

Sleep-wake cycle-involved hormone levels analysis based on sleep and activity data

1 Introduction

Sleep is an inevitable daily activity. Recently, I have become more and more aware of having a healthy and disciplined daily schedule. When trying to adapt with a natural sleep-wake schedule, I got to know that this rhythm is regulated by two key hormones: melatonin and cortisol. To be specific, the increase in cortisol level could result in sleep deprivation and disruption (Liu & Reddy, 2022). Cortisol is also influenced by the 24-hour circadian rhythm (Jiao et al., 2025). However, high level cortisol can result in chronic stress over along period of time (Knezevic et al., 2023). Allen et al. (2013) stated that cortisol level usually peaks during the morning, at around 30 minutes after wake up. Melatonin, also has a regulating effect on the sleep-wake cycle of mammals (Scheer & Czeisler, 2004). However, melatonin level is generally higher during the night and lower after we wake up.

To the best of my knowledge, the ability to predict hormone fluctuation based on objective metrics is not extensively implemented. In this project, I utilize the Multilevel Monitoring of Activity and Sleep in Healthy people (MMASH) dataset to investigate the relationship between daily lifestyle factors and physiological hormone levels. The project aims to study their relationship and find which features could be used in a Machine Learning model in order to predict changes in hormone levels.

In order to answer these questions, we will go through several statistical analyses, such as Pearson/Spearman correlations and t-tests, on data from 22 healthy male participants. Then, based on the data analysis result, we will select some specific features to implement two ML models to predict cortisol levels after the participant wakes up.

The result confirms a significant relationship between (1) hormone levels after sleep and (2) the cortisol level difference before and after sleep. It also brings idea about several possible correlation between variables in our high-dimensional data.

Based on this, ML task is introduced to predict the cortisol level after wake up based on the cortisol level before sleep, self-validated morning-eveningness score, sleep efficiency, daily stress index, and state anxiety value. Remarkably, the random forest model achieve a relatively promissing result with a average RMSE (Root mean square error) of 0.046 with a limited sample space.

2 Problem Formulation

In this project, my center object is hormones levels, specifically cortisol and melatonin. In the given dataset, these hormone levels are measured both before and after sleep.

I am attempting to find the possible answer to the following research questions can be:

- What are the relationship between cortisol and melatonin levels before and after sleep?
- Do melatonin and cortisol level have any correlation?
- Do these hormone levels have any influence on the sleep quality?
- Can melatonin and cortisol impact sleep measures (e.g., total sleep time, sleep schedule, sleep efficiency, fragmentation index)?
- Do activities during the day affect hormones levels?

Thus, the project mainly focus on sleep measures (provided in the files 'sleep.csv'), and activities of participant during the day (provided in the file 'Activity.csv').

Furthermore, as cortisol is usually related to stress level (Knezevic et al., 2023), I am also going to investigate if there is any relation between the cortisol levels (both before and after sleep) and self-reported stress index (STAI).

Additionally, after find out some features that is highly correlated to a specific hormone, ML models can be implemented to predict this target hormone level from the given numerical features.

These factors are motivated and supported by other scientific findings and would be discussed in further detail in the next section.

3 Dataset

3.1 Data introduction

The MMASH (Multilevel Monitoring Activity and Sleep in Healthy people) dataset is provided by Rossi et al. (2020). According to their publication, in the study, the data of 22 adults males participants are collected with an IoT wearable devices. The data collected for each participant is highly complex and is stored in separate folders. Data for each participant are stored in several files, i.e., 'user_info' (which stores basic information such as age, height, and weight), 'Actigraph' (which contains accelerometer and inclinometer data), 'Activity' (which includes a list of activities among 12 categories that the participant do throughout the day), 'questionnaire' (which combines the self-evaluated score for multiple aspects), 'RR' (which contains beat-to-beat interval data),

'saliva' (which includes hormones levels before and after sleep), and 'sleep' (which stores numerical representation of sleep quality and sleep duration). These values are recorded through two consecutive days. However, the sleep features are missing for participant is missed for participant number 11, and the hormone level is not recorded for participant number 21. Therefore, the combined data frame contains sleep and saliva biomakers features only have 20 datapoints.

