Mood Versus Identity: Studying the Influence of Affective States on Mobile Biometrics

Tempestt Neal[1] and Shaun Canavan[2]
Department of Computer Science and Engineering, University of South Florida, Tampa, Florida, USA
[1]tjneal@usf.edu, [2]scanavan@usf.edu

***Abstract*— Mobile device usage data such as mobile app use and acceleration measurements fluctuate often as individuals carry out their daily tasks. As these data have emerged in recent years as promising biometric identifiers, it is important to understand the many causes of these variations such that these systems can adapt without degradation in performance. In this paper, we seek to understand the impact of changes in a person's mood on the performance of a mobile biometric system using a publicly available dataset of 27 subjects. We explore the verification and identification tasks, along with mood prediction from smartphone data. We achieved an equal error rate of 3% and a d-prime value of 5.05 for the verification task, wherein experiments showed that verification is minimally influenced by an individual's mood, although negative arousal slightly degraded performance. We created a multi-class problem to study the identification task, achieving an average 83% $F1$-score. Here, we observed that subjects with lower identification accuracy (<70%) experienced fewer mood changes compared to the average (13.48), while all but one subject with high identification accuracy (>95%) experienced more mood changes compared to the average. Contrasting previous claims, our findings suggest that frequent changes in mood may have little negative impact performance. Finally, positive arousal and negative valence yielded the highest area under the curve (0.67) for mood prediction. This was also the class associated with the highest average genuine and lowest average imposter scores for verification experiments, suggesting a correspondence between recognition and mood prediction tasks that applications such as sensor-enhanced mHealth apps could leverage.**

I. Introduction

Research on behavioral biometrics for smartphone users has resulted in a diverse set of topics ranging from continuous authentication using touch and keystroke data (e.g., [30], [18]) to usability surveys on commercial face and fingerprint systems (e.g. [3], [15]). Behavioral mobile biometrics have gained attention considering the merits associated with transparent, continuous sensing in comparison to point-of-entry methods [21]. For example, point-of-entry methods may require nearly 50 unlock attempts per day [10]. For some, point-of-entry methods have been regarded as awkward [5] or inconvenient [10].

Behavioral biometrics pose challenges as well. Behavioral signals, such as tone in voice [8], touch location on the screen during swipes or flicks [9], and usage activity (e.g., app launches or calling patterns) change over time according to an individual's location, current task, etc. Exactly how these changes affect mobile, behavioral biometric systems, however, has yet to be clearly realized. Meanwhile, the requirement that a biometric modality "should be sufficiently invariant (with respect to the matching criterion) over a period of time", or *permanent*, is a critical component of performance [11], [12], [22].

This work was not supported by any organization.

In this paper, we seek to understand the impact of mood on performance in a mobile biometric system as mood is frequently claimed to be one of many factors influencing changes in mobile, behavioral biometric modalities [20]. Mood, less intense "affective states that are capable of influencing a broad array of potential responses, many of which seem quite unrelated to the mood-precipitating event" [19], has been extensively measured via physiological signals [14], [25], [27], [7]. However, activity data gathered from mobile devices have also proven useful for mood detection in a more passive manner [2], [28], [17], although none have evaluated the impact of mood on phone usage in the context of biometric recognition. One particular study evaluated the impact of stress on handwriting recognition [4], but most efforts have only *assumed* that mood may affect usage which in turns affects recognition performance [17].

Improved mood detection performance using person-specific classification models trained on smartphone activity data suggests that an individual's mood may alter their biometric template data [17], [1], [23]. These results motivate our current research to better understand the impact of variations in mood on the permanence of phone usage when considering this data as a biometric modality. We view the possible impact of mood comparable to occlusions or pose in face recognition [6]. By identifying *mobile occlusions*, future efforts can work toward developing algorithms which can control or adapt to these factors. We explore three research questions:

1) Could changes in an individual's mood have a significant impact on the performance of a mobile biometric system?
2) If so, do these changes affect verification, identification, or both recognition tasks?
3) What might mood prediction experiments using smartphone activity data imply about the recognition tasks?

