Smart Health 9–10 (2018) 219–231

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Improving momentary stress measurement and prediction with bluetooth encounter networks

Congyu Wu[a,b,*], Mehdi Boukhechba[a,b], Lihua Cai[a,b], Laura E. Barnes[a,b], Matthew S. Gerber[a,b]

[a] *Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22904-4259, USA*
[b] *Cyber-Physical Systems Link Laboratory, Charlottesville, VA 22904-4259, USA*

ARTICLE INFO

Keywords:
Bluetooth encounter network
Stress recognition
Mobile sensing

ABSTRACT

Medical research has found strong connections between cognitive stress and various physical and mental health conditions. This paper addresses the need for more timely, less obtrusive measurements of cognitive stress. Given recent advances in and uptake of smartphone technology, we hypothesize significant correlations between a subject's smart-phone Bluetooth encounter networks and their cognitive stress levels. A review of related research indicates a lack of systematic feature engineering for these networks as well as evidence of their predictive power in mental health outcomes. We base our feature engineering on structural, edge, and neighbor attributes that incorporate network analysis measures as well as social and temporal commonalities. These features offer significant improvement in goodness-of-fit and prediction performance (AUC) for momentary stress outcomes in multiple experimental and statistical evaluation settings. Our results indicate potential value of incorporating Bluetooth encounter data into mental health monitoring practice via mobile sensing technology.

1. Introduction

Medical research has found strong connections between cognitive stress and physical and mental health (Salleh, 2008). With stressful experiences common at school (Misra & McKean, 2000), at work (Colligan & Higgins, 2006), and at home (McCubbin et al., 1980), stress remains a barrier that prevents people from living well and reaching their health and lifestyle goals, especially among younger populations (Anderson et al.). Several stress management techniques (e.g., autogenic training) have been proposed and clinically validated (Varvogli and Darviri, 2011). A primary challenge in stress management is obtaining accurate measurements of stress levels in a timely and unobtrusive fashion. Individuals are reluctant to seek help with elevated stress symptoms (Eisenberg, Golberstein, & Gollust, 2007), making timely treatment challenging. Clinicians have developed multiple survey inventories (e.g., the Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983)) to solicit subjective stress measurements from patients. These inventories, although validated and reliable, are practitioner administered procedures that can be expensive and time intensive, and they are typically retrospective and suffer from various degrees of recall bias (Goyal, Singh, Vir, & Pershad, 2016). Computerized ecological momentary assessment (EMA) surveys have been deployed via mobile technology to assess momentary stress levels (Serino et al., 2013), thus reducing recall bias and increasing ecological validity. However, the burden of stress self-report lowers EMA

* Corresponding author at: Department of Systems and Information Engineering, University of Virginia, Charlottesville, VA 22904-4259, USA.
E-mail addresses: cw9dd@virginia.edu (C. Wu), mob3f@virginia.edu (M. Boukhechba), lc3cp@virginia.edu (L. Cai), lb3dp@virginia.edu (L.E. Barnes), msg8u@virginia.edu (M.S. Gerber).

https://doi.org/10.1016/j.smhl.2018.07.017

Available online 07 July 2018

compliance (Shiffman, Stone, & Hufford, 2008).

Recent improvements in and uptake of mobile sensing technology present opportunities to use passively collected data from smartphone-embedded sensors for objective and unobtrusive health state inferences. For example, studies have used passively collected data from smartphone-embedded accelerometers, GPS sensors, and Bluetooth radios to infer and predict mental health outcomes (LiKamWa, Liu, Lane, & Zhong, 2013; Saeb; Boukhechba et al., 2017). This approach mitigates cognitive biases and allows for continuous and unobtrusive data collection, thus enabling just-in-time interventions (Nahum-Shani, Hekler, & Spruijt-Metz, 2015). While most existing work focuses on characterizing a subject's personal behavior (e.g., place visits and physical activity levels), information on one's social interactions has been less explored. Given the ubiquity of Bluetooth encounter networks as proxies for social networks (Eagle, Pentland, & Lazer, 2009; Do & Gatica-Perez, 2013; Zheng & Ni, 2013), as well as the multifaceted effects of social interaction on personal stress (Bovard, 1959; Kissel, 1965; DeVries, Glasper, & Detillion, 2003), questions remain as to whether and to what extent Bluetooth encounter data can be used to predict stress and thus enhance real time mental health tracking.

This paper identifies four experimental settings for momentary stress recognition, differentiating stress estimation versus stress forecasting in practical applications. We conduct systematic feature engineering on Bluetooth encounter data based on network analysis as well as social and temporal commonalities. We evaluate the proposed features in correlation analyses and predictive modeling, achieving significant improvements in goodness-of-fit and prediction performance in multiple evaluation settings. Key contributions of this paper are twofold. First, to the best of our knowledge, this work is the first to evaluate the value of Bluetooth encounter data in real-time stress recognition. Second, we propose novel features designed from Bluetooth signals that not only demonstrate a significant correlation with stress outcomes but also provide sociological and psychological insights into the mental health implications of social networks. We anticipate that these features will be applicable to other aspects of mental health such as loneliness, depression, and social anxiety.

2. Related work

2.1. Stress recognition

The physiological and behavioral covariates of human stress responses have been extensively studied in recent years, with a body of research focusing on physiological and motor-sensory indicators. Ranabir and Reetu (2011) identified hormonal changes associated with stressful experiences. During sessions of human-computer interaction, human stress levels are correlated with gaze and click patterns (Huang, Li, Ngai, & Leong, 2016), touch intensity and duration (Carneiro, Castillo, Novais, FernaNdez´-Caballero, & Neves, 2012), as well as physiological measures including blood volume pulse, galvanic skin response, pupil diameter, and skin temperature (Zhai & Barreto, 2006). In real-time driving scenarios, the driver's stress level can be detected from facial expression features (Paschero et al., 2012), galvanic skin response, and photoplethysmography (Singh, Conjeti, & Banerjee, 2013). Other research has focused on detecting stress from human speech (Hansen & Patil, 2007). Sharma and Gedeon (2012) provide a comprehensive review of the physiological and motor-sensory indicators used for stress recognition.

The studies above were conducted in laboratory settings using special purpose equipment to collect data. Such measurements are impractical in natural settings outside of the laboratory. In light of this, researchers have attempted to perform stress recognition using data passively collected from mobile phones and wearable devices that are widely owned and carried around. Sano and Picard (2013) showed preliminary success using wearable motion sensors and mobile phone usage data to classify stressed versus non-stressed individuals. Maxhuni et al. (2017) used phone-embedded accelerometer readings before and after phone calls to detect stress levels of office workers. Bogomolov, Lepri, Ferron, Pianesi, and Pentland (2014) predicted an individual's daily stress levels using a range of predictors including personality type, weather conditions, SMS, phone call, and Bluetooth features. A similar study was conducted by Gjoreski, Gjoreski, Lutrek, and Gams (2015) targeting momentary stress detection using multiple data sources available from a smartphone, such as accelerometer, audio sensor, GPS, WiFi access, call log, and light sensor. However, despite evidence showing the stress-modulating effect of social stimuli (Bovard, 1959; Kissel, 1965; DeVries et al., 2003), existing work on automated stress recognition has made limited use of information regarding subjects' social interactions.

