

# Sleep Disorder Prediction using Machine Learning

**Ms. Archana N G**

Assistant Professor

Dept of Computer Science and Engineering  
ACS College of Engineering  
Bangalore, India

**Ms. Prathiksha A.H**

Dept of Computer Science and  
Engineering  
ACS College of Engineering  
Bangalore, India  
[prathiksha.a.h.8@gmail.com](mailto:prathiksha.a.h.8@gmail.com)

**Ms. Kusuma C**

Dept of Computer Science and  
Engineering  
ACS College of Engineering  
Bangalore, India  
[kusumakushi283@gmail.com](mailto:kusumakushi283@gmail.com)

**Ms. Priya P**

Dept of Computer Science and  
Engineering  
ACS College of Engineering  
Bangalore, India  
[priyakushi303@gmail.com](mailto:priyakushi303@gmail.com)

**ABSTRACT** – Sleep disorders have become increasingly common due to changes in lifestyle, work patterns, and prolonged screen exposure, affecting both physical health and cognitive performance. Early identification of such disorders is essential, as untreated conditions may lead to long-term complications including cardiovascular issues, obesity, and reduced productivity. This study presents a data-driven approach for the prediction of sleep disorders using structured lifestyle and physiological attributes that can be collected without specialized clinical equipment. The proposed framework involves systematic data preprocessing, feature scaling, and the application of supervised learning models for multi-class classification. In particular, a comparative analysis between a statistical classifier and an ensemble learning technique is carried out to understand their behaviour on tabular healthcare data. Instead of relying only on overall accuracy, the evaluation focuses on class-level performance, feature influence, and model stability to reflect real-world applicability. The experimental results indicate that ensemble learning provides better generalization and more reliable separation between normal sleep patterns, insomnia, and sleep related cases. The study also highlights the practical feasibility of integrating the trained model into a lightweight web-based interface for preliminary screening. By using non-invasive and easily obtainable parameters, the proposed system offers a cost-effective support tool for early risk identification and preventive healthcare. The findings demonstrate that machine learning can assist medical decision-making by enabling scalable and accessible sleep health assessment.

**KEYWORDS:** Sleep disorder prediction, ensemble learning, Gradient Boosting, Quadratic Discriminant Analysis, healthcare analytics, lifestyle data.

## I. INTRODUCTION

Sleep is a fundamental biological process that plays a vital role in maintaining physical health, emotional stability, and cognitive efficiency. In recent years, rapid changes in lifestyle, increased work pressure, irregular sleep schedules, and excessive use of digital devices have contributed to a noticeable rise in sleep-related disorders across different age groups. Conditions such as insomnia and Sleep Apnea not only reduce the quality of life but are

also closely associated with serious health complications including hypertension, diabetes, obesity, and cardiovascular diseases.

Despite their impact, these disorders often remain undiagnosed for long periods because conventional diagnostic procedures require specialized clinical infrastructure, expert supervision, and overnight monitoring, making them expensive and inaccessible for routine screening. With the growing availability of digital health records and lifestyle datasets, machine learning has emerged as a promising tool for developing intelligent and scalable healthcare solutions. Instead of relying solely on complex physiological signal acquisition, structured data such as sleep duration, stress level, body mass index, heart rate, and daily physical activity can be analysed to identify patterns that indicate potential sleep abnormalities. These attributes are comparatively easier to collect and enable the

development of cost-effective prediction systems that can assist in early-stage health assessment.

of their sleep condition without the need for complex clinical tests. This approach is intended to support preventive healthcare by enabling early detection and encouraging timely medical consultation for high-risk cases. Furthermore, the proposed framework is designed to be computationally efficient and scalable so that it can be integrated into real-time health monitoring applications in the future. Through these objectives, the study seeks to provide a cost-effective and accessible alternative to traditional diagnostic methods while maintaining reliable prediction performance.

Another important objective of the proposed work is to examine the overall behaviour of the developed models in terms of stability, generalization capability, and computational efficiency when applied to healthcare data. In real-world applications, a prediction system must not only provide high accuracy but also maintain consistent performance for different input conditions. Therefore, the study aims to analyse class-wise prediction behaviour using appropriate validation techniques in order to ensure that the model does not favour a particular class while misclassifying others.

