iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 14 **Issue:** III **Month of publication:** March 2026

**DOI:** https://doi.org/10.22214/ijraset.2026.78018

# Sleep Disorder Prediction Using Machine Learning

Ganesh Ishwarchandra Ghonsikar[1], Rucha Ravindra Galgali[2]
[1]*Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies*
[2]*Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies*

***Abstract: sleep disorder prediction involves identifying early signs that a person may develop problems with sleep, such as insomnia, sleep apnea, or restless leg syndrome. It relies on monitoring patterns in sleep duration, quality, and consistency. Factors like stress levels, irregular work schedules, lifestyle habits, and medical history can indicate higher risk. Physical symptoms such as excessive daytime sleepiness, difficulty falling asleep, or frequent awakenings are important clues. Regular tracking of these signals can help anticipate disorders before they become severe. Early prediction allows for timely lifestyle adjustments, medical consultations, and preventive measures to maintain healthy sleep. The results show that Random Forest, Gradient Boosting, and XGBoost all reached a peak accuracy of 0.947, with Random Forest chosen as the final model due to its stability, ability to prevent overfitting, and suitability for structured medical data.***
***Keywords: sleep disorder prediction, insomnia, sleep apnea, restless leg syndrome, sleep quality, early detection, preventive interventions, machine learning, Random Forest, XGBoost, Gradient Boosting, Decision Tree, K Nearest Neighbours, Support Vector Machine (RBF), Multilayer Perceptron (Neural Network), Logistic Regression, sleep health.***

## I. INTRODUCTION

Sleep plays a vital role in maintaining physical health, mental stability, and overall well-being. However, sleep disorders often remain undiagnosed due to lack of awareness, high medical costs, and limited access to specialized clinical testing. Traditional diagnostic methods such as polysomnography are time consuming, expensive, and unsuitable for large-scale or early screening. With the increasing availability of health and lifestyle data, machine learning offers an effective approach for predicting sleep disorders using parameters such as sleep duration, stress level, physical activity, body mass index (BMI), blood pressure, occupation, and quality of sleep. These data-driven techniques can identify hidden patterns that are difficult to detect through conventional analysis. This application was developed to provide an intelligent, affordable, and efficient system for early prediction of sleep disorders. By analyzing lifestyle and health-related attributes, the system aims to assist healthcare professionals and individuals in identifying potential sleep-related issues at an early stage. Multiple machine learning models are implemented and compared to determine the most accurate and reliable approach, demonstrating the practical use of machine learning in preventive healthcare and medical decision support.

## II. SYSTEM DEVELOPMENT

### *A. Dataset Used*

This dataset has been developed to support NLP driven diagnostic analysis within Hospital Network Information Management Systems (HNIMS). It represents simulated electronic health records (EHRs), incorporating digitized medical reports, text extracted through optical character recognition (OCR), and essential clinical indicators associated with sleep disorders.
Dataset link: https://www.kaggle.com/datasets/ziya07/sleep disorder-diagnostic-dataset

### *B. Implemented Algorithms*

1) XGBoost: XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm based on sequential decision trees, where each new tree aims to reduce the errors of the previous ones. It incorporates regularization to prevent overfitting, making it highly efficient, scalable, and popular in practical applications and data science competitions.
2) Gradient Boosting: Gradient Boosting is an ensemble method that sequentially builds weak models, typically decision trees, with each newmodel correcting the errors of the previous ones. By optimizing a loss function through gradient descent, it combines these models to deliver increasingly accurate predictions.
3) Random Forest: A Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting. Each tree is trained on a random subset of data and features, making the overall model more robust and less sensitive to noise. Essentially, it turns many “weak learners” into a strong, reliable predictor.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

