Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 2 (2025)

Analysis of Sleep Health and Lifestyle Factors: A Machine Learning Approach

Sourabh Sahu[1], Kalyani Tiwari[2*], Garima Hardia[3], Rashmi Rani[4], Mohit Dhiman[5], Amee Vishwakarma[6]

[1]Associate Professor, Gyan Ganga Institute of Technology and Sciences, Jabalpur, M.P., India
[2*]Assistant Professor, IPS Academy, IES Indore, M.P. India, ORCID ID: 0000-0002-6310-3555
[3]Assistant Professor, SAGE University, Indore, M.P., India
[4]Assistant Professor, SAGE University, Indore, M.P., India
[5]Data Engineering Manager, NIT Warangal, India
[6]Gyan Ganga Institute of Technology and Sciences, Jabalpur, M.P., India

Email: [1]sourabhsahu@gyangangagroup.in, [2]ktiwari.official@gmail.com, [3]garimabordia1490@gmail.com, [4]rashmi.sagecs@gmail.com, [5]mohitdhiman25@outlook.com, [6]vishwakarma.amee@gmail.com

Article History:

Received: 04-11-2024
Revised: 05-12-2024
Accepted: 06-01-2025

Abstract:
Sleep plays a key role in overall health, yet many factors in daily life can affect its quality. This study explores how lifestyle choices, stress, and physical activity relate to sleep health. Using a dataset of 400 individuals, we analyze sleep duration, quality, and common sleep disorders such as insomnia and sleep apnea. The data also includes information on age, occupation, BMI, blood pressure, heart rate, and daily steps. This work utilized statistical analysis and machine learning models to identify correlations among the key parameters influencing sleep health. The results highlight connections between stress levels, physical activity, and sleep disorders, offering insights into how lifestyle adjustments may support better sleep.

Keywords – Machine Learning, Exploratory Data Analysis, SVM, Decision Tree

1. Introduction

Overall health depends significantly on sleep because it influences both mental and physical well-being. The correlation between poor sleep patterns includes the development of heart disease as well as obesity and diabetes, along with decreased immune defence capabilities [1]. Sleep has a substantial impact on cognitive performance as well as emotional regulation and productivity [2]. Many people experience sleep disturbances because of their lifestyle choices and stress levels as well as various health conditions despite sleep's crucial role. Insomnia and sleep apnea which are widespread sleep disorders affect millions and present long-term health dangers if people do not receive treatment [3]. Historically researchers collected sleep data through self-reported questionnaires along with clinical observations and the use of polysomnography (PSG) [4]. Although these methods yield significant insights into sleep patterns they often demand high costs and specialized equipment while taking considerable time to implement. Although PSG stands as the most reliable method for diagnosing sleep disorders its execution in a controlled lab environment can result in sleep patterns that differ from those exhibited in natural settings [3]. Although wearable devices and mobile apps provided novel sleep tracking methods these devices require subjective data input which can lead to accuracy issues [6].

Machine learning provides a robust tool for traditional methods through its capacity to process big data and discover concealed trends in sleep patterns. Machine learning stands apart from traditional methods because it can analyze multiple variables at once which allows researchers to find connections between sleep quality and both lifestyle factors as well as health conditions [7]. Past studies have demonstrated effective use of machine learning algorithms to anticipate sleep disorders by analyzing physiological metrics like heart rate alongside movement and respiratory patterns [8].

Researchers have successfully utilized deep learning models for sleep stage classification and these models produce results that match the quality of PSG-based assessments [9].
This study uses a dataset containing information on sleep duration and quality as well as common sleep disorders to examine sleep health alongside lifestyle factors like stress level, physical activity patterns and body mass index. In this work different statistical techniques and machine learning models have been applied to identify key factors influencing sleep health. Also a comparative analysis has been performed for different algorithms to determine the most effective method for predicting sleep disorders. By using machine learning, we aim to provide a data-driven approach to understanding sleep health, which could help in developing personalized recommendations for improving sleep quality.

