International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1450-1456
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0197

SleepSight: Sleep Disorder Risk Assessment and Recommendations

Mrs. Sunayana Sutar[1], Ms. Ashwini Dhumal[2], Omkar Nimse[3], Gururaj Bhase[4], Om Patil[5], Akshit Garg[6]
[1,2]Professor, Dept. of AI&DS, Dr D Y Patil Institute of Engineering, Management and Research, Pune
[3,4,5,6]UG Scholar, Dept. of AI&DS, Dr D Y Patil Institute of Engineering, Management and Research, Pune
Emails: *sunayna.sutar@dypiemr.ac.in[1], ashwini.dhumal@dypiemr.ac.in[2], omkarvijaynimse@gmail.com[3], gururajbhase@gmail.com[4], 11ompatil2004@gmail.com[5], gargakshit556@gmail.com[6]*

Abstract
Sleep disorders such as insomnia, sleep apnea, narcolepsy and restless leg syndrome have been on the rise due to reasons like stress, irregular lifestyle habits and lack of public awareness. Early detection is the cure, but most contemporary solutions track only sleep patterns without providing accurate risk estimates, hence creating a vast chasm in preventive healthcare. SleepSight is a web application based on machine learning technology that estimates the likelihood of different sleep disorders based on demographic and lifestyle details like age, gender, sleep duration, stress and activity. The architecture encompasses data preprocessing, optimized multi-class classification and lean architecture to provide accurate and rapid predictions. A set of algorithms like Random Forest, XGBoost, Logistic Regression and Support Vector Machines will be tested and the most efficient model will be implemented. Users are provided with percentage-based risk scores and personalized improvement tips through an easy-to-use interface. The process facilitates proactive lifestyle intervention, encourages frequent self-checks and on-time medical consultations as and when needed. The target users are individuals, wellness professionals, telemedicine practitioners and corporate wellness programs with a larger goal of sensitizing the public, enabling early intervention and encouraging overall sleep wellness at community level
Keywords: *Logistic Regression; Random Forest; SleepSight; Support Vector Machines; XGBoost*

1. Introduction

1.1.Background and Problem Context

Sleep disorders constitute a growing global health concern due to their strong association with physical, cognitive, and psychological impairments. Conditions such as insomnia and obstructive sleep apnea often remain undiagnosed for extended periods, leading to reduced quality of life and increased risk of chronic diseases. Conventional diagnostic methods, including polysomnography, provide accurate clinical assessment but are expensive, time-consuming, and impractical for large-scale or early screening. Consequently, many individuals fail to receive timely evaluation and intervention, highlighting the need for accessible and preventive sleep health assessment approaches[1].

1.2. Related Work and Limitations

Recent studies have demonstrated the effectiveness of machine learning techniques in sleep analysis and sleep disorder detection. Both classical and deep learning models have been applied to physiological signals, questionnaire data, and lifestyle attributes to classify sleep stages and predict disorder severity (Sekkal et al., 2022; Xu et al., 2022). Although sensor-based deep learning approaches achieve high predictive accuracy, they often require complex data acquisition setups and substantial computational resources, limiting their deployment beyond controlled clinical environments. Other research has explored health-risk prediction using lifestyle and sleep-related features, emphasizing the feasibility of non-invasive assessment methods (Birari et al., 2023; Rajan et al., 2023). However, many existing systems focus on individual disorders or lack real-time, user-centric implementation[9].

1.3. Research Gap and Motivation

Despite significant progress in sleep analytics, there remains a gap in the development of lightweight, web-based systems that provide comprehensive sleep disorder risk assessment using easily obtainable

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1450-1456
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0197

inputs[13]. Current approaches either rely heavily on clinical instrumentation or lack scalability and interpretability for general users. Addressing this gap requires a balanced framework that combines reliable machine learning models with accessible deployment, enabling early awareness without replacing professional diagnosis[2].

