**Other Supplementary materials**

**Design of features**

Because mood episodes could be affected by misalignment or disturbance of CRs,1,2 we focused on identifying a set of features that would capture disruptions to CR organization. To achieve this, we first focused on basic features stemming from the four main categories: (1) light exposure, (2) step counts (Fitbit-inferred), (3) sleep parameters (Fitbit-inferred), and (4) CR of heart rate (Appendix Table 2). To calculate light exposure, we considered the average value of light exposure levels observed in the morning (from sunrise to noon). To measure activity levels, we collected step data that were calculated as total steps within each of the four timeslots: ‘morning,’ ‘afternoon,’ ‘evening,’ and ‘bedtime.’ The timeslots ‘morning,’ ‘afternoon,’ ‘evening,’ and ‘bedtime’ were respectively defined as the period from sunrise to noon, from noon to sunset, from sunset to 8 h before sunrise of the next day, and from 8 h before sunrise to sunrise of the next day. Sleep data, such as sleep length and efficiency, and sleep onset and offset, were also obtained. The usual daily pattern showed that the heart rate falls at night and rises during the day. CR estimated with heart rate showed a higher correlation with that of salivary cortisol than that estimated with step count.3,4 Therefore, cosinor analysis (cosine curve fitting) was performed of the hourly average heart rate during the previous 48 h, and four CR parameters were generated: CR amplitude, CR acrophase, CR mesor, and CR goodness of fit. Since the CR acrophase circulates for 24 h, according to the distribution of the variable, the lowest value was set at noon for the major depressive episode and hypomanic episode (HME) prediction model and 6 AM for the manic episode prediction model. The value increased as time increased. Those at noon and 6 AM are phase-time reference points expressed as 0 h in the CR. Finally, we extracted the extended features from the basic features. In constructing the prediction model, all features (basic plus extended features) were used as predictors for mood states or episodes. To predict mood in the near future, it can be helpful to consider snapshots of the past few days as people are probably affected by features of mood change from the past few days. Therefore, we extended the daily snapshot features to the past 3, 6, and 12 days. For example, if today’s date is d, then the mean value of the past 3 days would be from d-2 to d. In this way, the mean (‘m’), standard deviation (‘SD’), and gradient (‘gr’) coefficient (i.e., a parameter gained from linear regression; ‘gradient’ indicates a trend of increasing or decreasing values) can be computed for the extended features. The names of all features in the across-period perspective included one of the aforementioned four main categories. The suffix terms also included the three elements in the statistic perspective for the given period: ‘m,’ ‘SD,’ and ‘gr.’ We finally acquired 140 features (= 14 basic features for every day + [14 basic features $×$ 3 types of past periods $×$ 3 types of statistics for those periods]).

 Fitbit fails to collect data on heart rate at times when the band is removed from the wrist or loosely worn. Therefore, instead of processing the cosine fit with all data for 24 h minute by minute (i.e., up to 1440 measures per day), we used hour-based 24 samples per day, which are average values per hour, minimizing the possible effects of missing data on minute-based intervals on the curve fitting process.

**The referenced time of sunset and sunrise in the study**

Korea is not a big country geographically, so the time of sunrise and that of sunset are mostly same for different cities. Therefore, we used a time reference table of annual statistics from the observatory reported in Seoul, with the assumption that the reference table is static without any annual change. The following is the table we referenced in the study; an unknown date between the rows in the table was estimated with the interpolation method because time flows linearly.