2.2. Bluetooth as a social interaction sensor

Bluetooth sensors are widely available in smartphones and functional in both indoor and outdoor environments (Do & Gatica-Perez, 2013). Bluetooth encounters are triggered by physical proximity between two Bluetooth enabled devices with an expected detection range of 10 m (Chaffin et al., 2017). As many in-person social interactions require physical proximity, researchers have found that Bluetooth encounter data contains valuable information about an individual's social connections. Eagle et al. (2009) used Bluetooth encounter data to infer friendship networks. Do and Gatica-Perez (2013) proposed a generative probabilistic model to extract latent human interaction types based on Bluetooth encounters. Yan, Yang, and Tapia (2013) focused on classifying the context (e.g., in a meeting or at lunch) of Bluetooth encounters and clustering users based on their encounter patterns. In each of these studies, Bluetooth encounter data drives inferences regarding subjects' social connections.

Despite the growing body of research on sense-making of Bluetooth encounters, little has been done to correlate Bluetooth encounters with mental health outcomes. Boonstra et al. demonstrated the feasibility of collecting Bluetooth encounter data for depression recognition but offered no further findings on the relations between Bluetooth encounters and depression. A set of nine Bluetooth features covering encounter counts, entropy, and inter-encounter times were explored in a daily stress estimation problem (Bogomolov et al., 2014) but the extent to which the proposed Bluetooth features enhance performance was not assessed quantitatively. To the best of our knowledge, current research lacks systematic feature engineering from Bluetooth encounter data and evaluation of Bluetooth encounters as stress predictors. We seek to address these limitations in this study.

3. Hypotheses

Below, we hypothesize relationships between Bluetooth encounters and personal stress outcomes. Hypotheses 1 and 2 concern the statistical relationships between Bluetooth encounter networks and stress outcomes, whereas Hypotheses 3 and 4 concern predictive power. We further differentiate between the estimation of current stress levels and the forecasting of future levels. In addition to Bluetooth encounter data, we use data from GPS, accelerometer, audio sensors as well as sleep quality self-reports to drive baseline models for comparison.

Hypothesis 1. A subject's physical proximity based social environment, as measured through Bluetooth encounter networks, is more correlated with his or her *current* stress outcomes compared with prior probabilities of such outcomes.

Hypothesis 2. A subject's physical proximity based social environment, as measured through Bluetooth encounter networks, is more correlated with his or her *future* stress outcomes compared with prior probabilities of such outcomes.

Hypothesis 3. A subject's recent Bluetooth encounter network data, combined with baseline variables, estimates his or her *current* stress outcomes more accurately than only using the baseline variables.

Hypothesis 4. A subject's recent Bluetooth encounter network data, combined with baseline variables, forecasts his or her *future* stress outcomes more accurately than only using the baseline variables.

4. Approach

4.1. Experimental design

We identify two key distinctions in stress recognition experiments: (1) the goal being diagnostic (explaining current outcomes, which is the primary focus of existing stress recognition work) versus prognostic (predicting future outcomes); and (2) the response variable being the value of a mental health outcome (e.g., stress level) versus its change (e.g., increase/decrease in stress level). In practice, the diagnostic setting will inform reactive interventions that aim to mitigate negative consequences of the present outcome, whereas the prognostic setting will inform proactive interventions that aim to mitigate negative consequences of future outcomes. The four resulting designs are illustrated in Table 1 and further defined below.

- **Value Diagnosis**: to estimate at the current time t the current stress level, given mobile sensing data available by t: "What is the stress level of a user now, given our recent observations?"
- **Change Diagnosis**: to estimate at the current time t whether the current stress level has increased compared to the previous known level, given mobile sensing data available by t: "Has a user's stress level increased, given our recent observations?"
- **Value Prognosis**: to forecast at the current time t the next stress level, given mobile sensing data available by t: "What will the next stress level of a user be, given our recent observations?"
- **Change Prognosis:** to forecast at the current time t whether the next stress level will increase compared to the current level, given mobile sensing data available by t: "Will a user's stress level increase, given our recent observations?"

Table 1
Four experimental settings for momentary stress recognition. t represents the current time and Δt represents the feature extraction window. Flags represent momentary self-reports of stress levels. Value/Change indicates whether the predicted outcome is stress level itself or the increase and decrease therein. Diagnostic/Prognostic indicates whether the predicted outcome temporally coincides with or succeeds Bluetooth encounter observations.

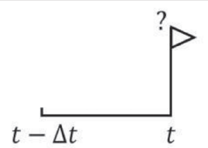
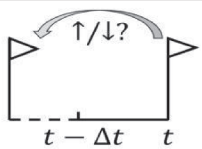
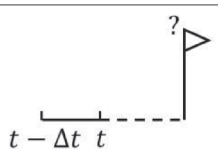
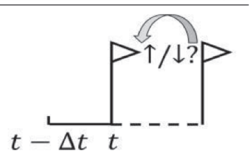
	Value	Change
Diagnostic		
Prognostic		

Table 2
A sample of Bluetooth encounter data.

Timestamp	Observer ID	Observed ID
2013-03-29 00:00:55	u56	E4:CE:8F:73:CE:65
2013-03-29 00:01:34	u39	u26
2013-03-29 00:02:42	u08	04:0C:CE:EB:14:7E
2013-03-29 00:02:42	u08	04:0C:CE:EC:5C:21

Each of the four settings includes a window size parameter Δt, a period of time preceding a stress self-report from which we extract features.

4.2. Data collection

We use the StudentLife dataset (Wang et al., 2014) to test our hypotheses within the designs described above. The StudentLife dataset records Bluetooth encounters among 49 participants over 66 days (2013-03-27 to 2013-05-31). Table 2 shows a sample of the Bluetooth encounter data, each row of which consists of a timestamp and two device identifiers. A study participant's device has an ID starting with "u" followed by a two-digit number whereas the ID of a non-participant is a 17-character string. The observer IDs are only from participant's devices whereas the observed IDs can be devices of both participants and non-participants. We add to the dataset an encounter event between two non-participant devices when both of them are found to be observed by a study participant's device at the same timestamp. For example, in Table 2, noting that the third and fourth row share the same timestamp ("2013-03-29 00:02:42") and the same observer ID ("u08") but different observed IDs ("04:0C:CE:EB:14:7E" and "04:0C:CE:EC:5C:21"), we add another row with "04:0C:CE:EB:14:7E" and "04:0C:CE:EC:5C:21" being the two encountering device identifiers.

Stress level readings are obtained through ecological momentary assessment surveys deployed on the subjects' smartphones multiple times per day at random times. The survey comprises a question text "Right now, I am…" and 5 response options "feeling great", "feeling good", "a little stressed", "definitely stressed", and "stressed out". We convert ordinal outcome values to binary values. Concretely, for value diagnosis and value prognosis, we assign self-reported stress "feeling great" and "feeling good" as negative (0) and "a little stressed", "stressed", and "stressed out" as positive (1); for change diagnosis and change prognosis, we treat increased and non-increased stress level as positive and negative respectively. By doing so we simplify modeling and attain larger sample size associated with each level of the response variable.

4.3. Feature engineering

We consider three aspects of an individual's social interaction network when conducting feature engineering using Bluetooth encounter data. The first aspect is structure. Some people have more social contacts than others. Some people prefer to distribute their social interactions more evenly among their social contacts whereas others prefer to focus on a small subset. Characteristics of an encounter event include time, location, and the content of verbal and non-verbal exchanges. Measuring these characteristics usually requires additional sensors (e.g., GPS and audio sensors). Moreover, an individual's social interaction experience can also be affected by the nature of his or her social contacts. We propose three feature classes—structural, edge, and neighbor—to capture these aspects of a subject's social interaction network. To the extent possible, we ground feature measurement in Bluetooth encounter data. We will define our baseline features following the discussion of our proposed Bluetooth features.