The main target of this project is melatonin and the hormone level. To reduce the complexity of the given dataset, I create two separate data frames to examine physiological relationship between hormones level and other relevant features.

The first data frame contain factors that are not time-dependent or not required Datetime-based pre-processed. Usually, each user have only one valued recorded for these features. The main representative columns and the corresponding explanation of this data frame is presented on Table 1 below.

Table 1: Columns in the first data frame (adapted from Rossi et al. (2020))

Column Name	Details
user_id	22 participants are ordered with simple numbering from 1 to 22.
Gender	Male or Female. In these cases, all users are male.
Age	How old is the participant?
Weight	Participant's weight.
Height	Participant's height.
Latency Efficiency	ratio of sleep time divided by in-bed time.
Number of Awakenings	Number of times the participant awake during their sleep.
Average Awakening Length	Time they stay awake during the night (measured in seconds).
Movement Index	ratio of time without movement and total duration of the movement phase.
Fragmentation Index	ratio of the time with movement and duration of the immobile phase.
Sleep Fragmentation Index	The movement index divided by the fragmentation index.
Total Minutes in Bed	Total time spent in bed (measured in minutes).
Total Sleep Time (TST)	Total time spent for sleep (measures in minutes).
Continued on next page...	

Table 1 – continued from previous page

Column Name	Details
MEQ	“Morningness-Eveningness Questionnaire value. The chronotype score is ranging from 16 to 86: scores of 41 and below indicate Evening types, scores of 59 and above indicate Morning types, scores between 42-58 indicate intermediate types.”
Pittsburgh	“PSQI: Pittsburgh Sleep Quality Questionnaire Index. It gives a score rating from 0 to 21, with values lower than 6 indicating good sleep quality.”
Daily_stress	DSI: “Daily Stress Inventory value is a 58-item self-reported measure that allows a person to indicate the events they experienced in the last 24 hours. After indicating which event occurred, they indicate the stressfulness of the event on a Likert scale from 1 (occurred but was not stressful) to 7 (caused me to panic). It gives a score between 0 and 406. The higher this values are, the higher is the frequency and degree of the events and the perceived daily stress.”
STAI1 and STAI2	“State Anxiety value obtained from State-Trait Anxiety Inventory. The results are range from 20 to 80. Scores less than 31 may indicate low or no anxiety, scores between 31 and 49 an average level of anxiety or borderline levels, and scores higher than 50 a high level of anxiety or positive test results.”
Cortisol/Melatonin Before and After Sleep	Level of cortisol and melatonin measured before participants go to bed and after they wake up.

In the second data frame, time-related factors in the 'Activity' data file. The study collect starting time and ending time for 12 types of activities that occur throughout the day (Rossi et al., 2020):

- 1. Sleeping
- 2. Laying down
- 3. Sitting, for example, when participants study or work
- 4. Light movement, such as slow walk or household chores.
- 5. Medium movement, e.g., faster walk.

- 6. Heavy movement, such as exercise.
- 7. Eating
- 8. Small screen usage through smartphone or a laptop.
- 9. Large screen usage, for instance, through TV.
- 10. Caffeine consumption.
- 11. Smoking
- 12: Alcohol consumption.

With this second data frame, only activities carried out before sleep are examined in comparison with the hormone level before sleep and sleep quality. The duration of activities in each group is calculated in minutes. From the 12 categories above, it is relatively clear that activities numbered from 1 to 6 measure daily physical movement; activities 8 and 9 represent digital usage, and the last three activities are related to addictive consumption. Thus, these three groups are evaluated separately against the melatonin/cortisol level before bed and the sleep quality index (PSQI).

