hypertension. 7. recommendations on stress management. canadian hypertension society, canadian coalition for high blood pressure prevention and control, laboratory centre for disease control at health canada, heart and stroke foundation of canada," *CMAJ: Canadian Medical Association Journal*, vol. 160, no. 9, p. S46, 1999.

[11] M. V. Rainforth, R. H. Schneider, S. I. Nidich, C. Gaylord-King, J. W. Salerno, and J. W. Anderson, "Stress reduction programs in patients with elevated blood pressure: a systematic review and meta-analysis," *Current hypertension reports*, vol. 9, no. 6, pp. 520–528, 2007.

[12] M. B. Ospina, K. Bond, M. Karkhaneh, L. Tjosvold, B. Vandermeer, Y. Liang, L. Bialy, N. Hooton, N. Buscemi, D. M. Dryden *et al.*, "Meditation practices for health: state of the research," *Evid Rep Technol Assess (Full Rep)*, vol. 155, no. 155, pp. 1–263, 2007.

[13] J.-M. Guise, "What is intermediate care?" *women*, vol. 15, pp. 232–40, 2001.

[14] J. Machajdik, A. Hanbury, A. Garz, and R. Sablatnig, "Affective computing for wearable diary and lifelogging systems: An overview," in *Machine Vision-Research for High Quality Processes and Products-35th Workshop of the Austrian Association for Pattern Recognition. Austrian Computer Society*, 2011.

[15] A. J. Sellen, A. Fogg, M. Aitken, S. Hodges, C. Rother, and K. Wood, "Do life-logging technologies support memory for the past?: an experimental study using sensecam," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2007, pp. 81–90.

[16] G. J. Wagner, "Predictors of antiretroviral adherence as measured by self-report, electronic monitoring, and medication diaries," *AIDS Patient Care and STDs*, vol. 16, no. 12, pp. 599–608, 2002.

[17] K. C. Farmer, "Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice," *Clinical therapeutics*, vol. 21, no. 6, pp. 1074–1090, 1999.

[18] B. A. Yon, R. K. Johnson, J. Harvey-Berino, B. C. Gold, and A. B. Howard, "Personal digital assistants are comparable to traditional diaries for dietary self-monitoring during a weight loss program," *Journal of behavioral medicine*, vol. 30, no. 2, pp. 165–175, 2007.

[19] E. Gjengedal, S. L. Storli, A. N. Holme, and R. S. Eskerud, "An act of caring–patient diaries in norwegian intensive care units," *Nursing in Critical Care*, vol. 15, no. 4, pp. 176–184, 2010.

[20] R. E. Knowles and N. Tarrier, "Evaluation of the effect of prospective patient diaries on emotional well-being in intensive care unit survivors: a randomized controlled trial," *Critical care medicine*, vol. 37, no. 1, pp. 184–191, 2009.

[21] P. M. Lewinsohn and G. E. Atwood, "Depression: A clinical-research approach." *Psychotherapy: Theory, Research & Practice*, vol. 6, no. 3, p. 166, 1969.

[22] C. S. Mackenzie, U. J. Wiprzycka, L. Hasher, and D. Goldstein, "Does expressive writing reduce stress and improve health for family caregivers of older adults?" *The Gerontologist*, vol. 47, no. 3, pp. 296–306, 2007.

[23] O. Dale and K. B. Hagen, "Despite technical problems personal digital assistants outperform pen and paper when collecting patient diary data," *Journal of clinical epidemiology*, vol. 60, no. 1, pp. 8–17, 2007.

[24] D. B. Sharp and M. Allman-Farinelli, “Feasibility and validity of mobile phones to assess dietary intake,” *Nutrition*, vol. 30, no. 11, pp. 1257–1266, 2014.

[25] I. Walker, C. Sigouin, J. Sek, T. Almonte, J. Carruthers, A. Chan, M. Pai, and N. Heddle, “Comparing hand-held computers and paper diaries for haemophilia home therapy: a randomized trial,” *Haemophilia*, vol. 10, no. 6, pp. 698–704, 2004.

[26] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, “Emotionsense: a mobile phones based adaptive platform for experimental social psychology research,” in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010, pp. 281–290.

[27] A. Madan, M. Cebrian, S. Moturu, K. Farrahi *et al.*, “Sensing the” health state” of a community,” *IEEE Pervasive Computing*, no. 4, pp. 36–45, 2012.

[28] L. Dennison, L. Morrison, G. Conway, and L. Yardley, “Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study,” *Journal of medical Internet research*, vol. 15, no. 4, p. e86, 2013.

[29] A. Pandey, S. Hasan, D. Dubey, and S. Sarangi, “Smartphone apps as a source of cancer information: changing trends in health information-seeking behavior,” *Journal of Cancer Education*, vol. 28, no. 1, pp. 138–142, 2013.

[30] M. Faurholt-Jepsen, M. Frost, M. Vinberg, E. M. Christensen, J. E. Bardram, and L. V. Kessing, “Smartphone data as objective measures of bipolar disorder symptoms,” *Psychiatry research*, vol. 217, no. 1, pp. 124–127, 2014.

[31] J. A. Healey and R. W. Picard, “Detecting stress during real-world driving tasks using physiological sensors,” *Intelligent Transportation Systems, IEEE Transactions on*, vol. 6, no. 2, pp. 156–166, 2005.

[32] B. Cinaz, R. La Marca, B. Arnrich, and G. Tröster, “Towards continuous monitoring of mental workload,” in *ACM UbiComp*, 2010.

[33] J. Hernandez, R. R. Morris, and R. W. Picard, “Call center stress recognition with person-specific models,” in *Affective computing and intelligent interaction*. Springer, 2011, pp. 125–134.

[34] A. Ghosh, M. Danieli, and G. Riccardi, “Annotation and prediction of stress and workload from physiological and inertial signals,” in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 1621–1624.

[35] J. J. Gross and O. P. John, “Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being.” *Journal of personality and social psychology*, vol. 85, no. 2, p. 348, 2003.

[36] J. J. Gross, “The emerging field of emotion regulation: an integrative review.” *Review of general psychology*, vol. 2, no. 3, p. 271, 1998.

[37] C. P. McLean, N. A. Miller, and D. A. Hope, “Mediating social anxiety and disordered eating: The role of expressive suppression,” *Eating Disorders*, vol. 15, no. 1, pp. 41–54, 2007.

[38] E. A. Butler, B. Egloff, F. H. Wlhelm, N. C. Smith, E. A. Erickson, and J. J. Gross, “The social consequences of expressive suppression.” *Emotion*, vol. 3, no. 1, p. 48, 2003.

[39] J. A. Soto, C. R. Perez, Y.-H. Kim, E. A. Lee, and M. R. Minnick, “Is expressive suppression always associated with poorer psychological functioning? a cross-cultural comparison between european americans and hong kong chinese.” *Emotion*, vol. 11, no. 6, p. 1450, 2011.

[40] J. L. Fleiss and J. Cohen, “The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability,” *Educational and psychological measurement*, vol. 33, no. 3, pp. 613–619, 1973.

[41] J. Cohen, “A coefficient of agreement for nominal scales. educational and psychosocial measurement, 20, 37-46,” 1960.

[42] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.

[43] J. W. Pennebaker, M. E. Francis, and R. J. Booth, “Linguistic inquiry and word count: Liwc 2001,” *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.

[44] F. Alparone, S. Caso, A. Agosti, and A. Rellini, “The italian liwc2001 dictionary,” *LIWC. net, Austin*, 2004.