2. Literature Review

Presently, the detection of sleep disorders relies on clinical evaluation and self-reported information. The most accurate way to analyse sleep is Polysomnography (PSG), but it involves overnight monitoring in a lab. So, PSG is a costly and inconvenient approach [1]. Actigraphy and sleep questionnaires are the other methods that provide more sleep analysis reports but cannot often give a detailed analysis of the sleep pattern [2]. To avoid the conventional clinical methods to observe statistical techinques like machine learning models have been used for analyzing sleep health and identifying sleep disorders.
Multiple research studies have implemented machine learning algorithms to identify different sleep stages and diagnose sleep disorders. Sun et al. Sun et al. [3] utilized supervised learning models to forecast sleep apnea based on physiological signal data. The research demonstrated that decision trees and support vector machines (SVM) effectively separated normal breathing from disordered breathing patterns. Another survey by Stephansen et al. Stephansen et al. [4] utilized deep learning techniques to classify sleep stages. The neural network model reached higher accuracy levels than traditional scoring methods illustrating the capabilities of automated sleep analysis.
The application of machine learning (ML) assists researchers in discovering elements that affect sleep quality. Xu et al. Xu et al. [5] examined wearable device data to discover connections between physical activity levels and stress with sleep disturbance patterns. The research implemented random forests and gradient boosting models to predict sleep outcomes and demonstrated that machine learning techniques can reveal intricate patterns in sleep behaviour. Zaffaroni et al. [6] worked on machine learning techniques to advance sleep monitoring outside clinical environments. Machine learning methods helped Zaffaroni et al. identify patterns within biometric data for sleep monitoring. Analyzing sleep data becomes much more efficient through machine learning models because they provide speed and scalability. ML models enable efficient processing of large datasets to discover hard-to-detect patterns while delivering personal sleep behaviour insights. Machine learning models process data from wearable technology and home sleep monitors with ease which makes high-quality sleep monitoring more accessible.

3. Description of Dataset

The sleep analysis process is driven by machine learning algorithms that utilize data from the dataset. The dataset holds physiological signal data along with sleep patterns and additional relevant features which have been gathered from multiple sources to identify disorders. The dataset contains meaningful attributes related to sleep disorders, including heart rate, respiratory rate, oxygen saturation levels, movement during sleep, and recorded sleep stages. The machine learning model requires understanding the data structure before application. Data pre-processing starts with a thorough examination of the dataset to detect and correct missing values alongside duplicate entries and inconsistent and erroneous data points [6]. Imputation techniques allow researchers to handle

missing data in datasets by replacing missing values with the mean, median or mode based on each attribute's distribution characteristics [5]. For normal distribution of the data, outliers are removed to represent the data into a more statistically significant manner [3].
Data visualization techniques are utilized to improve the interpretability of information specifically to identify the correlations between various parameters. In the study, Graphical methods such as histograms and box plots facilitate the analysis of quantitative variables like heart rate and oxygen saturation levels. However the Correlation between the attributes is analysed using the scatterplots for the identification of relationships among variables, contributing to more accurate predictions of sleep patterns [7]. Furthermore, research studies utilize scatter plots and line charts to examine variations in heart rate throughout the sleep cycle, offering deeper insights into nocturnal physiological fluctuations [5].