1.4. Objectives and Originality of Proposed Work

The primary objective of this work is to develop SleepSight, a machine learning–based sleep disorder risk assessment system that utilizes structured sleep and lifestyle data to provide early risk insights. The originality of the proposed approach lies in its integration of multiple machine learning models with a user-friendly web interface for real-time interaction and prediction. By emphasizing non-invasive data, interpretability, and accessibility, SleepSight advances the state of the art toward practical, scalable, and preventive sleep healthcare solutions[3].

2. Literature Review

Recent research in sleep analytics has increasingly focused on the use of machine learning techniques to automate sleep stage classification, sleep disorder detection, and risk prediction. Advances in sensing technologies and the availability of physiological, behavioral, and lifestyle data have enabled the development of computational models capable of identifying sleep-related abnormalities with improved accuracy[4]. These approaches aim to reduce dependence on traditional clinical diagnostics, such as polysomnography, which are costly and not scalable for population-level screening. Several studies have explored deep learning architectures using multimodal physiological signals, including electroencephalography, heart-rate variability, respiration patterns, and motion data. Convolutional neural networks and recurrent neural networks have been widely adopted to capture spatial and temporal patterns in sleep signals[14]. While such methods demonstrate high predictive performance, they often require large labeled datasets, specialized sensors, and significant computational resources, limiting their applicability in real-world and resource-constrained environments. In parallel, researchers have investigated classical machine learning and ensemble-based approaches using demographic, questionnaire-based, and lifestyle features. However, challenges remain in terms of dataset heterogeneity, reliance on self-reported inputs, and limited generalization across populations[15]. Overall, the existing literature highlights a trade-off between model complexity and real-world usability. While sensor-heavy deep learning systems offer high accuracy, lightweight and user-centric machine learning models provide better scalability and accessibility shown in Table 1.

Table 1: Literature Review

Paper Title	Dataset Type	Models / Algorithms Used	Limitations
Simultaneous Sleep Stage and Sleep Disorder Detection from Multimodal Sensors	Multimodal physiological data (EEG, PPG, respiration, accelerometer)	CNN–LSTM multi-task framework	Requires large labeled data; high computational cost; sensor variability
Automatic Sleep Stage Classification: From Classical ML to Deep Learning	Public PSG datasets (EEG, EOG, EMG)	SVM, Random Forest, k-NN, CNN, LSTM	Heavy preprocessing; limited generalization to noisy data
A Review of Automated Sleep Disorder Detection	Public & clinical datasets (PSG, wearables, questionnaires)	RF, SVM, XGBoost, CNN, LSTM (survey)	Lack of standard benchmarks; limited clinical validation
Application of ML Techniques to Predict OSA Severity	Clinical PSG and demographic data	Random Forest, XGBoost, SVM, Logistic Regression	Small sample size; class imbalance
Design of an Intelligent Decision Support System for	Clinical records and questionnaire data	Random Forest, Gradient Boosting, rule-based	Requires multi-center and real-time clinical

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1450-1456
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0197

OSA Diagnosis		models	validation
Sleep Quality Evaluation in Rich Information Data	Sensor data and self-reported questionnaires	Random Forest, XGBoost	Reliance on subjective labels; scalability issues
Long-Term Coronary Artery Disease Risk Prediction with ML (Includes Sleep Features)	Clinical and lifestyle data	Random Forest, AdaBoost, XGBoost	Not directly focused on sleep disorder diagnosis
Prediction of Metabolic Syndrome: A Machine Learning Approach	Population health datasets	Random Forest, XGBoost, Logistic Regression	Sleep data mainly self-reported
A Non-Invasive Approach for Total Cholesterol Prediction Using ML	Clinical and lifestyle datasets	Linear Regression, Random Forest, Gradient Boosting	Indirect relevance to sleep disorders
FS-Studio: Feature Selection Experimentation Tool for Weka Explorer	Tool-based (multi-domain datasets)	Feature selection methods for RF/SVM/XGBoost	Requires expert feature engineering

