Automatic Depression Prediction using Screen Lock/Unlock Data on the Smartphone

Jiwon Kim, Juyoung Hong, Yukyung Choi†

***Abstract*— As COVID-19 continues for a long time, more and more people feel psychologically anxious beyond stuffy. However, as people have high resistance to mental health treatment, there are many cases in which the treatment period with depression is missed and the symptoms are getting worse. In this paper, we study a technology to diagnose users' depression by using a smartphone that has become an indispensable item carried by the mass of modern people in everyday life. Most of the existing studies diagnosed depression by using the questionnaire responses from smartphone users directly, but this study intends to replace the questionnaire responses from only the mobile phone sensing data without the user's annoyingness. In particular in this paper, it shows that we can predict the user's sleep time using only lock/unlock data and detect changes in sleep patterns to predict the likelihood of depression in smartphone users.**

I. INTRODUCTION

As the prolonged corona 19 and social distancing continues, the number of severely depressed patients is rapidly increasing worldwide [1], [2]. Mental health problems such as depression, if left untreated, can lead to dementia, stroke, or even suicide in severe cases. Accordingly, the importance of monitoring and treatment for depression is growing day by day, and recent research on patient monitoring using smartphones is receiving a lot of attention from researchers.

Depression monitoring research using smart devices can be largely divided into studies using smartphones or wearable devices. Research using smartphones monitors depression levels by analyzing the results of survey responses related to mental health. The disadvantages of research using these survey responses are that the rate of response decreases over time and the accuracy of the responses is low, making accurate diagnosis difficult. In order to overcome these difficulties, researches for depression diagnosis are being actively conducted by using physical activity data by smart devices instead of survey responses from smartphone users. SmartWatch is the best choice for accurate physical activity data sensing. However, these wearable devices are expensive, making them unsuitable for solutions aimed at the general public. Therefore, for the popularization of depression diagnosis, it is a very necessary time to use smartphones to diagnose depression.

According to DSM-5 [3], a classification and diagnostic procedure for mental disorders published by the American Psychiatric Association (APA), depression is highly correlated with nine symptoms: depressed mood, diminished interest and pleasure in activities, fatigue, restlessness, sleep change, weight change, diminished ability to concentrate, feelings of worthlessness, and thoughts of death and suicide. Colbaugh et al. [4] use a smartphone's GPS and WIFI sensors to measure the decrease in human activity, and predict potential depression. R. Wang et al. [5] investigate the relationship between smartphone usage and diminished ability to concentrate. On top of that, a lot of researches [6], [7], [8], [9], [10], [11] demonstrated a correlation between passive sensors of smartphone and depression symptoms that is defined on DSM-5.

†Corresponding author
Jiwon Kim, Juyoung Hong, Yukyung Choi are with School of Intelligent Mechatronic Engineering, Sejong University, South Korea {jwkim, jyhong, ykchoi}@rcv.sejong.ac.kr

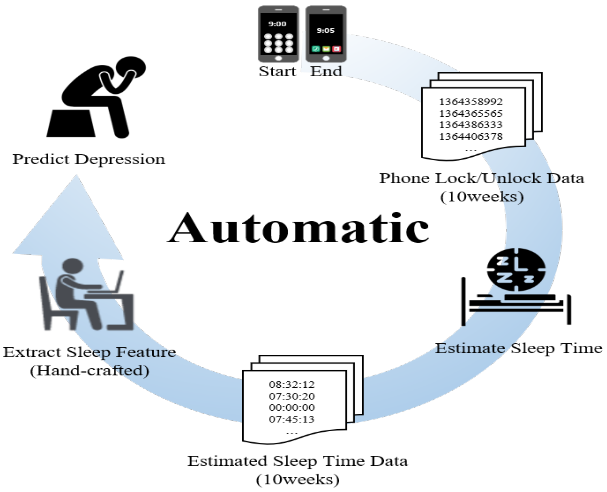

Fig. 1. **The circular flow diagram of automatic depression prediction** This circular flow diagram illustrates our proposed framework. Our proposed framework to predict depression uses only screen lock data, i.e. locking or unlocking, that is automatically collected when users utilize their smartphone.

