Mobile Sleep Health Analysis

AARUSHI SEHGAL, University of California, San Diego, Department of Computer Science and Engineering
VIVIAN LEE, University of California, San Diego, Department of Electronics and Computer Engineering
ZHEXU LI, University of California, San Diego, Department of Electronics and Computer Engineering

Sleep is crucial to our well being. Numerous studies have shown that one's sleep pattern is closely correlated with their physical and mental behaviors during the day, yet most of them relied on clinical data which might not be representative for everyday activities. In this project we investigate the associations between daily physical activities, psychological and emotional states and sleep pattern (sleep duration, sleep quality, etc) using a dataset collected with wearable devices in non-clinical everyday settings. The resulting correlations we found provided us insights into what might contribute to a healthy sleep.

ACM Reference Format:
Aarushi Sehgal, Vivian Lee, and Zhexu Li. 2022. Mobile Sleep Health Analysis. 1, 1, Article 1 (March 2022), 11 pages.

1 INTRODUCTION

1.1 Motivation

Physical activity and sleep are highly interrelated health behaviors. Our physical activity during the day influences our quality of sleep, and vice versa. Numerous studies have revealed strong associations between one's daily activities and sleep pattern, yet most of them rely on data collected in clinical settings, which might not be representative of the real-world experience for the general population. As it is hard to maintain the incentive for people to regularly report for the study and data collection, numerous people end up dropping out in between such studies, which leads to important data getting removed and the final data-set becoming biased.
The importance of sleep is paramount to health. Insufficient sleep can reduce physical, emotional, and mental well-being and can lead to a multitude of health complications among people with chronic conditions. The current popularity of wearables for tracking physical activity and sleep, including actigraphy devices, can foster the development of new advanced data analytics. We therefore focus on utilizing data collected using wearables for our analysis, accounting for people's behaviour in their day to day life.

1.2 Project Aim

Given that sleep is very crucial to our well being, we aim to investigate the associations between peoples' daily behaviors (physical and mental) and their sleep patterns (sleep duration and sleep quality). With the "Multilevel Monitoring of Activity and Sleep in Healthy People" dataset which includes sensors data collected from wearable devices in non-clinical environments, we aimed to find patterns and potential associations between daily activities and sleep in a day to day life setting.

Authors' addresses: Aarushi Sehgal, aasehgal@ucsd.edu, University of California, San Diego, Department of Computer Science and Engineering; Vivian Lee, University of California, San Diego, Department of Electronics and Computer Engineering, ; Zhexu Li, University of California, San Diego, Department of Electronics and Computer Engineering.

XXXX-XXXX/2022/3-ART1 $15.00
https://doi.org/

2 RELATED WORKS

Our project aims to contribute to mobile sleep health analysis, and there are many research in this area inspired our project. Thus, we will mention a couple notable related works regarding the topic in sleep quality analysis using wearable and mobile devices to demonstrate the potential in using wearable for deeper inspection of variables that affect sleep.

2.1 Sleep Quality Prediction From Wearable Data Using Deep Learning [2]

This study proposes a deep learning model to predict the quality of sleep based on data collected from wearable devices. Mentioned in the paper, the goal was to measure sleep quality without self reported data from the users. Overall the study contributed sleep quality assessments, a convolution neural network, and a multi-layer perceptron to predict sleep quality based on the indicators of daily sleep quality, weekly sleep quality, sleep consistencies. They analyzed data collected by both commercial smartwatches and clinical actigraph devices in National Sleep Research Resource. In their results, they found that CNN better predicts sleep quality than MLP on all three indicators. Furthermore, they concluded that bed awake percentage, sleep efficiency, and sleep onset latency are the most important features for daily sleep quality. In essence, there are studies analyzing different machine learning models to predict sleep quality. This gave inspiration for us to explore our own machine learning models.