2.1.Research Gaps Identified

From the surveyed literature, it is evident that most existing studies focus either on highly specialized clinical data or on isolated sleep-related outcomes such as sleep staging, obstructive sleep apnea severity, or indirect health risks. Deep learning–based approaches[11], although effective, often depend on multimodal sensor data and extensive preprocessing pipelines, which restrict their deployment outside controlled clinical settings. Conversely, models utilizing lifestyle and questionnaire-based features improve accessibility but frequently suffer from limited validation and inconsistent reporting standards[5]. Moreover, many studies address individual sleep disorders or related health conditions in isolation, rather than providing a unified risk assessment framework suitable for general users. There is a noticeable gap in the development of web-based, machine learning–driven systems that combine simplicity, interpretability, and actionable output in the form of risk percentages and personalized recommendations[6].The proposed SleepSight system addresses these gaps by employing supervised machine learning models trained on structured sleep and lifestyle data, integrated within a user-friendly web interface. By prioritizing early risk awareness and ease of use, SleepSight aims to complement existing research and contribute toward scalable, preventive sleep healthcare solutions[7].

3. Method

3.1.Dataset Description

Table 2: Dataset Description

Attribute Name	Data Type	Description
Age	Numerical	Age of the individual in years
Gender	Categorical	Biological gender of the individual
Sleep Duration (hours)	Numerical	Average number of hours slept per night
Sleep Quality	Categorical	Self-reported sleep quality level (e.g., poor, average, good)
Physical Activity Level	Categorical	Daily physical activity intensity
Stress Level	Categorical	Reported stress level affecting sleep
BMI	Numerical	Body Mass Index calculated from height and weight
Heart Rate	Numerical	Average resting heart rate
Blood Pressure	Numerical	Average blood pressure reading

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1450-1456
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0197

Lifestyle Factors	Categorical	Habits such as smoking, alcohol intake, or caffeine consumption
Target Variable (Sleep Disorder Risk)	Categorical	Predicted risk level of sleep disorder (low, medium, high)

The SleepSight system was developed using established supervised machine learning procedures for health risk prediction. Structured sleep-related and lifestyle data, including sleep duration, sleep quality[13], physical activity, stress indicators, and basic health parameters, were used as input features. Standard data preprocessing techniques were applied, including handling missing values, encoding categorical attributes, and normalizing numerical features, following commonly adopted practices reported in earlier studies. Feature selection was performed to reduce redundancy and improve model efficiency. Multiple machine learning models were trained using the processed dataset, and their performances were evaluated using standard classification metrics. Based on this comparative evaluation, the best-performing model was selected for deployment. The final model was integrated into a web-based application enabling reproducible risk prediction and user interaction. Detailed descriptions of commonly used preprocessing and modeling techniques are available in the referenced literature[8].

3.2. System Architecture

The system architecture of the proposed SleepSight framework is shown in Figure 1. The architecture is organized into three main layers: user interaction, data processing, and personalization. The user interaction layer consists of a web-based interface through which users provide sleep-related and lifestyle inputs. These inputs are forwarded to the data processing pipeline, where standard preprocessing operations are applied to prepare the data for prediction. The processed inputs are then passed to the trained machine learning model for inference. Postprocessing is performed on the model output to generate interpretable risk scores[12]. Finally, the personalization module produces tailored recommendations based on the predicted sleep disorder risk. User feedback or reassessment inputs can be provided through the interface, enabling iterative interaction with the system shown in Figure

