

A Study on ML-Based Sleep Score Model Using Lifelog Data

Jiyong Kim¹ and Minseo Park^{2,*} ¹ Department of Mathematics, Kwangwoon University, Seoul 01897, Republic of Korea² Department of Data Science, Seoul Women's University, Seoul 01797, Republic of Korea

* Correspondence: mpark@swu.ac.kr

Abstract: The rate of people suffering from sleep disorders has been continuously increasing in recent years, such that interest in healthy sleep is also naturally increasing. Although there are many health-care industries and services related to sleep, specific and objective evaluation of sleep habits is still lacking. Most of the sleep scores presented in wearable-based sleep health services are calculated based only on the sleep stage ratio, which is not sufficient for studies considering the sleep dimension. In addition, most score generation techniques use weighted expert evaluation models, which are often selected based on experience instead of objective weights. Therefore, this study proposes an objective daily sleep habit score calculation method that considers various sleep factors based on user sleep data and gait data collected from wearable devices. A credit rating model built as a logistic regression model is adapted to generate sleep habit scores for good and bad sleep. Ensemble machine learning is designed to generate sleep habit scores for the intermediate sleep remainder. The sleep habit score and evaluation model of this study are expected to be in demand not only in health-care and health-service applications but also in the financial and insurance sectors.

Keywords: sleep; health; sleep score; sleep dimension; lifelog data; stacking ensemble regression; credit rating



Citation: Kim, J.; Park, M. A Study on ML-Based Sleep Score Model Using Lifelog Data. *Appl. Sci.* **2023**, *13*, 1043. <https://doi.org/10.3390/app13021043>

Academic Editor: José Salvador Sánchez Garreta

Received: 14 December 2022

Revised: 2 January 2023

Accepted: 9 January 2023

Published: 12 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Good sleep is important for health. It is very important for improving the quality of life by enhancing physical recovery, strengthening memory and immunity, and protecting mental health [1–3]. However, the rate of sleep disorders is steadily increasing worldwide. According to previous research, about 10–30% of adults suffer from chronic insomnia [4]. Sleep disorders not only lower the quality of life of individuals but also increase social costs. In the United States, insufficient sleep is associated with economic losses estimated at more than \$411 billion [5].

Presently, various equipment such as wearable devices and smart scales are being used for sleep health [6–8]. In the past, only hospitals could test sleep through expensive polysomnography. Most previous studies for sleep quality scores use the Pittsburgh questionnaire [9,10], but have limitations due to reliance on interviewees' subjective responses. Research using objective data is lacking. However, using collected lifelog data, it is possible to track health signals daily as well as identify health trends by week and month.

To achieve good quality sleep scores, it is necessary to calculate sleep scores using multiple sleep dimensions such as sleep efficiency, regularity, duration, and timing [11,12]. This paper focuses on scoring sleep habits' healthiness by considering multiple dimensions of sleep and proposes a sleep habit score calculation methodology that considers objective data and various dimensions of sleep with a credit evaluation-based model and machine learning using data collected with a Samsung Galaxy 5.

The results of this study include an objective indicator of sleep health and are expected to be utilized in financial fields as well as digital health care. First, in the health-care industry, our scoring methodology is expected to be used as a comprehensive indicator of sleep health and to help improve sleep by checking and improving one's sleep habit score

every day. In addition, it is expected to help financial and insurance companies develop many insurance products linked to health indicators. In fact, a study by Moore (2002) [13] suggested that sleep health is related to financial information such as income.

This paper is structured as follows. Section 2 describes the proposed methodology. Section 2.1 describes data description and preprocessing methods. Section 2.2 describes the data, features, and target for generating primary sleep habits for good/bad sleep and explains the logistic regression model and methodology for generating scores. Section 2.3 describes the dataset and modeling methods for generating secondary sleep habit scores for intermediate sleep states as well as the methodology for generating these scores. Section 3 presents the overall sleep habit score results, which combine sleep habit scores for good/bad sleep and for intermediate sleep states. Section 4 summarizes this study, including interpretation of the results, considerations, limitations, and significance of the study.

2. Materials and Methods

Figure 1 shows the flow for proposed method.

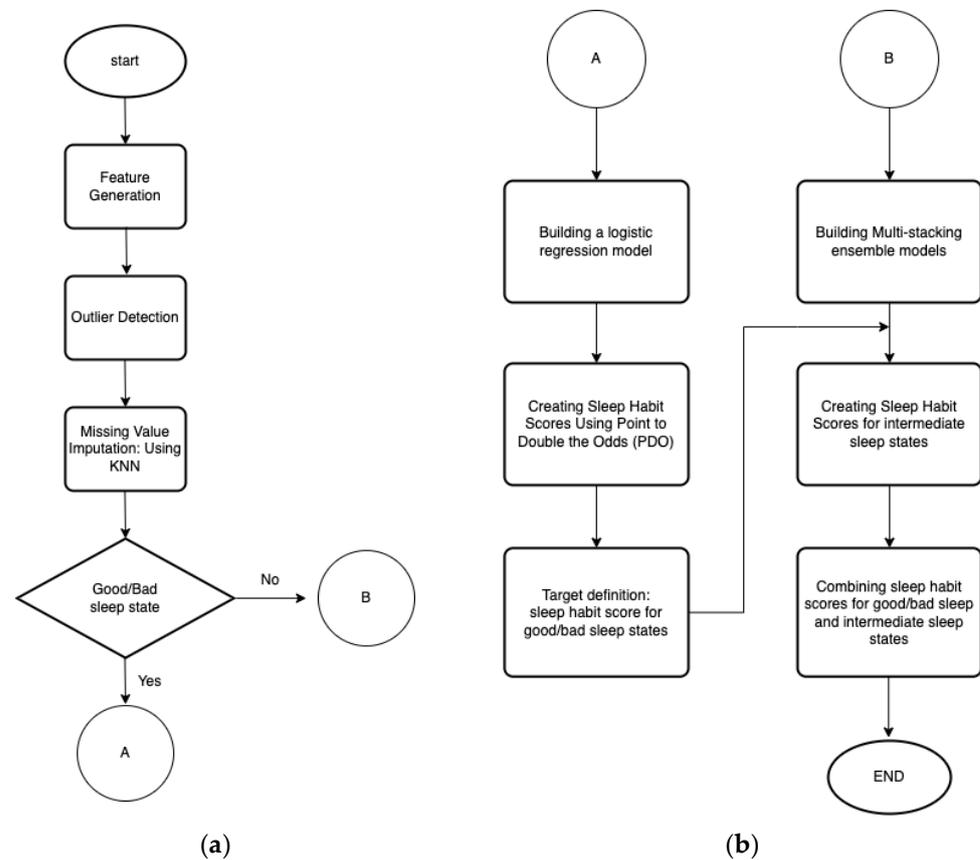


Figure 1. (a) Summary flowchart; (b) specific sleep score creation process.

Figure 1a shows the sleep habit score generation process for good/bad sleep states. The chart is divided into A and B according to the sleep state. Figure 1b illustrates the process of (A) and (B) which shows in Figure 1a.

As can be seen in Table 4 of Section 2.2.2, a simple but widely used logistic linear model for credit scoring classifies good/bad sleep habits and obtains intuitive weights. Based on a model that classifies good/bad sleep habits, it can be relatively ambiguous for intermediate sleep habit data with many factors mixed together, so a stacking machine learning model, a more complex model such as a nonlinear model, is used to classify intermediate sleep habits.

Each step in the summary flowchart of the proposed method is described in detail in subsequent sections. In brief, we perform feature generation on the collected raw data and then perform outlier removal and missing value imputation. The refined data are divided into two major categories according to the sleep state. For data on good/bad sleep states (A), a logistic regression model is used to generate a primary sleep habit score. The data on the intermediate sleep state (B) generate a sleep habit score using stacking models, where the target is defined as the sleep habit score obtained from A.

2.1. Data Preparation and Preprocessing

The data used in this study is set out in Table 1.

Table 1. Description of raw data collected from wearable devices (Samsung Galaxy watch).

