Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 33 No. 1s (2026)

Clustering and Classification Techniques for the Identification of Sleep Disorders Using the Sleep Health and Lifestyle of Worker

Mrs. Gundra S G E Sai Sree
Assistant Professor,Department of Artificial Intelligence and Data Science
Vignan Institute of Technology and Science
Deshmukhi, Yadadri Bhuvanagiri, Telangana, India
vignansatya@gmail.com

Trisha Erlapalli
Department of Artificial Intelligence and Data Science
Vignan Institute of Technology and Science
Deshmukhi, Yadadri Bhuvanagiri, Telangana, India
erlapallitrisha@gmail.com

Patla Harshitha
Department of Artificial Intelligence and Data Science
Vignan Institute of Technology and Science
Deshmukhi, Yadadri Bhuvanagiri, Telangana, India
harshithapatla21@gmail.com

Mogilagoni Eekshitha
Department of Artificial Intelligence
and Data Science
Vignan Institute of Technology and Science
Deshmukhi,Yadadri Bhuvanagiri, Telangana,India
eekshithayadav@gmail.com

Article History:

Received: 04-02-2026

Revised: 20-03-2026

Accepted: 10-04-2026

Abstract:
Sleep disorders including Insomnia and Obstructive Sleep Apnea have a major impact on physical and psychological health and quality of life. Precise and timely detection of these diseases is essential for successful diagnosis and cure. Conventional expert-diagnosing the sleep disorder is based on manual interpretation of experience and it is very slow and subjective. As more lifestyle and health data can be retrieved, machine learning methods present an effective substitute in automatic detection of sleep disorder. In this paper, we present a comparative approach that evaluates unsupervised and supervised machine learning based sleep disorder discovery using the Sleep Health and Lifestyle Dataset. In the system presented, clustering methodologies - K-Means and DBSCAN - were initially used to find subtle pattern, and group people by sleep and lifestyle. Confusion matrices and classification reports evaluate clustering findings by associating cluster labels with sleep disorder classes. Furthermore, supervised classification algorithms are applied in the same dataset to directly predict sleep disorders. The performance of clustering-based methods and classification-based approaches has been considered by comparing them using the usual measures such as accuracy, precision, recall, F1-score and confusion matrices. The experimental analysis reveals

the advantages and disadvantages of those approaches, showing that while clustering does find natural trends in the data, if labeled data are present, supervised classification methods can lead to a larger prediction.

Keywords— Sleep Disorder Detection, Machine Learning, Clustering Algorithms, Classification Algorithms, K-Means, DBSCAN, Sleep Health Dataset, Healthcare Analytics

Introduction

Sleep is a life-sustaining vital biological function that is known for its critical role in the preservation of human physical health, mental health, emotional health, and general wellness. However, with changes in lifestyle, stressors, irregular work schedules, and underlying health conditions, sleep disorders like Insomnia and Obstructive Sleep Apnea (OSA) have become more common in modern society.

These disorders can cause serious consequences, including fatigue, loss of productivity, cardiovascular complications, mental health disorders, and reduced quality of life. Therefore, early detection and accurate classification of sleep disorders is crucial for timely medical treatment and proper treatment planning.

Performance of clustering-based and classification-based approaches is examined with standard metrics including accuracy, precision, recall, F1-score, and confusion matrices. This comparison of machine learning techniques can be used to explore the efficacy of various models for healthcare analytics.

With the goal of assisting healthcare professionals, the proposed system is to generate an intelligent algorithm for timely decision support to assist individuals by identifying early sleep disorders. Finally, this study supports the process of working toward efficient, scalable, data-driven, and optimized healthcare by increasing diagnostic reliability and improving patient outcomes.

I. RELATED WORK

Recent developments in machine learning technologies have yielded drastic gains in sleep disorders detection accuracy and efficacy. Simple diagnostic techniques for sleep disorder diagnoses including subjective sleep diaries, clinical interviews and polysomnography tests often are not fast, cheap, and expandable.

Malik, Gupta and Sharma (2024) suggested a prediction analytics model using machine learning algorithm that is suitable to overcome the shortcomings of conventional sleep disorder detection. The study compared regression and classification algorithms for sleep related conditions. AdaBoost and XGBoost classifiers reached 93.49% and 92.73%, respectively, demonstrating their potential to optimize prediction accuracy on earlier-stage diagnosis and for timely intervention (Horton, 2009; Rizal et al.).

Rahman et al. (2025) proposed a framework for fine-tuning machine learning algorithms for sleep disorder classification with Sleep Health and Lifestyle Dataset. The study utilized a Gradient Boosting Regressor for feature selection and evaluated fifteen classifiers of machine learning. Our experimental data showed that ensemble methods including Gradient Boosting, Voting, CatBoost and Stacking had a good classification accuracy.