The feature data for the ML task will be described in the following sections when the data analysis results are explained.

3.2 Data preprocessing approach

3.2.1 Data frame 1

With the first data frame, I loop through 22 folders of participants and extract corresponding markers (as mentioned in Table 1) from the relevant files. The sleep file also contains some other columns, such as 'In Bed Time', 'In Bed Date', 'Out Bed Time', and 'Out Bed Date'. However, these features are not considered here as the sleeping schedule is already represented through other factors, e.g., MEG, and Total Sleep Time.

Since some participants (e.g., user 1) go to bed more than once during the two consecutive days, when extracting indicators from the 'sleep' file, we take the means of Latency Efficiency, Number of Awakenings, Average Awakening Length, Movement Index, Fragmentation Index, and Sleep Fragmentation Index; and take the sum of Total Time in Bed and Total Sleep Time, as representative values in this combined frame. Eventually, I notice that all participant are male, so the Gender column is dropped. Due to missing data, the row for user 11 and user 21 are removed, leaving 20 entries in this combined dataset.

3.2.2 Data frame 2

For the second part, I again loop through the folder of users. First, we can simply copy the extracted data related to the saliva biomarker and the corresponding user ID from the first data frame. Then some handle function are created to process the time. There is only one value for hormone levels recorded even though some user sleep several time during two days (e.g., user 1). Thus, for simplicity, I only evaluate the first time they sleep in this case. Some helper function is created to handle the time-series modification. Here, we only considered the activities that happen before user go to sleep. So that, I compared the time that those activity end with the first Onset time that participant fall asleep. Subsequently, we aggregate the total duration the users spent on a specific categories of activities (see the Appendix code for more detailed information).

For activities group related to addictive usages (caffeine, smoke, alcohol), we create a smaller dataset contains information related to saliva biomarkers before sleep, PSQI index, and binary values indicate whether participants spent time for consuming caffeine, smoking, or drinking alcohol. All

For activities related to screen time, we examine usage of large and small devices separately. We create a new data set with hormones indicators, PSQI, and duration spending on large screen. All participant that have no interact with large screen devices during these two days are removed, so that there is only 12 non-null value leave for this task. Similarly, the same process is repeated for the data related to small screen. In this case, we also have 12 valuable entries for small screen usage.

Respect to physical activities, the movement categories are clearly divided into 5 level of intensities: from Laying down, sitting, light movement, medium movement, to heavy movement. I calculate the total active time of all physical groups, and then take the percentage of each categories. I suppose that representing this in ratio could make it easier to compare between participants with different lifestyle.

4 Methods

4.1 For data analysis

4.1.1 Correlation analysis

Correlation matrix are extensively used in this project since it is relatively easy to interpreted and high-dimensional data can be represented in a straightforward way. Here, Pearson correlation coefficient is calculated so that the correlation between two variable is shown. Pearson correlation coefficient is the normalized version of the covariance, which is obtained by dividing the covariance of two variable by the multiplication of their standard deviation. The larger the absolute value of the Pearson coefficient, the

more dependent the two variable with each other. Positive value of Pearson coefficient shows positive relationship and vice verse. However, with our limited sample size, the data is potentially not adhere to normality. Therefore, in some cases, Spearman's rank correlation coefficient is also implemented.

4.1.2 t-test

t-test are selected here to find if some hypothesis is statistically significant or not. Since the sample size is remarkably small, I use the Bonferroni correction for the threshold of p-value. If we consider 95% significant level with alpha equal to 0.05, then when the correction is applied, the updated alpha value is 0.05 divided the number of data points.

4.1.3 Linear regression

Linear regression is used in some plot contains two variable to directly illustrate their relation in a more straightforward way.

4.2 For machine learning

4.2.1 Linear regression

Linear regression is used as the first basic ML model in this task mainly because its result is interpretable. This could be a remarkable factor in biological problems as the scientist needs to understand how the predicted value is created to, in some cases, make a clinical decision. Basically, the linear regression model assumes a linear relationship between features and tries to find the intercept and coefficients of the features in order to minimize the error between the predicted and the real target value on the training dataset.