Category	Value	Description
Quantity of sleep data collected by day	67,180 rows	Sleep data set collected by day with Samsung Galaxy 4 or 5
Quantity of sleep data collected by minute	2,494,862 rows	Sleep data set collected by day with Samsung Galaxy 4 or 5
Quantity of step (gait) data collected by day	78,643 rows	Step data set collected by day with Samsung Galaxy Watch 4 or 5
Quantity of step data collected by minute	18,710,423 rows	Step data set collected by day with Samsung Galaxy 4 or 5
Quantity of user information data	918 rows	User information such as height and age
Number of users	714	
Period of data collection	26 November 2020 to 1 January 2022	

The data were collected from 714 people from 26 November 2020 to 1 January 2022 by a Samsung Galaxy. Specifically, daily/minute sleep data, daily/by-minute step data, and user information (age, gender) were included. The collected data were preprocessed as daily data aggregation, sleep-related feature generation, outlier processing, and missing values. In daily data aggregation, features are generated by daily aggregation of sleep data collected per minute. Based on the study findings that sleep phase information for the initial 90 min of sleep indicates the quality of sleep, we created the sleep phase features for the first 90 min [14,15]. We also generated total daily sleep stage ratio features (REM stage, light stage, deep stage, awake stage) [16,17], sleep efficiency feature [18], SRI (sleep regularity index) [19,20], and so on. The SRI feature is calculated through the SRI calculation Equation (3) assuming Equations (1) and (2) for M daily epochs and N days [20]:

$$S_{i,j} = S_{i+1,j} \rightarrow \delta(S_{i,j}, S_{i+1,j}) = 1, \tag{1}$$

$$S_{i,j} \neq S_{i+1,j} \rightarrow \delta(S_{i,j}, S_{i+1,j}) = 0, \tag{2}$$

$$-100 + \frac{200}{M(N-1)} \sum_{j=1}^M \sum_{i=1}^{N-1} \delta(S_{i,j}, S_{i+1,j}). \tag{3}$$

where N and M are the number of days and the number of epochs per day, respectively. The function δ returns value 1 if the sleep occurrence (sleep-wake state) is the same at 24 h intervals and returns 0 otherwise. For example, if sleep occurred at 22:00 and ended at 06:00 on Friday and occurred at 22:30 and ended at 08:00 on Saturday, the function δ from 22:30 to 06:00 is 1 and the rest of the time zone is 0. Daily sleep data are used to generate total sleep time [21] and sleep midpoint features [22]. In addition, to generate features for the step information just before sleep, the step data collected per minute is preprocessed and used together with the sleep data to generate features. Some of the feature names and descriptions are summarized in Table 2, and the rest are shown in Appendix A Table A1 for readability.

Table 2. Feature names and descriptions as generated from raw data collected daily/per minute.

Feature	Meaning	Feature	Meaning
USER_CODE	User identification code	NAP_FLAG	Daily nap occurrence status
DATE	Data collection date	NAP_HOUR	Total sleep time from 12 noon to 3 p.m. (less than 3 h)
SLEEP_EFFICIENCY	Ratio of sleep time excluding awake time to total sleep time	WEEKLY_MEAN_SLEEP_MIDPOINT	The average time of the midpoint of sleep during the weekdays
DSR	Percentage of deep sleep phases per day	WEEKLY_MEAN_SLEEP_START_TIME	The average time of the sleep onset during the weekdays
RSR	Percentage of rem sleep phases per day	WEEKEND_MEAN_SLEEP_START_TIME	The average time of the sleep onset during the weekends
LSR	Percentage of light sleep phases per day	DIFF_WEEK_HOLI	Difference between average weekday onset sleep and average sleep onset on weekends
ASR	Percentage of awake sleep phases per day	WEEKLY_MEAN_TST	The average time of the total sleep time per day during the weekdays
TST	Total sleep time per day	DIFF_SLEEP_START_WEEKLY	The difference between the average weekly sleep onset time and daily average sleep onset time
SLEEP_START_H	Sleep onset time (hours) per day	DIFF_SLEEP_END_WEEKLY	The difference between the average weekly sleep offset time and daily average sleep offset time
SLEEP_END_H	Daily sleep offset time (hours) per day	WEEKLY_MIDPOINT_VAR	The variation of the midpoint of sleep during the weekdays
AWAKE_T	Total awake time per day	WEEKLY_TST_VAR	The variation of total sleep time during the weekdays
DEEP_T	Total time of deep sleep phases per day	GOOD_SLEEP_FLAG	Sleep quality status based on various sleep dimensions
REM_T	Total time of rem sleep phases per day	GENDER	User's gender
LIGHT_T	Total time of light sleep phases per day	AGE	User's age
SLEEP_EFFICIENCY_CAT	85% cutoff criterion flag for Sleep Efficiency Index	AGE_CATEGORY	User's age category
BED_TIME_VAR	sleep onset variability	FIRST_AWAKE_MIN	Total time of awake sleep phases in earlier (90 min) sleep cycles

Outlier processing proceeds as follows:

1. Data with a sleep stage value of 0 among the generated sleep stage features.
2. Less than 3 h of total sleep per day, since it does not record stages if the sleep is less than 3 h.
3. SRI index with negative values [23].

Missing-value processing based on sleep habit score will be described in detail after Section 2.2, but it is briefly described in this section as it is included in the overall preprocessing. Missing-value processing is organized into three steps as follows:

1. Set the sleep habit score derived by the logistic regression model as the target and set the related sleep variable as the explanatory variable. (This is detailed in Section 2.2.1.)
2. Process missing values based on the KNN (K-nearest neighbors) machine learning algorithm [24,25]. To derive the optimal k (number of neighbors), the support vector regression [26] and random forest models have been used for k evaluation [27]; k has been selected with the average number of neighbors yielding the best performance among the evaluations.
3. Fill in missing values using the KNN method with the derived optimal number of neighbors k. Specifically, the data set was divided into training data and test data at a ratio of 8:2, and the range of k was set from 2 to 15, and performance was measured with each k value. Support vector machine and random forest calculated the final performance with a weighted sum of the calculation results, calculated by applying a weighted sum of 0.5 each, that is, the result calculated by each classifier was multiplied by 0.5 to derive the result in an ensemble method. As a result of the experiment, it was confirmed that the performance was the best when k was 3, and imputation was performed with that value. As described above, a total of 67 variables, such as user identification ID value, date, and sleep characteristics, and 16,053 rows of data are used as analysis data through daily data aggregation, sleep feature generation, outlier processing, and preprocessing of missing values.

Sleep health is defined by information on sleep regularity, sleep duration, sleep timing, and sleep efficiency dimensions [11]. Sleep regularity, sleep duration, sleep efficiency,

and sleep timing are important indicators of sleep habits. Several recent studies have shown that sleep regularity is beneficial to physical and mental health and shown that irregular sleep increases the risk of developing cardiovascular disease [28,29]. As for the sleep duration indicator, many studies have found that sleep duration that is both too short and too long can negatively impact health and quality of life [28,29]. Additionally, both late sleep duration and large sleep variability are associated with poor sleep health, and regular sleep patterns have beneficial effects on health [28,29]. These are defined as follows through the sleep factors and cutoff values used by previous studies [28,29].

1. Sleep regularity: standard deviation of weekday sleep midpoint (variability), with a difference of less than 1 h defined as a good sleep state [28,29].
2. Sleep duration: the total daily sleep time, calculated as the difference between the daily sleep end time and sleep start time, where 7 to 9 h is defined as a good sleep state [28,29].
3. Sleep timing: the midpoint of sleep, calculated as the midpoint between the onset and the end of sleep, where between 2 and 4 a.m. is defined as a good sleep state [28,29].
4. Sleep efficiency: the ratio of total sleep time to total sleep time excluding waking time, where 85% or more is defined as a good sleep state [28,29].

Based on the cutoff values set above, data are defined as a good sleep state when all four conditions are satisfied, and as bad sleep when three or more of the four conditions are not satisfied. Bad sleep consists of five combinations: (1) bad-sleep regularity, duration, and efficiency, (2) bad-sleep regularity, duration, and timing, (3) bad-sleep regularity, efficiency, and timing, (4) bad-sleep duration, efficiency, and timing, and (5) bad-sleep regularity, duration, timing, and efficiency. The remaining combinations of conditions are taken to define the intermediate sleep state. The sleep habit score is derived using data for 326 good sleep states, 5168 bad sleep states, and 10,559 intermediate sleep states defined in this way.

2.2. Primary Habit Score: Good/Bad Sleep State

The number of classes of target used in this study is three, good/intermediate/bad. We first set the data consisting of good sleep and bad sleep as the analysis data set, excluding the data classified as intermediate sleep states. Based on the data with two target classes, the primary sleep habit score is derived by applying a traditional credit evaluation model and credit score generation method.

2.2.1. Setting Description Variables (Features) and Result Variables (Target)

The explanatory variables of the data set are divided into continuous variables and categorical variables as follows:

- Continuous variables: total sleep variability, SRI (Sleep Regularity Index) (2 days, 3 days, 4 days, 5 days, 6 days, 7 days), number and time of naps, sleep midpoint variability, day-of-week information, daily sleep start and end information, information on sleep stages within the first 90 min of sleep, information on steps 2 h before the first start of sleep.
- Categorical variables: 10~12 h sleep FLAG variable, sleep onset variability 1-h FLAG variable, total sleep time variability within 1 h FLAG variable.