4.3.1. Structural attributes

We create 16 features to characterize the topology of the Bluetooth encounter network surrounding a subject formed over a time window Δt. A complete list with definitions is provided in Table 3. The suffix of a feature name indicates the scale of the social interaction network the feature is extracted from. *ego* is an alias for ego-centric network, which indicates the network containing a subject and its neighbors (devices encountered) and Bluetooth encounter events (edges) only between a subject and a neighbor. *loc* refers to a local network, which refers to the network containing a subject and its neighbors and edges between a subject and a neighbor or between two neighbors. Lastly, *glo* indicates a global network encompassing all nodes detected through Bluetooth encounters given a period of time. Fig. 1 demonstrates the Bluetooth encounter network on the three network scales surrounding a subject (red node); grey weighted edges represent Bluetooth encounters and their volume. With the structural features we pay special attention to three network metrics: (1) degree, the number of edges connecting a subject with another, approximating level of social activeness; (2) betweenness centrality, defined by the proportion of shortest paths in a network that go through a vertex (Freeman, 1977) to describe how central in the Bluetooth encounter network a subject is, and; (3) transitivity, which as a global measure is defined by the proportion of closed connected triples (i.e., triangles) out of all connected triples in a network and as a local measure the proportion of closed connected triples connected to a vertex out of all connected triples centered on the vertex (Barrat, Barthelemy, & Vespignani, 2007). Transitivity quantifies the propensity for a network to exhibit (global) and a subject to be present in (local) triangular relations, which is an indicator of community formation.

Table 3
Structural features used to describe the topology of the Bluetooth encounter network surrounding a subject.

Feature	Description
ego_deg	Number of neighbors (devices encountered)
ego_avgWeight	Average weight (number of encounters) over all neighbors
ego_giniWeight	Gini inequality of weight distribution over all neighbors
loc_den	Density of the local network
loc_avgWeight	Average weight over all edges in the local network (sum of the number of Bluetooth encounters over the number of distinct edges)
loc_trans	Global transitivity of the local network
loc_avgCompSize	Mean group size (a group is defined as a connected component after ties with the subject node are removed)
loc_giniCompSize	Gini inequality of group size
glo_avgDegNb	Average degree of the neighbors
glo_giniDegNb	Degree inequality of the neighbors
glo_avgBetwNb	Mean betweenness centrality of the neighbors
glo_giniBetwNb	Gini inequality of betweenness centrality among neighbors
glo_avgTransNb	Mean local transitivity of the neighbors
glo_giniTransNb	Inequality of local transitivity of the neighbors
glo_betw	Betweenness centrality of the subject node in the global network
glo_trans	Local transitivity of the subject node in the global network

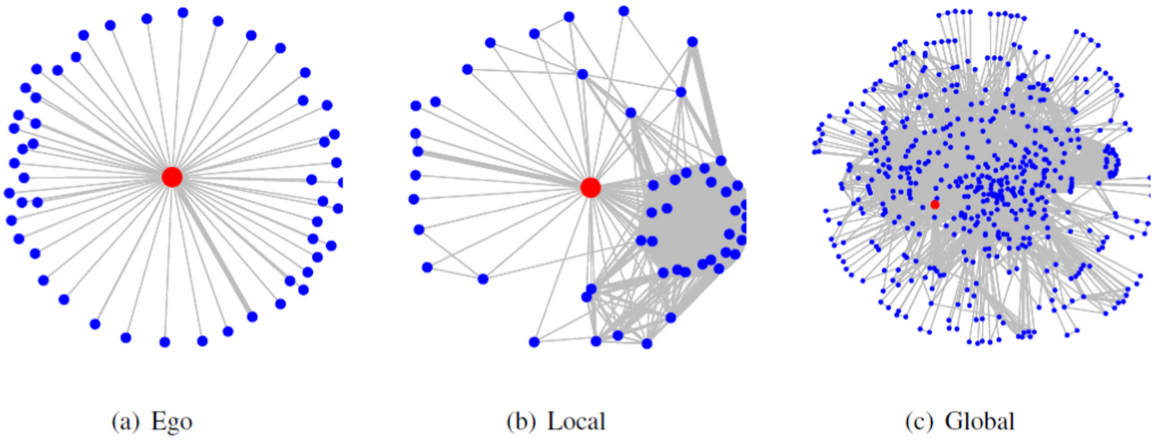


Fig. 1. An illustration of the three network scales from which structural attributes are extracted. Each red node represents a subject and blue nodes represent other devices; grey weighted edges represent Bluetooth encounters and their volume. Given all encounter events accumulated over a period of time, “ego” includes only the subject, its neighbors, and edges between the subject and a neighbor; “local” includes the identical set of devices as “ego” but also includes encounter events between pairs of neighbors; “global” includes all devices and all encounters within.

4.3.2. Edge attributes

Edge attributes are the characteristics of the encounter events. The only native edge attribute in Bluetooth encounter data is the timestamp of each encounter event (see Table 2). As this study focuses on Bluetooth data exclusively, we will use a daily epoch feature available from the timestamps as our only edge attribute. We define the daily epoch feature as a categorical variable with 4 levels *morning* (6am–noon), *afternoon* (noon–6pm), *evening* (6pm–midnight), *night* (midnight–6am), marking section of the day where the current time (the temporal upper bound for Bluetooth features extraction) resides.

4.3.3. Neighbor attributes

Named neighbor attributes, this class of features aims to characterize the social nature of the devices a subject has encountered. Further divided into three categories detailed below, each feature in this class is associated with a particular neighbor of a given subject.

4.3.3.1. Social commonality. An encounter between two individuals tends to have different significance to their respective social lives. For example, meeting with a familiar friend that one usually spends long durations with would likely incur different emotional responses than meeting with a less familiar acquaintance. Another example concerns the interactions between two pairs of individuals: if one pair has a large group of mutual friends and the other has none, the nature and content of their respective interactions are likely to be very different. Materializing these notions on familiarity and overlapping social circles, we design four social commonality features as listed in Table 4.

To compute these features, we first choose the $7 \times 24 = 168$ h leading up to $t - \Delta t$ as a base period to extract past social contacts. For a subject and for each neighbor encountered within $t - \Delta t$ to t we follow the following procedure: (1) compile a set of device IDs

Table 4
Social commonality features.

Feature	Description
usual	The proportion of the subject's past encounters with a neighbor out of the subject's past encounters with any device
usual_nb	The proportion of the subject's past encounters with a neighbor out of the neighbor's past encounters with any device
shared	The proportion of shared social contacts between the subject and a neighbor out of past social contacts of the subject
shared_nb	The proportion of shared social contacts between the subject and a neighbor out of past social contacts of the neighbor

the subject and the neighbor encountered respectively during the 168-h base period, producing a set $\boldsymbol{L}_s$ for the subject and $\boldsymbol{L}_n$ for the neighbor; note that $\boldsymbol{L}_s$ would include the neighbor and $\boldsymbol{L}_n$ would include the subject; (2) compute feature *shared* as $|\boldsymbol{L}_s \cap \boldsymbol{L}_n|/|\boldsymbol{L}_s|$ and feature *shared_nb* as $|\boldsymbol{L}_s \cap \boldsymbol{L}_n|/|\boldsymbol{L}_n|$; (3) for each device ID in $\boldsymbol{L}_s$ and $\boldsymbol{L}_n$ count the number of times (frequency) Bluetooth encounter events were recorded during the base period, producing a frequency vector $\boldsymbol{F}_s$ over $\boldsymbol{L}_s$ for the subject and $\boldsymbol{F}_n$ over $\boldsymbol{L}_n$ for the neighbor; (4) suppose during the 168-h period the subject and the neighbor encountered F_{sn} times, compute feature *usual* as $F_{sn}/\Sigma\boldsymbol{F}_s$ and *usual_nb* as $F_{sn}/\Sigma\boldsymbol{F}_n$. Note that all four social commonality features have a maximum of 1 and a minimum of 0.