2.2 SleepMiner [3]

With the aim to predict sleep quality, this study proposes a sleep quality prediction framework called Sleep Miner, a system to data collect and analyze user context data. Using data collected from Android phones, they attempt to analyze the relationship in sensor and communication data with sleep quality. After analysis, they then developed a learning model to predict sleep quality. The study also collects information through a questionnaire based on PSQI and utilizes a built in sleep quality monitoring software in Google Play Store. Several features that were extracted included daily activity, living environment, and social activity from the Android phone data. After training the model, the results indicate that the factor graph model improves sleep quality prediction. From their results, they demonstrate the possibility to accurately predict sleep quality at around 78% through context data. This study shows that mobile phone data that can help provide insight to sleep quality accurately. This study also reinforces the importance of mobile data for sleep health, supporting the reason for our work.

2.3 A multimodal analysis of physical activity, sleep, and work shift in nurses with wearable sensor data [4]

A recent study analyzed the correlation between physical activity, sleep, and circadian misalignment among day and night shift nurses. They aimed to compare day and night shift nurses to see if there were any differences in sleep patterns, physical activity levels and self reported behavioral variables. To carry out the study, they collected data from 113 nurses over a period of 10 weeks through Fit-Bit Charge 2, a wristband based wearable device. The device tracks heart rate, physical activities, and sleep quality both outside and during work. Sleep patterns were identified through the devices' features such as sleep onset time, wake up time, sleep duration and sleep efficiency. Furthermore, the study collected self assessments regarding behavioral variables. The takeaway from the work is they identified that day shift nurses have higher life satisfaction than night shift nurses. Night shift nurses reported poorer sleep quality (higher PSQI scores), lower level of positive affect, and were seen to walk less and have more physical rest compared to day shift nurses both during and outside work. In addition, the wearable also suggested that night shift nurses have more irregular sleep, less sleep on workdays, and larger circadian misalignment. This study demonstrates the wearable devices potential to provide us a fuller picture of factors that influence our sleep quality as well as mental health in nurses. Our work will also analyze the same variables such as STAI, PANAS, and PSQI, along with heart rate and sleep quality to attempt to find correlations among these variables in our sample of 22 young men.

Overall our goal is to further expand on these studies and see if we can further demonstrate insightful findings from data collected in BioBeat IoT wearable devices.

3 METHODOLOGY

3.1 Dataset

This data set was provided by BioBeat and University of Pisa, this data set contains heartbeat, accelerometer, sleep quality, physical activity, and psychological characteristic data from 22 participants. The data was collected with BioBeats IoT wearable devices and was recorded by scientists, psychologists, and chemists with the goal of assessing psycho-physiological responses to sleep and stress stimuli. All participants are males with age between 20 - 40, so it's certainly not diversified enough for demographic related analysis. More details about the features we ended up using are included in the later sections.

3.2 Terminology

(1) **PSQI** is a commonly used questionnaire based index to measure sleep quality and sleep disturbances over a month. The questionnaire comprises of 19 questions in 7 different components. The lower the values, the better the sleep quality.
(2) **STAI1** refers to state anxiety. State anxiety is anxiety about a particular event. It is more situation based rather than personality based. The higher the score, the greater the anxiety.
(3) **STAI2** refers to trait anxiety. Trait anxiety is a trait of personality, describing individual differences related to a tendency to present state anxiety. It is relatively stable within the individual and reflected in models of personality rather than situation. Again, the higher the score, the greater the anxiety.
(4) **Daily Stress (DSI)** is self reported measures based on events they experienced in the last 24 hours. The higher the values, the higher frequency and degree of events and the perceived stress.
(5) **PANAS** is the positive and negative affect schedule, rating from 5 to 50 in terms of positive and negative emotions. The higher the PANAS value, the higher level of positive or negative emotion.

3.3 Data Cleaning

This step involved writing python scripts for extracting data files of 22 users in the given data-set for different correlation analysis. We mainly focused on extracting files corresponding to activity, psychological and emotional state parameters, sleep and user information. The files utilized were as follows:-

(1) **Activity.csv**
Includes list of the activity categories throughout the day.
(2) **User_info.csv**
Includes anthropocentric characteristics (gender, age, height, weight, etc) of the participant.
(3) **Sleep.csv**
Includes information about sleep duration and sleep quality of the participant.
(4) **Questionnaire.csv**
Includes scores for all the questionnaires mentioned above (PSQI, STAI1, STAI2, DSI, PANAS).