|  |  |  |
| --- | --- | --- |
| Date (month-day) | Sunrise | Sunset |
| 01-01 | 7:47 | 17:25 |
| 01-16 | 7:45 | 17:38 |
| 02-01 | 7:36 | 17:55 |
| 02-16 | 7:21 | 18:11 |
| 03-01 | 7:03 | 18:26 |
| 03-16 | 6:43 | 18:39 |
| 04-01 | 6:19 | 19:26 |
| 04-16 | 5:57 | 19:39 |
| 05-01 | 5:38 | 19:21 |
| 05-16 | 5:23 | 19:34 |
| 06-01 | 5:13 | 19:47 |
| 06-16 | 5:10 | 19:55 |
| 07-01 | 5:14 | 19:57 |
| 07-16 | 5:23 | 19:53 |
| 08-01 | 5:35 | 19:41 |
| 08-16 | 5:48 | 19:24 |
| 09-01 | 6:01 | 19:02 |
| 09-16 | 6:14 | 18:40 |
| 10-01 | 6:27 | 18:15 |
| 10-16 | 6:40 | 17:55 |
| 11-01 | 6:56 | 17:35 |
| 11-16 | 7:12 | 17:21 |
| 12-01 | 7:27 | 17:14 |
| 12-16 | 7:39 | 17:15 |

**Model evaluation process**

In the machine learning evaluation process, a part of the data is used for training the model and the other part for testing the model. During the evaluation, the training data should not include unknown future measurements relative to the test data. To consider the temporal nature of the data and obtain a reliable evaluation statistic, we designed the model evaluation process as follows. First, the data were sorted over a timeline. For an arbitrary time t on the timeline, a prediction model was trained using data on days d[t-p, t] and tested using data on days d[t+1, t+q], where p and q are the periods of days for model training and model testing, respectively. In this study, we used the optimal parameters of p=18 and q=3. Our previous study5 demonstrated that a short period, such as 3 days, is the most reasonable and effective setting for predicting any distant future episode event in our experiment. Second, to obtain a reliable evaluation result, we needed to repeat the evaluation rounds (i.e., a round of model training and model test) so that we repeatedly measured performance metrics by moving t from the beginning to the end of the data over the timeline with the parameter setting. Thus, the reported figures of sensitivity, specificity, accuracy, and area under the curve in the paper are average statistics from repeated evaluation rounds.

**Contribution analysis process for the proposed features**

To investigate the contribution of our designed features to mood episode predictions, we used the Shapley value.6 The Shapley value, coined by Shapley in coalitional game theory, is a method for assigning payouts to players depending on their contribution to the total payout. Players cooperate in a coalition and receive a certain profit from this cooperation. In our context, the “game” is the prediction task for a single instance of the dataset. The “gain” is the actual prediction for this instance minus the average prediction for all instances. The “players” are the feature values of the instance that collaborate to receive the gain (= predict an episode occurrence probability). The Shapley value is the average marginal contribution of a feature value across all possible coalitions. We repeated the computation for all possible coalitions. The Shapley value is the average of all marginal contributions to all possible coalitions.

**Data imputation for missing values**

Hereafter, the term ‘record’ refers to a set of features observed for a given day, and the term ‘field’ refers to a feature (variable). Single or multiple features in a record can have missing values that occasionally fail to be obtained during the data collection process. It is considered a missing value when a part or the entire 24 h of the data on a variable is lost. When partial data were missed within 24 h in the process of plottinga wave of circadian rhythm (CR), the interpolation method7 was used for data imputation. However, day-based features were not computable when data for the entire 24 h were missing. In this case, the following data imputation process was conducted.

Missing values in a record were filled with valid ones from similar records. We adopted the approach named k-Nearest Neighbors (k-NN); Neighbors stand for similar records. Euclidean distance is used to find the nearest neighbors. Each missing field is imputed using values from the three nearest neighbors (k is set to 3) that have a value for the field. The field of the neighbors is averaged (uniformly weighted) and used to fill the missing one. If a record has more than one field missing, then the neighbors for that record can be different depending on the particular field being imputed. For more information about k-NN based data imputation, please refer to the reference.8

**Under-sampling process**

In the context of mood episode prediction modeling, a positive sample indicates a day with any episode event, and a negative sample indicates a day with no episode event (euthymic mood day). The model performance is affected by the number of biased training samples. Because the percentages of positive and negative samples were unbalanced (i.e., positives are usually fewer than negatives), we used the under-sampling method, which reduces the sample size in a dominant sample group and secures the balance of the sample size. For under-sampling, a fraction of the instances of the majority class are randomly deleted to form the new training set. Under-sampling is done on each training set. For more information about the under-sampling effect, please refer to the reference.9