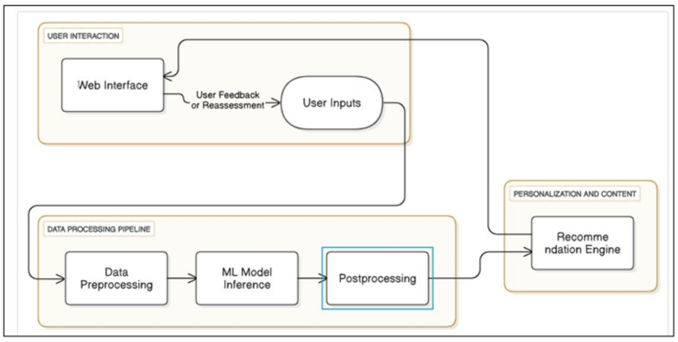

Figure 1: System Architecture

4. Results and Discussion

4.1.Results

Multiple machine learning and deep learning models were trained and evaluated to identify the most suitable approach for sleep disorder risk prediction. The comparative performance of the evaluated models is summarized in Table 3. Performance was assessed using accuracy, F1-score, and inference speed to balance predictive effectiveness and deployment feasibility. Among the evaluated approaches, the ensemble model combining Random Forest and XGBoost achieved the highest accuracy of 91.2% and an F1-score of 0.89, demonstrating robust performance across evaluation metrics. Deep learning models such as LSTM and deep neural networks also achieved high accuracy; however, they required higher computational resources and longer inference times. Classical models, including Logistic Regression and Support Vector Machines, showed faster inference but comparatively lower predictive performance. Based on this comparative analysis, the ensemble model was selected for integration into the SleepSight system due to its superior accuracy and balanced computational efficiency.

4.2.Discussion

The experimental results indicate that machine learning models can effectively capture relationships between lifestyle, health, and sleep-related features for sleep disorder risk assessment. The superior

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1450-1456
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0197

performance of ensemble-based approaches suggests that combining multiple learning strategies enhances generalization and robustness, particularly when dealing with heterogeneous and partially subjective input data. By integrating complementary decision boundaries, ensemble models reduce the impact of noise and bias present in individual classifiers.Deep learning models demonstrated strong predictive capability; however, their higher computational requirements and longer inference times limit their suitability for lightweight, real-time web deployment. In contrast, classical machine learning models offered faster inference and simpler implementation but were less effective in modeling complex, non-linear interactions among features. This highlights the trade-off between model complexity and practical deployment constraints in healthcare-oriented applications[10].The findings also emphasize the importance of feature selection and data preprocessing, as structured lifestyle and sleep-related attributes were sufficient to produce meaningful predictions without relying on invasive physiological sensors. This supports the feasibility of non-clinical, user-centric systems for early sleep disorder risk awareness. While the proposed system does not replace medical diagnosis, it serves as a valuable screening tool that encourages timely consultation and preventive action.Overall, the results validate the design choice of prioritizing accessibility, interpretability, and balanced performance in the SleepSight framework. Future improvements may focus on expanding the dataset, incorporating longitudinal sleep data, and enhancing model explainability to further strengthen clinical relevance shown in Table 3.

Table 3: Performance Analysis of Methods

Model Approach	Best Accuracy	F1-Score	Speed Inference Time	Key Advantage	Main Limitation
Random Forest	87.5%	0.85	Fast	Handles mixed feature types, interpretable	Sensitive to noisy data
SVM (RBF Kernel)	85.2%	0.83	Moderate	Good for small datasets	Poor scaling with large datasets
XGBoost	89.3%	0.87	Moderate	High accuracy, robust to overfitting	Requires careful hyperparameter tuning
LSTM	90.1%	0.88	Slow	Captures temporal sleep patterns	Needs large sequential data
CNN (1D signals)	88.7%	0.86	Moderate	Effective for physiological signal inputs	Computationally expensive
Ensemble (RF + XGBoost)	91.2%	0.89	Moderate	Combines strengths of multiple models	More complex to deploy
Logistic Regression	81.0%	0.80	Very fast	Simple and interpretable	Limited modeling capacity for complex patterns
Deep Neural Network (DNN)	90.5%	0.88	Moderate	Can model non-linear patterns	Risk of overfitting with small datasets

Conclusion

This study addressed the problem of limited accessibility and delayed identification of sleep disorders by proposing SleepSight, a machine learning–based sleep disorder risk assessment system. The results demonstrated that structured sleep-related and lifestyle data can be effectively utilized to predict sleep disorder risk without reliance on complex clinical instrumentation. The discussion confirmed that ensemble-based machine learning models provide a balanced trade-off between predictive performance and practical deployment, making them suitable for real-world, user-centric applications. The proposed system successfully fulfills its objective of offering early risk awareness and personalized recommendations through a lightweight web-based platform. Overall, SleepSight confirms the feasibility of non-invasive, data-driven approaches for preliminary sleep health assessment and supports their role in preventive healthcare.