After extracting the aforementioned files, these files were merged for all 22 users to get a unified representation of all the aforementioned data sets across all given users and later used for performing correlation analysis between all the different parameters.

In our final merged files, rows represent users while columns represent different parameters to be analyzed depending on the type of dataset. Merged files were then imputed for NaN values using mean of the given parameters. Rows containing NaN values above a threshold of 40% were removed and since none of our rows/users contained NaN values above 40%, we retained all our rows.

We utilized **pandas** *dataframes* along with **OS** libraries in python for the purpose of this section of analysis.

3.4 Distribution Analysis

We visualized the distribution of the features we interested in across different users. This was an essential first step because it provided valuable information about the variables that helped us better inform our downstream analysis on these parameters. For example, if certain features are not normally distributed, we can investigate the cause of that distribution, and if they are normally distributed, we can use Pearson Correlation coefficient analysis for finding correlation values and their p-values significance. We utilized **seaborn** *Distplots* and *Density* plots for plotting the distributions.

These distributions were plotted for both Activity based and Psychological and Emotional States data based parameters that we later worked with in our analysis.

3.5 Correlation Analysis

In order to investigate the correlation between different features and labels, we first utilized the density plot to make sure the distribution of the target variable is normal, then for every selected pair of variables, we calculate their Pearson's Correlation. The reason we chose Pearson Correlation over other metrics is that it is one of the best method of measuring association between variables of interest because it's based on the method of covariane, according to [1], and it's easy to interpret and visualize, and the relatively less strict assumptions makes it a powerful and popular metrics for correlation analysis. Pearson Correlation is defined by:

$$r = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2 \sum_i^n (y_i - \bar{y})^2}} \tag{1}$$

where r is the correlation coefficient, $\bar{x}$ is the mean of the variable x, and $\bar{y}$ is the mean of the variable y. Pearson Correlation assumes indepedence and linear relationship between variables, and homoscedasticity (residuals scatterplot are roughly rectangular-shaped) [1]. The linear relationship and homoscedasticity can be verified through regression plots, which are scatter plots but with regression lines to give a more direct visualization of the distribution correlation. Note that despite the Pearson Correlation doesn't assume the distribution of variables to be normal, it works the best if the variables are normally distributed [1]. And in our dataset, most variables we are interested in are indeed normally distributed. The p-value of the Pearson Correlation is defined using a T distribution with n - 2 degrees of freedom:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2}$$

It indicates the significance of the finding, which is the probability that the obtained correlation coefficient was caused by chance, more specifically, it indicates the probability that the one still find the current results even if the actual correlation coefficient is 0 (the null hypothesis). And considering our dataset is small, we use a significance level of 0.07 which slightly higher than conventional significance level of 0.05.

3.6 Regression Fitting

In order to make sure the variables we analyze meets the assumptions of Pearson Correlation mentioned above, and also to perform a better visualization of our results, we carried out regression analysis, after making scatter plots for all our Activity and Psychological, Emotional State parameters with sleep quality parameters. This was carried out using **seaborn** *scatter plot* function and **scikit-learn** library for regression analysis.

Before performing regression fitting, we also performed **Outlier Detection** and removal using **Z-score** for getting better fitting curves, since outliers were causing the curves to spread out and assume an unwanted shape that didn't deliver any information.

A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean. If Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 indicates a value that is one standard deviation from the mean.

$$Z = \frac{x - \mu}{\sigma} \tag{3}$$

The plots generated for the Psychological and Emotional State parameters with Sleep Quality Parameters, as well as plots for Activity data parameters with sleep quality parameters have been further discussed in Results section.