4.3.3.2. Temporal commonality. Similar to the concept of social commonality, the extent to which an encounter happened at a common time for the subject and for the neighbor could also be a telling sign of the nature of an encounter and potentially its effect on mental health. We construct two features describing such temporal commonality, *tempCom* and *tempCom_nb*, with definitions given in Table 5. A high temporal commonality indicates that the encounter transpired at a time that an individual (the subject or the neighbor) tends to be "encounterable" whereas a low value would reflect that an individual may be venturing out of his or her schedule for the encounter to be happening. Combining temporal commonality and social commonality we can get a read on the social nature of a Bluetooth encounter between two devices.

The process for computing temporal commonality is illustrated in Fig. 2 and defined as follows. We choose again the $7 \times 24 = 168$ h leading up to $t - \Delta t$ as a base period to extract evidence on usual encounter times. We (1) break down the encounter events by their timestamps into hourly blocks, compute event counts that fall within each block, and create a probabilistic distribution vector $TempDist_e$ over the blocks; (2) compile all encounter events involving the subject (regardless of who the neighbor is) detected during the same time period $t - \Delta t$ to t but on different days within the base period (which covers seven days) and create a similar event count distribution vector $TempDist_s$ using these past encounter events over the same hourly blocks as $TempDist_e$; (3) repeat step (2) for all encounter events involving the neighbor and create an analogous $TempDist_n$. The three *TempDist* vectors each sums up to 1 and all have the same support. Finally, the value feature *tempCom* is the inner product of $TempDist_e$ and $TempDist_s$ while the value of feature *tempCom_nb* is the inner product of $TempDist_e$ and $TempDist_n$, thus capturing the agreement between the distributions. Both temporal commonality features have a maximum of 1 and a minimum of 0.

4.3.3.3. Network structure. As the structural features described in Section 4.3.1 can be used to characterize the inter-action network structure surrounding any device in an Bluetooth encounter network, we build the same 16 features for each neighbor a subject encounters within Δt. Note that each neighbor attribute is associated with one neighbor of a subject's; and since a subject tends to encounter multiple devices during a given time period we compute four aggregation statistics mean, standard deviation, maximum, and minimum values of all neighbor attributes and use them all as features. This way, we designed in total 16 (structural attributes) + 3 (edge attributes: three dummy variables created for the four-level categorical variable that is daily epoch) + 22 (neighbor attributes) × 4 (aggregation statistics) = 107 Bluetooth features (later referred to as main features) that we will use for our further analysis.

4.3.4. Baseline features

As introduced in Section 3, we also build features measured from other standard mobile sensors as a baseline to evaluate the additional predictive power of our Bluetooth features when combined with them (Hypotheses 3 and 4). Our baseline features cover four groups: (1) semantic location features: a 7-level categorical variable indicating one of seven place types a subject is at at current time t (denoted "now" in later references) and their proportions within the feature extraction window Δt (denoted "past"). To learn place types, we perform clustering of GPS coordinate traces (Kang, Welbourne, Stewart, & Borriello, 2005), then obtain semantic annotations from Google Map API, and manually create seven place types as shown in Table 6; (2) activity level features: a 4-level categorical variable indicating the stationary/walking/running/unknown status of a subject at current time t and their proportions within Δt; (3) sound status features: a 3-level categorical variable indicating the silence/voice/noise status detected by a subject's mobile device at current time t and their proportions within Δt, and; (4) sleep quality: self-reported sleep quality of the previous night

Table 5
Temporal commonality features.

Feature	Description
tempCom	The degree to which the times when the subject encounters a neighbor agree with the usual times the subject encounters any device
tempCom_nb	The degree to which the times when the subject encounters a neighbor agree with the usual times the neighbor encounters any device

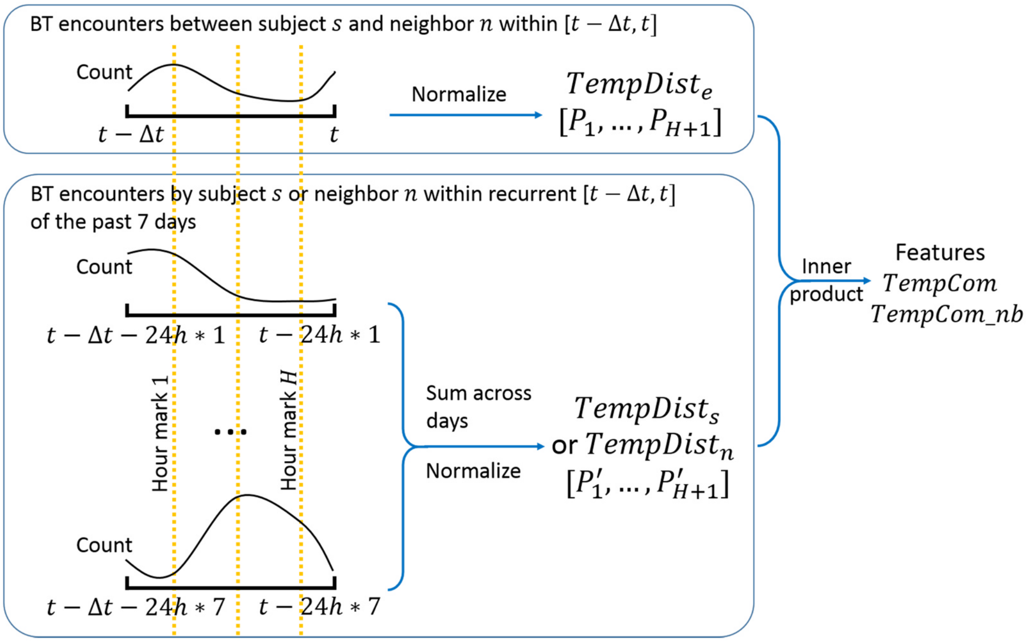


Fig. 2. Illustration of the process for computing temporal commonality features.

Table 6
Seven semantic labels of GPS locations.

Label	Criterion
Home	The place a subject spends the most time 12–6 a.m.
Education	University buildings
Food	Dining and food vending establishments
Health	Healthcare facilities and gymnasiums
Transit	Bus stops and stations, parking lots, airports
Religion	Places of worship (e.g., churches)
Other	Places not categorized above (e.g., post offices)

(5-scale ordinal value ranging from poorest to best); sleep quality is included as a baseline feature due to its positive correlation with next-day stress (Taylor et al., 2017). After encoding dummy variables, we create $[(7 - 1) + (4 - 1) + (3 - 1)] \times 2 + 1 = 23$ baseline features.