The categorization for continuous variables for scoring consists of two steps as follows:

- The first step, fine classing (Leung, 2008; Vejkanchana, 2019) [30,31], is carried out to improve consistency and explanatory power. Through this, a representative variable is selected in consideration of the correlation within the explanatory variable and the information value (Vejkanchana, 2019) [31] and a section for the variable is derived.
- The second step is coarse classing (Leung, 2008; Vejkanchana, 2019) [30,31]; based on the categorization in the first step, a new category is derived by checking the data state. Specifically, for a linear relationship with the occurrence of good sleep, adjacent categories with similar weight-of-evidence (WoE) values (Finlay, 2010; Zdravevski, 2011) [32,33] are integrated so that the WoE value increases or decreases monotonically

(Vanneschi, 2018) [34]. In this way, the amount of data on the number of occurrences and nonoccurrence of good sleep for each category is adjusted and categories are integrated based on the WoE value.

Features calculated based on WoE values for the target in this study are summarized in Table 3.

Table 3. Interval range information for each feature based on WoE values.

Feature	Calculation of Bin Interval Excluding Missing Values	Feature	Calculation of Bin Interval Excluding Missing Values
SRI_2 (Sleep regular index observed over 2 days)	[−inf, 52], [52, 72], [72, 86], [86, inf]	SRI_3 (Sleep regular index observed over 3 days)	[−inf, 56], [56, 70], [70, 82], [82, 90], [90, inf]
SRI_4 (Sleep regular index observed over 4 days)	[−inf, 52], [52, 62], [62, 70], [70, 82], [82, inf]	SRI_5 (Sleep regular index observed over 5 days)	[−inf, 52], [52, 58], [58, 82], [82, 88], [88, inf]
SRI_6 (Sleep regular index observed over 6 days)	[−inf, 56], [56, 80], [80, 86], [86, inf]	SRI_7 (Sleep regular index observed over 7 days)	[−inf, 56], [56, 78], [78, 84], [84, inf]
Daily Sleep offset time information (hour)	[−inf, 7.5], [7.5, inf]	Average weekend sleep onset information	[−inf, 2], [2, 12.5], [12.5, 23.5], [23.5, 24.5], [24.5, inf]
Daily Sleep onset time information (hour)	[−inf, 24.5], [24.5, inf]	Sleep midpoint variability	[−inf, 3], [3, 4], [4, 4.5], [4.5, inf]
Average weekly sleep onset information	[−inf, 2], [2, 6.5], [6.5, 11.5], [11.5, 16.5], [16.5, 24], [24, inf]	Daily total sleep time variance (HOUR)	[−inf, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 2.2], [2.2, 2.6], [2.6, 3.6], [3.6, inf]
REM sleep rate (%) in Initial 90 min	[−inf, 0.01], [0.01, 0.08], [0.08, 0.18], [0.18, inf]	LIGHT sleep rate (%) in Initial 90 min	[−inf, 0.54], [0.54, 0.62], [0.62, 0.84], [0.84, 0.9], [0.9, 0.96], [0.96, inf]
DEEP sleep rate (%) in Initial 90 min	[−inf, 0.01], [0.01, 0.05], [0.05, 0.09], [0.09, 0.15], [0.15, 0.23], [0.23, 0.32], [0.32, inf]	AWAKE sleep rate (%) in Initial 90 min	[−inf, 0.01], [0.01, 0.04], [0.04, 0.08], [0.08, 0.09], [0.09, 0.14], [0.14, 0.17], [0.17, 0.2], [0.2, inf]
Total REM sleep time (MINUTE) in initial 90 min	[−inf, 1], [1, 6], [6, 11], [11, inf]	Total LIGHT sleep time (MINUTE) in initial 90 min	[−inf, 29], [29, 34], [34, 37], [37, 39], [39, 50], [50, inf]
Total DEEP sleep time (MINUTE) in initial 90 min	[−inf, 1], [1, 7], [7, 12], [12, 16], [16, 22], [22, inf]	Total AWAKE sleep time (MINUTE) in initial 90 min	[−inf, 1], [1, 2], [2, 6], [6, 8], [8, 12], [12, 16], [16, inf]
Weekly total sleep time variance	[−inf, 0.8], [0.8, 1.2], [1.2, 2.8], [2.8, 3.7], [3.7, inf]	Total steps taken 2 h before sleep	[−inf, 10], [10, 720], [720, inf]
Total sleep stage time in initial 90 min	[−inf, 40], [40, 48], [48, 50], [50, 52], [52, 55], [55, 58], [58, 60], [60, inf]		

2.2.2. Defining Good Sleep Habit Labels Using a Logistic Regression Model for the Primary Sleep Habit Score

This study used a logistic regression model [35] to generate the primary sleep habit score. The reasons for this are: (1) ease of interpretation of regression coefficients; (2) since the model can estimate the probability of belonging to a class, it is often used for risk and credibility analysis required for probability calculation; (3) it can be used as a base model. For these reasons, this study uses the logistic regression model to score good and bad sleep habit status data. Good sleep habit level is expressed as the probability of developing a good sleep state that satisfies good sleep conditions. A model for the effect on the probability of good sleep occurrence has been created using logistic regression with various explanatory variables (Table 3). Logistic regression predicts the likelihood of an event using a linear combination of explanatory variables and is defined by Equation (4) [36]:

$$\ln(odds) = \ln \frac{p}{1 - p} = w_1 \times x_1 + w_2 \times x_2 + \dots + w_n \times x_n. \tag{4}$$

To evaluate the performance of the model, the training data and the verification data were first randomly extracted and divided, at a ratio of 7:3 and then three verification metrics commonly used in the credit evaluation model were used, specifically area under ROC (receiver-operating characteristic) curve [37], K–S (Kolmogorov–Smirnov) statistic [38], and Gini coefficient [39]. AUROC (area under ROC) means the area under the ROC curve: the closer the value is to 1, the higher the sensitivity and specificity, so the model can be called a good classification model. In the problem of generating scores, such as in the study of credit scoring, it is known that a model has good discriminating power when the value is 0.7 or more. The K–S statistic is an index that compares the difference in the cumulative distribution function between two groups (in our case, good sleep state and poor sleep state) and tests whether they come from the same distribution. Here, it refers to the maximum value of the difference between the cumulative good sleep incidence and the

cumulative bad sleep incidence. In general, if the K–S statistic is 0.5 or higher, the desired discriminatory power is judged to be secured. The Gini coefficient is used to determine the discriminatory power of the credit rating model using the cumulative defect distribution according to the credit score. Each metric calculated in this study is summarized in Table 4.

Table 4. Performance metric table for training data and validation data.

Category	AUROC	K–S	Gini Coefficient
Train data	0.9847	0.8912	0.9694
Validation data	0.9845	0.8882	0.969
Reference	>0.7	>0.5	>0.6

All index values are higher than the reference value. Therefore, the constructed model predicts the overall probability of occurrence of a good sleep state at an appropriate level.

2.2.3. Scoring for the Primary Sleep Habit Score

In this study, the sleep habit status is scored using points to double the odds (PDO) [40], a scoring methodology used in constructing a credit rating model [41]. If PDO is set to 20 or 50, it means that the odds double whenever the score increases by 20 or 50 points [42]. The higher the score, the lower the probability of satisfaction, focusing on the fact that good sleep habits are difficult to achieve. The standard value widely used in the credit evaluation model was applied. Specifically, the basic score was initialized to 100 and the PDO was set to 50, and the target odds for the initial score of 100 points were set at the level of 1:20. Specifically, for scoring, the score is calculated using (Equations (5)–(8)) [43]:

$$\text{Sleep Score} = \text{offset} - \text{factor} \times \ln(\text{odds}), \quad (5)$$

$$\text{factor} = \frac{50}{\ln(2)}, \quad (6)$$

$$\text{offset} = 100 - \frac{50}{\ln(2)} \times \ln(20). \quad (7)$$

$$\text{Sleep Score} = 100 - \frac{50}{\ln(2)} \times \ln(20) - \frac{50}{\ln(2)} \times \ln(\text{odds}). \quad (8)$$

2.2.4. Primary Sleep Habit Score Results

The distribution of the primary sleep score calculated in this study is shown in histogram form in Figure 2.

As can be seen from the graphs, the generated good sleep habit score (Figure 2a) is mostly distributed between 1400 and 1600 points, whereas the bad sleep habit score (Figure 2b) is distributed between 750 and 1000 points. The overall data distribution (Figure 2c) appears to follow a normal distribution, as expected. The basic statistical information of the primary sleep habit score generated in this study is summarized in Table 5.

Table 5. Statistical information of the sleep habit score obtained from the logistic regression model.