3.7 Machine Learning Model

After identifying the most relevant parameters which had the best associations and correlations with sleep quality, we extracted these for forming a feature set for our Machine Learning model. The parameters that we identified as having the best correlations with sleep quality are as follows:-

(1) Total Sleep Time (in minutes)
(2) Total Minutes in Bed
(3) STAI1 Data: represents the state anxiety mentioned above
(4) STAI2 Data: represents the trait anxiety mentioned above
(5) Daily Stress: stress level throughout the day
(6) Total Activity Time (in minutes)
(7) Heavy Activity Time: only counts heavy exercising time, in minutes.

We decided to perform a **Regression Analysis** using **XGBoost Regressor** model to analyze if the aforementioned parameters combined can help us predict the sleep quality of a person. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework. Since XGBoost gives one of the best performance among all prediction based machine learning models and doesn't require huge amounts of data to work with, we decided to utilize that for our analysis.

4 RESULTS

4.1 Distribution Analysis

For the plots below the black line corresponds to the Normal Distribution base curve and blue line depicts our parameter's actual distribution curve.

In Figure 1 and 2 below, we plotted the distributions of two main parameters, Trait and State Anxiety from our Questionnaire data corresponding to psychological and emotional states of users. Both Trait Anxiety and State Anxiety were found to follow a Normal Distribution over the users analysed. The distribution of state anxiety skewed left because of a participant reported 0 which indicates no anxiety caused by stressful event.

Figure 3 and 4 represents the Distributions of PSQI scores and Daily Stress scores from our Questionnaire data corresponding to psychological and emotional states of users. We can see that PSQI scores and Daily Stress scores both follow a Normal Distribution over the users analysed. These two distributions both skewed right, but only by a very small amount.

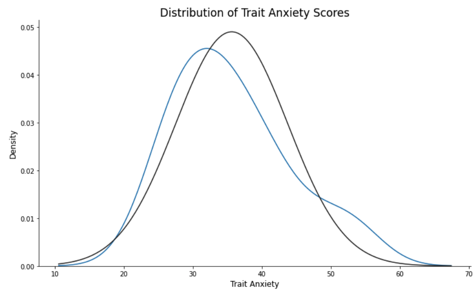

Fig. 1. Distribution of Trait Anxiety Scores

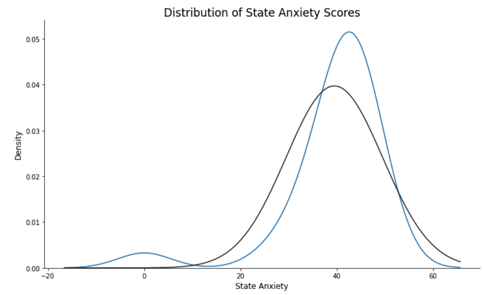

Fig. 2. Distribution of State Anxiety Scores

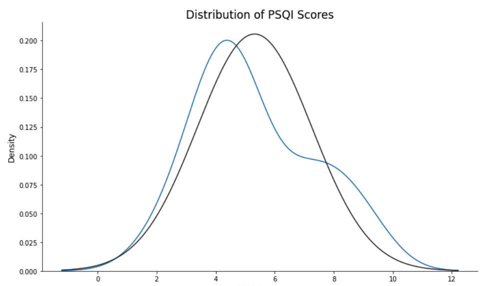

Fig. 3. Distribution of PSQI Scores

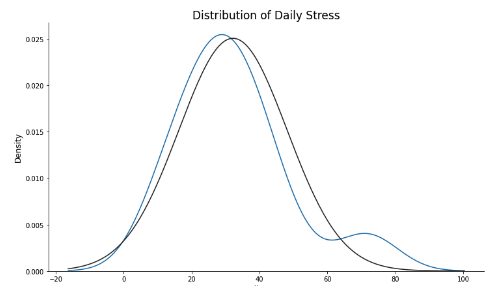

Fig. 4. Distribution of Daily Stress Scores

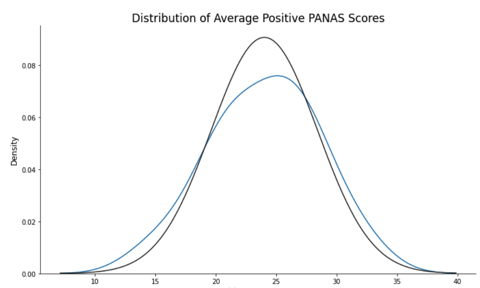

Fig. 5. Distribution of Average PANAS Scores over users

For Figure 5 above, we plotted the distribution of Average Positive PANAS scores another parameter from our Questionnaire data corresponding to psychological and emotional states of users. It was also found to follow Normal Distribution, and the skewness was minimal.