4.2.2 Random forest

Random forest (RF) is the second method being considered for this classification problem. In essence, RF combines two techniques, bagging and feature randomness, to generate several decision trees. Specifically, each tree is trained on a random subset (bagging), and each tree evaluates a random subset of features (feature randomness) so that the decision trees in the forest are uncorrelated (IBM, 2025). This method is robust and less sensitive to outliers and skewed distributions. In addition this approach does not assume a linear relationship between features. Thus, RF is an appropriate method to be compared with linear regression, and is potentially more suitable to evaluate complex biological correlations.

In this project, the Linear regression and Random forest models are simply imported

from the sklearn library.

Besides, as our dataset is considerably small (with 20 data points), k-fold cross-validation is implemented. For this method, the dataset is divided into k equal-sized subsets, or "folds." The model is iteratively trained k times; in each iteration, it is trained on k - 1 subsets (training sets) and validated on the single remaining subset (validation set) (Refaeilzadeh et al., 2009). In this case, $k = 5$ is selected so that each fold has only 4 value for testing and 16 values for training.

Root mean squared error is selected to evaluate the performance of the two models. RMSE preserves the original units of the target variables so that it is relatively easier to interpret, in comparison with MSE. Additionally, RMSE penalizes large errors more significantly than mean absolute error.

5 Results

In this section, I present the most interesting findings of the project.

5.1 Data frame 1

First, we use correlation matrix to obtain a broad observation of the high-dimensional data frame (Figure 1).

From this matrix, we clearly see that variables that are highly correlated to each other are:

- Efficiency and Movement Index (negative correlation)
- Fragmentation Index and Movement Index (positive correlation)
- Total Minutes In Bed and Total Sleep Time (positive correlation)
- Cortisol After Sleep and Melatonin After Sleep (positive correlation)

The correlation coefficient of two hormones level after sleep is 0.8567, which is considerably high. To confirm if their relation is statistically significant, I use both Pearson and Spearman tests with corrected $\alpha = 0.05/20 = 0.0025$. The p-values resulted from both tests are significant lower than 0.0025 ($1.42e-6$ and 0.00034 , respectively). Therefore, I suppose we can conclude that the cortisol and melatonin level after sleep are highly related to each other.

Subsequently, we further investigate the relationship between the two hormone levels before and after sleep with the pair plot and linear regression line (Figure 2). From this plot, it is visible that the cortisol level after sleep is consistently higher than that before