This is particularly significant in sleep disorder prediction, where the symptoms of different conditions often overlap and require careful differentiation. By focusing on balanced performance evaluation, the work attempts to improve the reliability of the system for practical deployment. In addition, the research intends to explore the feasibility of integrating the trained prediction model into a lightweight web-based environment so that the developed framework can be accessed without the need for high-end hardware or specialized software installation. Through this approach, the proposed work contributes toward building an intelligent, accessible, and preventive healthcare support system.

The primary objective of this research is to design a machine learning-based framework capable of classifying individuals into three categories: normal sleep, insomnia, and sleep apnea. To achieve this, both a statistical learning approach and an ensemble learning technique are implemented and evaluated under the same experimental conditions. The study not only compares their predictive performance but also examines their behaviour in terms of class-wise accuracy, feature contribution, and computational efficiency. Such an analysis is important in healthcare applications where model reliability and interpretability are as significant as overall accuracy.

The rest of the paper is structured to provide a clear understanding of the proposed work. A brief overview of the existing research in sleep disorder prediction and the role of machine learning in healthcare is presented in Section II. The characteristics of the dataset along with the preprocessing steps are discussed in Section III. Section IV explains the methodology adopted for model development, while Section V outlines the evaluation criteria used to measure performance. The experimental observations and their analysis are given in Section VI, followed by a detailed discussion in Section VII. Finally, the paper concludes by summarizing the key findings and highlighting possible directions for future research.

## II. RELATED WORKS

A detailed review of existing literature was carried out to understand the current progress in sleep disorder prediction using machine learning and data-driven healthcare systems. The following studies provide the research background and motivation for the proposed work:

**i. “Sleep Disorder Detection Using Machine Learning Techniques” — S. Patel and R. Shah (2023):** This study investigates the use of supervised learning algorithms for identifying sleep disorders from structured health data. The authors implemented Decision Tree, Support Vector Machine, and Random Forest models to classify sleep patterns based on attributes such as sleep duration, stress level, and body mass index. Their experimental results showed that ensemble methods achieved higher accuracy compared to individual classifiers due to their ability to handle non-linear relationships.

**ii. “Deep Learning Approaches for Automated Sleep Stage Classification” — M. Almutairi and K. Lee (2022):** This paper focuses on the application of deep learning models for analysing physiological signals such as EEG and ECG to classify sleep stages. Convolutional Neural Networks were used to extract temporal and spatial features automatically from raw signal data. The authors reported high classification accuracy; however, the

approach required complex data acquisition and high computational resources.

**iii. “Lifestyle-Based Health Risk Prediction Using Machine Learning” — R. Verma and P. Nair (2021):** This research proposes a predictive framework that uses lifestyle and behavioural attributes to assess health risks without relying on invasive clinical procedures. The authors demonstrated that parameters such as physical activity, heart rate, stress level, and sleep duration can be effectively used for classification tasks. Their work showed that structured datasets are suitable for real-time healthcare applications and can be integrated with web-based systems for instant prediction. The study directly influenced the dataset selection and preprocessing strategy used in the present work.

**iv. “Comparative Analysis of Statistical and Ensemble Learning Methods in Healthcare Data” — J. Kim and L. Park (2020):** The authors conducted a comparative evaluation of statistical models and ensemble learning techniques on medical datasets. Discriminant analysis provided faster training and better interpretability, while boosting-based models achieved higher prediction accuracy. The study concluded that ensemble learning reduces both bias and variance, making it suitable for multi-class classification problems. This comparison forms the basis for selecting Quadratic Discriminant Analysis and Gradient Boosting for the proposed system.

**v. “Web-Based Machine Learning Systems for Real-Time Health Monitoring” — D. Fernandes and T. Joseph (2023):** This paper explores the deployment of machine learning models using lightweight web frameworks for real-time health assessment. The authors integrated trained models into a browser-based interface that allowed users to input health parameters and obtain immediate prediction results. Their findings demonstrate that such systems can support preventive healthcare by enabling early risk identification without frequent hospital visits. This work influenced the implementation of the web-based prediction module in the present study.

## III. PROBLEM STATEMENT

Sleep disorders such as insomnia and sleep apnea have become increasingly prevalent due to modern lifestyle patterns, work-related stress, irregular sleep schedules, and reduced physical activity. These conditions often remain undiagnosed for long periods because their identification typically requires clinical procedures such as polysomnography, which are expensive, time-consuming, and not easily accessible to a large population. As a result, many individuals fail to receive timely medical attention, leading to severe long-term health complications including cardiovascular diseases, obesity, and decreased cognitive performance.

