# Predicting Sleep Disorders Using Lifestyle and Health Factors

1st Gatchalian, Marcelino III
*Department of Computer Science*
*National University - Manila*
Manila, Philippines
gatchalianmc@students.national-u.edu.ph

2nd Llenarez, Tehrence Joie
*Department of Computer Science*
*National University - Manila*
Manila, Philippines
llenareztc@students.national-u.edu.ph

3rd Villanueva, James Cedrick
*Department of Computer Science*
*National University - Manila*
Manila, Philippines
villanuevajp3@students.national-u.edu.ph

***Abstract*—Sleep disorders are prevalent issues that can significantly impact health and daily performance. This study explores the use of lifestyle and health metrics in predicting the likelihood of sleep disorders. By leveraging the Random Forest Algorithm, we analyze the relationship between variables such as stress level, pulse pressure, and sleep duration to predict sleep disorder status. Our approach provides insights into the most significant predictors of sleep disorders, supporting targeted interventions for improved sleep health.**

## I. Introduction

Sleep disorders impact millions of individuals worldwide, leading to significant public health concerns and economic implications. The interplay between lifestyle factors (e.g., stress, physical activity) and physiological conditions (e.g., BMI, blood pressure) often contributes to sleep disruptions. By utilizing machine learning to analyze these variables, we can provide valuable insights into modifiable risk factors that may mitigate sleep disorder prevalence and severity.

Machine learning models, particularly the Random Forest algorithm, offer robust classification capabilities that can handle diverse input features and reveal important predictors. This study leverages Random Forest to predict sleep disorder presence, aiming to enhance preventive care strategies through reliable, data-driven insights into lifestyle and health influences on sleep quality.

## II. Related Work

A growing body of research has utilized machine learning to predict health outcomes based on lifestyle and physiological data. Onargan et al. [1] applied logistic regression to assess sleep apnea risk factors, highlighting associations between sleep patterns and biometric indicators. Similarly, Widasari et al. [2] employed Support Vector Machines (SVM) for sleep disorder classification, demonstrating the efficacy of machine learning in identifying complex relationships within health data. This study expands on previous work by focusing on a comprehensive dataset incorporating lifestyle and health metrics, and by applying the Random Forest algorithm—a robust model known for its accuracy in classification tasks and interpretability via feature importance metrics.

## III. Data and Methodology

### A. Dataset

The dataset comprises various features such as Age, Gender, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Blood Pressure, Heart Rate, and Daily Steps.

### B. Preprocessing

Data preprocessing involved handling missing values, encoding categorical features, and normalizing numerical values. The dataset was then split into training and testing sets for evaluation.

### C. Algorithms and Evaluation Metrics

We evaluated six algorithms for this classification task:

- K-Nearest Neighbors (KNN): 0.87
- Logistic Regression: 0.87
- Gaussian Naive Bayes Classifier (GaussianNBC): 0.80
- Support Vector Classifier (SVC): 0.71
- Decision Tree Classifier (DTC): 0.88
- Random Forest Classifier: 0.88

Each algorithm was evaluated on accuracy, precision, recall, F1-score, and ROC AUC, with Random Forest selected as the final model due to its performance.

## IV. Results and Discussion

### A. Correlation Matrix

To understand the relationships between variables, we computed a correlation matrix (see Figure 1). This matrix shows the strength and direction of linear relationships between features. For instance, Sleep Duration and Quality of Sleep show a strong positive correlation, indicating that higher sleep quality is associated with longer sleep duration.

### B. Feature Importance

Feature importance was calculated for the Random Forest model to identify the most significant predictors of sleep disorders. The most influential features were Blood Pressure, BMI Category, and Sleep Duration, suggesting that these factors play critical roles in determining sleep health (see Figure 2).

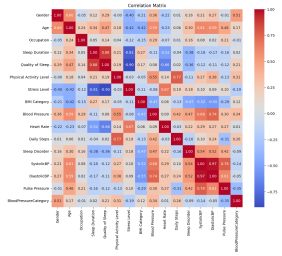

Fig. 1. Correlation Matrix of Features

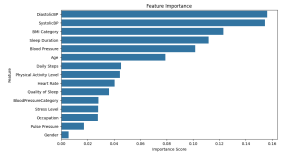

Fig. 2. Feature Importance in the Random Forest Model

### C. Model Performance

The Random Forest model achieved the following metrics:

- Accuracy: 0.88
- Precision: 0.8819
- Recall: 0.88
- F1-Score: 0.8785

### D. ROC AUC Analysis

To evaluate the model's ability to discriminate between classes, we analyzed the Receiver Operating Characteristic (ROC) curve and calculated the Area Under the Curve (AUC) score. The ROC AUC score for the Random Forest model was 0.88, indicating a strong ability to distinguish between individuals with and without sleep disorders.

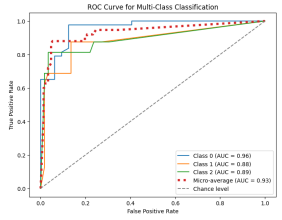

Fig. 3. ROC Curve for Random Forest Model

## V. Limitations and Future Work

One limitation of this study is the relatively small sample size of the dataset, which may limit the generalizability of the model's performance to larger, more diverse populations. Future studies could benefit from using a dataset with a larger sample size to improve the model's robustness and accuracy. Additionally, expanding the dataset to include more diverse demographic and lifestyle factors could further enhance the model's predictive capability.

## VI. Conclusion

This study demonstrates that lifestyle and health factors can effectively predict sleep disorder presence. By applying the Random Forest Algorithm, we identified critical predictors, including blood pressure and physical activity, that correlate strongly with sleep disorders. These findings can inform personalized health strategies aimed at mitigating sleep disorder risk, ultimately contributing to improved health outcomes.

## References


[1] A. Onargan, et al., "Prediction of Sleep Apnea Using EEG Signals and Machine Learning Algorithms," *IEEE Conference Publication*, 2021. Available: https://ieeexplore.ieee.org/document/9632895.

[2] E. R. Widasari, et al., "Automatic Sleep Disorders Classification Using Ensemble of Bagged Tree Based on Sleep Quality Features," *MDPI Electronics*, vol. 9, no. 3, pp. 512, 2020. Available: https://www.mdpi.com/2079-9292/9/3/512.