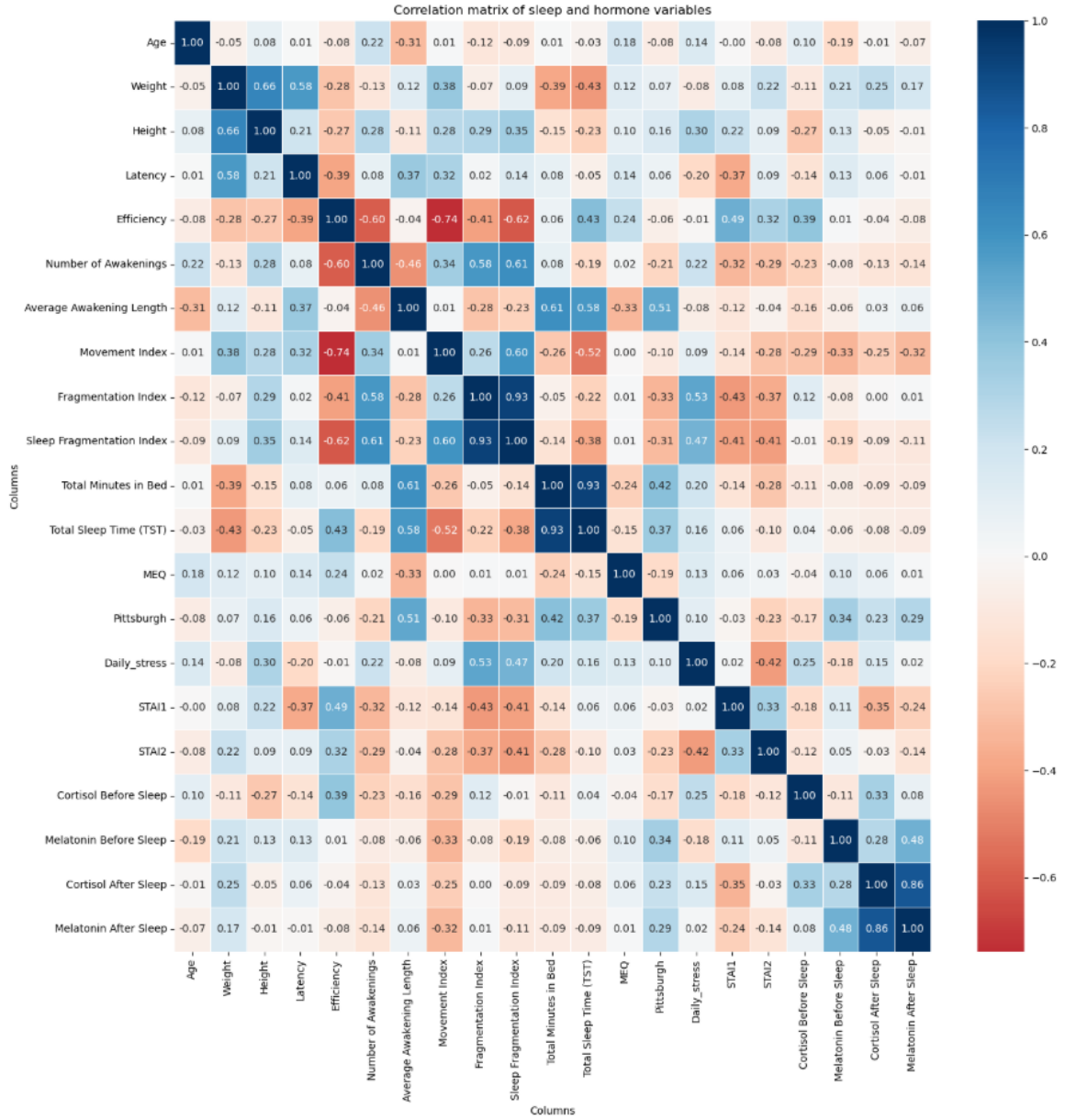


Figure 1: Correlation matrix between sleep measures, key user information, self-evaluated answers, and hormone levels.