**Supplementary Table 1.** Description and rationale for generating the basic features related to circadian rhythm obtained from the automatically measured passive digital log data of patients with mood disorders

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Feature name** | **Description** | **Rationale** |
| Light exposure | Morning light exposure† | Cumulative amount of light exposure in the morning is defined as the period ‘from sunrise to noon.’ | Light is the most important trigger that synchronizes the CR of all creatures on earth. Therefore, being exposed to light in the morning is highly recommended. |
| Step | Step count during the morning‡ | Cumulative counts of steps during the morning. | Walking enough in the morning or daytime and not much around bedtime is recommended to keep a sound CR. |
| Step count during the afternoon‡ | Cumulative counts of steps during the afternoon. |
| Step count during the evening‡ | Cumulative counts of steps during the evening. |
| Step count during the bedtime‡ | Cumulative counts of steps during bedtime. |
| Sleep | Sleep length‡ | Total length of sleep. | Sufficient and regular sleep are important factors in achieving a sound CR and relieving fatigue and stress. |
| Sleep efficiency‡ | The efficiency score of sleep is between 0 to 100. It is measured as ‘(sleep length - restless sleep length) /sleep length.’ |
| Deviation of sleep onset† | Deviation of sleep onset times is measured to check whether the sleep onset time is regular. The deviation is computed by measuring the time difference (absolute value) between sleep onset and bedtime (8 h before sunrise time). |
| Deviation of sleep offset† | Deviation of sleep offset (wake-up) times is measured to check whether the sleep offset time is regular. The deviation is computed by measuring the time difference between sleep offset and sunrise. We considered the seasonal sunrise time. |
| Circadian rhythm | CR amplitude† | The daily CR of the heart rate estimated from its fitted cosine curve has its amplitude. Large amplitudes imply a clear rhythm curve with a low heart rate during sleep and a high heart rate during physical activity. | The CR contains important rhythmic information. The heart rate decreases while sleeping and increases during physical activity. Therefore, it tends to have an S-shaped cosine curve on the CR graph as you sleep at night and perform activities during the day. From cosinor analysis for the daily CR, four parameters are produced: amplitude, acrophase, mesor, and R-squared.  |
| CR acrophase† | The daily CR rhythm of the cosine curve peaks at time t, where t is the acrophase. Acrophase is the time wherein the degree of misalignment of rhythm is indicated. |
| CR mesor†  | Mesor is the average heart rate per day. The more active the user, the higher the mesor. |
| CR goodness of fit† | The goodness of fit (R-squared) measures the fit of the user’s daily CR in the cosine curve. The better the fit, the better the cosine model fits the user’s data.  |
| Resting heartrate‡ | The average heart rate within timeslots where a user performs no activity (resting time). | Measuring the heart rate at rest is appropriate. It tends to rise when stressed, uneasy, or unhealthy, which might be correlated with the mood state. |

† indicates our estimated feature, and ‡ indicates the estimated feature that Fitbit inherently provides. CR: circadian rhythm

**Supplementary Table 2.** Comparison of characteristics of subjects included and excluded in the analysis

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable**  | **Subjects included in the analysis (n=270)** | **Subjects not included in the analysis (n=225)** | **p-value** |
| *Demographic variables* |  |  |  |
| Age at baseline, mean (SD), years | 23.3 (3.63) | 22.8 (3.10) | 0.106 |
| Sex |  |  | 0.317 |
|  Male | 123 (45.6) | 92 (40.9) |  |
|  Female | 147 (54.4) | 133 (59.1) |  |
| Education, years | 14.1(2.05) | 13.6(1.68) | 0.004 |
| *Clinical variables* |  |  |  |
|  Diagnosis |  |  | 0.669 |
|  MDD | 95 (35.2) | 71 (31.6) |  |
|  BD I | 78 (28.9) | 71 (31.6) |  |
|  BD II | 97 (35.9) | 83 (36.9) |  |
|  Age at onset, mean (SD), years | 17.8 (4.72) | 17.1 (4.14) | 0.116 |
| MDE/year | 1.15 | 1.66 |  |
| ME/year | 0.18 | 0.06 |  |
| HME/year | 0.24 | 0.25 |  |
| Follow-up duration, day | 580.1 (380.6) | 361.4 (94.3) | <0.001 |

