

Daytime Data and LSTM can Forecast Tomorrow’s Stress, Health, and Happiness

Terumi Umematsu^{1 2}, Akane Sano^{3 1}, Rosalind W. Picard¹

Abstract—Accurately forecasting well-being may enable people to make desirable behavioral changes that could improve their future well-being. In this paper, we evaluate how well an automated model can forecast the next-day’s well-being (specifically focusing on stress, health, and happiness) from static models (support vector machine and logistic regression) and time-series models (long short-term memory neural network models (LSTM)) using the previous seven days of physiological, mobile phone, and behavioral survey data. We especially examine how using only a portion of the day’s data (e.g. just night-time, or just daytime) influences the forecasting accuracy. The results show that accuracy is improved, across every condition tested, by using an LSTM instead of using static models. We find that daytime-only physiology data from wearable sensors, using an LSTM, can provide an accurate forecast of tomorrow’s well-being using students’ daily life data (stress: 80.4%, health: 86.0%, and happiness: 79.1%), achieving the same accuracy as using data collected from around the clock. These findings are valuable steps toward developing a practical and convenient well-being forecasting system.

I. INTRODUCTION

Early detection indicators that one’s well-being is getting worse can enable new kinds of interventions to potentially prevent a series of bad stress or bad mood days from taking a turn into depression. Stress is well-known to increase susceptibility to infection and illness [1]. Self-reported health strongly relates to actual health and all-cause mortality [2]. Self-reported happiness is strongly correlated to measures of depression [3]. The ability to forecast well-being levels, and identify what specifically changes them, could enable better self-management of one’s behavioral choices in ways that might prevent poor well-being and subsequently prevent its damage to physical and mental health. The ability to model and forecast well-being could be immensely beneficial, especially if such a forecast could be made using data collected in a privacy-sensitive, convenient and unobtrusive way.

Previous work has shown that tomorrow’s well-being (good/poor mood, high/low stress and good/poor health levels) can be predicted with 78-82% classification accuracy based on today’s physiological and behavioral data by using personalized machine learning models [4]. Other work has

shown that using seven days of time-series data with long short-term memory neural network models (LSTM) can give acceptable results in well-being prediction without the need to build personalized prediction models for forecasting the next-day’s stress [5]. These previous works assumed the use of a whole day of data (typically 20-23 hours of sensor and smartphone data, plus a few mins/day of daily surveys). However, if an accurate forecasting system could be built using only daytime data, then a person could be “sensor free” for the evening, which might afford them more privacy and comfort, while also using lower power and making batteries last longer. Depending on a person’s schedule, other intervals of the day may also be most informative and convenient.

Some previous work has focused on using skin conductance, skin temperature and acceleration data from only daytime working hours to forecast long-term stress, stress scores from Perceived Stress Scale questionnaire [6] [7]. However, for early detection of worsening well-being, higher granularity well-being prediction would be more effective. We desire to create daily well-being forecasting that is so unobtrusive and convenient that somebody finds it easy enough to participate for a month or longer, as long-term data can help a person learn patterns that may enable them to significantly lower their stress, decrease days of sickness, and increase days with positive mood.

In this paper, we evaluate the next-day’s well-being forecasting accuracies from static models (support vector machine (SVM) and logistic regression (LR)) and time-series models (LSTM) using the previous seven days of physiological, mobile phone, and behavioral survey data. We investigate whether well-being can be forecasted accurately using only a portion of the day’s data such as just daytime or night-time (defined in this paper’s analysis “daytime” as 10am-5pm and “night-time” as 0-10am, based on the sleep-wake patterns of the participating college students). Given that current and future human mental conditions are affected by the past days [1], we hypothesized that an LSTM model may achieve a sufficiently accurate well-being forecast using one week of a portion of the day’s data from a participant.