Existing diagnostic systems primarily depend on physiological signal acquisition using specialized medical equipment. While these methods provide high accuracy, they are not suitable for continuous monitoring or early-stage screening in everyday environments. In addition, most traditional approaches are designed for hospital settings and cannot be easily extended to real-time, user-friendly applications. This creates a significant gap between the availability of healthcare technology and its practical accessibility for common users.

With the growing availability of structured lifestyle and health-related data, there is an opportunity to develop intelligent systems that can identify sleep disorders at an early stage without relying on complex clinical infrastructure. However, several challenges still exist. Many of the current machine learning-based solutions focus on a single classification technique and do not analyse the comparative performance of different models on the same dataset. Some approaches achieve high accuracy but lack interpretability, while others are computationally efficient but fail to capture complex non-linear relationships in the data. Furthermore, limited work has been carried out on integrating such prediction models into real-time environments where users can obtain instant result.

#### IV. OBJECTIVES

The primary objective of this research is to develop an efficient and scalable machine learning-based framework for the early prediction of sleep disorders using structured lifestyle and physiological data. The study aims to analyse the influence of various health-related attributes such as sleep duration, stress level, body mass index, physical activity, and heart rate in identifying abnormal sleep patterns. In order to achieve reliable classification, the work focuses on implementing both a statistical learning approach and an ensemble learning technique and evaluating their performance under the same experimental conditions. A comparative analysis is carried out to determine the most suitable model for multi-class classification of normal sleep, insomnia, and sleep apnea. Another important objective is to understand the contribution of individual features in the prediction process so that the most significant parameters affecting sleep health can be identified.

In addition to model development, the research also aims to design a practical and user-friendly system that allows individuals to input their lifestyle parameters and obtain an instant assessment of their sleep condition without the need for complex clinical tests. This approach is intended to support preventive healthcare by enabling early detection and encouraging timely medical consultation for high-risk cases. Furthermore, the proposed framework is designed to be computationally efficient and scalable so that it can be integrated into real-time health monitoring applications in the future. Through these objectives, the study seeks to provide a cost-effective and accessible alternative to traditional diagnostic methods while maintaining reliable prediction performance.

Another important objective of the proposed work is to examine the overall behaviour of the developed models in terms of stability, generalization capability, and computational efficiency when applied to healthcare data. In real-world applications, a prediction system must not only provide high accuracy but also maintain consistent performance for different input conditions. Therefore, the study aims to analyse class-wise prediction behaviour using appropriate validation techniques in order to ensure that the model does not favour a particular class while misclassifying others. The long-term objective is to create a foundation that can be further extended by incorporating larger datasets, real-time data acquisition through wearable devices, and advanced learning techniques for continuous sleep health monitoring. Through this approach, the proposed work contributes toward building an intelligent works.

#### V. METHODOLOGY

The development of the proposed sleep disorder prediction system follows a structured and modular approach to ensure reliable classification, efficient data handling, and practical deployment. The methodology is divided into four major phases, where each phase addresses a specific stage in the overall prediction pipeline. The system is designed to process structured lifestyle data, perform intelligent analysis, and generate real-time prediction results through an integrated environment.

##### Phase 1: Data Acquisition and Preprocessing

The first phase focuses on collecting and preparing the dataset for model training. The sleep health dataset consists of demographic, behavioural, and physiological attributes such as age, sleep duration, stress level, body mass index, heart rate, and physical activity. Since the dataset contains both numerical and categorical features, preprocessing is carried out to convert categorical values into machine-readable form using encoding techniques. Feature scaling is applied to ensure that parameters with larger numerical ranges do not dominate the learning process. In addition, missing values and redundant attributes are handled carefully to improve data quality. The processed dataset is then divided into training and testing subsets to enable unbiased performance evaluation of the models.

##### Phase 2: Feature Engineering and Exploratory Analysis

In the second phase, an in-depth analysis of the dataset is performed to identify the most influential attributes affecting sleep disorders. Statistical observations and correlation analysis are used to understand the relationship between lifestyle factors and sleep conditions. This step helps in reducing dimensionality by eliminating irrelevant features and improving the learning efficiency of the model. The feature engineering process also ensures that the selected attributes contribute meaningfully to the classification task. By refining the input data in this phase, the system achieves better generalization and avoids unnecessary computational complexity during model training.