### C. System Architecture

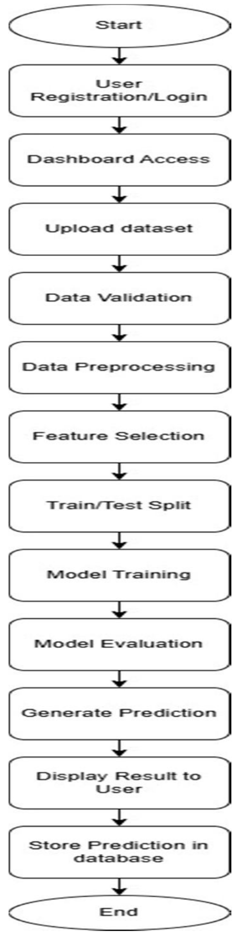



Fig.1. Block Diagram for Sleep Disorder Prediction Explanation

### D. Explanation

1) Start: The system begins when the user opens the web application in browser.
2) User Registration/ Login: New users can sign up, and existing users can log in. Authentication is handled by Django's built-in User model, ensuring secure access.
3) Dashboard access: After login, the user is redirected to the dashboard. The dashboard options are: Upload Dataset, View Previous Results, and Logout.
4) Upload Dataset: The user uploads a dataset in CSV format containing sleep-related parameters.
5) Data validation: The system checks for the correct file format, the presence of required columns, and the absence of corrupted data. If the file is invalid, an error message is displayed.
6) Data Preprocessing: Handle missing values, encode categorical variables, normalize numerical values, and remove duplicates.
7) Feature Selection: Relevant features are selected for model training, such as sleep duration, age, and stress level, while irrelevant attributes are removed.
8) Train/Test Split: The dataset is divided into training data (80%) and testing data (20%) to ensure unbiased evaluation.
9) Model Training: The model is trained using the following algorithms: Random Forest, Decision Tree, K-Nearest Neighbours, Support Vector Machine (SVM), Multilayer Perceptron (Neural Network), XGBoost, Gradient Boosting, Logistic Regression.
10) Model Evaluation: The model is tested using the test dataset. Evaluation metrics include precision, recall, F1-score, Support, Accuracy. If the accuracy is acceptable, the model proceeds; if not, it is retrained.
11) Generate Prediction: Using the trained model, predict sleep quality and classify it as Good Sleep, Moderate Sleep, or Poor Sleep.

12) Display result to user: The system shows the prediction result, accuracy score, and graph.
13) Store Prediction in database: Predictions are stored in SQLite, including user ID, dataset reference, prediction result, and timestamp, allowing result history tracking.
14) End: The user can upload another dataset, log out, or exit the application.

## III. EVALUATION METRICS

1) Precision: Indicates how many of the instances predicted as positive are actually positive, reflecting the model's exactness.

$$\text{Precision} = \frac{TP}{TP + FP}$$

2) Recall: Shows how well the model identifies all actual positive instances, measuring completeness.

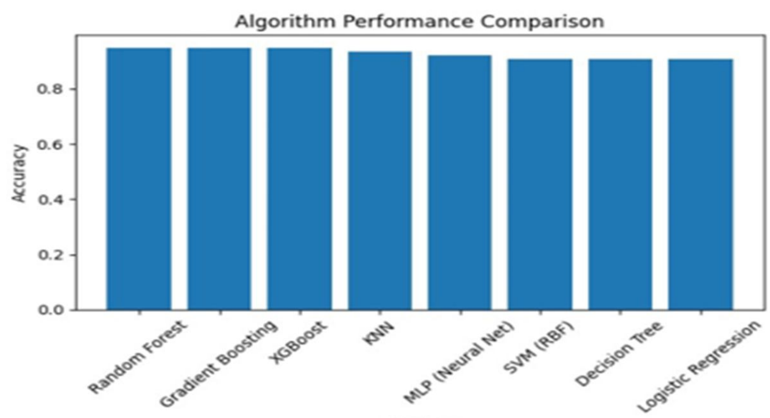
$$\text{Recall} = \frac{TP}{TP + FN}$$

3) F1-Score: Combines precision and recall into a single metric, giving a balanced measure of accuracy when classes are imbalanced.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4) Accuracy: Measures the overall correctness of a model by showing the proportion of correctly predicted instances out of all predictions.

## IV. RESULT




Graph 1: Algorithm Performance Comparison

| Algorithm | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Random Forest | 0.947 | 0.948 | 0.947 | 0.946 |
| Gradient Boosting | 0.947 | 0.948 | 0.947 | 0.946 |
| XGBoost | 0.947 | 0.948 | 0.947 | 0.946 |
| KNN | 0.933 | 0.936 | 0.933 | 0.932 |
| MLP (Neural Net) | 0.920 | 0.920 | 0.920 | 0.920 |
| SVM (RBF) | 0.907 | 0.914 | 0.907 | 0.908 |
| Decision Tree | 0.907 | 0.908 | 0.907 | 0.907 |
| Logistic Regression | 0.907 | 0.907 | 0.907 | 0.906 |