This paper makes new contributions that advance the capability of forecasting well-being in daily life. We provide significantly novel results showing that daytime physiology data from wearable sensors, using an LSTM, can provide an accurate forecast (79-86%) of next-day’s well-being ratings in students’ daily life. These performance is equivalent to the accuracy of prior work that required the data from around the clock. Furthermore, we will also show that the new LSTM method can achieve the high well-being forecasting accuracy

This work is supported by the National Institute of Health (R01GM105018), the National Science Foundation (#1840167), the MIT Media Lab Consortium, Samsung Electronics and NEC Corporation. We thank our collaborators and study participants (<https://snapshot.media.mit.edu/team/>).

¹Massachusetts Institute of Technology, Media Lab, 75 Amherst Street, Cambridge, MA, USA, 02139. {terumi, picard}@media.mit.edu

²NEC Corporation, Biometrics Research Laboratories.

³Rice University, Department of Electrical and Computer Engineering
akane.sano@rice.edu

using daytime physiology data plus a survey, if we can ask a few minutes of daily survey (stress: 82.3%, health: 87.5%, and happiness: 83.5%). These contributions make this work valuable toward developing the first practical system for forecasting and improving well-being, while letting people keep their night-time life private and unmonitored.

II. DATA AND METHODS

A. Dataset

The data in this experiment measured Sleep, Networks, Affect, Performance, Stress, and Health using Objective Techniques (SNAPSHOT) [8], which gathered 30-day multi-modal data, including physiological, mobile phone, and behavioral survey data from college students in one US university. The study participants obtained compensation based on their contribution to the study. Stress, health and happiness scores were collected every evening, using self-reported scores from 0 (stressed out) - 100 (calm), 0 (sick) - 100 (healthy), and 0 (sad) - 100 (happy), respectively. In this paper, for stress, we used a total of 1,231 periods of 8 consecutive days of data from 142 participants (these periods are overlapping, resulting in a total of 2,276 days of data being used). We used 1,246 periods for health, and 1,133 periods for happiness, respectively.

B. Feature Calculation

We computed 375 daily features including 37 behavioral survey (excluding self-reported well-being scores), 173 physiology, 150 mobile phone, and 15 mobility features. The feature modalities are explained in detail below.

(a) Survey:

Participants filled out a survey about their daily behaviors every evening. They self-reported the timing and duration of a variety of activities, including sleep, academic activities, extracurricular activities, and exercise. Whether the participant engaged in social activity before bed, amount of caffeine intake, and whether a positive or negative social interaction was experienced were also self-reported. The survey typically took a few minutes to complete. We defined 20 features as “sleep survey” related to sleep such as bedtime, sleeping duration, efficiency, pre-sleep activity, whether or not the participant slept, the number of minutes the participant spent awake after going to bed at night, and time spend napping.

(b) Physiology:

The physiological measurements were collected by wrist-worn Affectiva Q sensors at 8 Hz; they include 24-hour-a-day electrodermal activity (EDA) measured as skin conductance (SC), skin temperature (ST), and 3-axis accelerometer. Features such as step count, stillness, and SC responses were calculated — all of which relate to emotional arousal and stress. EDA, acceleration and ST were collected to measure sympathetic nervous activity, physical activity, sleep patterns, circadian rhythm, and stress responses [9]–[11]. Following [12] and [4], for each time period (0-24H, 0-3H, 3-10H, 10-17H, 17-24H) the following sets of features were computed: EDA Peak features (for all detected peak features and for only non-artifact peaks [13]), SC level

features, accelerometer features, temperature features, and various combinations of the three physiological data streams. We defined data during 10-17H (10am-5pm) as “daytime” data (most students were active then), data during 0-10H (0-10am) as “night-time” data (most student sleep took place during this interval), and data during 0-10 and 17-24H (5-11:59pm) as “other-time” data.