We then visualized the distribution of Exercise Time and Heartrate fluctuation using density plots. The fluctuation of heartrate is defined by the average difference between max and min heartrate over a period of time, which is usually their sleep time but different participants have different schedules. Results are shown in figure 6 and 7. And we can see they are roughly normally distributed (as expected) with minimal skewness, which suggests Pearson Correlation could give us good indication of their potential correlations.

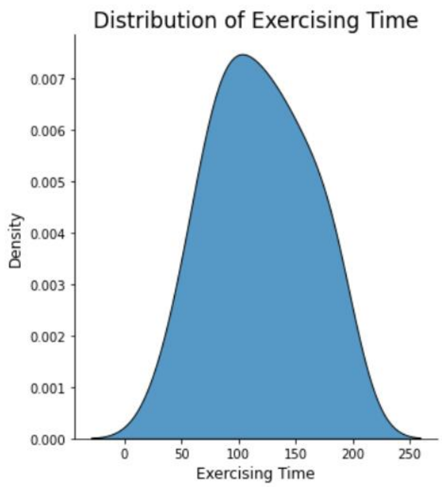

Fig. 6. Distribution of Exercise Time

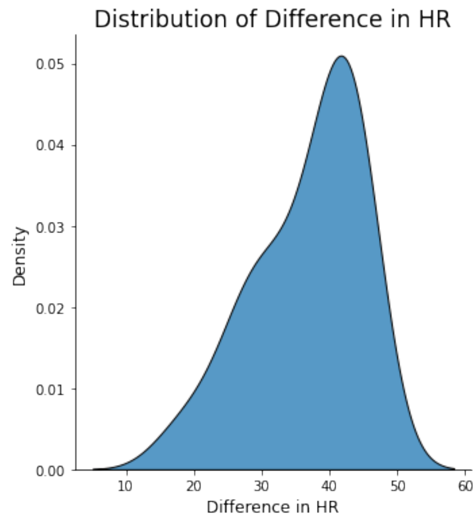

Fig. 7. Distribution of Heartrate fluctuation

4.2 Correlation Analysis

The scatter-plots obtained for Psychological and Emotional State parameters with Sleep Quality parameters, along with regression line fitting, are discussed below.

We first explored the correlation of five main Psychological and Emotional State parameters with two main Sleep Quality parameters. The Psychological and Emotional State parameters analyzed are:-

(1) **Trait Anxiety**: anxiety that shows up as part of one's personality, not just in stressful situations.
(2) **State Anxiety**: a natural human response when facing stressful situations.
(3) **PSQI scores**: measures sleep quality, lower PSQI indicates better sleep quality.
(4) **Daily Stress scores**: measures the level of stress, higher DSI indicates higher stress.
(5) **Average Positive PANAS Score**: measures negative or positive emotion, higher PANAS indicates higher level of (negative or positive) emotion.

Details about what these parameters signify have been discussed in section 3.1.

The Sleep parameters analyzed are:-

(1) Total Sleep Time (in minutes)
(2) Total Minutes in Bed

As can be observed from figure 8, 9, a positive correlation was discovered between State Anxiety scores and Total Sleep Time, and between State Anxiety scores and Total Minutes in Bed. The regression line fit along with a positive Pearson correlation coefficient supports our observations. A Pearson correlation coefficient of 0.5 was obtained for State Anxiety scores vs Total Sleep Time with a p-value significance of 0.034, implying that the correlation was significant and not observed by random chance. A similar trend was observed for State Anxiety score vs Total Minutes in Bed with Pearson correlation coefficient of 0.45 and p-value of 0.06, which again makes the correlation very significant. The observed results follows the logical reasoning, as State Anxiety is caused by a traumatic event and a user is likely to sleep more to get over such an event, which is why Sleep time will be more with more anxiety.