Abbreviations: MDD, major depressive disorder; BD I, bipolar I disorder; BD II, bipolar II disorder; MDE, major depressive episode; ME, manic episode; HME, hypomanic episode; SD, standard deviation

**Supplementary Table 3. The list of all the 140 features used in the model construction**

|  |  |  |
| --- | --- | --- |
| Sleep lengthSleep efficiencyDeviation of sleep onsetDeviation of sleep offsetCR amplitudeCR mesorCR goodness of fitCR amplitude\_sd\_3dCR amplitude\_m\_3dCR amplitude\_gr\_3dCR amplitude\_sd\_6dCR amplitude\_m\_6dCR amplitude\_gr\_6dCR amplitude\_sd\_12dCR amplitude\_m\_12dCR amplitude\_gr\_12dCR mesor\_sd\_3dCR mesor\_m\_3dCR mesor\_gr\_3dCR mesor\_sd\_6dCR mesor\_m\_6dCR mesor\_gr\_6dCR mesor\_sd\_12dCR mesor\_m\_12dCR mesor\_gr\_12dCR goodness of fit\_sd\_3dCR goodness of fit\_m\_3dCR goodness of fit\_gr\_3dCR goodness of fit\_sd\_6dCR goodness of fit\_m\_6dCR goodness of fit\_gr\_6dCR goodness of fit\_sd\_12dCR goodness of fit\_m\_12dCR goodness of fit\_gr\_12dSleep length\_sd\_3dSleep length\_m\_3dSleep length\_gr\_3dSleep length\_sd\_6dSleep length\_m\_6dSleep length\_gr\_6dSleep length\_sd\_12dSleep length\_m\_12dSleep length\_gr\_12dSleep efficiency\_sd\_3dSleep efficiency\_m\_3dSleep efficiency\_gr\_3dSleep efficiency\_sd\_6d | Sleep efficiency\_m\_6dSleep efficiency\_gr\_6dSleep efficiency\_sd\_12dSleep efficiency\_m\_12dSleep efficiency\_gr\_12dDeviation of sleep onset\_sd\_3dDeviation of sleep onset\_m\_3dDeviation of sleep onset\_gr\_3dDeviation of sleep onset\_sd\_6dDeviation of sleep onset\_m\_6dDeviation of sleep onset\_gr\_6dDeviation of sleep onset\_sd\_12dDeviation of sleep onset\_m\_12dDeviation of sleep onset\_gr\_12dDeviation of sleep offset\_sd\_3dDeviation of sleep offset\_m\_3dDeviation of sleep offset\_gr\_3dDeviation of sleep offset\_sd\_6dDeviation of sleep offset\_m\_6dDeviation of sleep offset\_gr\_6dDeviation of sleep offset\_sd\_12dDeviation of sleep offset\_m\_12dDeviation of sleep offset\_gr\_12dResting heartrateResting heartrate\_sd\_3dResting heartrate\_m\_3dResting heartrate\_gr\_3dResting heartrate\_sd\_6dResting heartrate\_m\_6dResting heartrate\_gr\_6dResting heartrate\_sd\_12dResting heartrate\_m\_12dResting heartrate\_gr\_12dStep count during morningStep count during afternoonStep count during eveningStep count during bedtimeMorning light exposureStep count during morning\_sd\_3dStep count during morning\_m\_3dStep count during morning\_gr\_3dStep count during morning\_sd\_6dStep count during morning\_m\_6dStep count during morning\_gr\_6dStep count during morning\_sd\_12dStep count during morning\_m\_12dStep count during morning\_gr\_12d | Step count during afternoon\_sd\_3dStep count during afternoon\_m\_3dStep count during afternoon\_gr\_3dStep count during afternoon\_sd\_6dStep count during afternoon\_m\_6dStep count during afternoon\_gr\_6dStep count during afternoon\_sd\_12dStep count during afternoon\_m\_12dStep count during afternoon\_gr\_12dStep count during evening\_sd\_3dStep count during evening\_m\_3dStep count during evening\_gr\_3dStep count during evening\_sd\_6dStep count during evening\_m\_6dStep count during evening\_gr\_6dStep count during evening\_sd\_12dStep count during evening\_m\_12dStep count during evening\_gr\_12dStep count during bedtime\_sd\_3dStep count during bedtime\_m\_3dStep count during bedtime\_gr\_3dStep count during bedtime\_sd\_6dStep count during bedtime\_m\_6dStep count during bedtime\_gr\_6dStep count during bedtime\_sd\_12dStep count during bedtime\_m\_12dStep count during bedtime\_gr\_12dMorning light exposure\_sd\_3dMorning light exposure\_m\_3dMorning light exposure\_gr\_3dMorning light exposure\_sd\_6dMorning light exposure\_m\_6dMorning light exposure\_gr\_6dMorning light exposure\_sd\_12dMorning light exposure\_m\_12dMorning light exposure\_gr\_12dCR acrophaseCR acrophase\_sd\_3dCR acrophase\_m\_3dCR acrophase\_gr\_3dCR acrophase\_sd\_6dCR acrophase\_m\_6dCR acrophase\_gr\_6dCR acrophase\_sd\_12dCR acrophase\_m\_12dCR acrophase\_gr\_12d |