II. PROPOSED METHODOLOGY

A. System Overview

The sleep disorder detection system is developed utilizing Python as it has plenty of support for Machine Learning applications, analysis of data, and healthcare analytics platforms. Data processing is performed by treating missing values and inconsistent entries with statistical imputation and scaling methods to guarantee a good model for the input. Outlier detection techniques are similarly used in order to maintain dataset consistency and improve model performance.

Feature engineering is carried out by encoding and normalizing important attributes that influence sleep quality. Additional features like sleep quality indices and lifestyle risk scores are derived using domain-based logic to enhance the predictive capability of the models.

Unsupervised learning models such as K-Means and DBSCAN clustering approaches are applied to detect latent patterns and natural clusters within the sleep health data, helping to uncover hidden structures. Cluster validation is accomplished by mapping cluster groups to previous sleep disorder classification systems, and performance is assessed using confusion matrices and classification reports.

Supervised learning models including Logistic Regression, Random Forest, Support Vector Machine (SVM), Gradient Boosting, and XGBoost are developed using scikit-learn and associated frameworks. These models are trained using labeled data to predict specific sleep disorders such as insomnia and sleep apnea. The performance of both clustering-based and classification-based methods is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrices.

Predicted sleep disorders and feature correlations are visualized using Matplotlib and Seaborn. Well-trained models are saved in .pkl or .sav formats using pickle or joblib for reuse and quick deployment. The prediction pipeline is integrated into a lightweight analytical interface for real-time usage. The system is implemented on a standard computing setup with an Intel i5 processor and 8GB RAM, demonstrating that the framework requires minimal computational resources and is efficient.

B. Algorithm

Step 1: System Initialization
Import required machine learning and data analysis libraries such as NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn.

Step 2: Sleep Health Dataset Acquisition
Load the Sleep Health and Lifestyle Dataset containing attributes related to sleep patterns, lifestyle habits, and health indicators.

Step 3: Pre-processing and Cleaning
Handle missing values, remove duplicate records, and normalize inconsistent entries in the dataset.

Step 4: Exploratory Data Analysis (EDA)
Compile statistics such as mean, variance, correlation, and feature distributions.

Step 5: Feature Selection and Extraction
Use significant sleep and lifestyle attributes for the detection of disorders.

Step 6: Identification of Clustering-based Sleep Disorders
Establish unsupervised learning algorithms like K-Means and DBSCAN to group people according to similar behavioral and health characteristics.

Step 7: Supervised Classification Model Development
Investigate and train classification models such as Logistic Regression, Random Forest, Support Vector Machine, or Decision Tree on labeled data.

Step 8: Model Evaluation and Comparing Performance
Evaluate performance using confusion matrices, accuracy, precision, recall, and F1-score to measure clustering and classification outcomes.

Step 9: Visualization and Interpretation
Produce performance comparison charts, cluster visualizations, and classification-based results graphs.

Step 10: Decision Support and System Integration
Conclude if the application of machine learning techniques is suitable for real-world sleep disorder prediction.

C. System Architecture

The system architecture for the Machine Learning-Based Sleep Disorder Detection Framework provides an intelligent automated framework to identify sleep disorders such as Insomnia and Obstructive Sleep Apnea. The architecture combines unsupervised clustering methods and supervised classification models for better analysis of sleep health patterns.

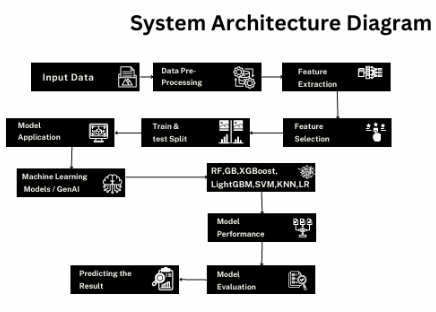
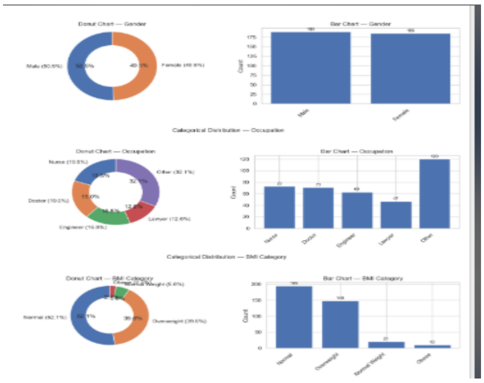


Fig. 1. System Architecture of Proposed Model

Data acquisition and preprocessing involve obtaining sleep health information using either user-defined attributes or a structured format such as the Sleep Health and Lifestyle Dataset.

