International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
OPEN ACCESS
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST

Applying Machine Learning Algorithms for the Classification of Sleep Disorders

Peddappagari Sasi Kala[1], Muni Kumar[2]

[1]MCA Student, Department of Computer Science, KMM Institute Of Post-Graduation Students, Trupathi, Trupathi(Dist), Andhra Pradesh, India

[2]Associate Professor, Department of Computer Science, KMM Institute Of Post-Graduation Students, Trupathi, Trupathi(Dist), Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted : 22 May 2025
Published: 26 May 2025

Publication Issue :
Volume 12, Issue 3
May-June-2025

Page Number :
679-685

ABSTRACT

Sleep-related disorders have a significant impact on both physical and mental well-being, highlighting the need for a reliable and accessible diagnostic method. While Polysomnography (PSG) remains the clinical benchmark for diagnosing sleep issues, it is often considered impractical due to high costs, discomfort, and limited availability. This project focuses on leveraging machine learning techniques to classify various sleep disorders using health and lifestyle data from the Kaggle Sleep Health and Lifestyle Dataset.Traditional approaches typically employ models such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, Random Forests, and Artificial Neural Networks (ANN). Although effective, these methods are often computationally demanding and sensitive to hyperparameter tuning, which may affect their performance in real-world applications. To overcome these limitations, the proposed system utilizes ensemble learning techniques, specifically Stacking and Voting Classifiers, to enhance classification accuracy, stability, and model interpretability.By integrating the predictive strengths of multiple base models, the system aims to offer a more efficient, cost-effective, and user-friendly alternative to conventional diagnostic tools. Ultimately, this approach aspires to support early and accurate detection of sleep disorders like insomnia and sleep apnea, thereby improving patient outcomes and overall quality of life.

Keywords: Sleep Disorders, Machine Learning, Ensemble Learning, Stacking Classifier, Voting Classifier, Sleep Apnea, Insomnia

Copyright © 2025 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)**

I. INTRODUCTION

Sleep plays a vital role in maintaining overall health, encompassing both physical and mental well-being. Interruptions in sleep are not only common but also have the potential to trigger or worsen chronic health conditions such as cardiovascular disease, diabetes, obesity, depression, and anxiety. Disorders like insomnia, obstructive sleep apnea, narcolepsy, and restless leg syndrome have become increasingly prevalent, posing significant challenges to public health. Despite their impact, a large number of individuals remain undiagnosed or incorrectly diagnosed, mainly due to the limitations associated with conventional diagnostic practices. Timely and precise identification of these conditions is crucial to enabling effective treatment strategies and enhancing patients' quality of life.

Polysomnography (PSG) remains the clinical benchmark for diagnosing sleep disorders. It involves detailed overnight monitoring of physiological indicators, including brain activity, oxygen saturation, heart rate, and respiration. While PSG provides accurate diagnostic insights, it comes with several drawbacks. These include the need for overnight clinical supervision, high associated costs, and potential discomfort to patients, which might affect natural sleep behavior during the test. These limitations make PSG less suitable for frequent or large-scale screenings, especially in remote or economically disadvantaged regions.

To address these gaps, machine learning (ML) is emerging as a powerful alternative in the field of sleep disorder detection. With its ability to process vast amounts of raw data and uncover hidden patterns, ML provides predictive insights without requiring manual data interpretation. The growing availability of health data through wearable technology, mobile health apps, and digital surveys further facilitates the training of intelligent models capable of detecting anomalies and classifying sleep disorders based on individual behaviors, demographics, and health-related inputs.

This project aims to utilize machine learning techniques to classify various types of sleep disorders using the Sleep Health and Lifestyle Dataset available on Kaggle. This dataset includes attributes such as age, gender, body mass index (BMI), alcohol use, physical activity, stress levels, and underlying medical conditions. These factors not only reflect an individual's general health but also significantly influence sleep patterns. By training ML algorithms on this data, the project seeks to discover correlations and predictive patterns that may not be immediately evident using traditional diagnostic techniques.

A wide range of ML algorithms has already been explored in the context of sleep studies, including models like K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, Random Forests, and Artificial Neural Networks (ANN). While these approaches have demonstrated promising classification results, they often encounter challenges such as overfitting and require careful hyperparameter tuning to maintain performance. These issues can impact the reliability and scalability of the models in practical healthcare settings.