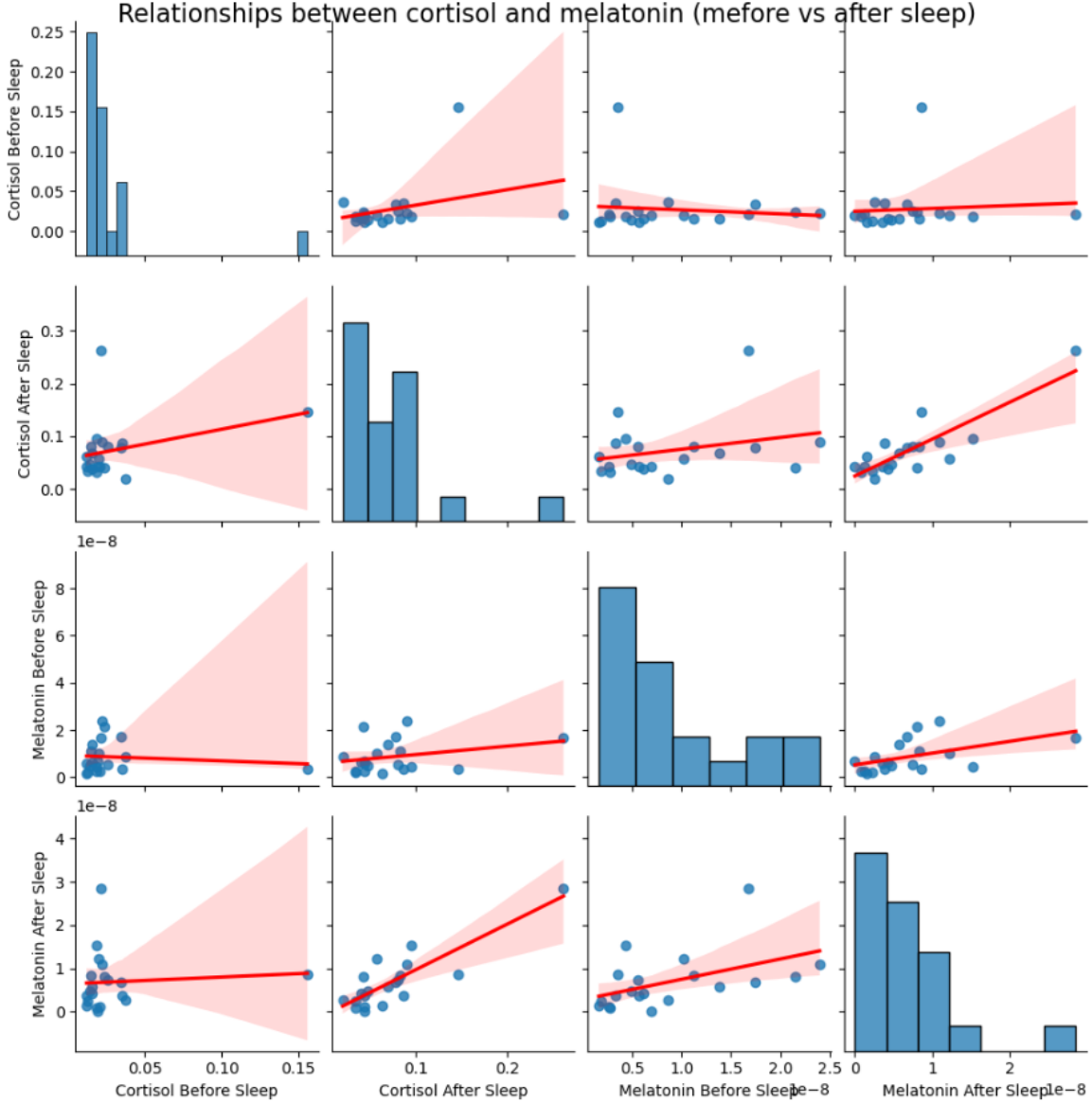


Figure 2: Results of the linear regression and random forest models.

sleep. In addition, this hypothesis is supported by Allen et al. (2013), as mentioned in the introduction. We test this hypothesis with a one-tailed t-test. The result p-value 0.00049 directly show that this difference is statistically significant in this case.

Subsequently, I also created some other plots to examine the relation of cortisol level with stress level and efficiency. These analysis bring me the idea to use some relevant features, e.g., MEQ, Efficiency, Cortisol Before Sleep, Daily stress, STAI1 to predict the level of Cortisol After sleep.