(c) Phone:

The phone log data consisted of information about the timing, type, and duration of phone calls and SMS messages, and times the screen was turned on and off. We assumed that there are two main mechanisms through which screen logs and communication information can affect well-being: (1) light from the screen can disrupt circadian rhythms and therefore sleep [14], and (2) the amount of social support in a person’s life is strongly linked to resilience to depression [15], [16]. As with physiology, the features were computed over the time intervals spanning the course of the day. We defined daytime data (10-17H) and other-time data (0-10, 17-24H) the same as physiology features.

(d) Mobility:

In addition to communication and screen events, the phone app logged the participants’ GPS coordinates throughout the day, as well as whether they were using Wifi or cellular data. Previous studies have shown that mobility patterns are linked with mental health states [17], [18]. We followed the method described in [19] to down-sample the signal and compute features such as the total distance traveled, statistical features about distance traveled in 5 minutes, and the amount of time spent on campus. These data were only included in two set of comparisons where they were used all features around the clock and just only this set.

III. EXPERIMENTS

We conducted a series of experiments to examine whether we could improve the well-being forecasting accuracy using collected various time data.

A. Classification Labels

We framed the problem as a binary classification: days on which a participant reported a stress-calm, a sick-healthy, and a sad-happy score in the top 40% of all scores are labeled as a low-stress, a low-sick, and low-sad day, and days in which participants reported a well-being score in the bottom 40% are labeled as a high-stress, a high-sick, and a high-sad day. We discarded only the middle 20% of scores, similarly to Taylor et al. [4].

B. Methods

(1) Long Short-Term Memory Networks (LSTM):

LSTM Networks [20] have the ability to learn long-term dynamics while avoiding vanishing and exploding gradient problems and have recently gained great success in sequence learning tasks such as speech recognition and machine translation. Fig. 1 presents an overview of our LSTM. We designed our LSTM as three layers with a single hidden-layer with 32 nodes and a dropout of 0.2. Drop-outs were

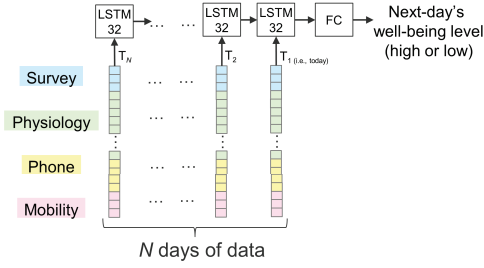


Fig. 1. Overview of our LSTM method

used between the LSTM and dense layers. The output of the last cell of the LSTM layer was connected to a dense layer. Finally, a sigmoid activation layer predicted the high/low well-being levels. We trained our LSTM using RMSprop [21] with binary cross-entropy loss.

(2) Static Methods:

For comparison to the LSTM, we used static models, including an SVM classifier with a radial basis function kernel and LR, as a baseline for learning algorithms because these methods were used in previous studies for mood prediction [4], [19], [22]. Because SVM and LR cannot directly handle time series data, we concatenated the time series feature values to create a single feature vector, allowing SVM and LR to learn a forecast model based on the same information.

C. Experimental conditions

We examine how accurately the previous seven days' multi-modal data using only a portion of the day's data can forecast a next-day's evening high/low well-being level. We used accuracy as evaluation metrics. The baseline accuracy (random classifier) was 54.1% for stress, 51.5% for health, and 50.0% for happiness, respectively. The full dataset was then used in a five-fold cross validation with 80% of the data for training and validating the models, and 20% for testing for each fold. While it's possible that days could have been repeated within the training/validation loops, the days in the test set were kept completely independent of the training and validation data (i.e., day-independent, but participant-dependent model). Specifically, within the training and validation set, we used 80% of the dataset for training and 20% as validation and selected the hyperparameters (LSTM: iteration number, SVM: C, LR: C) that yielded the highest accuracy on the validation set. For evaluating accuracy we computed the average and the standard deviation of the test set for the five folds.