In the following section, we detail the dataset used and our experimental approach. We provide results in Section 3 for all three research questions. We summarize and highlight key insights, future work, and limitations in Section 4.

II. METHOD

A. Dataset

We used a publicly available dataset consisting of the smartphone activity data collected from 27 subjects [1]. Examples of collected data include statistics on call events such as timestamp of the call and duration, text message information, and time and date of captured photos. Participants were selected if they were at least 18 years old and used an Android phone as their primary mobile device, while participants exhibiting depressive symptoms according to the Center for Epidemiologic Studies Depression Scale [16] were excluded. All captured variables used in our experiments are described in Table I. Accompanying these data were self-reported mood captured via ecological momentary assessment [26], [1]. Participants self-reported their mood up to five times a day over six weeks. Mood was assessed according to the Circumplex Model of Affect [24], where mood is represented on a two-dimensional scale of valence (goodness) and arousal (intensity).

Once data were collected, the researchers averaged device activity and self-reported mood entries per day. As a result, each sample in the post processed dataset (and the version released for public use) represents an individual's average device activities and their overall mood per day. The final dataset included mood and phone usage information for 27 participants, each having 35.5 days of data on average (totaling 959 days (samples) in the dataset).

TABLE I
DATASET FEATURES [1]

Features	**Description**	**# of Dimensions**
Valence	Daily average 2D level of valence from -2 to 2	1
Arousal	Daily average 2D level of arousal from -2 to 2	1
Images	Number of photos captured	1
SMS	Frequency of text messages sent to top 5 contacts	5
Call Frequency	Frequency of calls to top 5 contacts	5
Call Duration	Duration of calls to top 5 contacts	5
App Launches	Frequency of launch events for top 5 apps	5
App Duration	Duration of use for top 5 apps	5
App Category Frequency	Frequency of app categories (built-in, communication, entertainment, finance, game, office, social, travel, utilities, weather, other, unknown)	12
App Category Duration	Duration of app categories	12
Screen Frequency	Screen-on frequency	1
Screen Duration	Screen-on duration	1
Accelerometer	Average percentage of high acceleration	1
		Total: 55

B. Experiments

Verification (genuine versus imposter) and identification (one-to-many) experiments were run, both using a soft voting classifier (i.e., ensemble learning). A soft voting classifier fuses the matching scores of several individual classifiers. In our experiments, we used a random forest, a radial basis function kernel support vector machine (SVM), and a Gaussian naive Bayes classifier. We performed a grid search to optimize the parameters of the random forest (20 or 200 trees) and the SVM ($C = 1, 10, 100, 1000$ and $\gamma = 0.001, 0.0001$) with 5-fold cross validation for every training set. We employed leave-one-out cross validation for all experiments. For all experiments, we employed random under sampling to prevent a bias toward the majority class.

Recognition performance was evaluated using the $F1$-score (harmonic mean of precision and recall), Equal Error Rate (EER) (error rate at which the False Acceptance Rate and False Reject Rate are approximately equal), receiver operating characteristic (ROC) curve, and score distribution plots.

Finally, when studying mood, we considered four classes: positive arousal and positive valence (happiness or excitement), positive arousal and negative valence (tense or frustration), negative arousal and positive valence (relaxed or calm), and negative arousal and negative valence (sadness or depression).

III. RESULTS

A. Mood Versus Verification

We ran verification experiments using all features except valence and arousal. Treating a subject's ID as the classification label, Figures 1 and 2 show an achieved EER of 3% and d-prime of 5.05. Figure 2 shows satisfactory separation between genuine and imposter scores. We note that we are unaware of any previous work demonstrating that a single sample per day of smartphone activity is just as useful for verification as a series of samples.

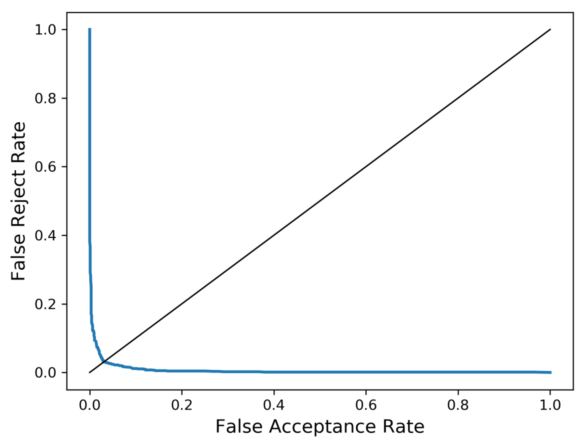

Fig. 1. DET curve of verification experiments ($EER = 0.03$).