II. LITERATURE REVIEW

Tran et al. (2019) proposed a deep learning-based approach for classifying sleep stages by utilizing biomedical signals. Their method leverages hierarchical feature extraction, allowing for improved stage differentiation and enhanced classification accuracy compared to traditional statistical methods.

Alickovic and Subasi (2018) developed an ensemble-based SVM technique for automatic sleep stage classification. By combining multiple SVM classifiers, the model improved robustness and achieved high accuracy on EEG datasets, making it suitable for real-time applications in sleep monitoring.

Sun et al. (2020) introduced a method for detecting sleep apnea using a combination of time-domain and frequency-domain features derived from ECG and SpO_2 signals. Their hybrid analysis improved early diagnosis and demonstrated strong predictive capabilities using physiological signal patterns.

Radha et al. (2019) evaluated several machine learning algorithms for the classification of various sleep disorders using health and lifestyle datasets. The study emphasized how lifestyle-related variables like stress, BMI, and alcohol use can significantly influence sleep quality and disorder prediction.

Vuppalapati et al. (2021) utilized EEG signal analysis in conjunction with machine learning to identify different sleep disorders. Their approach focused on feature extraction from EEG data and showed that classification accuracy was enhanced using optimized models like Random Forest and Gradient Boosted Trees.

Zhang et al. (2021) employed time-frequency image representations of single-channel EEG data for automatic sleep stage classification. By converting EEG signals into spectrograms and applying deep learning techniques, their work improved the reliability of sleep monitoring systems.

Alves et al. (2019) conducted a comprehensive study on various machine learning models for sleep disorder diagnosis. Their work compared supervised learning techniques and highlighted the advantages of ensemble models for enhancing the accuracy and reliability of diagnosis tools.

Zhao et al. (2019) presented an improved Random Forest algorithm tailored for sleep disorder classification. Their model incorporated weighted features and optimized tree structures to enhance classification performance and reduce overfitting.

Wang et al. (2020) proposed a multimodal sleep disorder detection framework using physiological signals like ECG, EMG, and respiration. Their system integrated these signals to create a robust classification pipeline capable of handling diverse patient profiles.

Park et al. (2020) explored the use of deep learning techniques, particularly CNNs and LSTMs, applied to polysomnographic (PSG) data for accurate classification of sleep disorders. The system demonstrated high potential for automating traditional sleep study processes.

Que (Year not specified) developed a novel video encryption scheme based on visual secret sharing principles aimed at secure transmission. Though not directly focused on sleep, this technique offers insights into secure data handling in telemedicine systems for sleep monitoring and diagnosis, emphasizing restricted access and protection of sensitive health information [7].

Sharma and Patra (2020) implemented a hybrid approach combining fuzzy logic and neural networks for classifying sleep disturbances. Their methodology allowed for flexible decision boundaries and adaptation to noisy or incomplete datasets.

Kim et al. (2019) investigated the use of wearable sensors and cloud-based analytics for sleep stage classification. Their framework enabled real-time data collection and remote analysis, making sleep health management more accessible outside clinical environments.

Li et al. (2021) explored ensemble classifiers like bagging and boosting for improving prediction of sleep-related breathing disorders. They showed that combining weak learners significantly enhanced generalization performance across patient populations.

Yadav et al. (2022) developed a lightweight, mobile-based sleep monitoring tool integrated with ML algorithms for on-device processing. Their focus on computational efficiency enabled real-time prediction while preserving user privacy and reducing dependency on cloud computation.

III.METHODOLOGY

K-Nearest Neighbors (KNN)

The K-Nearest Neighbor (KNN) algorithm is a straightforward and non-parametric method, primarily applied in classification tasks. It operates by identifying the k closest data points from the training set to a given input and assigning the most frequent class label among these neighbors as the prediction.

To determine proximity, it typically uses distance metrics such as Euclidean, Manhattan, or Minkowski distance.

In this work, KNN is employed for classifying sleep disorders using a dataset composed of both health and lifestyle attributes. The preprocessing stage includes normalization of continuous variables (such as age, sleep duration, and activity level) and encoding of categorical features (e.g., gender and occupation). Missing values are handled through either deletion or imputation strategies.