Using the set-up above, we compare the accuracies of the three machine learning methods. Specifically, we learn $p(y_{t+1}|x_t, \dots, x_{t-6})$, the probability of the person's well-being given the previous seven days' data, where x_t is all the data collected from behavioral surveys, wearable sensors, and mobile phones on day t , and y_{t+1} is the next-day self-reported binary well-being label. We also train models and compute metrics for each of the models when using a single data modality (e.g., physiology), and using only daytime, night-time, other-time data, and each combination. Table I describes each features' combination and the experimental conditions.

TABLE I
FEATURES (ABOVE) AND EXPERIMENTAL CONDITIONS (BELOW)

Survey		Physiology			Phone		Modality
Sleep	Other	Daytime (10-17H)	Night-time (0-10H)	Other (0-10, 17-24H)	Daytime (10-17H)	Other (0-10,17-24H)	Whole time (0-24H)
a	b	c	d	e	f	g	h
Features (#)							
All (375)							a+b+c+d+e+f+g+h
Survey (37)							a+b
Physiology (173)							c+d+e
Phone (150)							f+g
Mobility (15)							h
Sleep Survey (20)							a
Other Survey (17)							b
Daytime Physiology (44)							c
Night-time Physiology (87)							d
Other time Physiology (129)							e
Daytime Phone (21)							f
Other time Phone (129)							g
Daytime Physiology and Sleep Survey (64)							c+a
Daytime Physiology and Survey (81)							c+a+b
Night-time Physiology and Sleep Survey (107)							d+a
Night-time Physiology and Survey (124)							d+a+b
Other time Physiology and Sleep Survey (149)							e+a
Other time Physiology and Survey (166)							e+a+b
Daytime Physiology and Daytime Phone (65)							c+f
Other time Physiology and Other time Phone (258)							e+g

IV. RESULTS AND DISCUSSION

A. Stress-forecasting results

The stress-forecasting accuracies of the SVM, LR and LSTM models using seven days of data are shown in Fig. 2. The results further show that accuracy is improved by using an LSTM instead of the static SVM or LR, regardless of which feature sets are used. The best results obtained for each model were as follows: for the LSTM: 84.2% accuracy using all features, for the SVM: 77.7% using other time physiology and sleep survey features, and for LR : 66.4% using physiology features. Using the LSTM, the accuracy based on daytime physiology features (80.4%) showed no significant decrease compared to the best result using all features ($p > 0.0025$, one way analysis of variance, Tukey's honest significant difference test and Bonferroni correction). This result means we can automatically forecast next-day's stress levels with high accuracy using an LSTM processing passive daytime wearable sensor data. Furthermore, we also see using the combination of the daytime physiology data and survey data give strong results for forecasting stress (82.3%, $p > 0.0025$).

We computed the mean absolute weights of each feature across all connected nodes in the input layer of the LSTM in using daytime-only physiology features. Features with higher weights indicate a stronger influence on forecasting stress. The top five features were related EDA: (1) the median of 5-minute epochs of the number of EDA peaks normalized by the number of steps, (2) the percentage of signal containing 1 minute epochs with greater than five no-artifact EDA peaks, (3) SC level's percentage of period where sensor is off, (4) the median of the number of EDA peaks per 30 minute epoch, and (5) the number of no-artifact EDA peaks per 30 minute epoch. These weights are consistent with the results of prior works [9] [23] [24] that daytime EDA features are good indicators for stress.

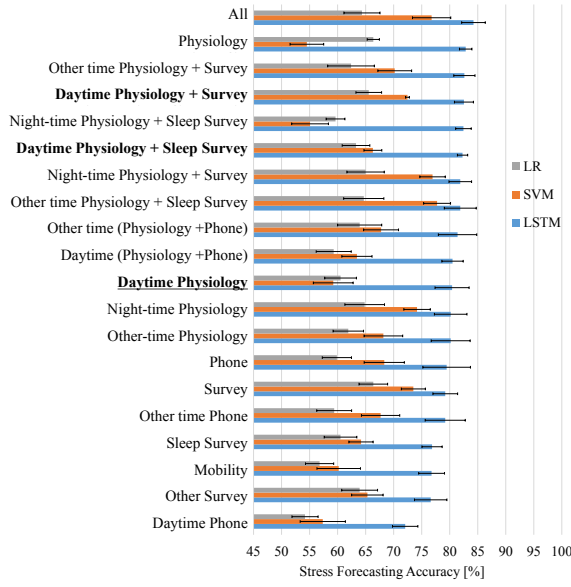


Fig. 2. Next-day's Stress Forecasting Accuracy (A)