Implementing Linear Regression and Random Forest with 5-fold cross-validation, we achieve the following result (Table 2). From the result, it is clear that the Random Forest outperforms the Linear Regression model (Table 2, Figure 3) as the average RMSE across 5 folds of Linear Regression model is nearly twice that of Random Forest model. The

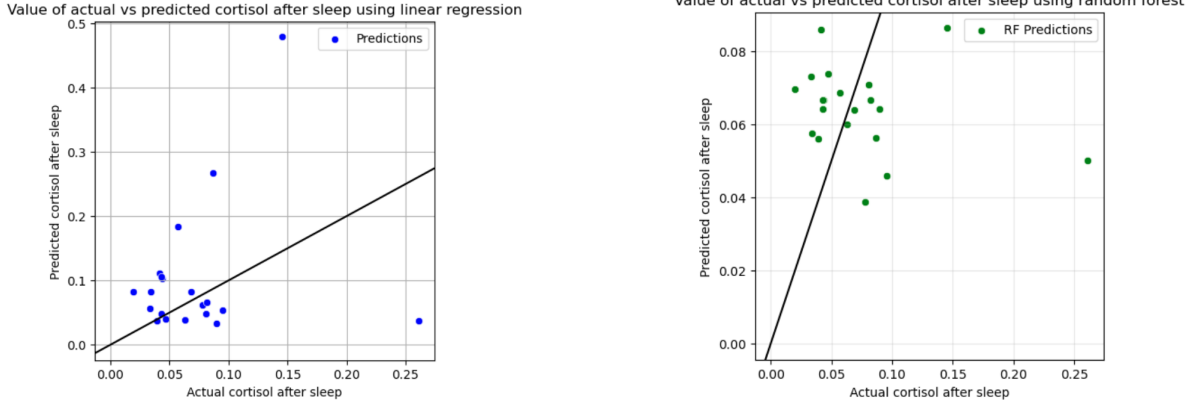


Figure 3: Results from two ML models

Table 2: Comparison of Root Mean Squared Error (RMSE) results for 5-fold cross-validation.

Fold	Linear Regression	Random Forest
1	0.115	0.110
2	0.046	0.027
3	0.030	0.021
4	0.044	0.035
5	0.202	0.035
Average	0.087	0.046

cortisol level after sleep ranged from 0.04 to 0.25 in our dataset, with the majority of values less than 1.0, and the RMSE error of Random Forest model is only around 0.02 in the third fold. The model therefore has the potential to achieve a better result on larger dataset. However, errors vary significantly in different folds (RMSE in the first fold is 0.11 which is approximately five times the error in the third fold). Thus, the models are not robust enough.

5.2 Data frame 2

Using the correlation analysis for three extracted data frames for physical activities and screen time, we got the matrices in Figure 4. The most outstanding result here is the relatively strong negative correlation between the level of melatonin before sleep, as well as the sleep quality index, with the duration spent on a large screen: correlation between large-screen time during the day and the melatonin level before bed is -0.71; and that with PSQI score is -0.62. When Pearson and Spearman test is conducted for screen time

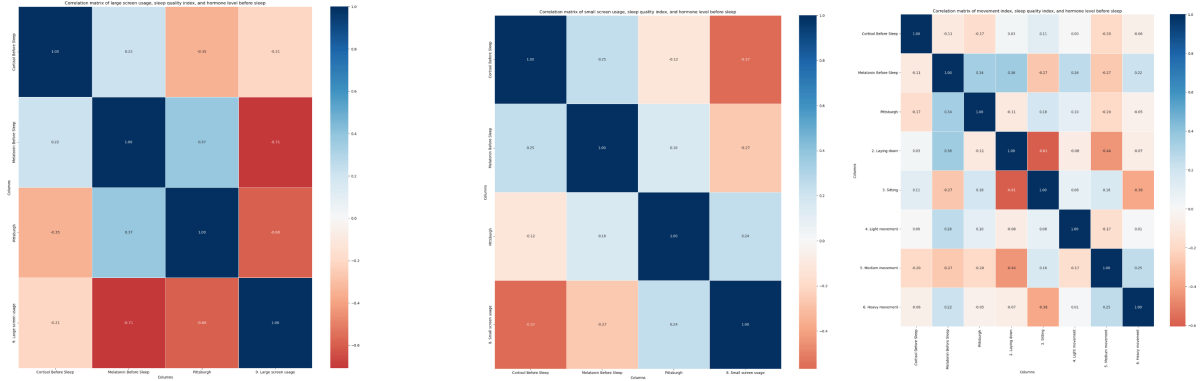


Figure 4: Correlation matrices between screen time on large/small devices, and movement activities with sleep quality index and hormone levels before sleep.

on large device and melatonin level, the p-values are 0.01 and 0.0026, respectively. These statistical values are smaller than the normal threshold 0.05 but is still large than the corrected p-value 0.0025. Therefore, I cannot surely confirm their significance in this case. However, the p-value for Spearman test is relatively close, so these factors could be worth considering with other more detailed sample.

