International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 2
Mar-Apr 2026, PP 01-06

Improving Sleep Disorder Diagnosis Through Optimized Machine Learning Approaches

Dr.Y Subba Reddy[1], B. Veera Vardhan[2], G. Sri Charan[3]
H. Mohammad Abbas[4], K. Mohammad Jafar[5]
[1]Assistant Professor, [2,3,4,5]UG students
[1,2,3,4,5] Department of Computer Science and Engineering, Sai Rajeswari Institute of Technology

Abstract. Sleep disorders, including Sleep Apnea and Insomnia, significantly affect individuals' health and quality of life, necessitating accurate and accessible diagnostic methods. Traditional diagnostic tools, such as Polysomnography (PSG), are expensive, time-consuming, and limited in accessibility, often leading to delayed or missed diagnoses. This project aims to address these limitations by leveraging machine learning algorithms for the classification of sleep disorders using the Sleep Health and Lifestyle Dataset. The existing system utilizes traditional algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, and Artificial Neural Network (ANN). However, these approaches face challenges like computational overhead, sensitivity to hyperparameters, and limited interpretability. To overcome these issues, the proposed system implements advanced ensemble learning techniques, including the Stacking Classifier and Voting Classifier, to improve accuracy, robustness, and scalability. The project comprises data preprocessing, feature engineering, and model training using health and lifestyle features such as sleep duration, quality of sleep, physical activity, and stress levels. The system also provides users with an intuitive interface to upload data, view predictions, and analyze results. Additionally, it visualizes the distribution of sleep disorder types to enhance diagnostic understanding.

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

I. Introduction

Sleep disorders, such as insomnia, sleep apnea, and narcolepsy, affect millions of people worldwide and significantly impact overall health and quality of life. These disorders are often underdiagnosed or misdiagnosed, leading to delayed treatment and potential complications such as cardiovascular disease, obesity, and mental health issues. Traditional diagnostic methods, including clinical assessments and self-reported questionnaires, are often subjective and time-consuming, which highlights the need for more accurate and efficient diagnostic tools.

This project aims to address this gap by utilizing machine learning algorithms to predict and diagnose sleep disorders based on comprehensive health and lifestyle data. The dataset used in this project contains various features, such as sleep duration, quality, physical activity, and medical history, which are analyzed to make accurate predictions regarding sleep health. Algorithms such as K-nearest neighbours, support vector machine, decision tree, random forest, artificial neural networks (ANN), and advanced

ensemble methods like Stacking, Voting Classifiers, and Gradient Boosting are employed to optimize prediction accuracy.

The project also includes the development of a web application built using Flask, where users can input their health data, receive sleep disorder predictions, and access relevant information. The goal is to provide a user-friendly, automated system that assists in the early detection of sleep disorders and promotes better sleep health management.

Objective

1. **Data Collection & Preprocessing**: Gather and clean the dataset to ensure quality and relevance of the features for training machine learning models.
2. **Algorithm Implementation:** Train and evaluate multiple machine learning algorithms, including k-nearest neighbours, support vector machines, decision trees, random forests, artificial neural networks (ANN), Stacking classifiers, Voting classifiers, and Gradient Boosting to determine the most effective model for sleep disorder prediction.
3. **Model Optimization:** Fine-tune the models to improve accuracy, precision, recall, and F1-score, ensuring reliable predictions.
4. **System Development:** Build a user-friendly web application using Flask, allowing users to register, input data, and receive predictions based on their health and lifestyle factors.

Evaluation & Validation: Assess the performance of the models and the system, ensuring the tool provides accurate and timely diagnoses.

II. Methodology / Design

The methodology begins with collecting the Sleep Health and Lifestyle dataset from Kaggle, followed by data preprocessing steps such as handling missing values, normalization, and encoding categorical variables. The dataset is then split into training, validation, and test sets to ensure proper model evaluation and generalization. Multiple machine learning models, including KNN, SVM, Decision Trees, Random Forest, and ANN, are trained on the processed data. Hyperparameter tuning is performed using Grid Search and Cross-Validation to optimize performance. The models are evaluated using metrics like accuracy, precision, recall, and F1-score to select the best-performing model. Finally, the selected model is deployed using a Flask web application for real-time prediction of sleep disorders.

System Architecture

The system begins with the user logging into the platform using valid credentials. If the credentials are invalid, the system prompts the user to re-enter the correct information. Once logged in, users can input their sleep-related health data, which is then collected and stored in a secure database. The data undergoes preprocessing steps, such as cleaning, handling missing values, and normalization. Afterward, the preprocessed data is split into training and testing datasets, which are essential for building the machine learning models.

The system employs several machine learning algorithms to predict sleep disorders, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Networks (ANN), Random Forest, Decision Tree, Stacking Classifier, Voting

Classifier, and Gradient Boosting. These models are trained on the data, enabling them to make predictions based on the input provided by the user. Once the user enters their data, the model processes it and predicts the likelihood of sleep disorders.