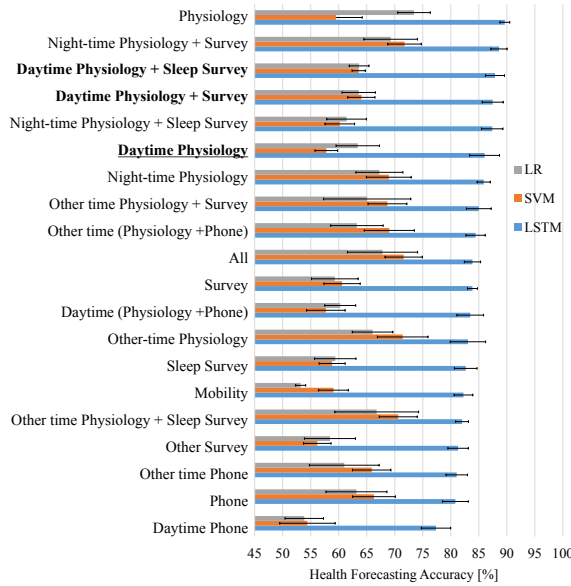


Fig. 3. Next-day's Health Forecasting Accuracy (B)

B. Health-forecasting results

The health-forecasting accuracies of the SVM, LR and LSTM models using seven days of data are shown in Fig. 3. The best results overall were obtained for the LSTM: 89.6% accuracy using physiology features, for the SVM: 71.7% using night-time physiology and survey, and for LR : 73.4% using physiology data. As with stress forecasting, using the LSTM, the accuracy with daytime physiology features data (86.0%) showed a similar accuracy to the best results (using physiology features, $p>0.0025$).

As with stress forecasting, we computed the mean absolute weights in daytime physiology data. The five top-weighted features were the same for forecasting stress, but just switched the order of the 2nd and 3rd features. These weights show that EDA features might be used as a bio marker for sickness.

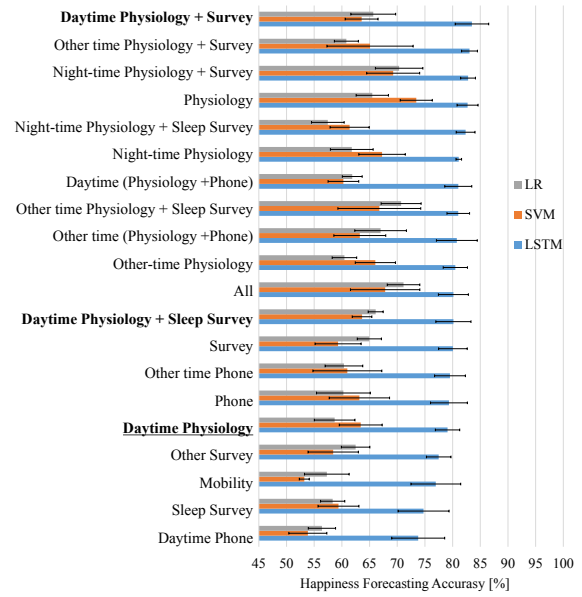


Fig. 4. Next-day's Happiness Forecasting Accuracy (C)

C. Happiness-forecasting results

The happiness-forecasting accuracies of the SVM, LR and LSTM models using seven days of data are shown in Fig. 4. The best results overall were obtained for the LSTM: 83.5% accuracy using daytime physiology features and survey, for the SVM: 73.4% using physiology features, and for LR : 71.1% using all features. As with stress and health forecasting, the daytime physiology showed a similar accuracy (79.1%) to the best result (using daytime physiology features and survey, $p>0.0025$).