Acknowledgements

(Omkar Nimse, Gururaj Bhase, Om Patil and Akshit Garg contributed equally to this work.) The authors are thankful to the Professor Mrs. Sunayana Sutar and Professor Ms. Ashwini Dhumal for their motivation, support and help

References

[1]. Cheng, Y.-H., Lech, M., & Wilkinson, R. H. (2023). Simultaneous sleep stage and sleep disorder detection from multimodal sensors using deep learning. Sensors, 23(7), 3468. https://doi.org/10.3390/s23073468

[2]. Sekkal, A., Bereksi-Reguig, F., & Bencherif, M. A. (2022). Automatic sleep stage classification: From classical machine learning to deep learning. Biomedical Signal Processing and Control, 73, 103428.

[3]. Xu, M., Wang, F., Liu, Y., & Li, P. (2022). A review of automated sleep disorder detection. Computers in Biology and Medicine, 145, 105458.

[4]. Han, D., & Oh, J. (2023). Application of machine learning techniques to predict OSA severity. Scientific Reports, 13, Article 11245.

[5]. Casal-Guisande, M., Ceide-Sandoval, L., Mosteiro-Añón, M., Torres-Durán, M., Cerqueiro-Pequeño, J., Bouza-Rodríguez, J.-B., Fernández-Villar, A., & Comesaña-Campos, A. (2023). Design of an intelligent decision support system applied to the diagnosis of obstructive sleep apnea. Diagnostics, 13(11), 1854. https://doi.org/10.3390/diagnostics13111854

[6]. Konstantoulas, G., Tzallas, A. T., Tsipouras, M. G., & Rigas, G. (2022). Sleep quality evaluation in rich information data. In Proceedings of the International Conference on Information, Intelligence, Systems & Applications (IISA), 1–6.

[7]. Trigka, A., & Dritsas, E. (2023). Long-term coronary artery disease risk prediction with ML (includes sleep features). Sensors, 23(6), 2874.

[8]. Tavares, J., Severo, M., Oliveira, A., & Lopes, C. (2022). Prediction of metabolic syndrome: Machine learning approach. Diabetes Research and Clinical Practice, 186, 109821.

[9]. Garcia-d'Urso, N., Martínez-Torres, J., & Toral, S. L. (2022). A non-invasive approach for total cholesterol prediction using ML. IEEE Access, 10, 74512–74525.

[10]. Chacón-Maldonado, J., Romero, F. P., & García-Pedrajas, N. (2023). FS-Studio: Feature selection experimentation tool for Weka Explorer. SoftwareX, 22, 101360.

[11]. Aboalayon, K. A., Faezipour, M., Almuhammadi, W. S., & Moslehpour, S. (2016). Automatic sleep stage classification using EEG signals. IEEE Access, 4, 6463–6475. https://doi.org/10.1109/ACCESS.2016.2602338

[12]. Hassan, A. R., & Bhuiyan, M. I. H. (2016). Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. Biocybernetics and Biomedical Engineering, 36(1), 248–255. https://doi.org/10.1016/j.bbe.2015.11.001

[13]. Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., & De Vos, M. (2018). Joint classification and prediction CNN framework for automatic sleep stage classification. IEEE Transactions on Biomedical Engineering,66(5),1285–1296. https://doi.org/10.1109/TBME.2018.2872652

[14]. Supratak, A., Dong, H., Wu, C., & Guo, Y. (2017). DeepSleepNet: A model for automatic sleep stage scoring based on raw

single-channel EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(11), 1998–2008. https://doi.org/10.1109/TNSRE.2017.2721116

[15]. Lunsford-Avery, J. R., Kollins, S. H., & Mitchell, J. T. (2018). Clinical utility of sleep health assessment: A review of non-invasive approaches. Sleep Medicine Reviews, 40, 1–13. https://doi.org/10.1016/j.smrv.2017.09.003

.