



A Comparative Machine Learning Method for Health Indicator-Based Sleep Disorder Prediction

K Naresh¹, Thangem Haritha²

¹Assistant Professor, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

²Postgraduate, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

How to Cite this Article:

Haritha, T. (2026). A Comparative Machine Learning Method for Health Indicator-Based Sleep Disorder Prediction. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04).
<https://doi.org/10.55041/ijcope.v2i4.069>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



OPEN ACCESS



<https://doi.org/10.55041/ijcope.v2i4.069>

Abstract

Due to shifting lifestyles, elevated stress levels, and inconsistent sleep cycles, sleep disorders have emerged as a major health concern in contemporary culture. In order to avoid major health issues including heart disease, emotional stress, and decreased productivity, early detection of sleep disturbances is crucial. Machine learning methods have demonstrated tremendous promise in recent years for the analysis of medical data and disease prediction. This study suggests a machine learning-based method for classifying sleep problems based on lifestyle and health-related data. The system makes use of a dataset that includes a number of physiological and behavioural characteristics, including age, heart rate, physical activity, stress level, sleep length, and sleep quality.

To address imbalances and enhance model performance, data preparation techniques are used. To create classification models, a number of machine learning techniques are used, such as Random Forest, K-Nearest Neighbours, Decision Tree, and Multinomial Logistic Regression. The best algorithm for predicting sleep disorders is determined by training and evaluating the models. Results from experiments show that machine learning models are capable of accurately and successfully classifying conditions related to sleep disorders. In order to offer user-friendly prediction capabilities, the suggested system also incorporates a web-based

interface. This study demonstrates how machine learning can help people and medical professionals recognise sleep-related health problems early on.

Keywords

Machine Learning, Sleep Disorder Classification, Health Data Analysis, Random Forest, K-Nearest Neighbors, Healthcare Prediction, Data Mining in Healthcare



I. Introduction

Sleep is essential for preserving one's physical and mental well-being. Getting enough sleep is crucial for enhancing general wellbeing, emotional stability, and cognitive function. However, due to unhealthy lifestyles, stress at work, erratic sleep patterns, and excessive use of digital gadgets, sleep disorders have become more prevalent in recent years. A person's health and everyday productivity can be greatly impacted by sleep problems such as insomnia, sleep apnoea, and restless sleep. These disorders can cause major health problems like obesity, depression, cardiovascular illnesses, and a lower quality of life if they are not detected early.

Traditionally, clinical assessments and specialised tests like polysomnography—which call for expert medical supervision and cutting-edge equipment—are used to detect sleep disorders. These diagnostic techniques can be expensive, time-consuming, and difficult for some people to use. As a result, there is an increasing need for intelligent systems that can use readily available health and lifestyle data to help with the early detection and classification of sleep problems. The development of artificial intelligence has led to the widespread usage of machine learning techniques in healthcare applications for health monitoring and disease prediction. Large amounts of health-related data can be analysed by machine learning models, which can find hidden patterns that conventional analytic techniques can miss.

These methods can be used to create predictive algorithms that can help medical professionals diagnose illnesses more accurately. This study proposes a machine learning-based method to categorise sleep problems based on lifestyle and health data. The study makes use of a dataset with several characteristics, including heart rate, physical activity, stress level, length of sleep, and quality of sleep. The dataset is analysed and sleep disorder classifications are predicted using a number of machine learning methods, such as Decision Tree, K-Nearest Neighbours (KNN), Random Forest, and Multinomial Logistic Regression. To determine the best algorithm for precise classification, the performance of different models is assessed and contrasted.

Additionally, a web-based interface that lets users enter health-related factors and get predictions about possible sleep disorders is connected with the suggested system. This strategy shows how machine

learning can be applied as a helpful tool in healthcare systems to facilitate early detection and raise people's awareness of their health.

II. Literature Review

Because sleep disorders have a major impact on daily functioning and human health, researchers are paying more and more attention to them. Numerous studies have investigated the application of machine learning approaches for forecasting and categorising sleep-related problems in light of the development of data analytics and artificial intelligence. Early detection of sleep disorders can be supported by machine learning algorithms, which can recognise intricate patterns in health-related data.