Count	Mean	Standard Deviation	25%	50%	75%	Max
5494	960.889516	285.114076	163	748	1162	1857

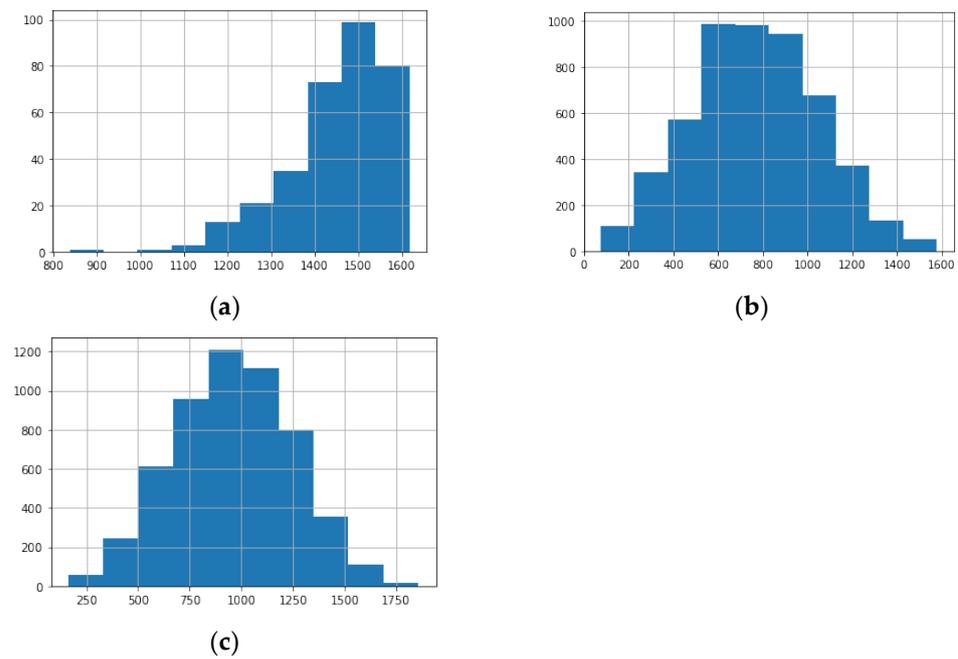


Figure 2. (a) Histogram of sleep habit scores for good sleep using a logistic regression model; (b) histogram of sleep habit scores generated for bad sleep data; (c) histogram of sleep habit scores for good and bad sleep data combined.

The scorecard for SRI (Sleep Regularity Index) is summarized in Table 6 as follows.

Table 6. Scorecard for each feature.

Feature	Interval Value	Score
SRI_2	~52	-2
	52~72	-1
	72~86	11
	86~	29
SRI_3	~56	-50
	56~70	-34
	70~82	6
	82~90	87
	90	180
SRI_4	~52	-45
	52~62	-19
	62~70	2
	70~82	15
	82~	29
SRI_5	~52	-17
	52~58	-14
	58~82	5
	82~88	34
	88~	67
SRI_6	~56	-19
	56~80	-13
	80~86	9
	86~	13
SRI_7	~56	-17
	56~78	-8
	78~84	0
	84~	6

The details for sleep duration, sleep timing, and sleep efficiency are described in the Appendix A.

2.3. Second Step: Intermediate Sleep Score

The sleep habit score was first generated using the good and bad sleep habit states. However, this excludes intermediate sleep habit states that can occur. Therefore, this study intends to generate a score for the intermediate sleep habit state using multi-stacking ensemble models that are effective in improving predictive performance. The machine learning and deep learning-based stacking ensemble learning model proposed in this study uses three data sets: training set and test set, plus a CV (cross-validation) set to prevent overfitting, which occurs mainly in the stacking method [44,45].

2.3.1. Data Preparation: Training and Test Data Set

The dataset, classified into good sleep and bad sleep data, is used as the training data, and the primary sleep habit score described in Section 2.2 is set as the training data's target. The second sleep habit score is derived by setting the data set classified as the intermediate sleep state as predictive (test) data, and the sleep habit score for the intermediate sleep state is predicted using machine learning and deep learning stacking models. Specifically, the stacking machine learning model trains with training data of 5494 data (good sleep: 326, bad sleep: 5168), and estimates sleep habit scores for 10,559 test data.

2.3.2. Modeling: Multi-Stacking Ensemble Models Based on Machine Learning and Deep Learning

Figure 3 shows in summary form the machine learning and deep learning-based stacking ensemble model construction and design used in this study.

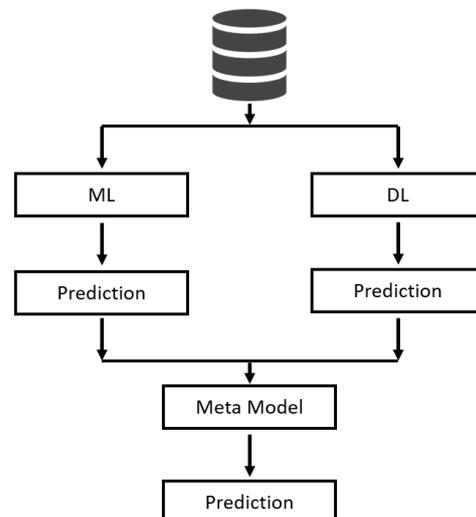


Figure 3. Summary graph for stacking technique: ML and DL indicate machine learning and deep learning, respectively.

Machine learning algorithms used for prediction are XGBoost [46], LightGBM [47], CatBoost [48], and the TabNet neural network model (a deep learning model) [49]. Meta-models used are linear regression, Bayesian Ridge Regressor [50], ElasticNet Regressor [51], and Ridge Regressor [52]. The stacking ensemble design method consists of three steps, presented in Figures 4–6. Figure 7 shows the operating process based on cross-validation within each individual model.

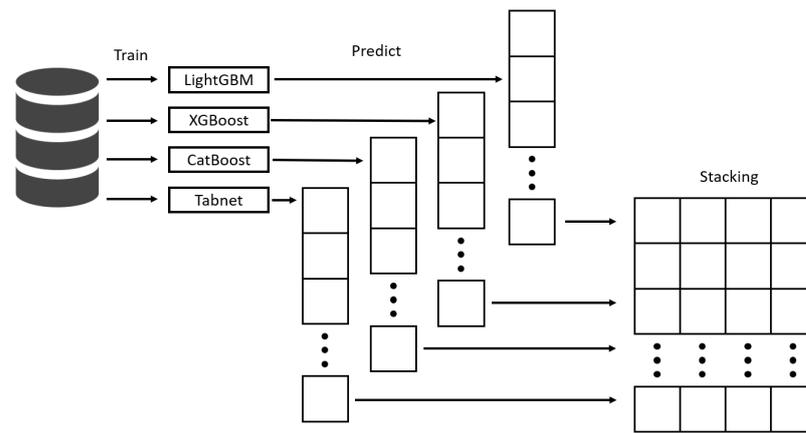


Figure 4. Summary graph for the first step of the stacking method: training dataset of metamodels is generated by machine learning models and a deep learning model in this step.

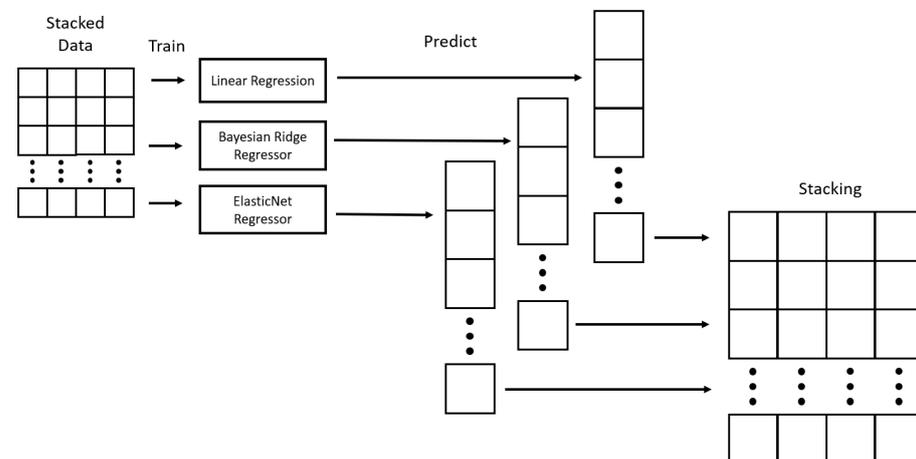


Figure 5. Summary graph for the second step of the stacking method: training dataset of final model is generated by metamodels in this step.

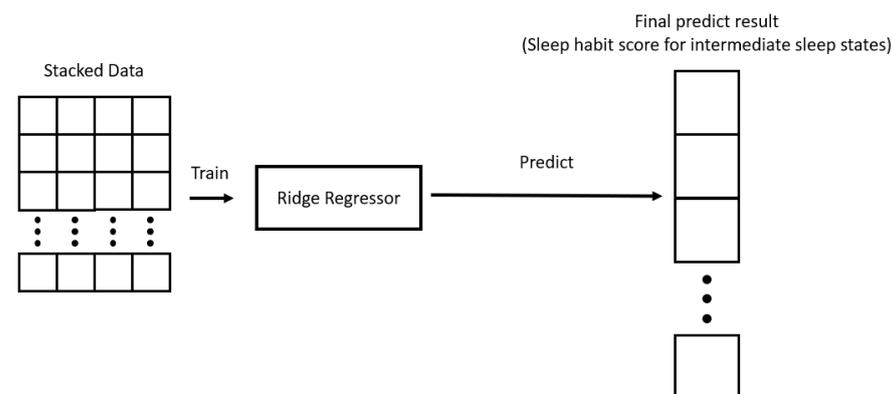


Figure 6. Summary graph for the final stage of the stacking method: in this step, a sleep habit score is calculated for the intermediate sleep state.