4.4. Correlation analysis

Our objective in this step is to evaluate how well our Bluetooth features account for the variance within an individual's momentary stress outcomes in the four experiment settings introduced in Section 4.1. The results from this analysis will allow us to test Hypotheses 1 and 2 and discover salient features that are responsible for the correlation. We use two model selection methods: (1) backward stepwise logistic regression with AIC (Akaike Information Criterion, defined as $2k - 2lnL$, where k is the number of estimated parameters and L is model likelihood) as the selection criterion and (2) 10-fold cross-validated logistic regression with LASSO regularization (least absolute shrinkage and selection operator (Tibshirani, 1996)) and model likelihood as the selection criterion on the following feature groupings: (1) a null model; (2) baseline features as defined in Section 4.3.4; (3) our Bluetooth features as described in Sections 4.3.1–4.3.3; and lastly (4) our Bluetooth features combined with baseline features. We choose Δt to be 6 h as it represents the length of one entire daily epoch (e.g., morning or afternoon). Response variables in each setting are as explained in Section 4.2. We conduct bootstrapping to break down potential dependencies between observations and further support the statistical relationships between our Bluetooth features and personal stress outcomes. Concretely, we draw 100 bootstrap samples on the original dataset and repeat stepwise and LASSO model selection on all four feature groupings. Then we compare the resulting AICs using Student's t-test.

4.5. Predictive modeling

The aim here is to evaluate the predictive power of our proposed Bluetooth features and test Hypotheses 3 and 4. We experiment with two different evaluation settings as discussed below.

4.5.1. Across-subject, leave-one-subject-out (LOSO)

In this setting we set aside one subject's data, train our models on data pooled together from all other subjects' data, and evaluate the performance on the one subject's data that was set aside. For each subject we obtain a sequence of predicted outcome values and a corresponding AUC (area under ROC curve) score. This experiment is applicable in practical cases where models can be built from available, existing subjects' data and performance on a new, unseen subject is desired.

4.5.2. Within-subject, leave-one-observation-out (LOOO)

In this experiment, for each subject, we set aside each observation of his or hers, train our models on the remaining data that belongs to the same subject, and evaluate on the one observation that was set aside. For each subject we also obtain a sequence of predicted outcome values and a corresponding AUC score. This experiment is applicable in cases where only a user's own historical data is available for insights regarding said user's future outcome values.

The critical difference between the across-subject LOSO and within-subject LOOO evaluation settings is that in the former, to predict an observation of a subject, only his or her peers' data and none of his or her own data is used; whereas in the latter, the situation is reversed: to predict an observation of a subject, only his or her own data and none of his or her peers' data is used. Note that in both experiments, we pool data from the entire available time period and thus ignore the temporal sequence of observations; this is justified by (1) our focus on non-temporal effect (own data versus peers' data) in these two experiments and (2) limited data size (1702 observations in total covering 49 subjects over a 66-day period).

As a preprocessing step, we discard (1) the first week's data as our social and temporal commonality features require a preceding seven days (168 h) to serve as a base period, and; (2) for all prediction settings except change diagnosis, which uses differenced feature values as predictors, observations that have a zero *ego_deg* as it indicates zero Bluetooth encounters. For all predictive experiments in this section, we choose Δt as 6 h and use a random forest learner with 3000 trees grown, 20 predictors randomly selected at each split, until reaching node size 5. We also explore the sensitivity of prediction performance to varied Δt sizes.

5. Results

5.1. Correlation analysis result

We list the AIC scores of selected models using stepwise and LASSO methods in Tables 7 and 8, respectively, both of which show that models using selected Bluetooth features (MAIN) achieve higher goodness-of-fit than null models when explaining momentary personal stress outcomes. Overall, stepwise logistic regression achieves lower AIC scores than LASSO as a feature selection method. Among the four experimental settings, the goodness-of-fit improvement is more pronounced in value diagnosis and value prognosis. Under all experimental settings, baseline features (BASE) and our Bluetooth features each outperformed the null models. Best performances are obtained when models are selected from Bluetooth and baseline features combined (BASE + MAIN), dominating baseline and main features individually, under all experimental settings except change prognosis, where using all features together seems to harm goodness-of-fit.

Mean and standard deviation of each group of bootstrapped AIC values are shown in Fig. 3 and clearly with strong confidence (p-value < 0.001) we achieve significantly better goodness-of-fit with selected Bluetooth features compared to null models, with either feature selection method. We successfully confirm Hypotheses 1 and 2 that a subject's Bluetooth encounter networks correlates with his or her current and future stress outcomes significantly better than null models.

Finally, we look into the important Bluetooth features that have driven the performance demonstrated above. As expected, LASSO and stepwise logistic regression selected moderately different but overlapping sets of features and LASSO resulted in more parsimonious models. We focus on the features in the BASE + MAIN models that are both significant in the stepwise models and selected by LASSO method under each experimental setting, for they are most likely important features. The selected features, the sign of their effect, and their p-values in the corresponding stepwise model are listed in Table 9, organized by the category (BASE and MAIN) and sorted in descending order of significance. We make several interesting observations on the correlations between our Bluetooth features and stress outcomes. First, *usual_max* is the number one significant and negative predictor associated with both present and future level of personal stress (value diagnosis and value prognosis). This suggests that interaction with a familiar and regular social contact has a potentially lasting stress relieving effect ("protective effect under stress" (Bovard, 1959)); in fact, it was confirmed that "the presence of a friend in a stressful situation reduces stress more than the presence of a stranger" (Kissel, 1965), which constitutes

Table 7
AIC scores of selected models using backward stepwise logistic regression on different feature groupings.

Stepwise	Value diagnosis	Change diagnosis	Value prognosis	Change prognosis
NULL	1,733	1,873	1,704	1,715
BASE	1,691	1,841	1,678	1,714
MAIN	1,656	1,828	1,626	1,701
BASE + MAIN	1,618	1,825	1,603	1,706

Table 8
AIC scores of selected models using LASSO logistic regression on different feature groupings.

Stepwise	Value diagnosis	Change diagnosis	Value prognosis	Change prognosis
NULL	1,733	1,873	1,704	1,715
BASE	1,662	1,853	1,689	1,713
MAIN	1,653	1,853	1,653	1,713
BASE + MAIN	1,626	1,850	1,652	1,714