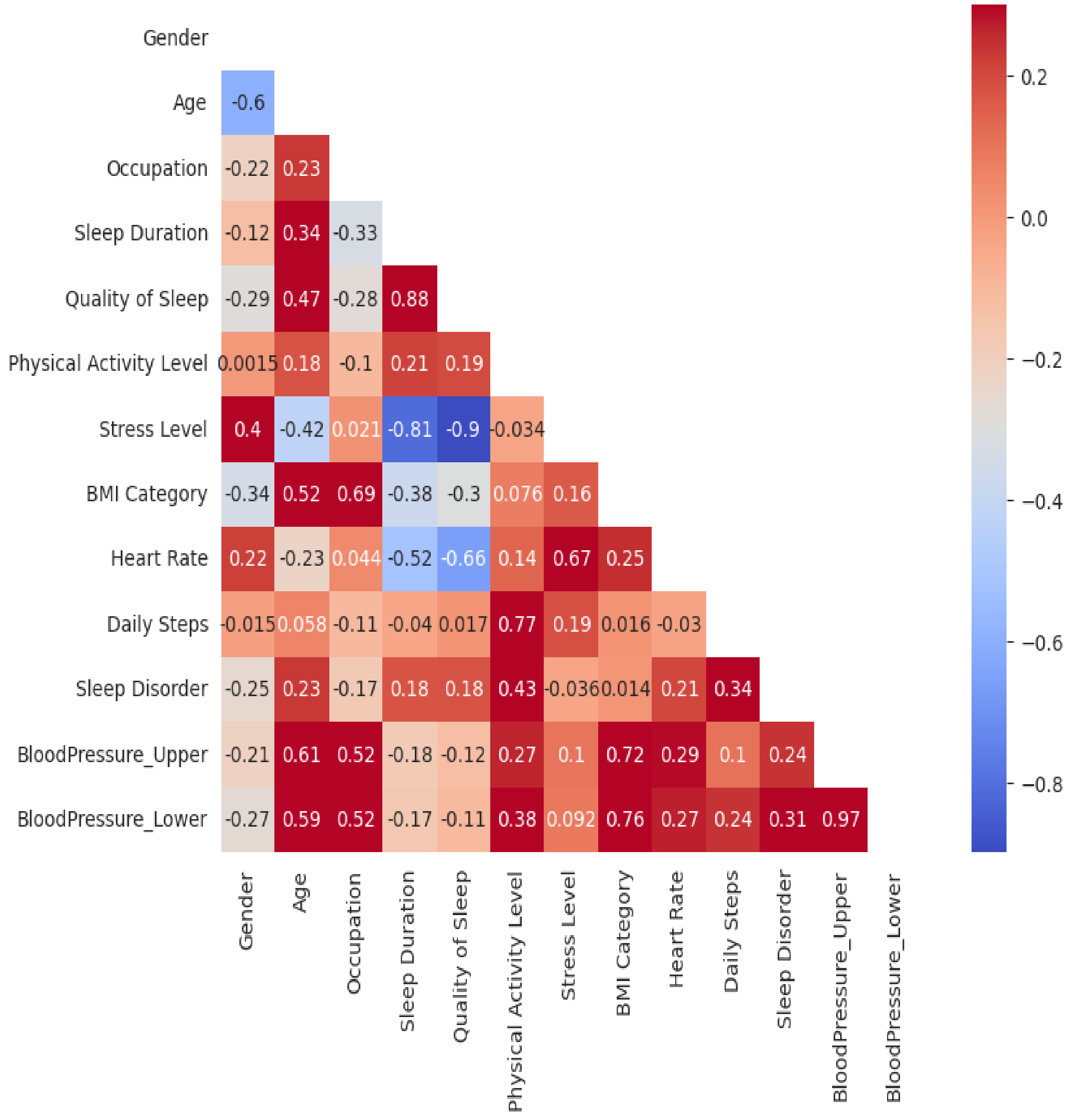

Figure 1. Correlation between different attributes of the dataset

In the analysis of sleep disorders such as sleep apnea and insomnia, data scientists assess class distributions to ensure balanced representation within the dataset [5]. The integration of resampling techniques with weighted modelling approaches has been shown to enhance predictive accuracy and, hence causes improvement in the overall performance metrics of the analysis.