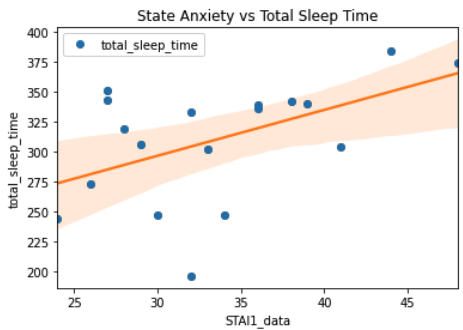

Fig. 8. State Anxiety Scores Vs Total Sleep Time

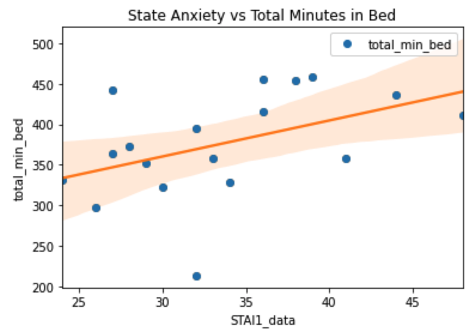

Fig. 9. State Anxiety Scores Vs Total Minutes in Bed

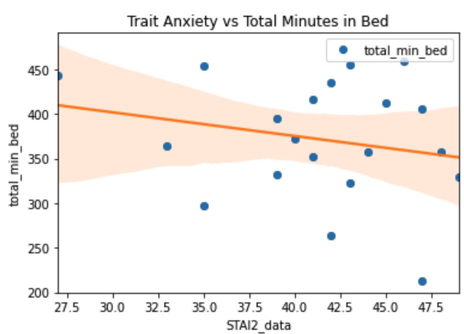

Fig. 10. Trait Anxiety Scores Vs Total Sleep Time

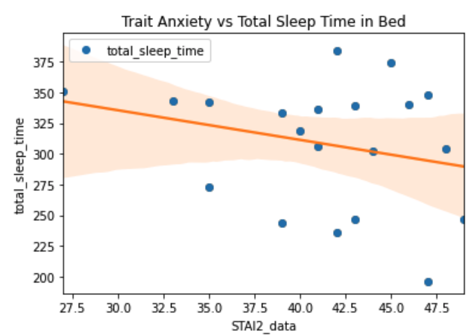

Fig. 11. Trait Anxiety Scores Vs Total Minutes in Bed

In comparison a negative correlation was discovered for Trait Anxiety scores with Total Sleep Time and Total Minutes in Bed (figure 10, 11). This was further confirmed by a negative regression fit and Pearson Correlation coefficient of -0.26 for Trait Anxiety Scores Vs Total Sleep Time with a p-value of 0.267, which isn't as significant as our correlations for State Anxiety. A Pearson correlation coefficient of -0.22 with p-value of 0.351 was found for Trait Anxiety Vs Total Minutes in Bed. Since Trait Anxiety relates to a person having a trait for being anxious in general, people with more anxious personality will have trouble falling asleep and hence less Total Sleep time or Total Minutes in bed, which aligns with our results.