In order to identify sleep disorders, a number of researchers have used machine learning algorithms to examine physiological and lifestyle data. For example, based on patient health information, classification algorithms like Decision Trees, Support Vector Machines, Random Forest, and Neural Networks have been used to find abnormalities associated to sleep.

To identify possible sleep problems, these models examine characteristics like heart rate, physical activity, stress levels, and length of sleep.

Because ensemble learning techniques like Random Forest can handle complicated datasets and minimise overfitting, they have demonstrated promising outcomes in healthcare prediction tasks in recent studies. Similarly, because to its ease of use and efficiency in finding similarities between data points, K-Nearest Neighbours (KNN) has been extensively applied to classification problems. In healthcare analytics, logistic regression models are frequently used to forecast categorical outcomes based on a variety of input variables.

Even though sleep problem prediction research is expanding, many current methods are mostly focused on clinical datasets that call for specialised medical equipment and supervision. Systems that use readily available health and lifestyle data to predict sleep disorders early are still needed. In order to choose the best model, this study evaluates the performance of many machine learning algorithms used to categorise sleep disorders using lifestyle and health data.

III. Problem Statement

A significant percentage of people worldwide suffer from sleep disorders, which, if untreated, can result in



major issues with both physical and mental health. Conventional diagnostic techniques can be costly and time-consuming because they frequently call for specialised medical testing and expert supervision. Because of this, many people do not receive an early diagnosis or treatment for sleep-related problems.

To help detect possible sleep disorders, an intelligent system that can evaluate lifestyle and health data is required. A predictive model that can categorise sleep disorder disorders can be created by applying machine learning techniques to readily available health markers, such as heart rate, physical activity, stress level, and sleep duration.

IV. Dataset

The Sleep Health and Lifestyle Dataset, which includes data on people's health issues, lifestyle choices, and sleep patterns, was utilised in this study. The connection between different physiological and behavioural parameters and the prevalence of sleep disorders is examined using this dataset. The dataset offers insightful information about how health markers and lifestyle decisions affect sleep quality and sleep-related diseases.

A person's daily routines and physical health metrics are described by a number of features in the dataset. Age, Sleep Duration, Physical Activity Level, Blood Pressure, Heart Rate, and Daily Steps are significant characteristics found in the dataset. These characteristics aid in spotting trends that could be linked to sleep disorders. The dataset's target variable, Sleep Disorder, reflects how each person's sleep circumstances are categorised. Every record in the collection reflects a specific person's health and lifestyle details. The Sleep Duration attribute shows how many hours a person sleeps each day, whereas the Age attribute shows the person's age. The quantity of daily physical activity that a person engages in is measured by their Physical Activity Level. The characteristics of heart rate and blood pressure are significant physiological markers of cardiovascular health. The Daily Steps feature also shows how much physical activity is done during the day.

In order to determine whether a person has a sleep-related problem, the target variable Sleep Disorder is divided into several classes. For machine learning processing, these classes are numerically encoded in the dataset. Machine learning algorithms can find

patterns that aid in the prediction of sleep problems by examining these features.

To enhance the machine learning models' performance, the dataset underwent additional processing and balancing. In addition to ensuring that the data is appropriate for training classification algorithms, proper preprocessing enhances the prediction system's accuracy and dependability.

Name	Age	Sleep Duration	Physical Activity Level	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder	
David	30	6.1	42	11	77	4200	1	
Machine Learning	1	28	6.2	60	9	75	3000	1
Matthew	2	28	6.2	60	9	75	3000	1
3	28	5.9	30	22	85	3000	2	
Prediction	4	28	5.9	30	22	85	3000	2
Logan	5	28	5.9	30	22	85	3000	0
6	28	6.3	40	22	82	3500	0	
7	28	7.8	75	6	70	8000	1	
8	28	7.8	75	6	70	8000	1	

Fig: Dataset View

V. Experimental Results and Discussion

Several machine learning algorithms were developed and tested using the processed dataset in order to assess the efficacy of the suggested sleep disorder classification method. Individual health and lifestyle characteristics, such as age, sleep duration, degree of physical activity, blood pressure, heart rate, and daily steps, were used to train the models. Standard evaluation criteria like accuracy, precision, recall, and F1-score were used to gauge the models' performance.