Table 1: Result

## V. CONCLUSION

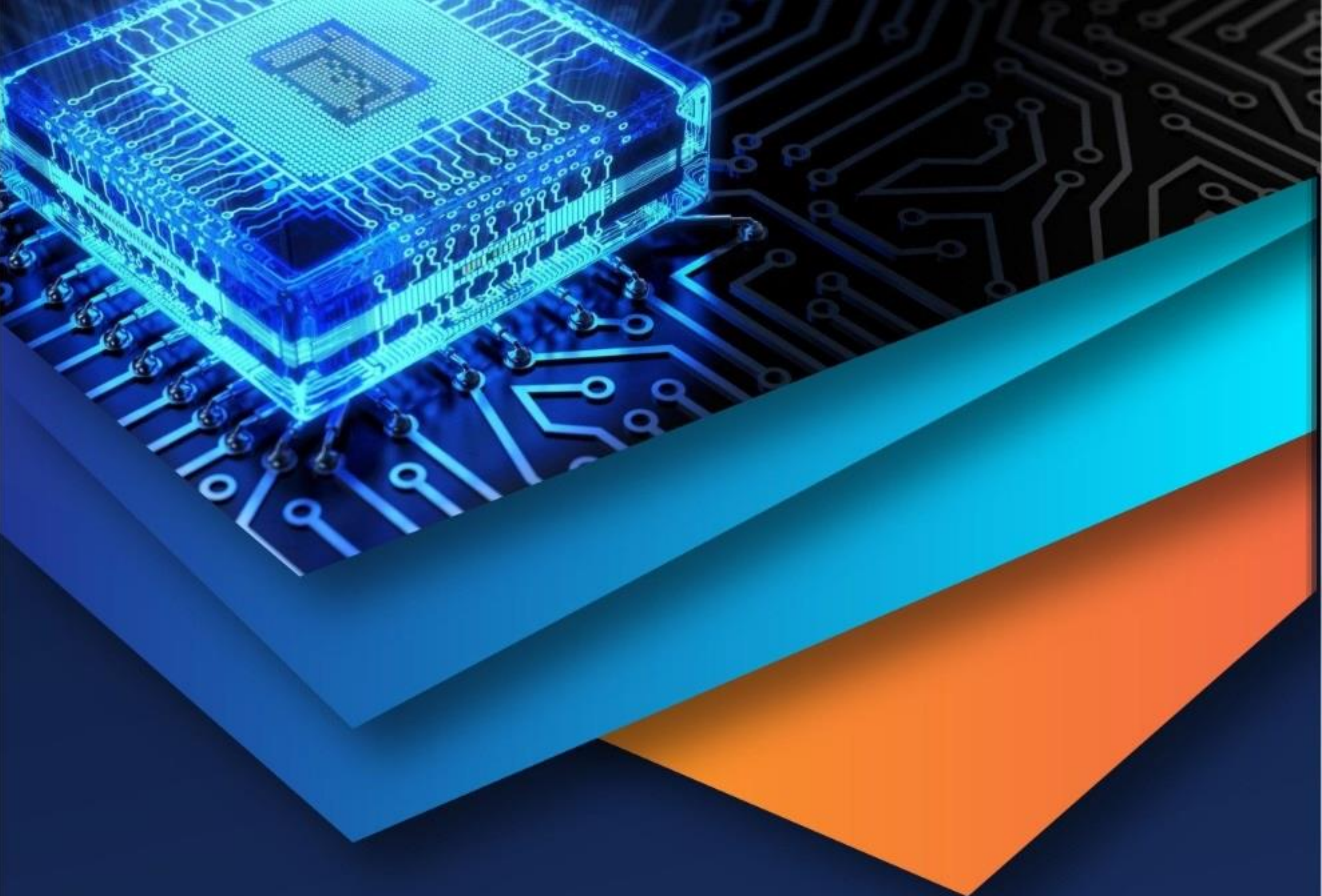
The experimental results show that the highest accuracy achieved is 0.947 and was obtained by Random Forest, Gradient Boosting, XGBoost. among these models Random Forest is selected as the final best model due to its high accuracy, robustness, resistance to overfitting, and suitability for structured medical datasets. This suggests that for structured health datasets tree-based ensemble learning models are highly effective.

## REFERENCES

[1] T. S. Alshammari, "Applying Machine Learning Algorithms for the Classification of Sleep Disorders," in IEEE Access, vol. 12, pp. 36110 36121, 2024, doi: 10.1109/ACCESS.2024.3374408.

[2] Locharla Ravikumar et al., (2025). SLEEP DISORDER PREDICTION USING MACHINE LEARNING. International Journal of Engineering Technology Research & Management (IJETRM), 09(07).

[3] Dr. V. Shanmugapriya et al., “Sleep disorder prediction using machine learning” International Journal of Scientific and Advanced Research in Technology, vol. 11, no. 3, 2025.




10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY