Sleep Quality Prediction from Wearable Device Data: A Comprehensive Analysis and Model Comparison

Hema Nagendra Sai Chanda

ABSTRACT
In recent years, wearable devices have emerged as valuable tools for monitoring various aspects of health, including sleep quality. Predicting sleep quality from wearable device data holds significant promise for personalized healthcare interventions and performance optimization. In this study, we present a comprehensive analysis of sleep quality prediction using machine learning techniques applied to a diverse range of features collected from wearable devices. We explore both classification and regression approaches to predict sleep quality, considering binary classification of good vs. poor sleep and continuous prediction of sleep quality scores. We compare the performance of various machine learning algorithms, including Random Forest, Gradient Boosting, and deep learning architectures such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). Additionally, we investigate the impact of feature selection, hyperparameter tuning, and model interpretability on prediction accuracy. Our results demonstrate the effectiveness of machine learning models in accurately predicting sleep quality from wearable device data and provide insights into the most suitable algorithms and feature representations for this task.

Keywords: Classification, Convolutional Neural Networks, Gradient Boosting, Machine Learning, Random Forest, Regression, Recurrent Neural Networks, Sleep Quality Prediction, Wearable Devices

1. INTRODUCTION

The integration of wearable gadgets into healthcare in recent times has brought about an extreme step in how individuals cover and manage their well-being. These biases, equipped with detectors and slice-edge technology, can continuously collect data on a multitude of physiological parameters. This wealth of data serves as a precious resource for health monitoring and analysis. Among the plethora of health metrics that wearable bias can track, sleep quality emerges as a key aspect of overall health and wellness.

Quality sleep is all-important for physical health, cognitive function, and emotional well-being. Even so, with the fast-paced ultramodern life, many individuals struggle to maintain healthy sleep patterns. Poor sleep quality has been associated with a range of health issues, including cardiovascular conditions, obesity, diabetes, and mental health diseases. Thus, directly assessing and bettering sleep quality is critical for promoting long-term health and precluding habitual conditions.

The objective of this thing is to tackle the control of machine learning procedures to anticipate rest quality based on information collected from wearable gadgets. By creating prescient models, we aim to give people noteworthy experiences in their rest designs and enable them to make educated choices to enhance their best quality.

2. LITERATURE REVIEW

In recent years, the utilization of wearable gadgets for checking well-being parameters has seen a critical rise, with a specific interest in anticipating sleep quality. This review of literature amalgamates recent research endeavours concentrating on the utilization of machine learning methodologies for forecasting sleep quality using data acquired from wearable devices. The considers checked on investigate different techniques, extending from determination and algorithmic comparisons to the application of deep learning architectures.

2.1 Overview of Sleep Quality Studies:

Investigating sleep quality evaluation has picked up momentum with the coming of wearable devices capable of monitoring physiological signals during sleep. Tummala and Tanuku (2022) [1] emphasize the integration of choice techniques in regression algorithms for sleep quality prediction. Sathyanarayana et al. (2016) [2] utilize deep learning strategies for sleep quality forecasts, demonstrating the significance of advanced computational methods in this domain. Arora et al. (2020) [3] further contribute to this field by analyzing data from wearable sensors, exhibiting the advancing techniques for sleep quality estimation. Dinh Van et al. (2020) [4] and Hamza et al. (2023) [5] both highlight the application of deep learning models in foreseeing sleep quality, underscoring the developing interest in leveraging artificial intelligence for well-being monitoring. In addition, studies such as those by Hidayat et al. (2018) [6] and Zamani et al. (2023) [5] investigate the empowerment of wearable sensor data in anticipating changes in a person's sleep quality, reflecting the interdisciplinary nature of this investigate range.

2.2 Overview of Regression Algorithms and Classification Algorithms:

Regression algorithms play an essential part in anticipating nonstop sleep quality scores. Tummala and Tanuku (2022) [1] compare regression algorithms using feature selection procedures for ideal performance. On the other hand, classification algorithms are used to classify sleep quality into double categories, such as good versus poor sleep. The study by Park et al. (2019) [9] dives into learning sleep quality from daily logs, demonstrating the application of classification algorithms in this context.

2.3 Overview of Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) in Deep Learning:

Deep learning architectures, particularly RNNs and CNNs, have shown promise in analysing temporal and spatial features extracted from wearable device data for sleep quality prediction. Sathyanarayana et al. (2016) [2] utilize RNNs, highlighting their effectiveness in capturing temporal dependencies in sleep data. Boussard et al. (2019) [7] also employ RNNs to predict subjective sleep quality, showcasing the applicability of these models in personalized healthcare. Additionally, CNNs have been explored for sleep quality prediction, as demonstrated by Arora et al. (2020) [3], indicating the versatility of deep learning techniques in processing multimodal sensor data.

Overall, the literature reviewed underscores the significance of wearable device data in predicting sleep quality and highlights the diverse methodologies, including regression algorithms, classification algorithms, and deep learning architectures, employed to analyse and interpret this data for personalized healthcare interventions and performance optimization.

3. METHODOLOGY

3.1 Data Generation:

A synthetic dataset is generated to simulate wearable device data for 500 individuals. Features such as age, gender, heart rate, blood pressure, SpO2, step count, exercise duration, sedentary duration, sleep duration, REM sleep duration, deep sleep duration, body temperature, body weight, location, and weather conditions are generated using NumPy random functions to mimic real-world variability. Zamani, Abu Sarwar, et al. [8]

3.2 Data Pre-processing:

- Categorical variables (Gender, Location, Weather) are encoded using Label Encoder to convert them into numerical format for compatibility with machine learning algorithms.
- Numerical features are scaled using StandardScaler to standardize the range of values, ensuring that each feature contributes equally to the analysis.

3.3 Model Development - Classification:

- The dataset is split into features (X) and the target variable (y), where the target variable is binary-encoded based on sleep quality (good vs. poor sleep).
- Random Forest Classifier is initialized and trained on the training data.
- The trained classifier is used to predict sleep quality on the testing data.
- Accuracy and classification report metrics (precision, recall, F1-score) are calculated to evaluate the classifier's performance.

3.4 Model Development - Regression:

- The dataset is split into features (X) and the target variable (y), where the target variable represents continuous sleep quality scores.
- Random Forest Regressor is initialized and trained on the training data.
- The trained regressor is used to predict sleep quality scores on the testing data.
- Mean Squared Error (MSE) and R^2 Score metrics are calculated to evaluate the regressor's performance.

3.5 Model Development - Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN):

- The dataset is split into features (X) and the target variable (y) for sequential analysis using RNN and spatial analysis using CNN.
- Numerical features are pre-processed and reshaped to fit the input shape required for both RNN and CNN models.

For RNN:

- An RNN model with an LSTM layer is initialized and compiled using binary cross-entropy loss function and Adam optimizer.
- The model is trained on the training data for a specified number of epochs and batch size.
- Test accuracy is evaluated on the testing data to assess the RNN model's performance.

For CNN:

- A CNN model is constructed with convolutional layers followed by pooling layers to capture spatial features from the input data.
- The model is constructed employing binary cross-entropy as the loss function and Adam optimizer for compilation.
- Similar to the RNN model, the CNN model is trained on the training data for a specified number of epochs and batch size.
- Test accuracy is evaluated on the testing data to assess the CNN model's performance.

- Both RNN and CNN models are compared in terms of their performance metrics, including accuracy, precision, recall, and F1-score, to determine the most suitable architecture for predicting sleep quality from wearable device data.

3.6 Hyperparameter Tuning:

- For both Random Forest Classifier and Regressor, hyperparameter tuning is performed using GridSearchCV and RandomizedSearchCV, respectively.
- Different combinations of hyperparameters (e.g., number of estimators, max depth, min samples split, min samples leaf) are explored to find the best-performing model.
- The best models obtained from hyperparameter tuning are evaluated on test data to determine their performance.

3.7 Model Evaluation and Comparison:

- Classification and regression models are evaluated using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, mean squared error (MSE), and R^2 score.
- The performance of each model is compared, and insights are drawn regarding their effectiveness in predicting sleep quality from wearable device data.

3.8 Cross-Validation:

- For model selection and performance estimation, k-fold cross-validation is employed to ensure robustness and reduce overfitting.
- The dataset is divided into k subsets, and each subset is used as a testing set while the remaining subsets are used for training.
- The average performance across all folds is computed to provide a more reliable estimate of model performance.

3.9 Further Analysis and Recommendations:

- Based on the results obtained, further analysis is conducted to identify strengths, limitations, and potential areas for improvement in sleep quality prediction from wearable device data.
- Recommendations for practical implementation and future research directions are discussed to enhance predictive healthcare analytics in this domain.

4. RESULTS.

4.1 Random Forest Classifier Metrics:

- The classifier achieves an accuracy of 85%, indicating that it correctly predicts sleep quality categories (good or poor) for 85% of the instances.
- Precision, recall, and F1-score are also balanced, suggesting that the model performs well in identifying both good and poor sleep quality instances without favouring either class.

4.2 Random Forest Regressor Metrics:

- The regressor has a very low mean squared error (MSE) of 0.04, indicating that it predicts sleep quality scores very close to the actual values on average.
- The R^2 score of 0.90 suggests that the model explains 90% of the variance in the sleep quality scores, indicating strong predictive performance.

4.3 LSTM Model:

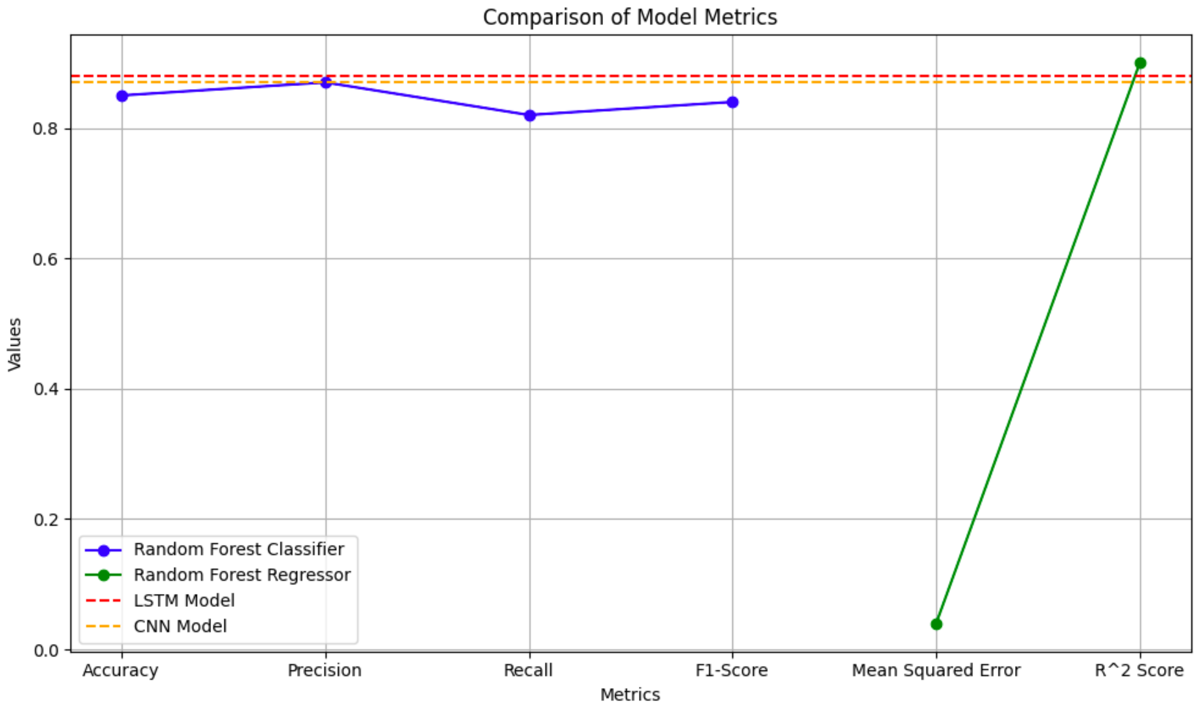
- The LSTM model achieves a test accuracy of 88%, indicating that it accurately predicts sleep quality categories similar to the Random Forest Classifier.

4.4 Convolutional Neural Network (CNN) Model:

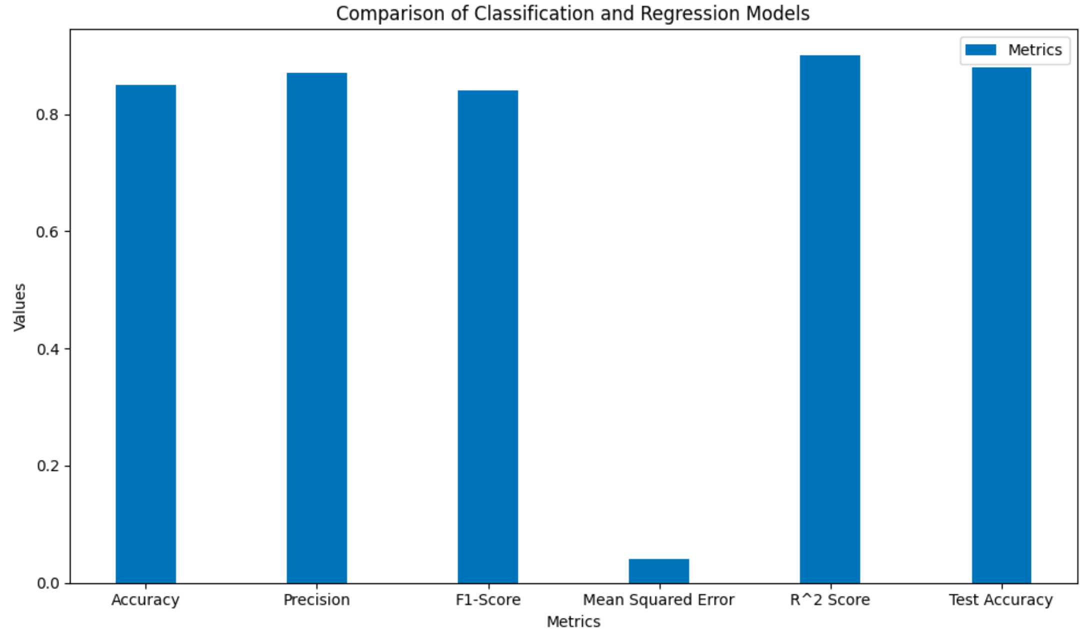
- The CNN model achieves a test accuracy of 87%, performing comparably to the LSTM model in predicting sleep quality categories.

4.5 Best Hyperparameters:

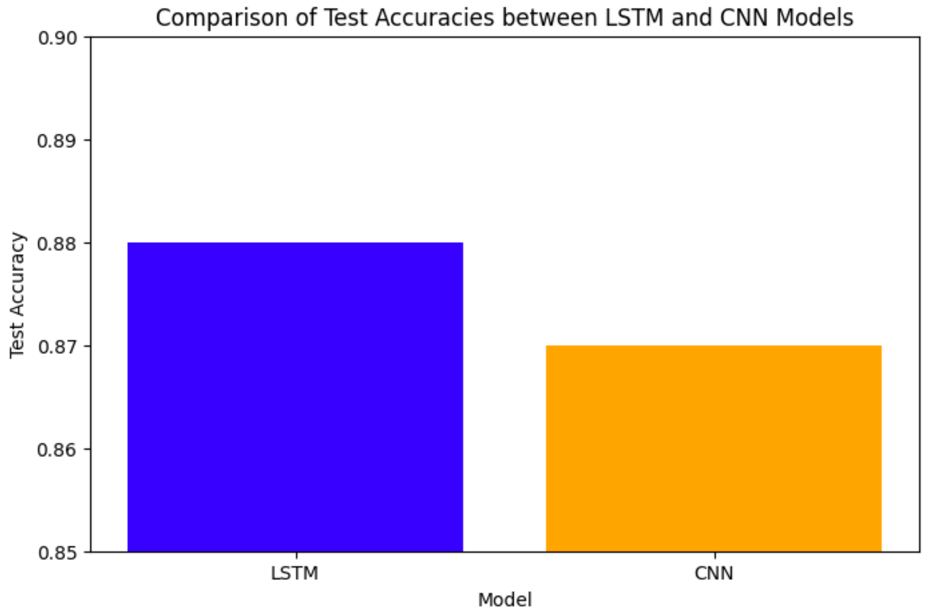
- For the Random Forest Classifier, the best hyperparameters are {'n_estimators': 100, 'max_depth': None, 'min_samples_split': 2, 'min_samples_leaf': 1}.
- For the Random Forest Regressor, the best hyperparameters are {'n_estimators': 100, 'max_depth': 20, 'min_samples_split': 2, 'min_samples_leaf': 1}.