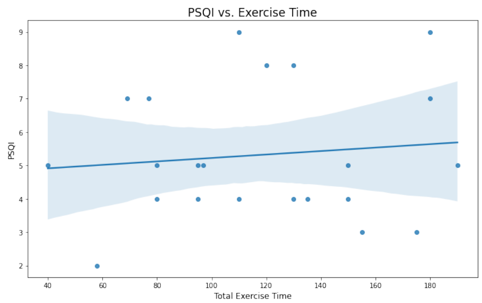

Fig. 12. Exercise Time vs. PSQI

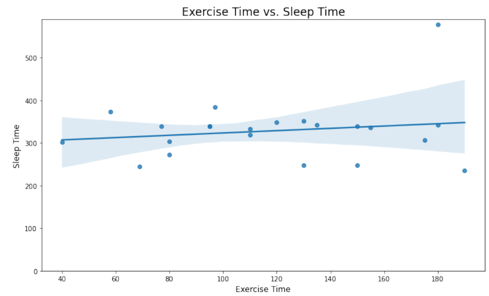

Fig. 13. Exercise Time vs. Sleep Time

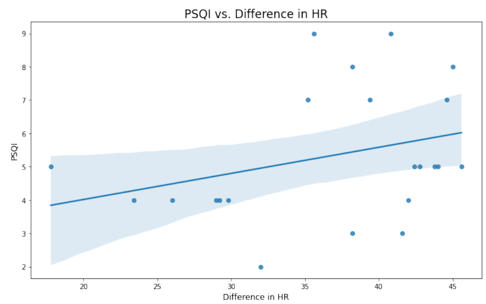

Fig. 14. Average Heart rate fluctuation vs. PSQI

We also analyzed how exercise time and heartrate are correlated with sleep time and sleep quality (measured by PSQI). The resulting regression plots are shown in figure 12, 13, and 14. We can verify the assumptions for Pearson Correlation were met based on these plots. We were not able to find any significant correlation between total exercise time and sleep time or quality, the p-values were 0.5 and 0.57, respectively. On the other hand, the p-value of the Pearson Correlation between heartrate fluctuation and PSQI is 0.157, while it's higher than our desired significance level (0.07), it's way more significant than other correlations, and worth mentioning considering the small size of our dataset. The correlation coefficient between these two variables is 0.31, which means higher fluctuation of heartrate (during bedtime) is associated with higher PSQI (worse sleep quality). This is roughly in line with a lot of clinical research, including a recent report from Havard which associates heartrate fluctuation with sleep phases, and suggests heartrate could fluctuate a lot during Rapid Eye Movement (REM) phase [5]. It possible that the high fluctuation of heartrate during sleep in our dataset was associated to REM phases, which could also explain the higher PSQI correlated, but we were unable to verify this association since there is no recorded indicator of REM phase in our dataset.

4.3 XG Boost Model

We built a XGBoost Regressor for modeling the sleep quality (measured by PSQI) with the features and parameters mentioned in section 3.7, the resulting model achieved an RMSE of 2.6 when excluding Total Activity and Heavy Activity from our feature set. The RMSE lowered down to 2.5 after including Total Activity and Heavy Activity in our dataset. The range of label PSQI is 0 to 21 in real world, and 0 to 9 in our dataset. Both these models were an improvement as opposed to an RMSE of 4.0 obtained using a naive baseline model which always predicted the average.

We also worked on finding the most important features that affect sleep quality prediction the most. This was carried out by calculating **SHAP** values on the top of XGBoost Regressor model that we built. In order to be able to visualize the importance values of different features, we plotted the SHAP value plots for different features in the form of Bar plot and Summary plot. The resulting generated plots can be found below.

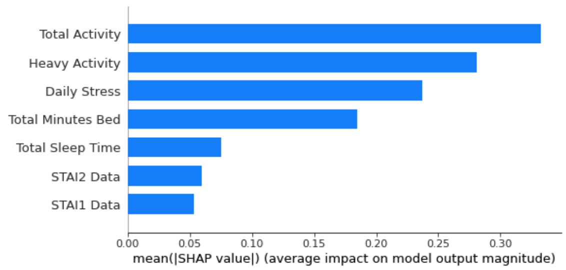

Fig. 15. Feature Importance calculated using mean SHAP

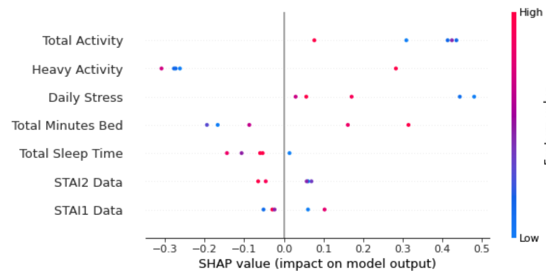

Fig. 16. Feature Importance plot using SHAP values

5 DISCUSSION

While our analysis revealed (insert meaningful findings), a lot of findings were insignificant because of the small dataset. The study population also lacks diversity which hindered demographic related analysis. (insert limitations). Furthermore, a small dataset made several analysis models not applicable, for example, we were not able to use two way ANOVA to analyze the difference in sleep time and quality by both level of stress and trait anxiety. Despite the predictive modeling we constructed brought improvements over the naive baseline model, it was not ideal, and possibly not meaningful because of the above limitations. Had we gained access to a more diverse and larger dataset, we could have produced more promising results. If presented a bigger and more diversified dataset collected in non-clinical settings, we would like to compare our results to existing research conducted in clinical settings, and experiment with different machine learning models to predict sleep duration and quality.

What we have learned from the **process** of this project is that mobile health dataset can be extremely dirty. We spent a large portion of our time to clean and impute the dataset even with relatively clear documentation provided. And it's really hard to find a desired dataset that is big, diversified, and clean. Which is the exact reason why mobile sensing is such a huge hotspot for medical research right now: mobile devices are able to collect valuable data in large scale, while being cheaper and faster than conventional approaches usually carried out in clinical settings.

6 CONCLUSION

In this project, we investigated how one's sleep duration and sleep quality are associated with their daily activities and psychological behaviors. We conducted distribution and correlation analysis on mobile health sensing dataset that consists of sleep, activity, and psychological data collected from 22 participants through wearable devices. Our analysis revealed that total sleep time had significant correlation with state anxiety (p-value = 0.034), and the positive correlation coefficient (0.5), which suggests that higher state anxiety (anxiety caused by stressful event) is associated with longer sleep time. Sleep duration was also positively correlated with sleep quality, daily exercising time, and positive PANAS, and negatively correlated with trait anxiety and daily stress level, but these correlations were found to be statistically insignificant. On the other hand, another indicator of sleep quality, PSQI, was positively correlated with heartrate fluctuations and total daily exercising time, which suggests higher heartrate fluctuation and more daily exercising time might be associated with worse sleep quality (higher PSQI indicates lower sleep quality), but the correlation between exercising time and PSQI was likely caused by chance because of its extremely high p-value (0.57).

Overall, despite the fact that a lot of correlations we obtained were insignificant, our discoveries are generally in line with a lot of existing research that has been done in this area in clinical settings. The analysis process definitely provided us valuable experience in working with mobile health datasets. Almost all analysis we carried out is easily scalable, which means we can easily deploy those on additional datasets after some cleaning. If we have access to more data collected from a more diverse population, we can quickly conduct more in-depth analysis on it using the techniques we learned, along with more advanced models, and potentially discover statistically significant correlations.

7 REFERENCES

[1] Statistics Solutions. (2021, June 9). Pearson's Correlation Coefficient. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/pearsons-correlation-coefficient/

[2] Arora, A., Chakraborty, P., Bhatia, M.P.S. Analysis of Data from Wearable Sensors for Sleep Quality Estimation and Prediction Using Deep Learning . Arab J Sci Eng 45, 10793–10812 (2020). https://doi.org/10.1007/s13369-020-04877-w

[3] Bai, Yin, Xu, Bin, Ma, Yuanchao, Sun, Guodong, Zhao, Yu. (2012). Will You Have a Good Sleep Tonight? Sleep Quality Prediction with Mobile Phone. BODYNETS 2012 - 7th International Conference on Body Area Networks. 124-130.10.4108/icst.bodynets.2012.250091.

[4] Feng, T., Booth, B.M., Baldwin-Rodríguez, B. et al. A multimodal analysis of physical activity, sleep, and work shift in nurses with wearable sensor data. Sci Rep 11, 8693 (2021). https://doi.org/10.1038/s41598-021-87029-w

[5] Corliss, J. (2021, January 29). How does sleep affect your heart rate? Harvard Health. https://www.health.harvard.edu/blog/how-does-sleep-affect-your-heart-rate-2021012921846