We also computed the mean absolute weights in using daytime-only physiology features. The top five features were (1) the median of 5-minute epochs of the number of EDA peaks normalized by the number of steps, (2) the percentage of signal containing 1 minute epochs with greater than five no-artifact EDA peaks, (3) the minimum of the raw temperature signal, (4) the sum of the area under the curve (AUC) of all EDA peaks for this period where the amplitude of peak was calculated as difference from base tonic signal, and (5) the sum of AUC of EDA peaks where the amplitude was calculated as difference from 0. These weights are consistent with the results of prior work [25] that EDA features might be an indicator of differentiating happy and neutral emotion.

D. Summary of results

To summarize the well-being ((A) stress, (B) health, and (C) happiness) results, we found that daytime-only physiology data from wearable sensors can provide an accurate forecast of next-day's well-being (bold letters in Fig. 2, 3, and 4), achieving the same accuracy as using data collected from around the clock. We also saw that only night-time physiology and other time physiology showed strong results for forecasting (A) stress and (C) happiness. For forecasting (B) health, the combination of night-time physiology and a survey (or a sleep survey) data was also effective. These

results might be desirable for some people who prefer wearing a sensor during night-time instead of daytime. It is especially effective using physiological data for forecasting (B) health. We think one of the reason is that physiology data were more directly measured from participants' body than a phone or survey, and self-reported health strongly relates to actual health [2].

V. CONCLUSIONS AND FUTURE WORKS

In this work, we tackled the hard problem of forecasting next-day's well-being levels of stress, happiness, and health, using data provided by college students. We compared the use of static models (SVM and LR) and time-series models (LSTM), each using the previous seven days of data – including physiological, mobile phone, and behavioral survey data. Different from previous work, we also built models that could operate using only daytime, night-time, other-time data, or their combinations. In some cases (e.g. children at school, or a team in an office) people may desire to wear sensors for only a limited portion of their day. We found that the LSTM model we trained was able to take daytime-only physiology data from wearable sensors, and accurately forecast the next-day's well-being in students' daily life, achieving accuracies of stress: 80.4%, health: 86.0%, and happiness: 79.1%. The model trained here, with these extensive comparisons, demonstrates a valuable step towards developing a practical well-being forecasting system that is more convenient than requiring continuous all-day wearing of sensors and collecting of data.

This work has several limitations. The dataset is from New England college students and might not generalize to other student populations or to other groups such as older office workers. We plan to collect daytime-only data from office workers, which they can use (privately) to get feedback that may improve their well-being. Further, we plan to examine not only classification of next-day's anticipated state of health (limited in this work to high/low values of the stress, health, and happiness) but also allow for the model to perform regression to anticipate more subtle changes. This can be done using other LSTM structures and deep learning structures to improve future well-being forecasting accuracies.

Finally, this work has focused on the hard problem of the forecasting and has not addressed several other very important problems, such as what to do with the forecasts. In particular, if a person gets a forecast saying tomorrow they are likely to have worse stress and worse health than today, then they will want to know what they can do to change the forecast. We are working on models that provide evidence-based recommendations to improve your well-being tomorrow. Thus, the ideal model will not only make it easy (passive data, minimal interaction) to let you know if tomorrow your well-being is on track to be the same or better, but it will also help you identify what you can most easily change to have a better outcome tomorrow (e.g. go to bed earlier, call a good friend or do something that elicits a positive social interaction). The work here does not

yet help people identify the best behaviors to change, nor does it support them in making those changes, both of which will be important parts of a complete system to help people significantly improve their future well-being.