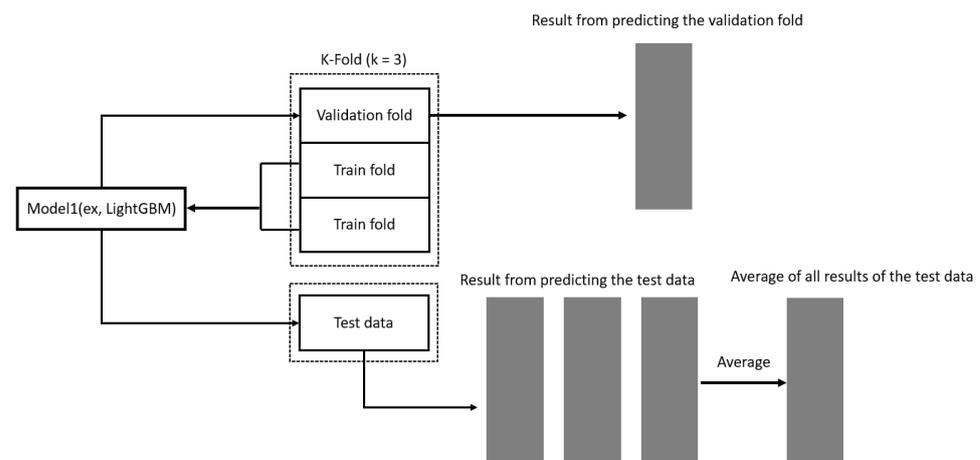


Figure 7. Summary graph of the process of CV stacking in each model.

In summary, in the first step, data is predicted using ML and DL models (LightGBM, XGBoost, CatBoost, TabNet) for good/bad sleep state data (feature) and sleep habit score (target). Stacking the output data by ML and DL models composes the data for the metadata (Figure 4). In the second step, three metamodels, linear regression, Bayesian Ridge Regressor, and Elastic-Net Regressor, are trained on the data constructed in the first step. Stacking the predicted data by metamodels composes the data for the final model. (Figure 5). In the last third step, the final prediction model, the Ridge Regressor algorithm, is used to predict the intermediate sleep habit score, and the performance error is measured by the mean squared error [53] (Figure 6). Specifically, in the first step, XGBoost, LightGBM, and CatBoost models derive optimal hyperparameters using the Optuna hyperparameter tuning framework [54]. The hyperparameters for each model are summarized in Table A3.

In Figure 7, to improve overfitting that may occur in the process shown in Figures 4–6, each model generates stacking data for metamodel training and testing through cross-validation. Based on the generated data, the metamodel then yields the training and prediction performance.

3. Results

Second Step: Intermediate Sleep Score

Figure 8 shows the distribution of the final sleep habit score calculated in this study, which is the sum of the first-generated (primary sleep habit) score and the intermediate sleep habit score. It is evenly distributed with an approximately normal distribution with a mean of 850. This is similar to the characteristics of a general scorecard in which scores are concentrated in the middle (average). It can be confirmed that the distribution of the calculated scores is close to a normal distribution, so that the data are not concentrated in a specific score range and are almost symmetrically distributed with no skew. This suggests that the score was well calculated without distortion. In addition, since it is an approximately normal distribution, it is possible to estimate the population by comparing various groups through inferential statistics, and it becomes possible to derive several kinds of statistical tests. Finally, it makes it easy to use and interpret scores.

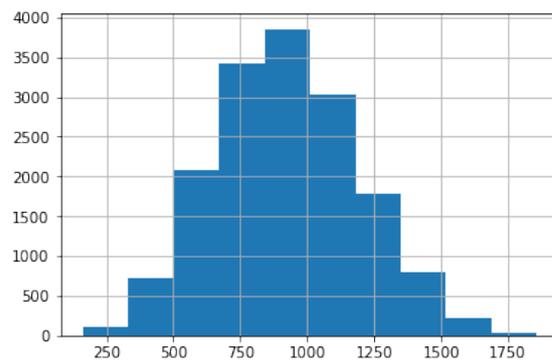


Figure 8. Sleep habit score distribution for all data.

Statistical values of sleep characteristics for each sleep state are as follows.

1. For sleep midpoint between 02:00 a.m. and 04:00 a.m. on weekdays, good sleep was 47.12%, bad sleep 12.25%, and intermediate sleep 21.39%.
2. For weekend sleep time between 10:00 p.m. and 12:00 a.m., good sleep was 43.25%, bad sleep 22.50%, and intermediate sleep 29.91%.
3. For steps within 2 h, good sleep averaged 813 steps, bad sleep 35, and moderate sleep 156.
4. For the SRI (Sleep Regularity Index) index (2 days), mean values were 87.21 for good sleep, 73.26 for bad sleep, and 80.58 for intermediate sleep.

Figure 9 shows the sleep state probabilities for each section for good and bad sleep states.

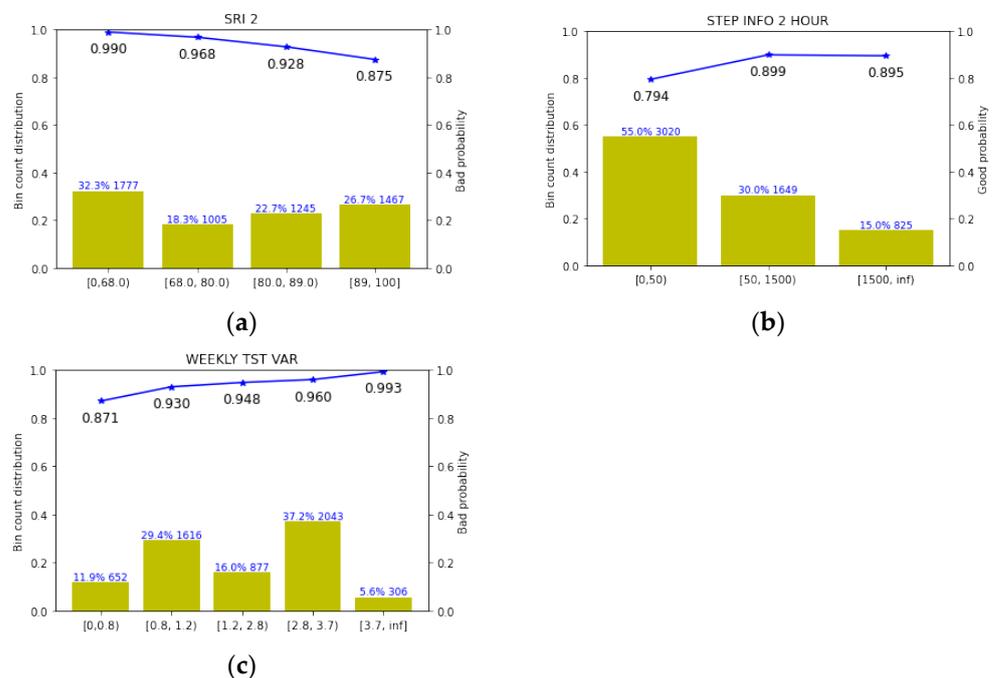


Figure 9. (a) Sleep state probabilities for SRI interval; (b) sleep state probabilities for STEP_INFO_2 interval; (c) sleep state probabilities for WEEKLY_TST_VAR interval.

Figure 9 shows the following:

- The higher the SRI value, the higher the probability of a good sleep state.
- The higher the gait (step) counts within the first 2 h before sleep, the higher the probability of a good sleep state.
- The greater the weekly total time variability, the higher the probability of a bad sleep state.

Overall, good sleep was mainly distributed in the range 1400–1600 points, intermediate sleep was distributed over 700–900 points, and bad sleep was mainly distributed over less than 700 points (Refer to Table 7). According to the method proposed in this study, the higher the sleep habit score, the more data classified as good sleep state, while the lower the score, the worse the sleep state. Therefore, it is expected that good sleep guides can be elaborated according to the proposed sleep habit score.

Table 7. Table of data distribution and ratio of good sleep states by score.