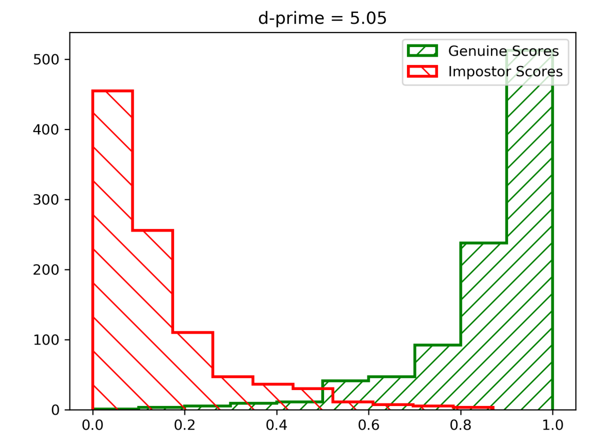

Fig. 2. Score distribution plot of verification experiments (EER=0.03).

In our analysis of the impact of mood on verification performance, we studied the mean and standard deviation of genuine and imposter scores for each of the four mood states previously detailed. Table II shows a slight increase in performance when individuals are in a tensed or stressed state, with improving accuracy from states of happiness, relaxed, and sadness. The spread of genuine and imposter scores also increases across these three. However, as these means and deviations remain fairly close and there are very few data sources to allow a broad evaluation of our research questions, we simply state our findings as observations instead of conclusions. In summary, our experiments show that the verification task is minimally influenced by an individual's mood, although negative arousal appears to have a slightly more negative impact on performance.

TABLE II
SCORE DISTRIBUTIONS PER MOOD STATE.

	Pos. Arousal, Pos Valence	Neg. Arousal, Pos. Valence	Pos. Arousal, Neg. Valence	Neg. Arousal, Neg. Valence
Genuine Score (μ)	0.87	0.85	0.89	0.84
Imposter Score (μ)	0.13	0.15	0.11	0.16
Standard Deviation	0.13	0.15	0.10	0.19

B. *Mood Versus Identification*

Using the same features as verification, we create a multi-class problem to study the identification task. Table III shows the achieved precision, recall, and $F1$-scores per subject. On average, an 83% $F1$-score was achieved. Again, to our knowledge, this is the first demonstration of the use of a single, daily sample of smartphone activity for identification. In fact, most mobile biometric work is focused on the verification task. In this table, light gray rows correspond with $F1$-scores above 95%, and dark gray rows correspond with $F1$-scores below 70%. We highlighted these rows to allow better visualization of potential trends that might be associated with each person's mood.

As most subjects have the same mood for most of their samples (see column Typical Mood), we found no significant correlation between a person's most common mood state and the ability to identify them. We also list the number of mood changes, going from one mood state to another in consecutive samples. The average number of mood changes was 13.48. Regarding this statistic, we found that subjects with lower identification accuracy (dark gray rows) had fewer mood changes than the average, while all but one subject with high identification accuracy (light gray rows) had more mood changes than the average. Finally, we also identified the number of extreme mood changes between consecutive samples, switching from positive valence and arousal to negative valence and arousal. There were very few of these occurrences (less than one per subject on average) as shown in Table III, such that no conclusions could be drawn.

In summary, our major insight regarding the identification task is that lower identification accuracy was associated with fewer changes in mood, while high performance was associated with more mood changes. Contrasting previous claims, our findings suggest that frequent changes in mood may not negatively impact performance, but may, in fact, improve performance.

C. *Mood Recognition via Activity Data*

We continued our analysis of mood by running an additional set of experiments to determine if smartphone activity data could be used to predict an individual's mood state. Using the same features as verification and identification, the classification label was changed to one of the four mood states. Figure 3 plots the ROC curve for each class. Interestingly, the average area of these curves (AUC) is 0.60. Positive arousal and negative valence yielded the highest AUC (0.67). This was also the class associated with the highest average genuine and lowest average imposter scores for verification experiments (see Table II). We also found a greater difference in performance for these experiments compared to the verification and identification tasks. Specifically, we found that the level of valence played a greater role in the utility of activity data to predict mood than it did for recognition. This could provide an indication that feeling down (stressed or depressed) alters the way a person uses their device (though not to a significant enough degree that they deviate from their usual usage patterns). Consequently, this observation is likely a key factor in the emergence of mHealth applications which leverage smartphone sensors to learn and respond to users' moods in real-time [29].

IV. SUMMARY

It is well known that smartphone activity data, an emerging and promising biometric identifier, changes over time as an individual's experiences or contexts change. Exactly how these changes affect mobile, behavioral biometrics, however, requires further investigation. We sought to understand the impact of mood on performance in a mobile biometric system as mood is frequently claimed to be one of many factors influencing changes in mobile, behavioral biometric

TABLE III

PRECISION AND RECALL FOR IDENTIFICATION EXPERIMENTS PER SUBJECT WITH CORRESPONDING MOOD STATISTICS. LIGHT GRAY ROWS CORRESPOND WITH $F1$-SCORES ABOVE 95%, AND DARK GRAY ROWS CORRESPOND WITH $F1$-SCORES BELOW 70%.

ID	Typical Mood	# Mood Changes	# Extreme Mood Changes	Precision	Recall	$F1$-Score
1	Pos. Valence, Neg. Arousal	15	0	0.92	0.6	0.73
2	Pos. Valence, Neg. Arousal	21	3	0.68	0.79	0.73
3	Pos. Valence, Pos. Arousal	4	0	0.95	0.97	0.96
5	Pos. Valence, Neg. Arousal	19	1	0.9	0.9	0.9
6	Pos. Valence, Neg. Arousal	23	0	1	0.97	0.99
7	Pos. Valence, Neg. Arousal	17	2	0.86	0.69	0.77
8	Pos. Valence, Neg. Arousal	12	0	0.59	0.76	0.67
9	Pos. Valence, Pos. Arousal	17	3	0.92	0.94	0.93
12	Pos. Valence, Pos. Arousal	21	1	0.85	0.89	0.87
13	Pos. Valence, Neg. Arousal	9	0	0.97	0.83	0.89
14	Pos. Valence, Neg. Arousal	13	0	0.76	0.94	0.84
15	Pos. Valence, Pos. Arousal	8	0	0.64	0.6	0.62
16	Pos. Valence, Pos. Arousal	11	0	0.78	0.39	0.52
17	Pos. Valence, Pos. Arousal	4	0	0.94	0.8	0.86
19	Pos. Valence, Neg. Arousal	20	0	0.94	0.78	0.85
20	Pos. Valence, Pos. Arousal	9	1	0.45	0.86	0.59
23	Pos. Valence, Pos. Arousal	13	0	0.94	0.89	0.91
24	Pos. Valence, Pos. Arousal	12	1	0.96	0.74	0.83
25	Pos. Valence, Neg. Arousal	16	2	0.95	0.77	0.85
26	Pos. Valence, Pos. Arousal	15	0	0.8	0.86	0.83
27	Pos. Valence, Pos. Arousal	20	0	0.92	0.89	0.9
28	Pos. Valence, Neg. Arousal	13	0	0.8	0.85	0.82
29	Pos. Valence, Neg. Arousal	15	0	0.94	0.97	0.96
30	Pos. Valence, Neg. Arousal	8	0	0.92	0.9	0.91
31	Pos. Valence, Pos. Arousal	2	0	0.59	0.84	0.69
32	Pos. Valence, Pos. Arousal	11	1	0.96	0.89	0.93
33	Pos. Valence, Neg. Arousal	16	2	1	0.91	0.95
		13.48	**0.63**	**85%**	**82%**	**83%**

modalities. We explored three aspects to this problem. (1) Could changes in an individual's mood have a significant impact on the performance of a mobile biometric system? (2) If so, do these changes affect verification, identification, or both recognition tasks? (3) Does the use of smartphone activity data for mood prediction correlate with biometric recognition?

