# Related work

The prime objective of sleep quality evaluation module is to allow real time sleep quality evaluation based on sensors’ data. This section reviews the related methods in sleep quality evaluation and their deliverables to our final implementation.

There are three methods which are widely used in clinical sleep quality assessment: Pittsburgh Sleep Quality Index(PSQI), Polysomnography and Actigraphy [4]. First, PSQI is a questionnaire based assessment focusing on subjective feedback on medium to long term sleep quality [5]. Based on subjects’ answers, it generates a score that is inversely proportional to sleep quality. Due to the limitation of long assessment interval, PSQI is not suitable for direct implementation into this module. However, this method is used to evaluate the general performance of our system on sleep quality enhancement and the result will be discussed in the Evaluation section.

Secondly, Polysomnography analyses sleep quality by using electroencephalograms(EEG), electro-oculograms(EOG) mad electromyograms of the mentalis and libs [7]. It reflects the precise proportion of each sleep stages during the 24-hours assessment interval and hence provides the most accurate sleep quality evaluation. Despite of its accuracy, it has a few critical disadvantages that prevent its application into our system. First, the sensors required are extremely intrusive to users and all signals required intensive processing algorithms to analyze. Secondly, the data collection process for complete analysis required at least 12 hours. Therefore, this method is not applicable in this case.

Finally, Actigraphy monitors the sleep quality by estimating ratio between ‘sleep’ and ‘awake’ patterns. Conventionally, ‘sleep’ and ‘awake’ patterns are defined as minor and intense body movements during sleep by using motion sensing device Actometer. The principle behind is that body muscle is completely paralyzed during deep sleep stage but not in others. By extending this principle further, redefining sleep-awake patterns and combining more sensors, several Actigraphy sleep quality evaluation methods have been invented. Mobile application such as iSleep [8], Sleep as android [9] and Toss ‘N’ Turn [10] uses mobile phone as the main sensor to collect data reflecting the sleeping noise, body movement, background light intensity. They determine the sleep-awake ratio in each night to evaluate sleep quality with mean accuracy over 80%. Furthermore, research by Ya-Ti Peng et.al has also shown that introducing heart rate data into normal motion tracking can improve sleep-awake pattern classification [11]. These applications act as a proof of concept for actigraphy validity and correlation between sleep quality and biometrics including body movement, heart rate, etc. Moreover, they demonstrate the method’s compatibility with mobile phone. As a result, we decided to utilize actigraphy as our detection principle. Nevertheless, we leverage machine learning approach for the implementation to provide short term sleep quality evaluation continuously.

# Implementation

### Hardware

The first design choice is to determine the right sensor to collect sleep quality related biometrics while accomplishing the heating system feedback. First, body temperature is found to decrease when sleep stage transits from NREM to REM in the third cycle [15]. Secondly, a trend towards a decrease of the low- to high-frequency ratio (LF/HF) derived from the power spectral analysis of Heart Rate Variability is expected to associate with the transition from wakefulness to NREM sleep. Meanwhile, the LF/HF ratio increases during the rapid-eye-movement (REM) sleep[12],[13]. Nevertheless, James W. Mold et.al have concluded that night sweating is associated with several sleep symptoms[16]. Most importantly, body movement monitoring data should be captured as conventional actigraphy method is proved to reach 90% agreement with traditional PSG for nocturnal sleep period [14]. Hence, Microsoft Band 2 is chosen as our main biometrics sensor as it contains skin temperature sensor, heart rate sensor, galvanic skin response sensor and accelerometer. Another advantage of using Microsoft Band 2 is that it supports heart rate sampling rate at 1 Hz, which is more than twice larger than the highest frequency that heart rate can normally achieve. This ensure the capability to calculate heart rate variability without using interpolation techniques.