##### Phase 3: Model Development and Comparative Analysis

The third phase involves the implementation of machine learning models for multi-class classification of sleep conditions. In this work, a statistical learning approach and an ensemble learning technique are trained and tested under identical conditions. The statistical model provides a mathematically interpretable baseline, while the ensemble model improves prediction performance by handling complex non-linear relationships in the dataset. Hyperparameter tuning is carried out to obtain optimal model configurations and to avoid overfitting. The performance of both models is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. A confusion matrix is also used to analyse class-wise prediction behaviour and to ensure balanced performance across all sleep disorder categories. In order to improve the robustness of the classification process, cross-validation was carried out during the training stage so that the models could be evaluated on multiple data partitions rather than a single split. This helped in obtaining a more reliable estimate of their generalization capability statistical models and ensemble learning techniques on medical datasets. Discriminant analysis provided faster training and better interpretability, while boosting-based models achieved higher prediction accuracy. The study concluded

that ensemble learning reduces both bias and variance, making it suitable for multi-class classification problems. This comparison forms the basis for selecting Quadratic Discriminant Analysis and Gradient Boosting for the proposed system.

#### Phase 4: System Integration and Real-Time Prediction

The final phase focuses on integrating the trained model into a lightweight and user-friendly prediction environment. A web-based interface is developed in which users can enter their lifestyle and health-related parameters to obtain an instant prediction of their sleep condition. The backend processes the input data, applies the trained model, and returns the classification result in real time. This phase demonstrates the practical feasibility of the proposed framework as a preliminary screening tool for preventive healthcare. The modular design of the system also allows future integration with cloud platforms and wearable health monitoring devices for continuous sleep assessment.

## VI. RESULT

The performance of the developed models was evaluated using the testing dataset to measure their ability to classify sleep conditions into normal, insomnia, and sleep apnea categories. The experimental results indicate that both approaches were capable of learning meaningful patterns from the structured lifestyle data; however, a noticeable difference was observed in their prediction behaviour and generalization capability. The statistical model produced stable and consistent results with comparatively lower computational complexity, making it suitable as a baseline for classification. On the other hand, the ensemble learning technique demonstrated superior predictive performance due to its ability to model complex non-linear relationships among the input features.



A detailed class-wise analysis revealed that the ensemble model achieved better separation between the three sleep condition categories, particularly in cases where the symptoms showed overlapping characteristics. This improvement can be attributed to its iterative learning mechanism, which minimizes the error of the previous stage and refines the decision boundaries. The confusion matrix obtained from the experiment showed a reduction in misclassification of minority classes, indicating balanced performance across all categories. In addition to overall accuracy, precision and recall values confirmed that the proposed approach was able to identify high-risk cases more effectively without significantly increasing false predictions.

Feature influence analysis highlighted that parameters such as stress level, sleep duration, and body mass index played a major role in determining the output class.

Individuals with higher stress levels and irregular sleep duration were more likely to be classified under insomnia, whereas increased body mass index and abnormal heart rate patterns showed stronger association with sleep apnea cases. This observation aligns with real-world medical findings and validates the relevance of the selected dataset for sleep disorder prediction.

## VII. DISCUSSION

The experimental results demonstrate that machine learning can effectively identify sleep disorder patterns using structured lifestyle and physiological attributes without the need for complex clinical data. The comparative analysis between the statistical model and the ensemble learning approach highlights a clear trade-off between interpretability and predictive performance. While the statistical classifier provided stable and computationally efficient results, the ensemble model achieved better classification accuracy and improved class separation. This behaviour can be attributed to its iterative learning mechanism, which refines the decision boundaries by minimizing the errors of the previous stage. Such an approach is particularly useful in healthcare datasets where multiple attributes interact in a non-linear manner.

**1) Model Performance Comparison:** The comparative evaluation shows that the ensemble learning model achieved better classification accuracy and class separation than the statistical approach. This improvement is mainly due to its ability to capture complex non-linear relationships among multiple health parameters, which are common in sleep disorder datasets.

**2) Impact of Lifestyle Attributes:** The analysis indicates that stress level, sleep duration, and body mass index are the most influential features in determining sleep health. This confirms that sleep disorders are strongly associated with daily behavioural patterns rather than a single medical factor.

**3) Balanced Multi-Class Classification:** The confusion matrix demonstrates that the proposed system maintains balanced prediction across normal sleep, insomnia, and sleep apnea categories. This is important for healthcare applications, as biased predictions toward a majority class can lead to incorrect risk assessment.

**4) Practical Deployment Capability:** The integration of the trained model into a web-based interface enables real-time prediction with minimal response time. This makes the system suitable for preliminary screening and supports preventive healthcare by allowing users to assess their sleep condition without clinical testing.