Abbreviations: CR, circadian rhythms; m, mean; SD, standard deviation; gr, gradient; d, days

**Supplementary Figure 1.** Model performance evaluation for each hospital.

Test results of the model constructed for predicting whether patients have any mood episodes. The group ‘All’ indicates the results of all patients without distinguishing site groups.

**References**

1. Moon JH, Cho CH, Son GH, et al. Advanced circadian phase in mania and delayed circadian phase in mixed mania and depression returned to normal after treatment of bipolar disorder. *EbioMedicine.* 2016;11:285-295. doi:10.1016/j.ebiom.2016.08.019.

2. Cho CH, Moon JH, Yoon HK, et al. Molecular circadian rhythm shift due to bright light exposure before bedtime is related to subthreshold bipolarity. *Sci Rep.* 2016;6;31846. doi:10.1038/srep31846.

3. Bowman C, Huang Y, Walch OJ, et al. A method for characterizing daily physiology from widely used wearables. *Cell Rep Method.* 2021;1(4):100058 doi:10.1016/j.crmeth.2021.100058.

4. Jeong S, Seo JY, Jeon S, et al. Circadian rhythm of heart rate assessed by wearable devices tends to correlate with the circadian rhythm of salivary cortisol concentration in healthy young adults. *Chronobiol Med.* 2020;2:109-114. doi:10.33069/cim.2020.0022.

5. Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. *J Med Internet Res.* 2019;21(4):e11029. doi:10.2196/11029.

6. Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW, eds. *Contributions to the Theory of Games (AM-28), Volume II*. New Jersey, NY: Princeton University Press; 1953:307-318.

7. Linear interpolation, https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html

8. Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein and Russ B. Altman, Missing value estimation methods for DNA microarrays, BIOINFORMATICS Vol. 17 no. 6, 2001 Pages 520-525.

9. Chris Drummond and Robert C. Holte, “C4.5, Class Imbalance, and Cost Sensitivity: Why Under-sampling beats Over-sampling,” Proceedings of the International Conference on Machine Learning (ICML 2003) Workshop on Learning from Imbalanced Data Sets II.