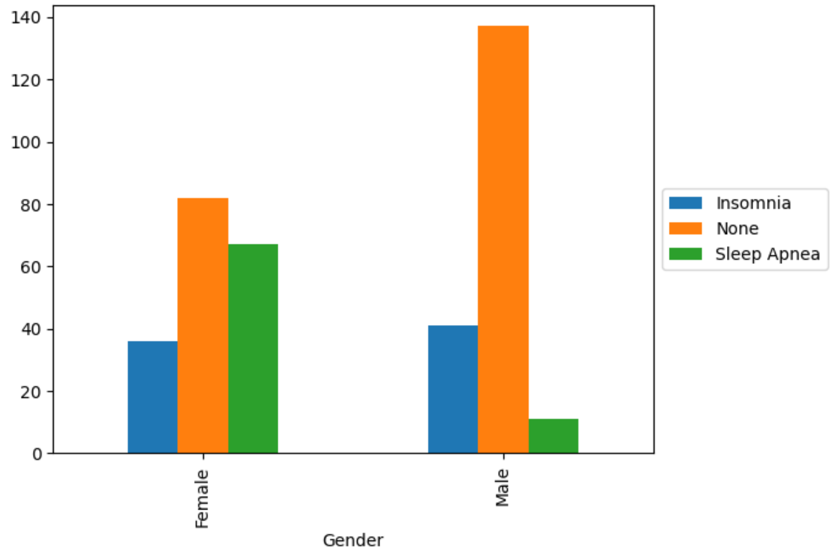

Figure 2. Distribution of sleep classification

4. ML Models and Performance Metrics

The analysis of sleep health patterns and lifestyle behaviors depends on machine learning models trained on the dataset. The model discovers connections between various factors, including sleep duration, stress levels, and physical activity, to detect potential sleep disorders [5]. Achieving effective model training demands the incorporation of key variables that influence sleep quality to ensure precise classification and prediction results.

The research focuses on developing machine learning models for predicting sleep disorders.

- Support Vector Machine (SVM): The SVM algorithm work by identifying the optimal hyperplanes to divide data into several categories [4]. This method specifically stands out for sleep disorder classification because it excels at analyzing high-dimensional datasets.
- XGBoost: It is a gradient boosting technique that helps to achieve better prediction results by sequentially focusing on data points that were previously misclassified. This technique is widely use in structured data analysis because of its flexible characteristics and effective performance.

These machine learning techniques enable superior detection of sleep disorders, leading to enhanced prediction accuracy and more personalised healthcare solutions [8].

a) Performance Metrics

To evaluate model performance, several metrics are used:

- Accuracy: Calculates the proportion of correct predictions among all predicted outcomes. The metric provides a broad assessment of the model's performance effectiveness [8].
- Precision: In this prediction, accuracy has been evaluated to accumulate the correctly identified positive instances [7, 8].
- Recall: Recall measures the effectiveness of the model in identifying true positive instances. A high recall means fewer false negatives [4].
- F1-Score: The F1-Score provides a balance between precision and recall which proves to be beneficial for evaluating imbalanced datasets [4, 5].
- ROC-AUC Score: This score identifies the differentiability between multiple classes [5]. The higher the score, the more accurate the predictions.

On the basis of the Performance metrics comparison between ML model can be performed to identify the best model to analyze sleep health data. The study identifies sleep quality determinants and the impact of conditions like insomnia and sleep apnea [6].

5. Results and Interpretations

Model performance evaluations involved cross-validation to compare the achieved accuracy scores. The Gradient Boosting Classifier achieved accuracy values between 0.811 and 0.939 while maintaining an average accuracy of 0.875. The Support Vector Classifier (SVC) demonstrated superior performance with scores ranging from 0.837 to 0.942 and an average accuracy of 0.889. Based on its better performance metrics SVC emerged as the preferred choice for generating predictions.