Sleep Score	Number of Data Points	Proportion of Data	Frequency of Good Sleep Status	Good Sleep Status Ratio
50 < ss ≤ 100	6	0.037%	0	0.00%
100 < ss ≤ 150	21	0.131%	0	0.00%
150 < ss ≤ 200	89	0.554%	0	0.00%
200 < ss ≤ 250	140	0.872%	0	0.00%
250 < ss ≤ 300	214	1.333%	0	0.00%
300 < ss ≤ 350	186	1.159%	0	0.00%
350 < ss ≤ 400	255	1.588%	0	0.00%
400 < ss ≤ 450	321	2.000%	0	0.00%
450 < ss ≤ 500	423	2.635%	0	0.00%
500 < ss ≤ 550	583	3.632%	0	0.00%
550 < ss ≤ 600	739	4.604%	0	0.00%
600 < ss ≤ 650	742	4.622%	0	0.00%
650 < ss ≤ 700	704	4.385%	0	0.00%
700 < ss ≤ 750	819	5.102%	0	0.00%
750 < ss ≤ 800	933	5.812%	0	0.00%
800 < ss ≤ 850	1099	6.846%	1	0.31%
850 < ss ≤ 900	1085	6.759%	0	0.31%
900 < ss ≤ 950	848	5.283%	0	0.31%
950 < ss ≤ 1000	901	5.613%	0	0.31%
1000 < ss ≤ 1050	872	5.432%	1	0.61%
1050 < ss ≤ 1100	938	5.843%	2	1.23%
1100 < ss ≤ 1150	788	4.909%	1	1.53%
1150 < ss ≤ 1200	676	4.211%	7	3.68%
1200 < ss ≤ 1250	544	3.389%	12	7.36%
1250 < ss ≤ 1300	432	2.691%	13	11.35%
1300 < ss ≤ 1350	448	2.791%	14	15.64%
1350 < ss ≤ 1400	477	2.971%	41	28.22%
1400 < ss ≤ 1450	319	1.987%	44	41.72%
1450 < ss ≤ 1500	167	1.040%	57	59.20%
1500 < ss ≤ 1550	160	0.997%	68	80.06%
1550 < ss ≤ 1600	117	0.729%	59	98.16%
1600 < ss ≤ 1650	7	0.044%	6	100.00%
Total	16,053	100.00%	326	2.03%

4. Discussion

This study presented a model for grading sleep habit level considering various sleep dimensions. First, the quality of sleep was defined as an index indicating the level of sleep habits, and data for good sleep, intermediate sleep, and bad sleep were classified according to the cutoffs of previous studies.

Based on the logistic regression model used in the credit rating model, a model for estimating the likelihood of occurrence was derived using lifelog factors that affect good and bad sleep. Specifically, the process of categorizing various sleep features generated from lifelog datasets, estimating probability of occurrence of each sleep state with the logistic regression model and evaluating the predictive power of the model were discussed. The primary sleep habit score was derived by grading and classifying sleep habit levels based on the PDO (points to double the odds) concept using the derived model. This study aimed to derive the sleep habit level for all sleep states by learning the primary sleep habit

score derived using a machine learning algorithm to generate the sleep habit index for intermediate sleep states.

Summarizing the characteristics of the sleep habit score derived from this study, the midpoint of sleep is between 2 a.m. and 4 a.m., the start time of sleep is between 10 p.m. and 12 a.m., and the walking activity in the evening increases the probability of receiving a high score. Also, the higher the Sleep Regularity Index (SRI), the higher the probability of good sleep.

Previous studies were reviewed to verify the validity of the methodology. It was confirmed that the results of this study were consistent with the results of previous studies. Halson (2022) [55] claimed that the average SRI value was 81.4 to 88.8, and Windred (2021) [56] found that the higher the SRI (94 points), the more regular the sleep state, and the lower the SRI (34 points), the more irregular the sleep. It is similar to the results of this study that the higher the SRI, the higher the probability of being in a good sleep state. Makarem (2020) [57] investigated the correlation between sleep variability and health, and confirmed that high sleep variability has a negative effect on health. Baron (2017) [58] found that higher sleep variability can negatively affect sleep quality, which is consistent with the results of this study. Buman (2014) [59] suggested that there was no relationship between evening exercise and sleep quality. Stutz (2019) [60] found that vigorous exercise one hour before bedtime could negatively affect sleep onset, total sleep duration, and SE, but found no evidence that evening exercise negatively affects sleep, in fact rather the opposite. Frimpong (2021) [61] argued that activity 2 to 4 h before bedtime does not affect sleep quality in healthy young and middle-aged adults. This is similar to the result of this study that walking activity for 2 h before sleep increases the probability of being a good sleep state. In addition, this study generated various features through gait and sleep data, and in the study of Kim (2022) [62], various step and sleep features were generated through lifelog data and body weight were predicted through these features. Liang (2019) [63] also generated various sleep features, and medical-grade sleep/wake classification was predicted with a tree-based model. In the study of Han (2018) [42], the PDO was set at 58.43994, which is similar to this study. A study on the optimized PDO setting will be conducted in the future. Studies using stacking machine learning algorithms to improve performance were presented (Jiang, 2020; Pavlyshenko, 2018) [64,65], and Yu (2022) [66] added CV (cross-validation) to the stacking technique to prevent overfitting, which is similar to the method proposed in this study. As a result of comparing and reviewing the results and methodology of previous studies with this study, most of the results were consistent. Therefore, it is recommended to measure sleep quality and generate an objective score using lifelog data and machine learning algorithms. Since this method is based on the data of the user's life pattern, it is expected that the more data that are accumulated over time, the more accurate the quality of sleep can be predicted and the more accurate the sleep habit score can be generated.

The limitations of this study are as follows. Since we created the sleep score by focusing on sleep habits and behaviors (sleep hygiene) rather than sleep quality itself, even if the definition of good sleep presented in previous studies is not met, expert review shows that good sleep quality can occur or vice versa.

In the future, this research will go beyond the rating of sleep habit level to evaluate overall lifestyle, including walking habits and weight habits. We also plan to conduct simulations using an optimization algorithm that goes beyond simple ratings to perform additional analysis of optimal combinations and factors to increase sleep scores. This study used a linear model logistic model, but future research will study a new technique that calculates weights with a nonlinear model and scores them. In addition, good/bad sleep habits were classified with a simple linear model, and intuitive weights were obtained. Based on this, intermediate sleep habit data in which various factors were mixed were classified using a more complex model—a stacking machine learning model. However, as the efficiency can be increased as the number of steps is reduced, a study on a model that can be solved end to end in a single step will be conducted in future work. Lastly, instead of

using all of the various indicators discovered in previous studies, we can consider and study regularization models such as LASSO that can identify features that are actually important and those that can be discarded. Through these studies, it is expected that our research will contribute to private medical insurance and comprehensive health management more substantially.

Author Contributions: Conceptualization, J.K. and M.P.; data curation, J.K.; formal analysis, J.K.; funding acquisition, M.P.; methodology, J.K. and M.P.; supervision, M.P.; validation, J.K.; visualization, J.K.; writing—original draft preparation, J.K. and M.P.; writing—review and editing, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a research grant from Seoul Women’s University (2021-0423).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Feature names and descriptions as generated from raw data collected daily/per minute.

Feature	Meaning	Feature	Meaning
BED_TIME_VAR_FLAG_1	Sleep onset fluctuation within 1 h (FLAG)	FIRST_LIGHT_MIN	Total time of light sleep phases in earlier (90 min) sleep cycles
BED_TIME_VAR_FLAG_2	Sleep onset fluctuation within 2 h (FLAG)	FIRST_DEEP_MIN	Total time of deep sleep phases in earlier (90 min) sleep cycles
BED_TIME_10_TO_12_FLAG	Sleep onset time between 10 p.m. and 12 a.m. status (FLAG)	FIRST_REM_MIN	Total time of rem sleep phases in earlier (90 min) sleep cycles
TST_VAR	Total sleep time variability per day	STAGE_SUM_MIN	Total time of all sleep phases in earlier (90 min) sleep cycles
SRI_2	Sleep regularity index observed on 2 days	FIRST_ASR	Percentage of awake sleep phases in earlier (90 min) sleep cycles
SRI_3	Sleep regularity index observed on 3 days	FIRST_LSR	Percentage of light sleep phases in earlier (90 min) sleep cycles
SRI_4	Sleep regularity index observed on 4 days	FIRST_DSR	Percentage of deep sleep phases in earlier (90 min) sleep cycles
SRI_5	Sleep regularity index observed on 5 days	FIRST_RSR	Percentage of rem sleep phases in earlier (90 min) sleep cycles
SRI_6	Sleep regularity index observed on 6 days	STEP_INFO_BEFORE_SLEEP_2	Number of steps taken 2 h before sleep
SRI_7	Sleep regularity index observed on 7 days	WEEKDAY	Week information expressed as an integer (0: Sunday)
SLEEP_START_MIN	Daily sleep onset time (minute info) per day	WEEKEND_FLAG	Weekend status (flag)
SLEEP_END_MIN	Daily sleep offset time (minute info) per day	HOLIDAY_FLAG	Holiday status (flag)
SLEEP_MIDPOINT	Midpoint between the onset and offset of sleep		

Table A2. Scorecard for each feature.

Feature	Interval Value	Score
DECIMAL_END_HOUR_MINUTE	~7.5	5
	7.5~	-4
DECIMAL_START_HOUR_MINUTE	~24.5	4
	24.5~	-116
WEEKLY_MIDPOINT_VAR	0~4	118
	4~5	-94
	5~5.5	-59
	5.5~	-155
WEEK_INFO	~2	37
	2~6.5	-10
	6.5~11.5	-49
	11.5~16.5	-12
	16.5~24	-47
	24	-96
HOLI_INFO	~2	0
	2~12.5	3
	12.5~23.5	1
	23.5~24.5	-4
	24.5~	-2
FIRST_LIGHT_MINUTE	0~29	31
	29~34	-28
	34~37	23
	37~39	-38
	39~46	-5
	46~	9
FIRST_AWAKE_MINUTE	~1	10
	1~3	9
	3~8	0
	8~17	-1
	17~	-12
STAGE_SUM_MINUTE	0~31	30
	31~48	-11
	48~56	2
	56~58	25
	58~60	-23
	60~	15
FIRST_DSR	~0.005	13
	0.005~0.05	-47
	0.05~0.145	-13
	0.145~0.275	33
	0.275~	-50
FIRST_ASR	~0.01	22
	0.01~0.04	21
	0.04~0.13	7
	0.13~0.2	-5
	0.2~	-34
FIRST_LSR	~0.5	50
	0.5~0.64	-21
	0.64~0.8	10
	0.8~0.86	-4
	0.86~0.9	-32
	0.9~	16

Table A2. *Cont.*

Feature	Interval Value	Score
FIRST_RSR	~0.01	−6
	0.01~0.05	28
	0.05~0.18	55
	0.18~	−10
TOTAL_SLEEP_TIME_VAR	~0.4	0
	0.4~0.6	34
	0.6~0.8	−16
	0.8~2.2	1
	2.2~2.6	31
	2.6~3.6	−26
	3.6~	2
STEP_INFO_2	~50	−31
	50~1500	184
	1500~	145

Table A3. Table of hyperparameters for each ML model.

Model	Hyperparameter	Value	Hyperparameter	Value
LightGBM	reg_alpha	1.5486	subsample	0.5
	reg_lambda	4.5005	learning_rate	0.008
	colsample_bytree	0.7	max_depth	10
	num_leaves	470	min_child_samples	47
	min_data_per_group	100	n_estimators	2000
XGBoost	lambda	0.008	alpha	3.818
	colsample_bytree	0.4	subsample	0.7
	learning_rate	0.02	min_child_weight	39
	n_estimators	2000	max_depth	7
CatBoost	bagging_fraction	0.7723	l_leaf_reg	1.629
	max_bin	235	learning_rate	0.0155
	min_data_in_leaf		n_estimators	2000
	max_depth	7	task_type	GPU
Tabnet	max_type	Entmax	n_da	64
	n_steps	2	gamma	1
	n_shared	3	lambda_sparse	9.07×10^{-5}
	patienceScheduler	9	epochs	15

In order to confirm that the stacking model has better performance than other single ML models, classification performance was performed on 15,727 total data (good sleep: 326, bad sleep: 5168, medium sleep: 10,559). First, in order to go through the same process as score generation, only good sleep and bad sleep were included in the learning data, i.e., 80% of good sleep + bad sleep was used as training data, and the remaining 20% was used as test data. Then, 80% of the 10,559 middle sleeps were randomly extracted and added to the test data. Then, the proposed stacking machine learning model was compared with XGBoost, LightGBM, CatBoost, and Tabnet models, known as SOTA. The compared performances are summarized in Table A4. The F1 score is out of 100. The decimal point is discarded since it is only necessary to check which model has the highest performance.

Table A4. Table of F1 score for each ML model.

ML Model	F1 Score
XGBoost	89
LightGBM	87
CatBoost	88
Tabnet	85
Stacking Method	90

References

- Lee, H.; Kim, J.; Moon, J.; Jung, S.; Jo, Y.; Kim, B.; Ryu, E.; Bahn, S. A study on the changes in life habits, mental health, and sleep quality of college students due to COVID-19. *Work* **2022**, *73*, 777–786. [\[CrossRef\]](#)
- Heuse, S.; Grebe, J.L.; Esken, F. Sleep Hygiene Behaviour in Students: An Intended Strategy to Cope with Stress. *J. Med. Psychol.* **2022**, *24*, 23–28. [\[CrossRef\]](#)
- Freeman, D.; Sheaves, B.; Waite, F.; Harvey, A.G.; Harrison, P.J. Sleep disturbance and psychiatric disorders. *Lancet Psychiatry* **2020**, *7*, 628–637. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bhaskar, S.; Hemavathy, D.; Prasad, S. Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities. *J. Family Med. Prim. Care* **2016**, *5*, 780–784. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hafner, M.; Stepanek, M.; Taylor, J.; Troxel, W.M.; Van Stolk, C. Why sleep matters—The economic costs of insufficient sleep: A cross-country comparative analysis. *Rand Health Q.* **2017**, *6*, 11.
- Estrada-Galiñanes, V.; Wac, K. Collecting, exploring and sharing personal data: Why, how and where. *Data Sci.* **2020**, *3*, 79–106. [\[CrossRef\]](#)
- Nyman, J.; Ekbladh, E.; Björk, M.; Johansson, P.; Sandqvist, J. Feasibility of a new homebased ballistocardiographic tool for sleep-assessment in a real-life context among workers. *Work* **2022**. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wei, Q.; Lee, J.H.; Park, H.J. Novel design of smart sleep-lighting system for improving the sleep environment of children. *Technol. Health Care* **2019**, *27*, 3–13. [\[CrossRef\]](#)
- Smyth, C. The Pittsburgh sleep quality index (PSQI). *J. Gerontol. Nurs.* **1999**, *25*, 10. [\[CrossRef\]](#)
- Carpenter, J.S.; Andrykowski, M.A. Psychometric evaluation of the Pittsburgh sleep quality index. *J. Psychosom. Res.* **1998**, *45*, 5–13. [\[CrossRef\]](#)
- Buysse, D.J. Sleep health: Can we define it? Does it matter? *Sleep* **2014**, *37*, 9–17. [\[CrossRef\]](#) [\[PubMed\]](#)
- Morrissey, B.; Taveras, E.; Allender, S.; Strugnelli, C. Sleep and obesity among children: A systematic review of multiple sleep dimensions. *Pediatr. Obes.* **2020**, *15*, e12619. [\[CrossRef\]](#) [\[PubMed\]](#)
- Moore, P.J.; Adler, N.E.; Williams, D.R.; Jackson, J.S. Socioeconomic status and health: The role of sleep. *Psychosom. Med.* **2002**, *64*, 337–344. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nishino, S. *The Stanford Method for Ultimate Sound Sleep*; Sunmark Publishing: Tokyo, Japan, 2017.
- Patel, A.K.; Reddy, V.; Araujo, J.F. *Physiology, Sleep Stages*; StatPearls [Internet]: Florida, FL, USA, 2021.
- Beattie, Z.; Oyang, Y.; Statan, A.; Ghoreyshi, A.; Pantelopoulos, A.; Russell, A.; Heneghan, C.J.P.M. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. *Physiol. Meas.* **2017**, *38*, 1968–1979. [\[CrossRef\]](#) [\[PubMed\]](#)
- Slyusarenko, K.; Fedorin, I. Smart alarm based on sleep stages prediction. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2020**, *2020*, 4286–4289.
- Reed, D.L.; Sacco, W.P. Measuring sleep efficiency: What should the denominator be? *J. Clin. Sleep Med.* **2016**, *12*, 263–266. [\[CrossRef\]](#)
- Phillips, A.J.; Clerx, W.M.; O'Brien, C.S.; Sano, A.; Barger, L.K.; Picard, R.W.; Czeisler, C.A. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Sci. Rep.* **2017**, *7*, 3216. [\[CrossRef\]](#)
- Lunsford-Avery, J.R.; Engelhard, M.M.; Navar, A.M.; Kollins, S.H. Validation of the sleep regularity index in older adults and associations with cardiometabolic risk. *Sci. Rep.* **2018**, *8*, 14158. [\[CrossRef\]](#)
- Rosenthal, L.; Roehrs, T.A.; Rosen, A.; Roth, T. Level of sleepiness and total sleep time following various time in bed conditions. *Sleep* **1993**, *16*, 226–232. [\[CrossRef\]](#)
- Randler, C.; Vollmer, C.; Kalb, N.; Itzek-Greulich, H. Breakpoints of time in bed, midpoint of sleep, and social jetlag from infancy to early adulthood. *Sleep Med.* **2019**, *57*, 80–86. [\[CrossRef\]](#)
- Cohen, S.; Fulcher, B.D.; Rajaratnam, S.M.; Conduit, R.; Sullivan, J.P.; St Hilaire, M.A.; Phillips, A.J.K.; Loddenkemper, T.; Kothare, S.V.; McConnell, K.; et al. Sleep patterns predictive of daytime challenging behavior in individuals with low-functioning autism. *Autism Res.* **2018**, *11*, 391–403. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, M.L.; Zhou, Z.H. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognit.* **2007**, *40*, 2038–2048. [\[CrossRef\]](#)
- Rashid, W.; Gupta, M.K. A Perspective of Missing Value Imputation Approaches. In *Advances in Computational Intelligence and Communication Technology*; Springer: Singapore, 2021; pp. 307–315.
- Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst.* **1998**, *13*, 18–28. [\[CrossRef\]](#)