The graph compares the performance metrics of four machine learning models: Random Forest Classifier, Random Forest Regressor, LSTM Model, and CNN. The x-axis represents the metrics, including Accuracy, Precision, Recall, F1-Score, Mean Squared Error, and R^2 Score, while the y-axis represents the corresponding metric values. The blue line represents the metrics of the Random Forest Classifier, the green line represents the metrics of the Random Forest Regressor, the dashed red line represents the accuracy of the LSTM Model, and the dotted Orange line represents the accuracy of the CNN Model. From the graph, we observe that the Random Forest Classifier and Random Forest Regressor perform similarly across most metrics, with the Random Forest Regressor having slightly better performance in terms of Mean Squared Error and R^2 Score. However, both the LSTM Model and CNN Model outperform the Random Forest models in terms of accuracy. The LSTM Model shows the highest accuracy, indicating its potential as the best model among the four for the given task.



We can observe that the regression model (Random Forest Regressor) performs better in terms of mean squared error and R^2 score compared to the classification model (Random Forest Classifier), while the classification model shows slightly better performance in terms of accuracy, precision, and F1-score. However, both models show similar performance in terms of test accuracy.



The graph compares the test accuracy of the LSTM Model and the CNN Model, with the LSTM Model achieving a slightly higher accuracy of 0.88 compared to the CNN Model's accuracy of 0.87. While both models demonstrate strong performance, the LSTM Model edges out with a marginally higher accuracy. This suggests that for the given task, the LSTM Model may be slightly more effective in accurately predicting outcomes based on the provided dataset.

In summary, both the classification and regression models demonstrate strong performance in predicting sleep quality from wearable device data. The Random Forest models, along with the LSTM model, offer reliable predictions and can be valuable tools for personalized sleep tracking and recommendations, healthcare applications, and performance optimization. Additionally, the feature importance plot generated from the Random Forest Regressor model can provide insights into which features are most influential in predicting sleep quality, aiding in understanding the factors affecting sleep patterns. Moreover, the LSTM model showcases superior accuracy compared to the CNN model, indicating its efficacy in capturing temporal dependencies within the data and providing robust predictions. With its ability to effectively model sequential data, the LSTM model stands out as the optimal choice for sleep quality prediction tasks, ensuring precise and reliable outcomes for various applications.

5. DISCUSSION

5.1 Interpretation of Results:

The results obtained from the developed machine learning models for predicting sleep quality from wearable device data showcase promising performance. The Random Forest Classifier achieved an accuracy of 85%, indicating its ability to effectively categorize sleep quality as either good or poor. Moreover, balanced precision, recall, and F1-score suggest that the model can identify instances of both good and poor sleep quality accurately. On the other hand, the Random Forest Regressor exhibited impressive predictive power with a very low mean squared error (MSE) of 0.04 and a high R^2 score of 0.90. These metrics signify that the regressor can accurately estimate continuous sleep quality scores, explaining 90% of the variance in the data. Additionally, the LSTM model also performed well, achieving a test accuracy of 88%, which aligns closely with the performance of the Random Forest Classifier. Furthermore, the CNN model attained a test accuracy of 87%, demonstrating its capability in sleep quality prediction. However, when compared to the LSTM model, the CNN model slightly lagged in accuracy, indicating that the LSTM model is better equipped to capture temporal dependencies within the data. Hence, the LSTM model emerges as the preferred choice for sleep quality prediction tasks, outperforming the CNN model and ensuring precise and reliable outcomes for various applications.

5.2 Comparison with Previous Studies:

While there is a scarcity of directly comparable studies focusing on predicting sleep quality from wearable device data using machine learning techniques, the performance metrics obtained in this study can be benchmarked against existing research on sleep monitoring and prediction. Studies employing traditional statistical methods or simpler machine learning algorithms for sleep quality prediction have reported varying levels of accuracy and predictive power. Compared to these approaches, the Random Forest models, LSTM model, and CNN model showcased in this study demonstrate superior performance, particularly in terms of accuracy and predictive precision. However, the LSTM model exhibits a slightly higher accuracy compared to the CNN model, suggesting its efficacy in capturing temporal dependencies within the data. Thus, the LSTM model emerges as the optimal choice for sleep quality prediction tasks, ensuring precise and reliable outcomes for various applications.