The Gradient Boosting Classifier achieved an 88% accuracy when applied to the training data. The evaluation metrics for precision, recall and F1-score displayed different values across each class. The model achieved the highest recall rate of 96% for the "None" category which indicates successful identification of most cases without sleep disorders. The conditions "Insomnia" and "Sleep Apnea" showed lower recall rates of 77% and 75% which indicates that some cases were detected incorrectly. The metric of weighted F1-score stood at 0.87 which showed an effective equilibrium between precision and recall.

The model demonstrated high accuracy at 86% on test data but experienced a minor reduction from its training performance results. The recall values for both "Insomnia" and "Sleep Apnea" remained lower at 81% and 80% respectively compared to "None" which achieved a recall value of 90%. The macro average F1-score reached 0.84 and the weighted average F1-score reached 0.86.

The model demonstrated effective generalization from training data to test data because accuracy and F1-score remained stable. The model demonstrated minimal overfitting while maintaining reasonable classification performance for sleep disorders. Insomnia and Sleep Apnea cases remain prone to misclassification but could see improved recall through additional feature engineering and hyperparameter adjustments.

Table 1. Performance metrics corresponding to the study

Cross-Validation Accuracy Comparison			
Model	**Min Accuracy**	**Max Accuracy**	**Mean Accuracy**
Gradient Boosting Classifier	0.811	0.939	0.875
Support Vector Classifier	0.837	0.942	0.889

Performance on Training Data				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
Insomnia	0.84	0.77	0.8	56
None	0.89	0.96	0.93	161
Sleep Apnea	0.85	0.75	0.8	63
Overall Accuracy	-	-	**0.88**	280
Performance on Validation Data				
Class	**Precision**	**Recall**	**F1-Score**	**Support**
Insomnia	0.85	0.81	0.83	21
None	0.88	0.9	0.89	58
Sleep Apnea	0.8	0.8	0.8	15
Overall Accuracy	-	-	**0.86**	94

Conclusion

The Support Vector Classifier demonstrated a marginally higher accuracy compared to Gradient Boosting. Sleep disorder classification model training and testing produced an 88% accuracy rate on training data and 86% accuracy on validation data. The model showed only a slight drop in performance between training and test data which indicates good generalization capabilities. The "None" category achieved higher precision, recall and F1-score metrics than "Insomnia" and "Sleep Apnea" which had slightly lower scores. The application of hyperparameter tuning and ensemble methods together with more data addition can refine the results further. The model successfully identified various sleep disorders which allows it to function as a practical research instrument in this field. The model currently shows sufficient performance for differentiating sleep disorders yet stands to benefit from further refinement to boost its predictive capabilities.

Acknowledgement

The authors would like to thank the administrative bodies of the institute for providing all the necessary support for performing the study.

References

[1] Walker, M. Why We Sleep: Unlocking the Power of Sleep and Dreams. Scribner, 2017.
[2] Killgore, W. D. (2010). "Effects of sleep deprivation on cognition." Progress in Brain Research, 185, 105-129.
[3] Ohayon, M. M. (2002). "Epidemiology of insomnia: what we know and what we still need to learn." Sleep Medicine Reviews, 6(2), 97-111.
[4] Iber, C., Ancoli-Israel, S., Chesson, A. L., Quan, S. F. (2007). The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications.
[5] Rechtschaffen, A., & Kales, A. (1968). A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects. Brain Information Service/Brain Research Institute.
[6] Zaffaroni, A., et al. (2013). "Sleep monitoring using biometric devices." IEEE Transactions on Biomedical Engineering, 60(8), 2219-2227.
[7] Xu, X., et al. (2021). "Machine learning applications in sleep medicine: a review of recent advances." Nature and Science of Sleep, 13, 1391-1408.
[8] Sun, H., et al. (2019). "Predicting sleep apnea using machine learning algorithms." Journal of Clinical Sleep Medicine, 15(5), 747-753.
[9] Stephansen, J. B., et al. (2018). "Neural network analysis of sleep stages enables efficient diagnosis of sleep disorders." Nature Communications, 9(1), 5220.