27. Biau, G. Analysis of a random forests model. *J. Mach. Learn. Res.* **2012**, *13*, 1063–1095.
28. Dong, L.; Martinez, A.J.; Buysse, D.J.; Harvey, A.G. A composite measure of sleep health predicts concurrent mental and physical health outcomes in adolescents prone to eveningness. *Sleep Health* **2019**, *5*, 166–174. [[CrossRef](#)] [[PubMed](#)]
29. Brindle, R.C.; Yu, L.; Buysse, D.J.; Hall, M.H. Empirical derivation of cutoff values for the sleep health metric and its relationship to cardiometabolic morbidity: Results from the Midlife in the United States (MIDUS) study. *Sleep* **2019**, *42*, zsz116. [[CrossRef](#)] [[PubMed](#)]
30. Leung, K.; Cheong, F.; Cheong, C.; O'Farrell, S.; Tissington, R. Building a Scorecard in Practice. In Proceedings of the 7th International Conference on Computational Intelligence in Economics and Finance, Taoyuan, Taiwan, 5–7 December 2008.
31. Vejkanchana, N.; Kuacharoen, P. Continuous Variable Binning Algorithm to Maximize Information Value Using Genetic Algorithm. In *International Conference on Applied Informatics*; Springer: Cham, Switzerland, 2019; pp. 158–172.
32. Finlay, S. Data Pre-Processing. In *Credit Scoring, Response Modelling and Insurance Rating*; Palgrave Macmillan: London, UK, 2010; pp. 144–159.
33. Zdravevski, E.; Lameski, P.; Kulakov, A. Weight of evidence as a tool for attribute transformation in the preprocessing stage of supervised learning algorithms. *IJCNN* **2011**, 181–188.
34. Vanneschi, L.; Horn, D.M.; Castelli, M.; Popovič, A. An artificial intelligence system for predicting customer default in e-commerce. *Expert Syst. Appl.* **2018**, *104*, 1–21. [[CrossRef](#)]
35. Dastile, X.; Celik, T.; Potsane, M. Statistical and machine learning models in credit scoring: A systematic literature survey. *Appl. Soft Comput.* **2020**, *91*, 106263. [[CrossRef](#)]
36. Peng, C.Y.J.; Lee, K.L.; Ingersoll, G.M. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* **2002**, *96*, 3–14. [[CrossRef](#)]
37. Obuchowski, N.A. Receiver operating characteristic curves and their use in radiology. *Radiology* **2003**, *229*, 3–8. [[CrossRef](#)] [[PubMed](#)]
38. Zeng, G. A comparison study of computational methods of Kolmogorov–Smirnov statistic in credit scoring. *Commun. Stat. Simul. Comput.* **2017**, *46*, 7744–7760. [[CrossRef](#)]
39. Abdou, H.A.; Pointon, J. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intell. Syst. Account. Financ. Manag.* **2011**, *18*, 59–88. [[CrossRef](#)]
40. Woo, H.S.; Lee, S.H.; Cho, H. Building credit scoring models with various types of target variables. *J. Korean Data Inf. Sci. Soc.* **2013**, *24*, 85–94.
41. Park, I. Developing the osteoporosis risk scorecard model in Korean adult women. *J. Health Inform. Stat.* **2021**, *46*, 44–53. [[CrossRef](#)]
42. Han, J.T.; Park, I.S.; Kang, S.B.; Seo, B.G. Developing the High-Risk Drinking Scorecard Model in Korea. *Osong Public Health Res. Perspect.* **2018**, *9*, 231–239. [[CrossRef](#)]
43. Siddiqi, N. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*; Wiley & Sons.: Hoboken, NJ, USA, 2012; Volume 3.
44. Divina, F.; Gilson, A.; Gómez-Vela, F.; García Torres, M.; Torres, J.F. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies* **2018**, *11*, 949. [[CrossRef](#)]
45. Wang, T.; Zhang, K.; Thé, J.; Yu, H. Accurate prediction of band gap of materials using stacking machine learning model. *Comput. Mater. Sci.* **2022**, *201*, 110899. [[CrossRef](#)]
46. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
47. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.
48. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363.
49. Arik, S.Ö.; Pfister, T. Tabnet: Attentive interpretable tabular learning. *AAAI* **2021**, *35*, 6679–6687. [[CrossRef](#)]
50. Rasifaghihi, N.; Li, S.S.; Haghghat, F. Forecast of urban water consumption under the impact of climate change. *Sustain. Cities Soc.* **2020**, *52*, 101848. [[CrossRef](#)]
51. Hans, C. Elastic net regression modeling with the orthant normal prior. *JASA* **2011**, *106*, 1383–1393. [[CrossRef](#)]
52. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [[CrossRef](#)]
53. Gunst, R.F.; Mason, R.L. Biased estimation in regression: An evaluation using mean squared error. *JASA* **1977**, *72*, 616–628. [[CrossRef](#)]
54. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; ACM: New York, NY, USA, 2019.
55. Halson, S.L.; Johnston, R.D.; Piromalli, L.; Lalor, B.J.; Cormack, S.; Roach, G.D.; Sargent, C. Sleep Regularity and Predictors of Sleep Efficiency and Sleep Duration in Elite Team Sport Athletes. *Sport. Med. Open* **2022**, *8*, 79. [[CrossRef](#)]
56. Windred, D.P.; Jones, S.E.; Russell, A.; Burns, A.C.; Chan, P.; Weedon, M.N.; Rutter, M.K.; Olivier, P.; Vetter, C.; Saxena, R.; et al. Objective assessment of sleep regularity in 60 000 UK Biobank participants using an open-source package. *Sleep* **2021**, *44*, zsab254. [[CrossRef](#)]

57. Makarem, N.; Zuraikat, F.M.; Aggarwal, B.; Jelic, S.; St-Onge, M.P. Variability in sleep patterns: An emerging risk factor for hypertension. *Curr. Hypertens. Rep.* **2020**, *22*, 19. [[CrossRef](#)]
58. Baron, K.G.; Reid, K.J.; Malkani, R.G.; Kang, J.; Zee, P.C. Sleep variability among older adults with insomnia: Associations with sleep quality and cardiometabolic disease risk. *Behav. Sleep Med.* **2017**, *15*, 144–157. [[CrossRef](#)]
59. Buman, M.P.; Phillips, B.A.; Youngstedt, S.D.; Kline, C.E.; Hirshkowitz, M. Does nighttime exercise really disturb sleep? Results from the 2013 National Sleep Foundation Sleep in America Poll. *Sleep Med.* **2014**, *15*, 755–761. [[CrossRef](#)]
60. Stutz, J.; Eiholzer, R.; Spengler, C.M. Effects of evening exercise on sleep in healthy participants: A systematic review and meta-analysis. *Sport. Med.* **2019**, *49*, 269–287. [[CrossRef](#)] [[PubMed](#)]
61. Frimpong, E.; Mograss, M.; Zvionow, T.; Dang-Vu, T.T. The effects of evening high-intensity exercise on sleep in healthy adults: A systematic review and meta-analysis. *Sleep Med. Rev.* **2021**, *60*, 101535. [[CrossRef](#)]
62. Kim, J.; Lee, J.; Park, M. Identification of Smartwatch-Collected Lifelog Variables Affecting Body Mass Index in Middle-Aged People Using Regression Machine Learning Algorithms and SHapley Additive Explanations. *Appl. Sci.* **2022**, *12*, 3819. [[CrossRef](#)]
63. Liang, Z.; CHAPA-MARTELL, M.A. Predicting Medical-Grade Sleep-Wake Classification from Fitbit Data Using Tree-Based Machine Learning. *Rep. Number IPSJ SIG Tech. Rep.* **2019**, 2019, 14.
64. Jiang, M.; Liu, J.; Zhang, L.; Liu, C. An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. *Phys. A Stat. Mech. Appl.* **2020**, *541*, 122272. [[CrossRef](#)]
65. Pavlyshenko, B. Using Stacking Approaches for Machine Learning Models. In Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 255–258.
66. Yu, W.; Li, S.; Ye, T.; Xu, R.; Song, J.; Guo, Y. Deep ensemble machine learning framework for the estimation of PM 2.5 concentrations. *Environ. Health Perspect.* **2022**, *130*, 037004. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.