To explore these questions, we first ran verification experiments, achieving an EER of 3% and d-prime of 5.05. Overall, our experiments showed that the verification task is minimally influenced by an individual's mood, although negative arousal appears to have a slightly more negative impact on performance. We then created a multi-class problem to study the identification task. On average, an 83% $F1$-score was achieved. In general, we found no significant correlation between a person's mood and the ability to identify them, but there was an indication that subjects with lower identification accuracy had fewer mood changes compared to the average.

We continued our analysis of mood by running an additional set of experiments to determine if smartphone activity

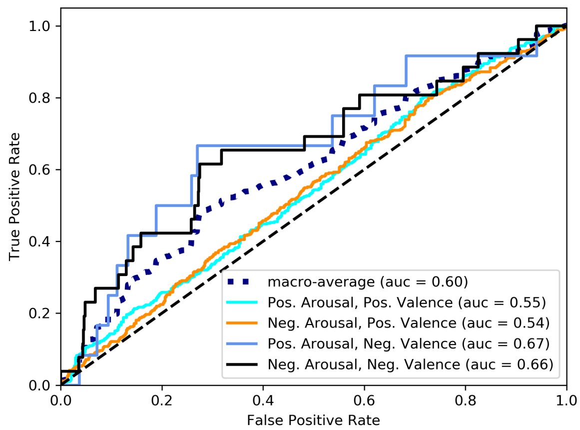

Fig. 3. ROC curve for smartphone activity recognition according to the four quadrants of the Circumplex mood model.

data could be used to predict an individual's mood state, and if outcomes of these experiments would correlate with our findings from studying verification and identification. Using the same features as verification and identification, the classification label was changed to one of the four mood states. Positive arousal and negative valence (i.e., tense, stressed, or frustration) yielded the highest AUC (0.67). This was also the class associated with the highest average genuine and lowest average imposter scores for verification experiments. Consequently, this observation is likely a key factor in the emergence of sensor-enhanced mHealth applications.

One limitation of our study is the total number of samples per subject, covering approximately only one month of smartphone activity. In addition, due to the limited number of publicly available datasets with samples annotated with mood, our experiments were only evaluated with a single dataset. Further, although our experiments were evaluated against a voting classifier, future work should also consider a wide range of classifiers, including deep neural networks. Because there were very few samples, we were not able to leverage the learning capabilities of deep neural networks. Future work should also consider the impact of emotion; whereas mood lasts much longer and is more difficult to explain, emotions are brief episodes with concrete causes. As a consequence, people may be much more reactive to sudden emotions compared to mood, and it would be worth exploring the impact of emotion on changes in behavior. Considering this, an interesting application of this work is a real-time system that both recognizes the emotion of an individual, as well as uses this information to help further identify them. It has been shown that video and audio can be fused to recognize emotion in a ubiquitous environment (e.g. mobile device) [13]. An ensemble-based approach, such as through the fusion of facial data, audio, and smartphone activity, could have the potential to offer new insight into the impact of emotion on mobile biometrics.

References

[1] J. Asselbergs, J. Ruwaard, M. Ejdys, N. Schrader, M. Sijbrandij, and H. Riper. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. *Journal of medical Internet research*, 18(3), 2016.

[2] J. E. Bardram, M. Frost, K. Szántó, and G. Marcu. The monarca self-assessment system: A persuasive personal monitoring system for bipolar patients. In *Proceedings of the 2Nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, pages 21–30, New York, NY, USA, 2012. ACM.

[3] R. Bhagavatula, B. Ur, K. Iacovino, S. M. Kywe, L. F. Cranor, and M. Savvides. Biometric authentication on iphone and android: Usability, perceptions, and influences on adoption. In *In Proc. USEC*, 2015.

[4] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and E. Bella-Pulgarin. Automatic usability and stress analysis in mobile biometrics. *Image and Vision Computing*, 32(12):1173 – 1180, 2014.