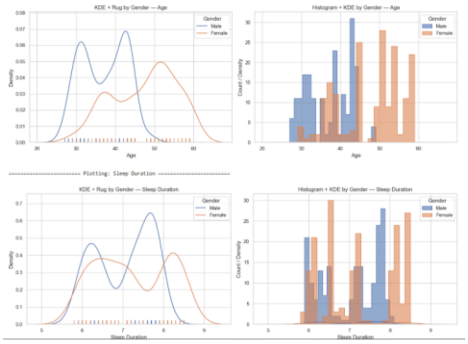
Pattern discovery by clustering techniques is performed by first analyzing the preprocessed dataset using unsupervised learning techniques like K-Means or DBSCAN. These clustering methods detect hidden patterns as well as naturally occurring groupings among people based on sleep activities and lifestyle-related information.

Supervised classification and comparative analysis are then carried out, where supervised methods are used to classify sleep disorder categories directly using labeled data after clustering analysis.



Visualization and the decision support interface present the final prediction results and analytical insights through an interactive visualization interface.

The data flow and integration of the sleep disorder detection system follow a structured analytical pipeline. Input consists of sleep health and lifestyle parameters derived from user input or existing datasets representing real-world sleep activity and health conditions. Processing involves preprocessing the input data, applying clustering algorithms to identify natural groupings, and subsequently using supervised classification to predict sleep disorder categories. Output includes sleep disorder prediction results, clustering insights, and performance evaluation reports presented through the visualization interface.



III. RESULTS AND DISCUSSION

A. *Testing Methodology*

The Machine Learning-Based Sleep Disorder Detection System follows a systematic and structured testing methodology to ensure that the proposed model fulfills reliability, analytical, and scalability requirements, as well as healthcare decision support capabilities for sleep disorder detection. Test case development involves creating an extensive array of test cases representing variations in sleep duration, stress levels, physical activity, heart rate, body mass index, and daily life habits to examine sleep health conditions and lifestyle characteristics. These tests provide insights into how sleep and mental health differ.

User testing is conducted with healthcare learners, research students, and general users who evaluate the analytical outputs produced by the machine learning models. Automated evaluation is performed using bulk records from the Sleep Health and Lifestyle Dataset to assess clustering and classification models under large sample sizes, ensuring robustness and consistency of the methodologies. During the testing phase, extensive experimental logs are maintained to monitor cluster behavior, classification results, and AI model inference responses.

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 33 No. 1s (2026)

Evaluation of the sleep disorder detection framework is carried out analytically using multiple performance metrics to measure prediction effectiveness and model interpretability.

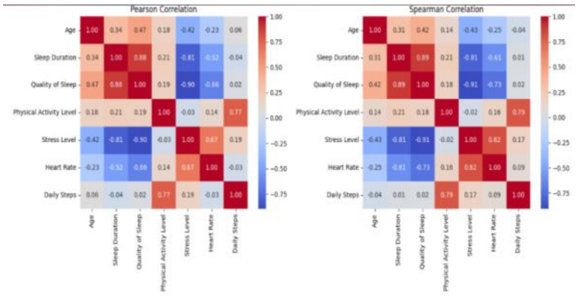


TABLE I. SYSTEM PERFORMANCE OBSERVATION

Clustering Pattern Consistency	Stable grouping across repeated runs
Classification Reliability	Strong predictive decision support
Model Response Time	Fast analytical processing
Insight Interpretability	Clear and understandable results
Precision	Effective disorder differentiation
Recall	Reliable identification of sleep risk patterns
F1-Score	Balanced predictive performance
Overall System Robustness	Consistent analytical outcomes

B. *Performance Evaluation*

The performance assessment concentrates on key metrics that determine the healthcare analytics system’s suitability for practical use and automatic monitoring of sleep disorders. Results of analytical processing speed are evaluated by measuring the time taken to process input sleep and lifestyle parameters, conduct clustering analysis, perform classification predictions, and generate comparative evaluation reports.

The effectiveness of disorder prediction is assessed by comparing cluster group interpretations with classification-based disorder predictions using standard analytical validation tools such as confusion matrices and classification reports.

Interpretability and practical considerations are also examined by evaluating the usefulness of the analytical insights generated based on clarity, contextual relevancy, and their significance in enhancing healthcare awareness.