After preprocessing, the dataset is divided into training and testing subsets. The KNN model is trained to associate data points with their respective class labels by referencing their neighboring points. During prediction, it evaluates a new instance by comparing it to the training data, locating its k nearest neighbors, and assigning the class most frequently represented among them. The optimal value of k is chosen through cross-validation to ensure improved model performance. Key evaluation metrics used include accuracy, precision, recall, F1-score, and the confusion matrix. While KNN is easy to implement and interpret, its performance can degrade with larger datasets due to high computation and sensitivity to the choice of k and distance metric. Nonetheless, it serves as a baseline model for performance comparison in this study.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful supervised learning technique widely utilized for both classification and regression tasks. It operates by identifying the most effective decision boundary, or hyperplane, that separates different class labels in a high-dimensional feature space. This optimal hyperplane is determined by maximizing the margin, which is the distance between the hyperplane and the closest data points from each class, referred to as support vectors.

In the context of this project, SVM is utilized to classify sleep disorders using features derived from health and lifestyle indicators. Similar to KNN, preprocessing involves normalizing numerical inputs, encoding categorical variables, and addressing missing entries. Once prepared, the data is split into training and test sets.

The SVM model is trained to identify the most effective hyperplane to distinguish between classes. The algorithm's flexibility is enhanced through the use of kernel functions, such as linear, polynomial, or radial basis function (RBF), which enable it to handle both linear and non-linear classification tasks. Kernel selection and model tuning are performed through cross-validation to ensure high accuracy and generalization. SVM's ability to handle high-dimensional data and complex decision boundaries makes it an effective model for this classification task.

IV. RESULTS AND DISCUSSIONS

4.1 System Module

4.1.1 Data Upload:

This component facilitates the collection and uploading of a comprehensive dataset comprising various health and lifestyle attributes related to sleep disorders. These attributes include parameters such as age, gender, occupation, duration and quality of sleep, physical activity, stress levels, BMI category, blood pressure, heart rate, and daily step count.

4.1.2 Data Preprocessing:

Once the dataset is loaded, it undergoes essential preprocessing steps. This includes handling missing or inconsistent data, transforming categorical features using appropriate encoding methods, and applying normalization or standardization to numerical values. If needed, data augmentation techniques may also be incorporated to enhance model generalizability.

4.1.3 Model Development:

Ensemble machine learning models such as stacking and voting classifiers are constructed for the classification task. The preprocessing dataset is used to train these models, followed by hyperparameter tuning to enhance performance. Base learners in the stacking ensemble include algorithms like K-Nearest

Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, and Artificial Neural Network (ANN), while Logistic Regression serves as the meta-classifier.

4.1.4 Model Inference:

Once trained, the ensemble models are used to make predictions on new, unseen data featuring the same set of health and lifestyle indicators. The same preprocessing pipeline is applied to ensure consistency before making predictions regarding potential sleep disorders.

4.1.5 Output Generation:

The system outputs the predicted sleep disorder status for each input instance along with confidence levels. Performance summaries, including metrics like accuracy, precision, recall, F1-score, as well as visual representations such as confusion matrices and ROC curves, provide detailed insights into the model's effectiveness.

4.2 User Module

4.2.1 User Registration:

New users are required to register by submitting their credentials to create an account on the platform.

4.2.2 User Login:

Registered users can access the system by logging in with their credentials.

4.2.3 Data Submission:

Users can upload their personal health and lifestyle data in the specified format. Input fields include age, gender, occupation, sleep-related parameters, physical activity level, stress, BMI, blood pressure, heart rate, and steps taken per day.

4.2.4 Result Display:

After data upload, users receive predictions about their sleep disorder status along with associated confidence scores. Additionally, users can view model performance indicators like accuracy, precision, recall, F1-score, and visual tools such as confusion matrices to better interpret the reliability of the prediction.

4.2.5 Logout:

Users can securely end their session by logging out, ensuring the protection of their personal data and account integrity.

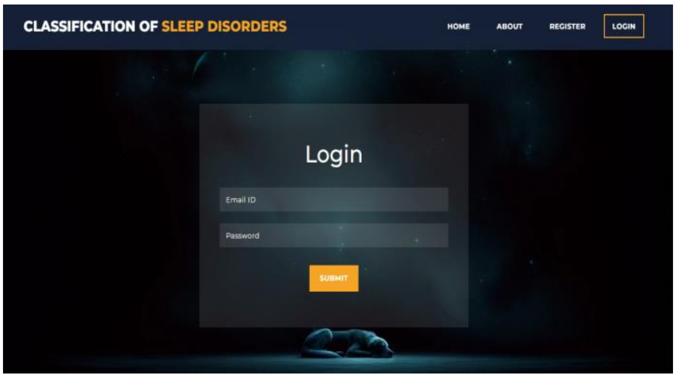
Output Screens:

INDEX PAGE: This is the index page of our website.