[5] A. De Luca, A. Hang, E. von Zezschwitz, and H. Hussmann. I feel like i'm taking selfies all day!: Towards understanding biometric authentication on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1411–1414, New York, NY, USA, 2015. ACM.

[6] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama. 3d face recognition under expressions, occlusions, and pose variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2270–2283, Sep. 2013.

[7] D. Fabiano and S. Canavan. Emotion recognition using fused physiological signals. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 42–48, Sep. 2019.

[8] H. Feng, K. Fawaz, and K. G. Shin. Continuous authentication for voice assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, MobiCom '17, pages 343–355, New York, NY, USA, 2017. ACM.

[9] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, Jan 2013.

[10] M. Harbach, E. von Zezschwitz, A. Fichtner, A. D. Luca, and M. Smith. It's a hard lock life: A field study of smartphone (un)locking behavior and risk perception. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 213–230, Menlo Park, CA, 2014. USENIX Association.

[11] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, Jan 2004.

[12] A. K. Jain, A. A. Ross, and K. Nandakumar. *Introduction*, pages 1–49. Springer US, Boston, MA, 2011.

[13] R. Jannat, I. Tynes, L. L. Lime, J. Adorno, and S. Canavan. Ubiquitous emotion recognition using audio and video data. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 956–959, 2018.

[14] N. Jaques, S. Taylor, A. Azaria, A. Ghandeharioun, A. Sano, and R. Picard. Predicting students' happiness from physiology, phone, mobility, and behavioral data. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 222–228. IEEE, 2015.

[15] Y. Javed, M. Shehab, and E. Bello-Ogunu. Investigating user comprehension and risk perception of apple's touch id technology. In *Proceedings of the 12th International Conference on Availability, Reliability and Security*, ARES '17, pages 35:1–35:6, New York, NY, USA, 2017. ACM.

[16] P. M. Lewinsohn, J. R. Seeley, R. E. Roberts, and N. B. Allen. Center for epidemiologic studies depression scale (ces-d) as a screening instrument for depression among community-residing older adults. *Psychology and aging*, 12(2):277, 1997.

[17] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, pages 389–402, New York, NY, USA, 2013. ACM.

[18] W. Meng, Y. Wang, D. S. Wong, S. Wen, and Y. Xiang. Touchwb: Touch behavioral user authentication based on web browsing on smartphones. *Journal of Network and Computer Applications*, 117:1 – 9, 2018.

[19] W. N. Morris. *Mood: The Frame of Mind*. Springer, New York, NY, 1989.

[20] T. J. Neal and D. L. Woodard. Surveying biometric authentication for mobile device security. *Journal of Pattern Recognition Research*, 1(74-110):4, 2016.

[21] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello. Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine*, 33(4):49–61, July 2016.

[22] L. Peel and A. Clauset. Detecting change points in the large-scale structure of evolving networks. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[23] A. Pratap, D. Atkins, B. Renn, M. Tanana, S. Mooney, and P. Anguera, J .and Areán. The accuracy of passive phone sensors in predicting daily mood. *Depression and anxiety*, 36(1):72–81, 2019.

[24] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[25] A. Sano, S. Taylor, A. W. McHill, A. J. Phillips, L. K. Barger, E. Klerman, and R. Picard. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. *Journal of medical Internet research*, 20(6), 2018.

[26] S. Shiffman, A. A. Stone, and M. R. Hufford. Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4(1):1–32, 2008. PMID: 18509902.

[27] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss. Activity-aware mental stress detection using physiological sensors. In *International conference on Mobile computing, applications, and services*, pages 282–301. Springer, 2010.

[28] J. Torous, P. Staples, and J.-P. Onnela. Realizing the potential of mobile mental health: New methods for new data in psychiatry. *Current Psychiatry Reports*, 17(8):61, Jun 2015.

[29] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. 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*, pages 3–14. ACM, 2014.

[30] H. Xu, Y. Zhou, and M. R. Lyu. Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In *10th Symposium On Usable Privacy and Security (SOUPS 2014)*, pages 187–198, Menlo Park, CA, 2014. USENIX Association.