6 Conclusion & Discussion

In conclusion, this project analyzes the complex relationship between cortisol and melatonin, sleep quality, and daily activities using the MMASH dataset.

The results of this study should be interpreted carefully since there is a significant limitation in the sample size of 22 males. Furthermore, since data is collected for only two days, even though the data set is already highly complex, it may not be able to capture long-term behavioral patterns of participants.

The machine learning model has the potential to be improved with more data points. In addition, other advanced methods, such as Deep Neural Networks, can be considered to further study the complex relationship between these features.

Some possible relationships should be further examined with a larger sample size, for example:

- Relationship between screen time on large devices and hormone level or sleep quality.
- Melatonin level before sleep and laying down pattern of the users.
- Caffeine, alcohol, and cigarette impact on hormone levels and sleep quality
- Relation between cortisol and stress indices.

From our limited sample, there could be some hints for their correlation, and unfortunately these correlation are not statistically significant enough to draw a meaningful conclusion.

Moreover, the MMASH also contains a range of other features (i.e., data in the ACtigraph and RR files) that I, unfortunately, did not able to fully investigate in this project. In the future, I will definitely examine the impact of heart rate or heart rate variability with the hormone level. Furthermore, since melatonin and cortisol rely on the circadian cycle of the body, a time series analysis would potentially provide some interesting results.

References

- Allen, A. P., Kennedy, P. J., Cryan, J. F., Dinan, T. G., & Clarke, G. (2013). Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. *Neuroscience Biobehavioral Reviews*, *38*, 94–124. <https://doi.org/10.1016/j.neubiorev.2013.11.005>
- IBM. (2025, June). Random Forest. <https://www.ibm.com/think/topics/random-forest>
- Jiao, Y., Butoyi, C., Zhang, Q., Adotey, S. A. A. I., Chen, M., Shen, W., Wang, D., Yuan, G., & Jia, J. (2025). Sleep disorders impact hormonal regulation: unravelling the relationship among sleep disorders, hormones and metabolic diseases. *Diabetology Metabolic Syndrome*, *17*(1), 305. <https://doi.org/10.1186/s13098-025-01871-w>
- Knezevic, E., Nenic, K., Milanovic, V., & Knezevic, N. N. (2023). The role of cortisol in chronic stress, neurodegenerative diseases, and psychological disorders. *Cells*, *12*(23), 2726. <https://doi.org/10.3390/cells12232726>
- Liu, P. Y., & Reddy, R. T. (2022). Sleep, testosterone and cortisol balance, and ageing men. *Reviews in Endocrine and Metabolic Disorders*, *23*(6), 1323–1339. <https://doi.org/10.1007/s11154-022-09755-4>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). *Cross-Validation*. https://doi.org/10.1007/978-0-387-39940-9_{_}565
- Rossi, A., Eleonora, D. P., Menicagli, D., Tremolanti, C., Priami, C., Sirbu, A., Clifton, D., Martini, C., & Morelli, D. (2020, June). Multilevel Monitoring of Activity and Sleep in Healthy people. <https://doi.org/10.13026/cerq-fc86>
- Scheer, F. A., & Czeisler, C. A. (2004). Melatonin, sleep, and circadian rhythms. *Sleep Medicine Reviews*, *9*(1), 5–9. <https://doi.org/10.1016/j.smr.2004.11.004>