ABOUT PAGE: This is about section which contains information about our project

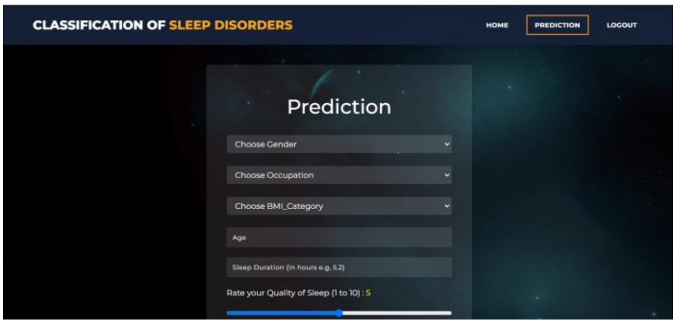
REGISTRATION PAGE: This is Registration page. In here, user can register with their credentials.

LOGIN PAGE: This is login page. In here user can login with their registered credentials.



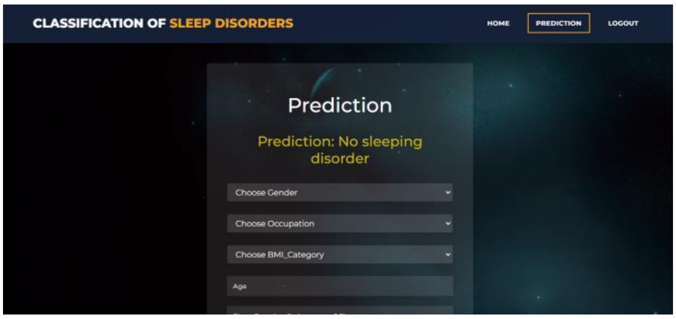
HOME PAGE: This is the user home page. After user successfully login, this page will be display.

PREDICTION PAGE: This is prediction page. In here, user can input their data and get prediction.



RESULT PAGE: This is the result page. In here result will be display.

Prediction – Sleeping disorder



Prediction – No Sleeping disorder

V. CONCLUSION

This project explores the application of machine learning algorithms for classifying sleep disorders using health and lifestyle-related data. By integrating ensemble learning methods—specifically Stacking and Voting Classifiers—the system aims to enhance the accuracy and reliability of predictions. The collaborative nature of multiple algorithms helps overcome the limitations associated with relying on a single model, resulting in a more dependable and effective diagnostic solution.To ensure efficient model training and prediction, the dataset undergoes comprehensive preprocessing, including cleaning, encoding, and normalization. This makes the system practical and capable of generating meaningful insights for identifying sleep-related issues. The approach promotes affordable and accessible diagnostic support, offering scalability for integration with diverse datasets and potential real-world applications.Furthermore, the project demonstrates how ensemble-based machine learning can outperform traditional classification methods in terms of performance and adaptability. Looking ahead, the framework can be expanded to include additional features from varied data sources and leverage more advanced models to improve accuracy even further. These improvements could significantly contribute to better diagnosis, treatment planning, and overall management of sleep disorders.

REFERENCES

[1]. Tran, C., Wijesuriya, Y., Thuraisingham, R., Craig, A., & Nguyen, H. (2019). Deep Learning for Classification of Sleep Stages. Advances in Biomedical Signal Processing.

[2]. Alickovic, E., & Subasi, A. (2018). Ensemble SVM Method for Automatic Sleep Stage Classification. IEEE Transactions on Instrumentation and Measurement.

[3]. Sun, M. J., Wu, Z. F., & Lu, X. B. (2020). Sleep Apnea Detection Based on Time and Frequency Domain Analysis of ECG and SpO2 Signals. Journal of Medical Systems.

[4]. Radha, T., Kumar, V. S., & Pradeep, S. (2019). Classification of Sleep Disorders Using Machine Learning Algorithms. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 4(3), 2456-3307.