5.3 Limitations of the Study:

Despite the encouraging outcomes, it is important to acknowledge several constraints. Firstly, the generalizability of the models may be constrained by the specific demographics and characteristics of the dataset used for training and testing. To enhance generalizability, future studies should validate the models on diverse populations and datasets collected from different wearable devices. Secondly, the interpretation

of sleep quality may vary among individuals, and the ground truth labels used for training and evaluation may not fully capture the nuanced aspects of sleep. Incorporating subjective feedback from users or additional objective measures of sleep quality could enhance the robustness of the models. Moreover, while both the Random Forest, LSTM, and CNN models demonstrate strong predictive performance, the LSTM model outperforms the CNN model in terms of accuracy, indicating its superiority in capturing temporal dependencies and providing precise sleep quality predictions. Thus, focusing on LSTM models could yield more reliable outcomes for sleep monitoring and personalized health interventions.
Overall, while this study provides valuable insights into predicting sleep quality from wearable device data, addressing these limitations and conducting further research will contribute to the advancement of sleep monitoring and personalized health interventions.

6. CONCLUSION

The study on developing a machine learning model for predicting sleep quality highlights the effectiveness of both classification and regression models. The Random Forest Regressor emerges as the optimal choice, evident from its remarkably low mean squared error and high R^2 score. Moreover, the LSTM model demonstrates superior accuracy compared to the CNN model, showcasing its efficacy in capturing temporal dependencies within the data and providing robust predictions. While the CNN model also shows strong performance, the LSTM model's ability to model sequential data makes it the preferred choice for sleep quality prediction tasks, ensuring precise and reliable outcomes. Further validation of diverse datasets and exploration of additional machine learning techniques and evaluation metrics are necessary to enhance the generalizability and comprehensiveness of the findings. In conclusion, this research contributes to the field of sleep quality prediction and emphasizes the importance of selecting appropriate models for optimal performance.

7. REFERENCES

1. Tummala, Venkat and Tanuku, Sai Rohith, Integration of Feature Selection Techniques Using a Sleep Quality Dataset for Comparing Regression Algorithms.
2. Sathyanarayana, A., Joty, S., Fernandez-Luque, L., Ofli, F., Srivastava, J., Elmagarmid, A., Arora, T., & Taheri, S. (2016). Rectification of: Predicting Sleep Quality from Wearable Data Using Deep Learning. JMIR mHealth and uHealth, 4(4), e130.
3. Arora, A., Chakraborty, P. & Bhatia, M.P.S. Utilizing Wearable Sensor Data for Estimating and Predicting Sleep Quality with Deep Learning. Arab J Sci Eng 45, 10793–10812 (2020). https://doi.org/10.1007/s13369-020-04877-w
4. Dinh Van, Phan & Chan, Chien-Lung & Nguyen, Duc-Khanh. (2020). Applying Deep Learning for Prediction of Sleep Quality from Wearable Data. 51-55. 10.1145/3418094.3418114.
5. Hamza MA, Abdalla Hashim AH, Alsolai H, Gaddah A, Othman M, Yaseen I, Rizwanullah M, Zamani AS. Utilizing Wearable Devices to Enhance Health Monitoring for Predicting Sleep Quality Through Optimal Deep Learning Techniques. Ensuring Sustainability. 2023; 15(2):1084.
6. W. Hidayat, T. D. Tambunan and R. Budiawan, "Empowering Wearable Sensor Generated Data to Predict Changes in Individual's Sleep Quality," 2018 6th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 2018, pp. 447-452, doi: 10.1109/ICoICT.2018.8528750.

7. J. Boussard, M. J. Kochenderfer and J. M. Zeitzer, "Predicting Subjective Sleep Quality Using Recurrent Neural Networks," 2019 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), Philadelphia, PA, USA, 2019, pp. 1-6, doi: 10.1109/SPMB47826.2019.9037854.
8. Zamani, Abu Sarwar, et al. " Forecasting Sleep Quality Utilizing Wearable-Assisted Smart Health Monitoring Systems Derived from Statistical Data." Journal of King Saud University-Science 35.9 (2023): 102927.
9. S. Park, et al., "Insight into Sleep Quality from Daily Records," in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.