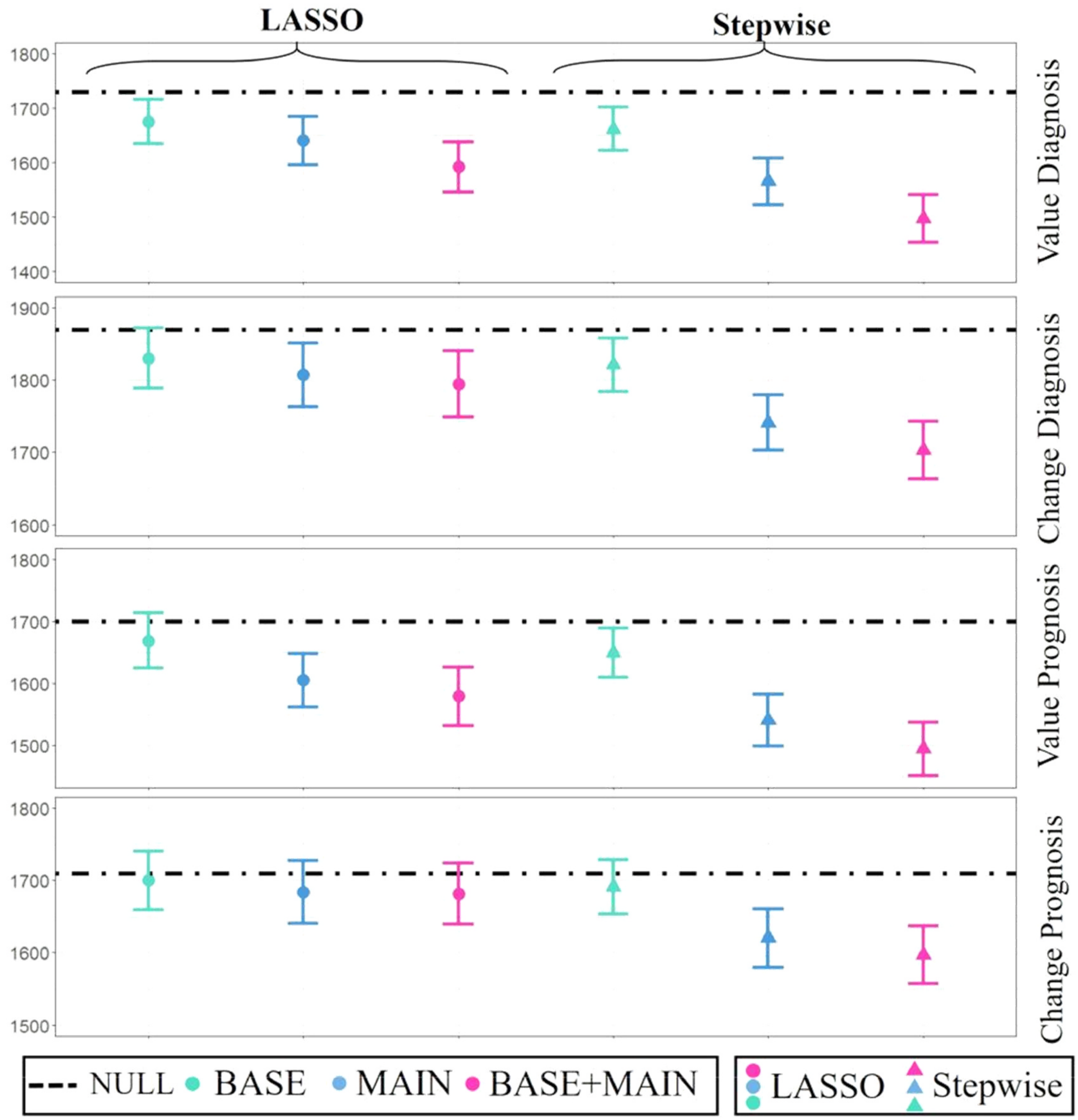


Fig. 3. Bootstrap result: mean (round and triangular dots) and standard deviation (upper and lower error bars) of AIC scores achieved by LASSO and stepwise logistic regression on each bootstrap sample compared against the corresponding null models (dashed horizontal lines).

the social buffering theory (Kikusui, Winslow, & Mori, 2006). Second, an increased value of *usual_sd* appears to be positively correlated with increased stress level in the change diagnosis setting. This indicates that interacting with people of different familiarity levels may comprise a stressful experience. Such effect is mentioned in Bogomolov et al. (2014), Alshamsi, Pianesi, Lepri, Pentland, and Rahwan (2016) and agrees with the role strain theory (Goode, 1960). The significance of feature *loc_trans* mean in change diagnosis indicates that increased transitivity within a subject's local encounter network, which can be a result of real life scenarios like gatherings and group study sessions, is likely to accompany decreasing personal stress. Last, for change prognosis only one feature made the list, indicating it is less well-modeled by our Bluetooth features than other settings; *shared_nb_max*, representing the

Table 9
Important features in the MAIN + BASE models selected by both stepwise and LASSO logistic regression under each experimental setting; following each feature is the sign of their effect and p-value in the corresponding stepwise models.

	Value diagnosis		Change diagnosis		Value prognosis		Change prognosis	
BASE	sleepQ	− 0.000	religion_now	− 0.000	stationary_now	+ 0.008		
	walking_past	− 0.002	noise_now	+ 0.003	transit_past	− 0.009		
	stationary_now	+ 0.003	noise_past	− 0.046	religion_now	− 0.009		
	home_now	− 0.006			edu_now	+ 0.016		
	health_now	− 0.007			noise_past	− 0.030		
	transit_now	− 0.011			health_now	− 0.035		
MAIN	usual_max	− 0.000	loc_trans_mean	− 0.000	usual_max	− 0.000	shared_nb_mean	− 0.019
	ego_deg	+ 0.000	usual_sd	+ 0.001	loc_avgWeight_mean	− 0.000		
	shared_nb_mean	− 0.004	ego_avgWeight_min	− 0.008	evening	− 0.002		
	shared_mean	+ 0.004			tempCom_mean	− 0.005		
	tempCom_max	+ 0.005			loc_giniCompSize_mean	+ 0.012		
					usual_nb_min	− 0.022		

highest proportion of shared social circle with a subject among his or her encounter neighbors, exhibited a similar effect as *usual_max* which is also a social commonality feature.

As for baseline features, sleep quality of the previous night is found to have a significant correlation with low current stress level, which confirms findings in Taylor et al. (2017). Among semantic location features, *religion_now* shows strong correlation with low or lowered stress level, suggesting mental health benefit of worship places. The contrast between the positive correlation of feature *stationary_now* and the negative correlation of features like *walking_past* and *transit_past* in both value diagnosis and value prognosis in both suggest the benefit of traveling and not staying still in stress management. These interpretations provide motivation and evidence for future studies and smart health applications to further investigate and utilize their effect.

5.2. Predictive modeling result

Table 10 shows the mean and standard deviation of area under ROC (AUC) values obtained for all subjects in the across-subject LOSO and within-subject LOOO experiments using different groupings of features. One evident pattern we found is that regardless of feature groups and experimental settings, prediction performance is significantly higher under within-subject LOOO than across-subject LOSO. This observation suggests that historical Bluetooth encounters are more predictive of stress than encounter networks measured from other individuals. This can be explained by the fact that an encounter of the same particular characteristics can impact different individuals differently, due to personal differences in the reaction to affective stimuli (Larsen & Ketelaar, 1991; Clark, Woodley, & De Halas, 1962). As such, in real world applications, caution should be taken applying existing models to new subjects. Calibration with personal data may be important when incorporating Bluetooth based features in stress recognition tasks.

Our Bluetooth features prove predictive: all MAIN models achieved AUC values significantly greater than the random guess baseline of 0.5. However, when comparing performance achieved by our Bluetooth features combined with baseline features (BASE + MAIN) versus by baseline features only (BASE), we discovered mixed results. For value diagnosis, incorporating our Bluetooth features evidently improves prediction performance than baseline, confirming Hypothesis 3: mean AUC rose from 0.67 to 0.73 in the within-subject LOOO experiment (p-value 0.0564 in a one-tailed *t*-test). However, our results do not support Hypothesis 4 as performance change is minimal for the prognosis settings. Moreover, value diagnosis enjoys the highest AUC regardless of the prediction approach, regardless of feature groups. The relatively stronger performance under value diagnosis indicates more promising value of our proposed features in stress estimation applications, compared to forecast-oriented settings.

Our motivation to perform stress recognition using passively sensed proximity network data is rooted in the effect of social interaction on personal stress level studied in psychology research (DeVries et al., 2003). However, questions remain regarding the inertia and decay of the effect over time, which governs the extent to which past social interaction episodes affect current stress level and in turn concerns the choosing of feature extraction window in our predictive modeling tasks. We attempt to answer this question

Table 10
Prediction performance (Area under ROC curve) achieved by different feature groupings under the across-subject leave-one-subject-out and the within-subject leave-one-observation-out evaluation settings.