Fig. 2. Sleep Health Web Interface

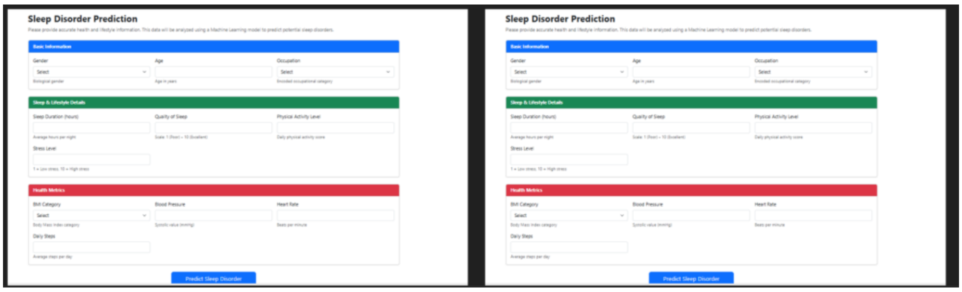


Fig. 3. Sleep Disorder Classification Output

IV. CONCLUSION

Towards a global understanding of sleep disorder as sleep disturbances such as Insomnia and OSA is also very prominent in contemporary health system The identification and diagnosis of sleep disturbances such as Insomnia and Obstructive Sleep Apnea constitute fundamental problems in the current healthcare system. Conventional diagnosis is mostly manual, requiring human interpretation based on professional judgment as well as evidence from clinical observation which are slow and subjective in nature and susceptible to subjective judgment/biases. To mitigate such issues, in this work, a new methodologies of comparative machine learning has been proposed to attempt to improve the accuracy and efficiency of automated sleep disorder detection in automated sleep disorder detection using machine learning techniques to address these limitations. The proposed system used the Sleep Health and Lifestyle Dataset to analyse different lifestyle and physiological traits around sleep conditions. Clustering methods such as K-Means and DBSCAN were deployed as first methods to get insights into latent patterns and natural groups in the dataset. Unsupervised methods such as those used here helped finding hidden structures that might not be easily seen on manual inspection. Clustering outputs were organized and evaluated systematically by mapping labels to known disorder classes through confusion matrices and classification reports.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to **Mrs. GUNDRA S G E SAI SREE**, Department of Artificial Intelligence and Data Science, for his valuable guidance, support, and encouragement throughout the completion of this work.

REFERENCES

[1] F. Mendonça, S. S. Mostafa, F. Morgado-Dias, and A. G. Ravelo-García, "A portable wireless device for cyclic alternating pattern estimation from an EEG monopolar derivation," *Entropy*, vol. 21, no. 12, p. 1203, Dec. 2019.

[2] Y. Li, C. Peng, Y. Zhang, Y. Zhang, and B. Lo, "Adversarial learning for semi-supervised pediatric sleep staging with single-EEG channel," *Methods*, vol. 204, pp. 84–91, Aug. 2022.

[3] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.

[4] V. Singh, V. K. Asari, and R. Rajasekaran, "A deep neural network for early detection and prediction of chronic kidney disease," *Diagnostics*, vol. 12, no. 1, p. 116, Jan. 2022.

[5] J. Van Der Donckt, J. Van Der Donckt, E. Deprost, N. Vandenbussche, M. Rademaker, G. Vandewiele, and S. Van Hoecke, "Do not sleep on traditional machine learning: Simple and interpretable techniques are competitive to deep learning for sleep scoring," *Biomedical Signal Processing and Control*, vol. 81, Mar. 2023, Art. no. 104429.

[6] Y. Yang, Z. Gao, Y. Li, and H. Wang, "A CNN identified by reinforcement learning-based optimization framework for EEG-based state evaluation," *Journal of Neural Engineering*, vol. 18, no. 4, Aug. 2021, Art. no. 046059.

[7] Y. J. Kim, J. S. Jeon, S.-E. Cho, K. G. Kim, and S.-G. Kang, "Prediction models for obstructive sleep apnea in Korean adults using machine learning techniques," *Diagnostics*, vol. 11, no. 4, p. 612, Mar. 2021.

[8] Z. Mousavi, T. Y. Rezaii, S. Sheykhivand, A. Farzamnia, and S. N. Razavi, "Deep convolutional neural network for classification of sleep stages from single-channel EEG signals," *Journal of Neuroscience Methods*, vol. 324, Aug. 2019, Art. no. 108312.

[9] S. Djanian, A. Bruun, and T. D. Nielsen, "Sleep classification using consumer sleep technologies and AI: A review of the current landscape," *Sleep Medicine*, vol. 100, pp. 390–403, Dec. 2022.

[10] N. Salari, A. Hosseinian-Far, M. Mohammadi, H. Ghasemi, H. Khazaie, A. Daneshkhah, and A. Ahmadi, "Detection of sleep apnea using machine learning algorithms based on ECG signals: A comprehensive systematic review," *Expert Systems with Applications*, vol. 187, Jan. 2022, Art. no. 115950.