In this paper, we show the advantage of predicting depression by using screen lock/unlock data independent of physical activity and only using automatically measured sleep changes. There was an attempt to integrate various sensors to extract features correlated with sleep change in previous researches [4]. However, protecting personal information is becoming more important in the smartphone industry. Therefore, the access to sensor data has been getting more restricted. Thus, we estimate sleep changes without various sensors, using only screen lock/unlock data taken from the sensor in a smartphone when the screen is unlocked or locked. We predict depression by extracting features based on the estimated sleep changes.

II. METHOD

In this section, we explain in more detail the method to predict depression by using only screen lock data. As shown

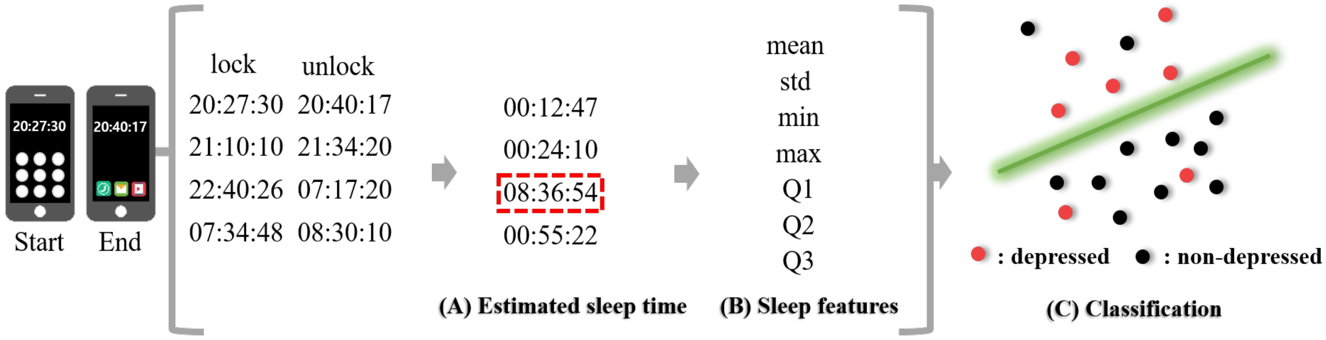

Fig. 2. **Schematic of our proposed framework** Our proposed framework predicts the depression by using the sleep feature which is extracted from estimated sleep time. We use the machine learning classifiers such as Support Vector Machine(SVM), Random Forest(RF), ans Extreme Gradient Boosting(XGBoost) to classify depressed or non-depressed.

in Figure 2, our proposed framework was divided into three-step: (A) estimating the sleep time from screen lock/unlock data, (B) extracting the statistical features using estimated sleep time, (C) classifying users of smartphone as depressed or non-depressed.

A. Estimating the sleep time

Some researchers [12], [13], [14] proposed methods to estimate sleep time by using smartphone sensors. Sano et al. [12] found that the screen lock/unlock information on the smartphone is highly correlated with the Pittsburgh Sleep Quality Index (PSQI). Cuttone et al. [13] proposed a Bayesian model to extract sleep patterns from smartphone events such as screen on/off based on [12]. Saeb et al. [14] use all available smartphone sensors, such as Light, Sound, Screen, Battery, GPS and WIFI, to estimate sleep time. In this paper, we used lock/unlock data to estimate sleep times by using a simple algorithm. This is detailed in Experiment B.

B. Extracting sleep features

There are various feature extraction methods for using sensor data. Quiroz et al. [15] designed features that consist of 17 characteristics for activity recognition, such as energy, kurtosis, skewness, root mean square and root sum square. Min et al. [16] used statistical features including minimum, maximum, the first quartile (Q1), the second quartile (Q2), the third quartile (Q3), average, and standard deviation of the sensor value to estimate sleep patterns. In this paper, we use statistical features based on [16] since the higher degree features can cause overfitting. In statistics, a quartile is a type of quantile that divides the number of data points into four parts.

C. Depression prediction

We classify users into two groups, depressed and non-depressed, by using the extracted sleep features. To find the best classifier for sleep features which is defined above, we used a commonly used machine learning classifier such as SVM [17], Random Forest [18], and XGBoost [19]. To optimize hyperparameters, we applied a Grid-search method that selects optimal hyperparameters, automatically. The method is for searching optimal hyperparameters for a learning algorithm. Typical examples are C and gamma for SVM.

TABLE I
DISTRIBUTION OF SCORES ABOUT PRE-POST PHQ-9 DEPRESSION SEVERITY FOR A GROUP OF STUDENTS

Depression severity	Cut-off score	pre survey	post survey
minimal	1∼4	17	19
minor	5∼9	15	12
moderate	10∼14	6	3
moderately severe	15∼19	1	2
severe	20∼27	1	2
Total	27	40	38

III. EXPERIMENT

A. Dataset

We conducted experiments on the publicly available StudentLife [20] dataset. The dataset was collected from 48 students over 10 weeks at Dartmouth College. The StduentLife dataset provides automatic sensing data from smartphones and the sensing data consists of accelerometer, microphone, light sensor, GPS, Bluetooth, screen lock/unlock, etc. Furthermore, the dataset also includes the results of a mental health survey to measure individual levels of depression, stress, flourishing, and loneliness.

In order to predict depression, we used Patient Health Questionnaire(PHQ)-9 [21] scores as ground truth. The PHQ-9 is the major depressive disorder module of the full PHQ module, which scores each of the 9 DSM-IV criteria as zero to three. PHQ-9 scores of 5, 10, 15, and 20 represented mild, moderate, moderately severe, and severe depression, respectively. Table I shows the distribution of scores about pre-post PHQ-9 depression severity for a group of students in the StudentLife dataset.

In this paper, we only used the result of post PHQ-9 survey that consisted of 38 students for experiments since the PHQ-9 is designed to capture depressive symptoms based on what they have done over the past 2 weeks.

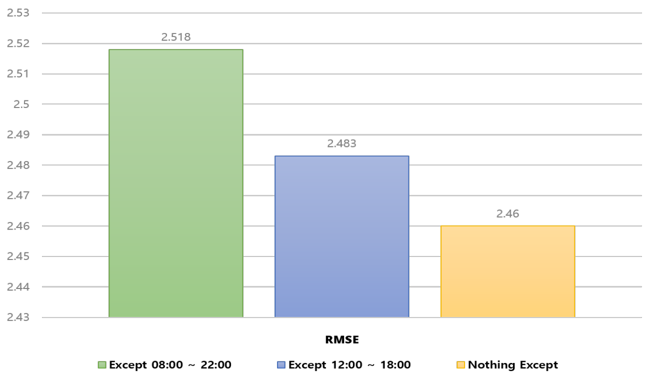

Fig. 3. **Comparison of three assumptions** We compare the root mean square error (RMSE) of assumptions. RMSE is calculated by taking the difference between the sleep time estimated from the lock/unlock data with assumptions and the sleep time obtained from responded through the Ecological Momentary Assessment(EMA).

B. Evaluation Metric

Two evaluation metric are used in depression prediction. Some studies [8], [9], [10], [11] use F-score metric to evaluate their methods. F-score is the harmonic mean of precision and recall. F-score is defined as follows :

$$F\text{-}score = \frac{precision * recall}{precision + recall} \tag{1}$$

Other studies [10], [7], [5], [11] apply Area under the Curve(AUC). AUC is a performance measurement for the classification problems at various threshold settings. In this paper, we utilize F-score for evaluating our methods and use AUC for comparing other methods which offer the result of their methods on StudentLife dataset.

C. Feature extraction

The lock/unlock data which provided in the StudentLife dataset contains time information about the screen lock state (i.e. start, end) in UNIX Epoch format. We designed a simple algorithm to estimate sleep. Firstly, we collect the lock/unlock data within one day before the users respond through the Ecological Momentary Assessment(EMA) about the sleeping hours. Secondly, we calculate smartphone usage as follows:

$$T_{Usg(i)} = T_{start(i)} - T_{end(i-1)} \tag{2}$$

where i represents the order index of lock/unlock data. T and Usg are the abbreviation for Time and Usage. Lastly, we estimate the sleep time with three assumptions: (a) Aside from the time between 8 in the morning and 10 in the afternoon, the longest time the users do not use the mobile phone is the sleeping hours, (b) Aside from the time between 12 noon and 6 in the afternoon, the longest time the users do not use the mobile phone is the sleeping hours, (c) For a day, the longest time the users do not use the mobile phone is the sleeping hours. Figure 3 shows the result of comparison that the assumption (c) has the lower root mean squared error(RMSE) by comparing with sleep time from EMA. Therefore, the sleep time is calculated as follows:

$$T_{sleep(i)} = T_{end(i)} - \{T_{end(i-1)} + T_{Usg(i)}\} \tag{3}$$

TABLE II

EXPERIMENT RESULTS ON THE STUDENTLIFE DATASET

	SVM	Random Forest	XGBoost
EMA	85%	74%	74%
EMA*	81%	74%	74%
Screen lock-(a)	85%	65%	65%
Screen lock-(b)	78%	79%	79%
Screen lock-(c)	**88%**	**88%**	86%

$$T^{*}_{sleep(d)} = \max_{i=1}^{n} T_{sleep(i)} \tag{4}$$

where n refer to the number of screen lock data which is collected within one day before the users respond through the EMA, and d refer to Day Index which is defined by the specified date range. (i.e. the first day in the date range is 0, for the second day 1). The percentage of missing data varies from person to person. Therefore, we use hand-crafted features based on statistical calculation and minimize the degree of features so that we can use them to overcome overfitting. Consequently, our sleep features are composed of minimum, maximum, the first quartile(Q1), the second quartile(Q2), the third quartile(Q3), average, and standard deviation of sleep time.

D. Evaluation

We predict depression through sleep features extracted from sleep time by using the lock/unlock data. Previous studies [10], [4] predict depression into binary classification that predicts Depressed(1) or Non-depressed(0). The threshold of PHQ-9 scores that determine either depression or non-depression is 10 based on [22]. When the threshold is applied to post-survey as ground truth, 7 students are depressed and the other 31 students are non-depressed. We use leave-one-out cross validation procedure which is commonly applied in the research [10], [4] for evaluation. Additionally, we conduct experiments to observe the performance difference caused by missing data. EMA and EMA* are the results of depression prediction by features extracted from the sleeping hours the users responded through the Ecological Momentary Assessment(EMA). Unlike EMA, EMA* eliminated the data on the day the lock/unlock data was missed. The lock/unlock data is missed 14% for each student on average. Experiment result can be found in Table II. The table shows that our experiments demonstrated that prediction using only lock/unlock data is as accurate as prediction based on EMA (or EMA*) which is data responded by the subjects. Therefore, using sleep time estimated to the lock/unlock data could be more meaningful for depression prediction than the users responded through the EMA. Moreover, we examine three assumptions to check whether the assumption (c) is correct selection. Table II shows difference in performance between the assumptions, which is explained at Feature extraction.

Furthermore, we conduct additional experiments to demonstrate the competitiveness of the proposed method on the StudentLife dataset. Experiment result can be found in Table III. Figure 4 shows the AUC of our method. The table

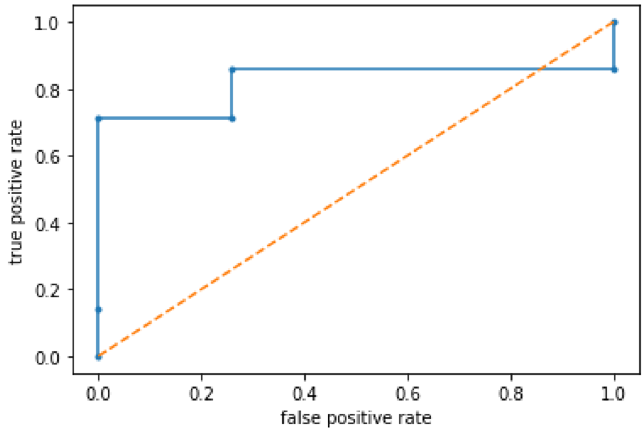

Fig. 4. **ROC curve** The ROC curve of using SVM to predict depression. The area under the ROC curve(AUC) is 0.820

TABLE III
COMPARISION WITH OTHER METHODS ON THE STUDENTLIFE DATASET

Method	AUC	Passive sensors
Gerych et al. [11]	0.800	GPS, accelerometer
Colbaugh et al. [4]	**0.975**	GPS, accelerometer, Screen lock/unlock, light sensor, microphone
Ours	0.820	Screen lock/unlock

shows that the AUC of our method is higher than the AUC of Gerych's method [11] which use GPS to extract features. In addition, Screen lock data can be easily collected regardless of smartphone type so that our method is more useful than a multi-sensor data fusion method [4].

IV. CONCLUSION

In this paper, we only used the screen lock/unlock data for predicting depression. The proposed approach achieves 88% F-score on the publicly available StudentLife dataset. Furthermore, we demonstrate that passive sensor data can be more accurate than self-reported information. We argue that our research has tremendous potential when the smartphone has become a ubiquitous presence in our lives. In addtion, our research can also be applied in various robotics field, such as robots for observing human action. We expect the result of this study will be a milestone for pre-diagnosing a depression by just simply using a few sensors in a smartphone without worrying about leakage of personal information.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT). (No.2020-0-01969, Development of HW/SW platform for interactive self-monitoring in real-time addressing the mental health problems based on brain-cognitive behavioral data)

REFERENCES

[1] C. K. Ettman, S. M. Abdalla, G. H. Cohen, L. Sampson, P. M. Vivier, and S. Galea, "Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic," *JAMA Network Open*, vol. 3, no. 9, p. e2019686, 2020.

[2] Bahk, Yong-Chun, et al., "Psychological Impact of COVID-19 in South Korea: A Preliminary Study," *Korean Journal of Clinical Psychology*, vol. 39, no. 4, pp. 355–367, 2020.

[3] American Psychiatric Association., "Diagnostic and statistical manual of mental disorders," The American Psychiatric Publishing Textbook of Geriatric Psychiatry, Fifth Edition, 2013.

[4] Colbaugh, R and K Glass, "Detecting and monitoring brain disorders using smartphones and machine learning," Technical Report, Volv Global, Lausanne, Switzerland, 2020.

[5] R. Wang et al., "Tracking depression dynamics in college students using mobile phone and wearable sensing," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–26, 2018.

[6] S. Saeb et al., "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study," *The Journal of Medical Internet Research*, vol. 17, no. 7, p. e175, 2015.

[7] S. Saeb, E. G. Lattie, K. P. Kording, and D. C. Mohr, "Mobile phone detection of semantic location and its relationship to depression and anxiety," *JMIR MHealth UHealth*, vol. 5, no. 8, p. e112, 2017.

[8] N. Narziev, H. Goh, K. Toshnazarov, S. A. Lee, K.-M. Chung, and Y. Noh, "STDD: Short-term depression detection with passive sensing," *Sensors*, vol. 20, no. 5, p. 1396, 2020.

[9] S. Ware et al., "Large-scale automatic depression screening using meta-data from WiFi infrastructure," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, pp. 1–27, 2018.

[10] S. Ware et al., "Predicting depressive symptoms using smartphone data," *Smart Health*, vol. 15, no. 100093, 2020.

[11] W. Gerych, E. Agu, and E. Rundensteiner, "Classifying depression in imbalanced datasets using an autoencoder- based anomaly detection approach," in *Proceedings of the IEEE International Conference on Semantic Computing*, 2019.

[12] A. Sano et al., "Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones," in *Proceeding of International Conference on Wearable and Implantable Body Sensor Networks*, 2015.

[13] A. Cuttone, P. Bækgaard, V. Sekara, H. Jonsson, J. E. Larsen, and S. Lehmann, "SensibleSleep: A Bayesian model for learning sleep patterns from smartphone events," *PLoS ONE*, vol. 12, no. 1, p. e0169901, 2017.

[14] S. Saeb, T. R. Cybulski, K. P. Kording, and D. C. Mohr, "Scalable passive sleep monitoring using mobile phones: Opportunities and obstacles," *The Journal of Medical Internet Research*, vol. 19, no. 4, p. e118, 2017.

[15] J. C. Quiroz, M. H. Yong, and E. Geangu, "Emotion-recognition using smart watch accelerometer data: Preliminary findings," in *Proceedings of the ACM International Symposium on Wearable Computers on UbiComp*, 2017.

[16] J.-K. Min, A. Doryab, J. Wiese, S. Amini, J. Zimmerman, and J. I. Hong, "Toss 'n' turn: Smartphone as sleep and sleep quality detector," in *Proceedings of the ACM conference on Human factors in computing systems*, 2014.

[17] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.

[18] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[19] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[20] R. Wang et al., "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014.

[21] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ-9: validity of a brief depression severity measure," *The Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[22] M. V. Costa et al., "Accuracy of three depression screening scales to diagnose major depressive episodes in older adults without neurocognitive disorders," *The Revista Brasileira de Psiquiatria*, vol. 38, no. 2, pp. 154–156, 2016.