AUC		Value diagnosis		Change diagnosis		Value prognosis		Change prognosis	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Across-subject LOSO	BASE	0.6393	0.1229	0.6263	0.1110	0.6284	0.1355	0.6389	0.1269
	MAIN	0.6532	0.1173	0.6025	0.1220	0.6093	0.1099	0.6356	0.1347
	BASE + MAIN	0.6595	0.1202	0.6018	0.1102	0.6112	0.1215	0.6243	0.1362
Within-subject LOOO	BASE	0.6706	0.1240	0.6754	0.1374	0.6966	0.1540	0.6562	0.1333
	MAIN	0.7115	0.1578	0.6720	0.1620	0.6660	0.1432	0.6628	0.1355
	BASE + MAIN	0.7262	0.1507	0.6759	0.1594	0.6695	0.1521	0.6619	0.1359

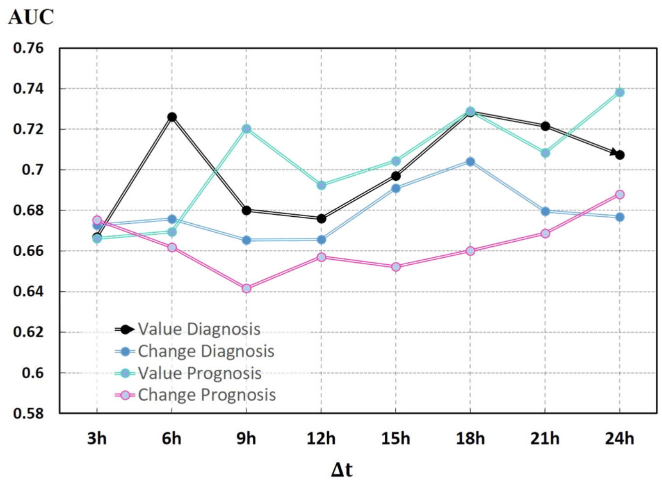


Fig. 4. Sensitivity of prediction performance to the choosing of Δt, in the within-subject, leave-one-observation-out experiments using BASE + MAIN features. Shown in figure are mean AUC values for each feature extraction window size and prediction setting.

empirically by exploring the sensitivity of prediction performance to the size of feature extraction window Δt. Specifically, we repeat our within-subject leave-one-observation-out prediction experiments using MAIN + BASE features built with different sizes of Δt, namely 3 h, 9 h, 12 h, 15 h, 18 h, 21 h, and 24 h; and compare the resulting performances with that obtained with $\Delta t = 6$ h as shown in Table 10 and Fig. 4.

The pattern shown in Fig. 4 indicates that prediction performance does vary with different feature extraction window sizes but does not exhibit a monotonous or unimodal trend. For all experimental settings except change prognosis, we observe a bimodal shape in the trend of prediction performance: performance is relatively low at $\Delta t = 3$ h; there seems to be a local maximum at 6 h (value diagnosis and change diagnosis) or 9 h (value prognosis) early on and a later local maximum that comes quite unanimously at 18 h. Although, for value prognosis prediction performance reached global maximum with $\Delta t = 24$ h, unlike value diagnosis and change diagnosis for which performance reached maximum at $\Delta t = 18$ h. In contrast, performance under change prognosis appears to have a V-shaped trend with 24 h being the optimal Δt. From our observations it is difficult to draw a general conclusion on the optimal choice of feature extraction window but we recommend examination of various window sizes up to greater than 15 h and be cautious using shorter-than-6-h window sizes.

6. Conclusions

We have aimed to address the need for accurate, timely, and unobtrusive stress monitoring technology through passively sensed Bluetooth encounter networks. Bluetooth is a widely available mobile sensor that can detect an individual's social environment based on physical proximity. Systematic feature engineering and examination of predictive value using Bluetooth encounter data is lacking in existing literature. We have investigated Bluetooth encounter data in terms of structural attributes, edge attributes, and neighbor attributes, and built features accordingly incorporating measures from social network analysis and concepts of social and temporal commonalities. Our correlation analysis, involving Bluetooth encounters among 49 student subjects over 66 days, suggests that our features extracted from Bluetooth encounter data account for the variance of stress outcomes significantly more effectively than null models. In doing so, we supported our Hypotheses 1 and 2, confirmed existing findings (e.g., sleep quality as a predictor of next day stress), and called attention to salient features (e.g., *usual_max*) for consideration in real world stress recognition applications as well as their interpretations in social psychology.

Moreover, we tested the predictability of momentary stress levels from Bluetooth encounter data using random forest under four different prediction settings (value diagnosis, change diagnosis, value prognosis, change prognosis) and two evaluation settings (within-subject leave-one-observation-out and across-subject leave-one-subject-out). The results presented in the previous sections provide preliminary but promising evidence for the performance boosting effect of our Bluetooth features in momentary stress recognition applications. When incorporated with baseline features built with data collected from other standard mobile sensors and sleep quality self-reports, our Bluetooth features achieved performance improvement ($0.7262 - 0.6706 = 0.0556$ in AUC) for value diagnosis but did not for other experimental settings. As such, we supported Hypothesis 3 and did not support Hypothesis 4. Further evidence is needed to evaluate the utility of our Bluetooth features in stress forecasting tasks. In practice, we recommend incorporation of our proposed features with those extracted from other mobile data sources (e.g., GPS, accelerometer, phone usage) as input for predictive modeling to enhance stress recognition performance.

This work has several limitations. The primary one lies in our ability to capture in-person social interaction, which is limited by existing sensing technologies. A Bluetooth encounter, triggered by physical proximity, is not necessarily produced by an in-person interaction. Because of this, the correlation discovered and predictability achieved should be treated as evidence for incorporating our proposed features in future stress recognition tasks, as opposed to causal relations between certain characteristics of in-person social interaction and mental health. To accomplish the latter we would need more advanced sensing technology (e.g., audio sampling of verbal exchanges) and quasi-experimental designs. Second, as the study subjects in the StudentLife dataset are all college students, the generalizability of our results to other demographic groups awaits further evidence. Finally, as in many mental health monitoring tasks, ground truth curation is subject to the availability of self-report. Cognitive biases and low self-report compliance from the subjects could negatively impact the validity of response variable data, especially for change-based and forecast-oriented experimental settings. Future work may involve further evaluation of our proposed features in prediction tasks targeting other health outcomes (e.g., affect recognition). Another important task is to incorporate comprehensive contextual information obtained from other data sources such as GPS and accelerometer to more effectively and concisely represent social interaction. Evidence of the value Bluetooth encounter data provides for personal stress estimation presented in this paper motivates the incorporation of passively sensed in-person social network data with other sensor data sources in automated mental health monitoring.

Conflict of interest statement

None.

References

Alshamsi, A., Pianesi, F., Lepri, B., Pentland, A., & Rahwan, I. (2016). Network diversity and affect dynamics: The role of personality traits. *PLoS One, 11*(4), e0152358.

Anderson N.B., Belar C.D., Breckler S.J., Nordal K.C., Ballard D.W., Bufka L.F., & Wiggins K. *Stress in america: Paying with our health*. American Psychological Association.