REFERENCES

- [1] S. Cohen *et al.*, "Psychological Stress and Susceptibility to the Common Cold," *New England Journal of Medicine*, vol. 325, no. 9, pp. 606–612, 1991.
- [2] K. Abiola *et al.*, "Does the perception that stress affects health matter? the association with health and mortality," *Health Psychology*, vol. 31, no. 5, pp. 677–684, 2012.
- [3] H. Cheng *et al.*, "Personality, self-esteem, and demographic predictions of happiness and depression," *Personality and individual differences*, vol. 34, no. 6, pp. 921–942, 2003.
- [4] S. A. Taylor *et al.*, "Personalized Multitask Learning for Predicting Tomorrow's Mood, Stress, and Health," *IEEE Transactions on Affective Computing*, no. 99, pp. 1–14, 2017.
- [5] T. Umematsu *et al.*, "Improving stress forecasting using lstm neural networks," *EMBC*, 2018.
- [6] S. Cohen *et al.*, "A global measure of perceived stress," *Journal of health and social behavior*, pp. 385–396, 1983.
- [7] Y. Nakashima *et al.*, "Improvement in chronic stress level recognition by using both long-term and short-term measurements of physiological features," *EMBC*, 2018.
- [8] A. Sano *et al.*, "Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: Observational study," *J Med Internet Res*, vol. 20, no. 6, p. e210, 2018.
- [9] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.
- [10] C. G. Scully *et al.*, "Skin surface temperature rhythms as potential circadian biomarkers for personalized chronotherapeutics in cancer patients," *Interface Focus*, vol. 1, no. 1, pp. 48–60, 2011.
- [11] K. A. Herborn *et al.*, "Skin temperature reveals the intensity of acute stress," *Physiology & Behavior*, vol. 152, pp. 225–230, 2015.
- [12] S. A. Taylor, "Characterizing Electrodermal Responses during Sleep in a 30-day Ambulatory Study," *MIT, Master's Thesis*, 2016.
- [13] S. Taylor *et al.*, "Automatic identification of artifacts in electrodermal activity data," *EMBC*, pp. 1934–1937, 2015.
- [14] C. A. Czeisler *et al.*, "Bright light resets the human circadian pace-maker independent of the timing of the sleep-wake cycle," *Science*, vol. 233, no. 4764, pp. 667–671, 1986.
- [15] M. E. Seligman, "Flourish: a visionary new understanding of happiness and well-being," *Policy*, vol. 27, no. 3, pp. 60–61, 2011.
- [16] R. S. Peirce *et al.*, "A longitudinal model of social contact, social support, depression, and alcohol use," *Health Psychology*, vol. 19, no. 1, pp. 28–38, 2000.
- [17] L. Canzian *et al.*, "Trajectories of depression," *UbiComp*, pp. 1293–1304, 2015.
- [18] S. Saeb *et al.*, "Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study," *Journal of Medical Internet Research*, vol. 17, no. 7, p. e175, 2015.
- [19] N. Jaques *et al.*, "Predicting students' happiness from physiology, phone, mobility, and behavioral data," *ACII*, pp. 222–228, 2015.
- [20] S. Hochreiter *et al.*, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [21] T. Tieleman *et al.*, *Lecture 6.5-rmsprop*. COURSERA: Neural networks for machine learning Technical report, 2012.
- [22] Y. Suhara *et al.*, "DeepMood: Forecasting Depressed Mood Based on Self-Reported Histories via Recurrent Neural Networks," *World Wide Web Conference*, pp. 715–724, 2017.
- [23] C. Setz *et al.*, "Discriminating Stress From Cognitive Load Using a Wearable EDA Device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2010.
- [24] J. Hernandez *et al.*, "Call Center Stress Recognition with Person-Specific Models," *ACII*, pp. 125–134, 2011.
- [25] D. Schneider *et al.*, "Empathic behavioral and physiological responses to dynamic stimuli in depression," *Psychiatry Research*, vol. 200, no. 2-3, pp. 294–305, 2012.