The Random Forest classifier demonstrated good predictive potential for detecting sleep disorder classes, with an overall accuracy of almost 0.89 (89.33%) among the deployed models. With a weighted average precision of 0.90, recall of 0.89, and F1-score of 0.89, the model demonstrated balanced performance across several classes.

According to the classification report, the model maintains decent prediction performance for other classes while doing exceptionally well for the majority class.

In a similar vein, the Decision Tree classifier also attained an accuracy of roughly 0.89, demonstrating the efficacy of tree-based models in spotting trends in the sleep health dataset. Based on feature circumstances like heart rate, sleep length, and stress-related indicators, decision tree models can learn decision boundaries. They are especially helpful in prediction tasks relating to healthcare because of their interpretability.

The experimental investigation demonstrates that the



merging of several decision trees lowers the danger of overfitting, allowing ensemble-based models like Random Forest to produce consistent and trustworthy results.

On the other hand, when the dataset is properly preprocessed, single decision tree models can still function well even though they are occasionally more sensitive to changes in the dataset.

All things considered, the findings show that machine learning algorithms can efficiently categorise sleep disorder disorders according to lifestyle and health factors. By enabling users to enter health-related data and receive predictions about possible sleep disorders, the incorporation of these predictive models into a web-based system further improves usability.

Confusion Matrix Analysis:

By contrasting the predicted labels with the actual labels in the dataset, the confusion matrix is used to assess how well classification models perform. It offers a thorough comprehension of the model's accuracy in classifying each category as well as the locations of misclassifications. Confusion matrices for the machine learning models utilised in the sleep disorder categorisation system were produced in this study. For each class, the matrices show the distribution of accurate and inaccurate predictions. The actual class is represented by each row in the matrix, while the anticipated class is represented by each column.

The confusion matrix for the Support Vector Classifier (SVC) model indicates that most samples in class 1 were correctly classified by the model, with many of the diagonal elements of the matrix having accurate predictions. Nonetheless, a small number of samples from classes 0 and 2 were incorrectly categorised as class 1, suggesting that these groups in the dataset had comparable traits.

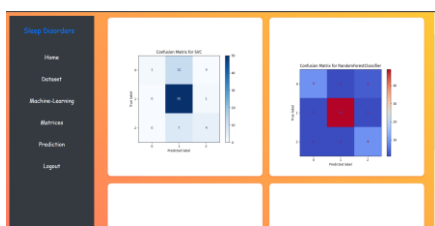


Fig: Matrix Analysis

In a similar vein, the Random Forest Classifier exhibits excellent classification performance, with the majority of predictions occurring along the confusion matrix's

diagonal. This shows that the model effectively uses the available health and lifestyle factors to differentiate between various categories of sleep disorders. There aren't many misclassifications, indicating that the model can identify significant patterns in the dataset.

VI. Output

Fig: Screenshot-1

Fig: Screenshot-2

VII. Conclusion

This study introduces a machine learning-based method that uses lifestyle and health data to classify sleep disorders. Sleep patterns were examined using a variety of characteristics, including age, sleep duration, physical activity, blood pressure, heart rate, and daily steps. For prediction, a number of machine learning techniques were used, such as Random Forest, Decision Tree, and Support Vector Classifier. The experimental findings demonstrate that these models are capable of accurately and successfully classifying situations related to sleep disorders. The suggested models' dependability is validated by the evaluation metrics and confusion matrix. All things considered, the created web-based solution shows how machine learning can help with the early identification and evaluation of sleep-related health problems.



References

- [1] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, USA, 2018.
- [2] T. Mitchell, Machine Learning, McGraw-Hill Education, New York, USA, 1997.
- [3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, Cambridge, MA, USA, 2016.
- [4] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” *Informatica*, vol. 31, no. 3, pp. 249–268, 2007.
- [5] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [6] C. Cortes and V. Vapnik, “Support-Vector Networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.