Barrat, A., Barthelemy, M., & Vespignani, A. (2007). The architecture of complex weighted networks: Measurements and models. *Large scale structure and dynamics of complex networks: From information technology to finance and natural science* (pp. 67–92). World Scientific.

Bogomolov A., Lepri B., Ferron M., Pianesi F., & Pentland A.S. (2014). Daily stress recognition from mobile phone data, weather conditions and individual traits. In *Proceedings of the 22nd ACM international conference on multimedia* (pp. 477–486). ACM.

Boonstra T.W., Werner-Seidler A., O'Dea B., Larsen M.E., & Christensen H. Smartphone app to investigate the relationship between social connectivity and mental health. arXiv preprint arXiv:1702.02644.

Boukhechba M., Huang Y., Chow P.I., Fua K., Teachman B.A., & Barnes L.E. (2017). Monitoring social anxiety from mobility and communication patterns. In *Proceedings of the 2017 ACM international joint conference on pervasive and Ubiquitous Computing UbiComp 17*. ACM Press. 〈http://dx.doi.org/10.1145/3123024.3125607〉.

Bovard, E. W. (1959). The effects of social stimuli on the response to stress. *Psychological Review, 66*(5), 267.

Carneiro, D., Castillo, J. C., Novais, P., FernaNdez´-Caballero, A., & Neves, J. (2012). Multimodal behavioral analysis for non-invasive stress detection. *Expert Systems with Applications, 39*(18), 13376–13389.

Chaffin, D., Heidl, R., Hollenbeck, J. R., Howe, M., Yu, A., Voorhees, C., & Calantone, R. (2017). The promise and perils of wearable sensors in organizational research. *Organizational Research Methods, 20*(1), 3–31.

Clark, T., Woodley, R., & De Halas, D. (1962). Gas-graphite systems. In R. Nightingale (Ed.). *Nuclear graphite* (pp. 387). New York: Academic Press.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior,* 385–396.

Colligan, T. W., & Higgins, E. M. (2006). Workplace stress: Etiology and consequences. *Journal of Workplace Behavioral Health, 21*(2), 89–97.

DeVries, A. C., Glasper, E. R., & Detillion, C. E. (2003). Social modulation of stress responses. *Physiology Behavior, 79*(3), 399–407.

Do, T. M. T., & Gatica-Perez, D. (2013). Human interaction discovery in smartphone proximity networks. *Personal and Ubiquitous Computing, 17*(3), 413–431.

Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences, 106*(36), 15274–15278.

Eisenberg, D., Golberstein, E., & Gollust, S. E. (2007). Help-seeking and access to mental health care in a university student population. *Medical Care, 45*(7), 594–601.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry,* 35–41.

Gjoreski M., Gjoreski H., Lutrek M., & Gams M. (2015). Automatic detection of perceived stress in campus students using smartphones. In *Intelligent environments (IE), 2015 international conference on* (pp. 132–135). IEEE.

Goode, W. J. (1960). A theory of role strain. *American Sociological Review,* 483–496.

Goyal, A., Singh, S., Vir, D., & Pershad, D. (2016). Automation of stress recognition using subjective or objective measures. *Psychological Studies,* 1–17.

Hansen, J. H., & Patil, S. (2007). Speech under stress: Analysis, modeling and recognition. *Speaker classification I* (pp. 108–137). Berlin, Heidelberg: Springer.

Huang M.X., Li J., Ngai G., & Leong H.V. (2016). Stressclick: Sensing stress from gaze-click patterns. In *Proceedings of the 2016 ACM on multimedia conference* (pp. 1395–1404). ACM.

Kang, J. H., Welbourne, W., Stewart, B., & Borriello, G. (2005). Extracting places from traces of locations. *ACM SIGMOBILE Mobile Computing and Communications Review, 9*(3), 58. https://doi.org/10.1145/1094549.1094558.

Kikusui, T., Winslow, J. T., & Mori, Y. (2006). Social buffering: Relief from stress and anxiety. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 361*(1476), 2215–2228.

Kissel, S. (1965). Stress-reducing properties of social stimuli. *Journal of Personality and Social Psychology, 2*(3), 378.

Larsen, R. J., & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology, 61*(1), 132.

LiKamWa R., Liu Y., Lane N.D., & Zhong L. (2013). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on mobile systems, applications, and services* (pp. 389–402). ACM.

Varvogli, L., & Darviri, C. (2011). Stress management techniques: Evidence-based procedures that reduce stress and promote health. *Health Science Journal, 5*(2).

Maxhuni, A., Hernandez-Leal, P., Morales, E. F., Sucar, L. E., Osmani, V., Munoz´-Melendez, A., & Mayora, O. (2017). Using intermediate models and knowledge learning to improve stress prediction. *Applications for future internet* (pp. 140–151). Cham: Springer.

McCubbin, H. I., Joy, C. B., Cauble, A. E., Comeau, J. K., Patterson, J. M., & Needle, R. H. (1980). Family stress and coping: A decade review. *Journal of Marriage and the Family,* 855–871.

Misra, R., & McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. *American Journal of Health Studies, 16*(1), 41.

Nahum-Shani, I., Hekler, E. B., & Spruijt-Metz, D. (2015). Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. *Health Psychology, 34*(S), 1209.

Paschero M., Del Vescovo G., Benucci L., Rizzi A., Santello M., Fabbri G., & Mascioli F.F. (2012). A real time classifier for emotion and stress recognition in a vehicle driver. In *Industrial electronics (ISIE), 2012 IEEE international symposium on* (pp. 1690–1695). IEEE.

Ranabir, S., & Reetu, K. (2011). Stress and hormones. *Indian Journal of Endocrinology and Metabolism, 15*(1), 18.

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *Journal of Medical Internet Research, 17*(7).

Salleh, M. R. (2008). Life event, stress and illness. *The Malaysian Journal of Medical Sciences: MJMS, 15*(4), 9.

Sano A., & Picard R.W. (2013). Stress recognition using wearable sensors and mobile phones. In *Proceedings of the affective computing and intelligent interaction (ACII),*

2013 Humaine association conference on (pp. 671–676). IEEE.
Serino, S., Cipresso, P., Tartarisco, G., Baldus, G., Corda, D., Pioggia, G., Gaggioli, A., & Riva, G. (2013). Computerized experience-sampling approach for realtime assessment of stress. *EAI Endorsed Transactions on Ambient Systems, 1*(2), 1–8.
Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine, 108*(3), 1287–1301.
Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.
Singh, R. R., Conjeti, S., & Banerjee, R. (2013). A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control, 8*(6), 740–754.
Taylor, S., Jaques, N., Nosakhare, E., Sano, A., Klerman, E., & Picard, R. (2017). Importance of sleep data in predicting next-day stress, happiness, and health in college students. *Journal of Sleep and Sleep Disorders Research, 40*(Suppl_1), SA294–SA295.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological),* 267–288.
Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev D., Campbell, A. T. (2014). Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* (pp. 3–14). ACM.
Yan Z., Yang J., & Tapia E.M. (2013). Smartphone bluetooth based social sensing. In *Proceedings of the 2013 ACM conference on pervasive and ubiquitous computing adjunct publication* (pp. 95–98). ACM.
Zhai J., & Barreto A. (2006). Stress recognition using non-invasive technology. In *Proceedings of the FLAIRS conference* (pp. 395–401).
Zheng J., & Ni L.M. (2013). An unsupervised learning approach to social circles detection in ego bluetooth proximity network. In *Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing* (pp. 721–724). ACM.