**5) Limitations and Future Scope:** The current model performance depends on the size and diversity of the dataset, and the use of larger real-time data can further improve accuracy and generalization. Future work can focus on integrating wearable sensor data and cloud-based monitoring for continuous sleep health assessment.

The obtained results also indicate that the proposed system is capable of handling real-world data variations without significant performance degradation. The consistency observed across different evaluation metrics suggests that the model is not overfitted

to a specific data split and can generalize well to unseen samples. Another important observation is that the use of structured lifestyle data makes the prediction process more practical and user-friendly when compared to approaches that rely on complex physiological signals. This improves the accessibility of the system for individuals who do not have access to specialized medical facilities. The study further highlights the importance of data quality in healthcare-based machine learning applications. Even a well-performing model can produce unreliable predictions if the input data is inconsistent or incomplete. Therefore, proper preprocessing and feature selection play a critical role in maintaining the stability of the system. Overall, the discussion confirms that the proposed approach provides a practical balance between prediction accuracy, computational efficiency, and real-time usability for early sleep disorder screening.

## VIII. LUSION

The study presented a machine learning-based framework for the early prediction of sleep disorders using structured lifestyle and physiological attributes. The proposed approach demonstrates that reliable classification of normal sleep, insomnia, and sleep apnea can be achieved without relying on complex clinical procedures or physiological signal acquisition. A comparative analysis between the statistical learning model and the ensemble learning technique revealed that the ensemble approach provides improved predictive performance and better class separation, while the statistical model offers computational simplicity and interpretability. The experimental observations also confirmed that factors such as stress level, sleep duration, and body mass index play a significant role in determining sleep health, which aligns with real-world medical findings.

## REFERENCES

- [1] -H. Liu, S.-Y. Chien, Y.-L. Wu, T.-H. Sun, C.-S. Huang, K.-C. Hsu, and L.-W. Hang, "Efficient Net-based machine learning architecture for sleep apnea identification," 2024.
- [2] Alazaidah, G. Samara, M. Alijaidi, M. H. Qasem, A. Alsarhan, and M. Alshammari, "Potential of machine learning for predicting sleep disorders: A comprehensive analysis of regression and classification models," 2024.
- [3] A. Hassan and R. Patel, "Using gradient boosting and random forest for predicting sleep quality," 2024.
- [4] Thomas and L. Martin, "ECG-based sleep apnea detection using convolution neural networks," 2024.
- [5] Y. S. Taspinar and I. Cinar, "Prediction of sleep health status, visualization and analysis of data," 2023.
- [6] . Lakshmi and M. Raghuvver, "Classification of sleep disorders using random forest on sleep health and lifestyle dataset," 2023.
- [7] ] R. Das and S. Tiwari, "Machine learning models for lifestyle-based sleep disorder prediction," 2023.
- [8] R. Shalini and G. Kumar, "Wearable-sensor-based sleep monitoring using unsupervised learning techniques," 2022.
- [9] L. Zhao and M. Wu, "Machine learning approaches for sleep stage classification using heart rate variability," 2022.
- [10] P. Tripathi, M. A. Ansari, "Ensemble computational intelligence for insomnia sleep stage detection via sleep ECG signal," 2022.

The integration of the trained model into a lightweight web-based interface highlights the practical feasibility of the system for real-time preliminary screening. This enables users to obtain an instant assessment of their sleep condition using easily obtainable health parameters, thereby supporting preventive healthcare and reducing the dependency on traditional diagnostic methods. The proposed framework is cost-effective, scalable, and suitable for deployment in environments where access to specialized medical infrastructure is limited.

Although the current work provides promising results, the performance of the system can be further enhanced by using larger and more diverse datasets. Future research can focus on incorporating real-time data collected from wearable devices and exploring advanced learning techniques for continuous sleep health monitoring. With these improvements, the proposed model has the

potential to evolve into a comprehensive decision-support tool for both individuals and healthcare professionals.

The proposed framework also demonstrates the potential of data-driven healthcare systems in promoting awareness about sleep health among the general population. By transforming routinely

available lifestyle information into meaningful clinical insight, the system encourages individuals to monitor their daily habits and make informed decisions to improve their sleep quality. The modular nature of the developed model allows it to be extended for integration with mobile health applications and cloud-based platforms, enabling continuous and remote health assessment. In the long term, such intelligent and accessible solutions can contribute to reducing the burden on healthcare institutions by supporting early risk identification and timely medical intervention.