### Features Extraction Analysis

Given the sensors provided by Microsoft Band 2, the features generated covers 3 modalities which are summarized in Table 1. All sensor data is in temporal manner and hence a time interval for feature generation is required. Our final windowing size is 10 minutes of data sampled at 1Hz. This decision provides the best compromise between mobile phone hardware capabilities and feature validity. Typical range of sleep stage transition time varies from 7 to 45 minutes and thus it would be optimal that to generate feature every 45 minutes. However, the window width is constrained by the memory available on mobile phone. Storing 7 sensors reading at 1 Hz for 45 minutes exceeds the available memory and it also causes long delayed in data sending to server. Therefore, a smaller 10 minutes’ interval is chosen as it is proved to be effective in Toss ‘N’ Turn [10].

The feature for each modality are chosen by referencing existing signal processing techniques. Previous research has shown that the heat exposure to extreme heat and humid environment can affect sleep quality and mean skin temperature can capture this exposure effect on body core temperature<http://europepmc.org/abstract/med/10505822>. Moreover, temperature tends to decrease at night time sleep onset but increase when awake [15] and standard deviation can be used to capture the fluctuation of body temperature within each time window. Therefore, mean and standard deviation of skin temperature were extracted.

Accelerometer reading is used to reflect users’ arm movement.

Instead of using interval average acceleration over time suggested in [18], we extract the data in terms of mean and standard deviation of squared amplitude shown in (1).

The motivation behind is to capture movement information every 1 second. The suggested method is effective when the time interval is around 100 second, which is much smaller than the chosen 10 minutes’ interval. If similar approach is used, the excessive smoothing on 10 minutes of data can remove acute magnitude fluctuations caused by sudden arm movements. Therefore, mean and standard deviation on mean squared amplitude are used to extract overall movement intensity and frequency.

Heart rate variability is proved to be higher in rapid eye movement sleep stage than others < Zemaityte D, Varoneckas G, Sokolov E. Heart rhythm control during sleep Psychophysiology 1984;21:279e89> while mean and standard deviation on RR intervals are shown to be adequate measures of HRV during sleep stages transition < Heart rate variability, sleep and sleep disorders>. Additionally, instantaneous heart rate data is also analyzed in similar manner. Apart from using mean and standard deviation, kurtosis is also used to analyze the extremes. From < "The meaning of kurtosis: Darlington reexamined". *The American Statistician>*, kurtosis is a measure on outlier’s population out of the total samples. The higher the kurtosis is, the sample is less distributed around the statistical median. Hence, it can be a useful tool to quantify the chronic changes in RR interval and heart rate that is removed under mean and standard deviation. Lastly, we have excluded the introduced Galvanic Skin Response reading as input features because of three reasons. First, the sweat production is proven to be independent with galvanic skin response amplitude < INDEPENDENCE OF GALVANIC SKIN RESPONSE AMPLITUDE AND SWEAT PRODTJCTION\* >. Secondly, the galvanic skin response(GSR) sensor required locking system to provide accurate data and the acquisition-locking cycle is difficult to be controlled as it is done by Microsoft Band 2 API. Thirdly, the variance of GSR across different sleep quality is found to be nearly zero from data collected. This demonstrates its insignificance in sleep quality evaluation and hence GSR is excluded in the final feature set.

|  |  |  |
| --- | --- | --- |
| Modality | Sensor Data | Features |
| Temperature | Temperature readings in Celsius | Mean, standard deviation(STD) |
| Movement | 3-axes Accelerometer readings | Mean, STD of root mean squared amplitude |
| Heart rate | Optical Heart Rate readings and RR interval | Mean, STD, Kurtosis |

Table 1: Initial Feature Set

### Model Selection

We have defined our sleep quality evaluation problem as a binary classification problem after learning from various sleep-awake pattern classifiers < Comparison of Sleep-Wake Classification using Electroencephalogram and Wrist-worn Multi-modal Sensor Data >< https://infoscience.epfl.ch/record/135627/files/EPFL\_TH4391.pdf >. Given the final feature set, it is necessary to select an optimal classifier based on the obtained data. Matlab classification learner is used to perform cross-model benchmarking. From previous testing result as shown in Table 2, a user specific classifier performs better than a unified classifier and hence our final model selection process only focus on optimizing model which is trained on a specific subject. To prevent loss of generality, we carefully selected a subject that has the most uniformly distributed sleep quality. The result is shown in Table 3.

|  |  |  |
| --- | --- | --- |
| Models | All Data Accuracies | Personalized Data Accuracies |
| Best Tree | 78.1% | 90.5% |
| Logistic Regression | 78.1% | 85.7% |
| Best SVM | 75% | 95.2% |
| Best KNN | 84.4% | 90.5% |
| Best Random Forest | 75% | 95.2% |
| Boosted Tree | 81.3% | 85.7% |
| Subspace Discriminant | 60% | 100% |

Table 2: Feature Analysis Results

|  |  |
| --- | --- |
| Models | Personalized Data Accuracies |
| Decision Tree | 95.8% |
| Logistic Regression | 85.7% |
| Best SVM | 96.4% |
| Best Random Forest | 96.3% |
| Boosted Tree | 95.6% |

Table 3: Model Selection Results

From Table 3, it can be observed that either support vector machine or random forest should be chosen as the implementation model. Apart from model accuracy, training and testing time are also considered as model selection criteria. Thus, random forest is chosen to be implemented due to its efficient training and testing principle. Another drawback from this testing result is that it is based on feature set generated from one specific subject over 1 week due to limited resources such as time and available Microsoft Band 2. Moreover, the limited subject diversity also reduced the available sample size. This is because most of data collected are from good sleepers, which causes imbalanced sleep quality distribution. Nevertheless, the sleep quality labelling on training data is based on assumption that overall night sleep quality can be interpolated into individual interval sleep quality. This assumption should be abandoned if more time and resource are given to perform clinical testing on overnight sleep quality monitoring with the device.

### Machine Learning Model Server Deployment

To integrate machine learning module into the system smoothly, an implementation of random forest classifier is done on server side in order to provide online estimation upon requests from Sleepify mobile clients.

As we have chosen Python Django for the server development platform, Scikit-Learn [20] is used for our machine learning implementation. Model persistence on server is achieved by binary serialization and recovering using Pickle, a Python object serialization tool. From the preliminary study results, it is necessary that each user requires a specific classification model which is trained by personalized data. Therefore, these binary models files are linked to user entries in database to allow user specific mapping.

To communicate with the mobile application, a RESTFUL API is created to offer three functions; sensor data storage, machine learning model retraining and prediction. When the mobile application sends a packet of new sensor data to the server, this interface will first extract the date, time and user details of the packet and store the data into database under specific user accounts. After that, the packet will be preprocessed into a feature sample. Then, binary model files under the specific user is recovered and prediction method will be called with the generated feature vector. Finally, the prediction outcome is sent back to mobile application in the form of JSON. Moreover, continuous learning is supported by the use of existing models provided by Scikit-Learn. A user specific data tracker is implemented to monitor the accumulated count of untrained data packet. When it exceeds a threshold, the model will be recovered from the binary file and perform online learning. The block diagram for this machine learning server infrastructure is shown in Figure 2.



Figure 2: Server Architecture for ML model

### Testing

To evaluate the model implementation, an offline model testing is done to demonstrate the practical performance of random forest model provided by Scikit learn. To compare with testing performed by Matlab classification learner, identical hyperparameters such as tree depth and tree numbers are used. Moreover, a similar testing and training set splitting ratio is used which are 0.2 and 0.8 respectively. The result shows that implemented classifier can achieve 90% accuracy, which is similar to Matlab implementation.

# References

[1] H. Schulz, ‘Rethinking Sleep Analysis’, *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.*, vol. 4, no. 2, pp. 99–103, Apr. 2008.

[2] K. N. Mims and D. Kirsch, ‘Sleep and Stroke’, *Sleep Med. Clin.*, vol. 11, no. 1, pp. 39–51, Mar. 2016.

[3] S. S. Gilbert, C. J. van den Heuvel, S. A. Ferguson, and D. Dawson, ‘Thermoregulation as a sleep signalling system’, *Sleep Med. Rev.*, vol. 8, no. 2, pp. 81–93, Apr. 2004.

[4] S. M. G. P. Togeiro and A. K. Smith, ‘[Diagnostics methods for sleep disorders]’, *Rev. Bras. Psiquiatr. Sao Paulo Braz. 1999*, vol. 27 Suppl 1, pp. 8–15, May 2005.

[5] D. J. Buysse, C. F. Reynolds III, T. H. Monk, S. R. Berman, and D. J. Kupfer, ‘The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research’, *Psychiatry Res.*, vol. 28, no. 2, pp. 193–213, May 1989.

[6] T. Blackwell *et al.*, ‘Comparison of Sleep Parameters from Actigraphy and Polysomnography in Older Women: The SOF Study’, *Sleep*, vol. 31, no. 2, pp. 283–291, Feb. 2008.

[7] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. L. Marcus, and B. V. Vaughn, ‘The AASM manual for the scoring of sleep and associated events’, *Rules Terminol. Tech. Specif. Darien Ill. Am. Acad. Sleep Med.*, 2012.

[8] T. Hao, G. Xing, and G. Zhou, ‘iSleep: unobtrusive sleep quality monitoring using smartphones’, in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, 2013, p. 4.

[9] ‘Sleep as Android’. .

[10] J.-K. Min, A. Doryab, J. Wiese, S. Amini, J. Zimmerman, and J. I. Hong, ‘Toss “n” turn: smartphone as sleep and sleep quality detector’, 2014, pp. 477–486.

[11] null Ya-Ti Peng, null Ching-Yung Lin, null Ming-Ting Sun, and C. A. Landis, ‘Multimodality sensor system for long-term sleep quality monitoring’, *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 3, pp. 217–227, Sep. 2007.

[12] M. Malik, ‘Heart rate variability’, *Ann. Noninvasive Electrocardiol.*, vol. 1, no. 2, pp. 151–181, 1996.

[13] E. Vanoli, P. B. Adamson, null Ba-Lin, G. D. Pinna, R. Lazzara, and W. C. Orr, ‘Heart rate variability during specific sleep stages. A comparison of healthy subjects with patients after myocardial infarction’, *Circulation*, vol. 91, no. 7, pp. 1918–1922, Apr. 1995.

[14] A. Sadeh, K. M. Sharkey, and M. A. Carskadon, ‘Activity-based sleep-wake identification: an empirical test of methodological issues’, *Sleep*, vol. 17, no. 3, pp. 201–207, Apr. 1994.

[15] H. J. Burgess, A. L. Holmes, and D. Dawson, ‘The relationship between slow-wave activity, body temperature, and cardiac activity during nighttime sleep’, *Sleep*, vol. 24, no. 3, pp. 343–349, May 2001.

[16] J. W. Mold, J. H. Woolley, and Z. Nagykaldi, ‘Associations between night sweats and other sleep disturbances: An OKPRN study’, *Ann. Fam. Med.*, vol. 4, no. 5, pp. 423–426, Oct. 2006.

[17] B. Pomeranz *et al.*, ‘Assessment of autonomic function in humans by heart rate spectral analysis’, *Am. J. Physiol.*, vol. 248, no. 1 Pt 2, pp. H151-153, Jan. 1985.

[18] M. J. B. Rogers, K. Hrovat, K. McPherson, M. E. Moskowitz, and T. Reckart, ‘Accelerometer Data Analysis and Presentation Techniques’, Sep. 1997.

[19] I. Guyon and A. Elisseeff, ‘An Introduction to Variable and Feature Selection’, *J Mach Learn Res*, vol. 3, pp. 1157–1182, Mar. 2003.

[20] ‘scikit-learn: machine learning in Python — scikit-learn 0.18.1 documentation’. [Online]. Available: http://scikit-learn.org/stable/. [Accessed: 22-Feb-2017].