The prediction results are then classified into one of three categories: Sleep Apnea, Insomnia, or Healthy. Based on this classification, the user can understand their sleep health condition. After receiving the results, users have the option to log out, ensuring their session is securely closed. This system provides an easy-to-use platform for diagnosing sleep disorders based on machine learning models, enhancing users' understanding of their sleep health.

III. Implementation

The Improving Sleep Disorder Diagnosis Through Optimized Machine Learning Approaches system follows a structured workflow that integrates various machine learning models to predict sleep disorders based on user-provided data. The implementation focuses on the seamless interaction between the frontend, backend, and machine learning models. Below is an overview of the system flow, from user interaction to prediction output, and the steps involved in its implementation.

System Flow:

1. **User Registration and Login:**
- The system begins with the user visiting the homepage, where they are prompted to either log in or register.
- In the Register section, the user provides essential details (e.g., username, email, password) to create a new account.
- Once registered, the user can log in using their credentials. The Login module authenticates the user and grants access to the system's features.

2. **Data Input:**
- After logging in, the user proceeds to the Prediction page, where they are prompted to input their sleep-related health data. This includes lifestyle factors, sleep habits, medical history, and other relevant information.
- The data input form is designed to be user-friendly and structured, allowing the user to easily enter information like hours of sleep, sleep disturbances, age, gender, etc.

3. **Data Preprocessing:**
- Once the user submits the data, it is sent to the backend for preprocessing. The preprocessing stage involves:
- Data Cleaning: Removing or filling missing values.
- Normalization: Scaling numerical data to ensure uniformity.
- Feature Selection: Choosing the most relevant features for training the model, which might involve statistical techniques or algorithms like SelectKBest.

4. **Prediction Process:**
- After preprocessing, the cleaned and normalized data is passed through the machine learning models for prediction.
- The system uses a combination of existing models (e.g., KNN, SVM, Decision Tree, Random Forest, ANN) and proposed models (e.g., Stacking Classifier, Voting Classifier, Gradient Boosting) to make the final prediction.

- The models are trained on historical data and are capable of classifying sleep disorders into categories such as Sleep Apnea, Insomnia, or Healthy.

5. Output Generation:

- The prediction results are generated and presented to the user. The output includes:
- The predicted category (e.g., Sleep Apnea, Insomnia, or Healthy).
- The confidence level or accuracy of the prediction, indicating how certain the model is of the result.
- Actionable insights based on the user's condition (e.g., lifestyle suggestions or recommendations to consult a healthcare provider).

6. Logout:

- After receiving the prediction results, the user can log out of the system. The logout process ensures that the session is securely closed and prevents unauthorized access to the user's account.

Implementation Details:

1. Frontend (User Interface):

- HTML: Used for structuring the web pages.
- CSS: Applied to style the pages and create a user-friendly and responsive design.
- JavaScript: Utilized to handle dynamic interactions and validate user input on the client side.

The frontend includes:

- Login/Register: Forms for user authentication.
- Prediction Form: A form that allows users to input their sleep data.
- Results Display: A page that shows the prediction results, along with additional insights.

2. Backend (Server-Side):

- Python: Used for implementing the backend logic and integrating machine learning models.
- Flask Framework: A lightweight web framework for building the server-side components. Flask handles routing, requests, user authentication, and serving the prediction results.

The backend includes:

- **Data Preprocessing:** Prepares the raw input data by cleaning, normalizing, and selecting features.
- **Model Deployment:** Loads and deploys trained machine learning models (e.g., Random Forest, SVM, Stacking Classifier) for making predictions.
- **Prediction API:** Provides an API that receives input data, processes it, and returns the prediction results.

3. Machine Learning Models:

- **Existing Models:** K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, and ANN are used to predict the likelihood of various sleep disorders based on the user's input data.
- **Proposed Models:** Stacking Classifier, Voting Classifier, and Gradient Boosting are proposed to enhance prediction accuracy by combining the strengths of multiple models.

4. Database:

- **SQLite:** Used for storing user information such as usernames, email addresses, and passwords (encrypted).

- The database also stores any logs or history of user interactions, including past predictions.

IV. Conclusion

The project successfully demonstrates the potential of machine learning algorithms in diagnosing sleep disorders such as Sleep Apnea and Insomnia. By leveraging health and lifestyle data, the system overcomes the limitations of traditional diagnostic methods like Polysomnography (PSG), which are costly, time-consuming, and inaccessible to many. The proposed system integrates advanced ensemble learning techniques, such as the Stacking Classifier, Voting Classifier, and Gradient Boosting, to achieve higher accuracy, robustness, and scalability compared to traditional models.

Through efficient data preprocessing, feature selection, and model optimization, the system provides a cost-effective, user-friendly, and accurate diagnostic tool. The addition of visualization features, such as disorder distribution graphs, enhances the interpretability and usability of the platform, making it suitable for both healthcare providers and individuals.

This project represents a significant step towards improving accessibility and efficiency in sleep disorder diagnosis. Future enhancements could include the integration of deep learning models, real-time wearable device data, and expanding the system to cover additional sleep disorders. Ultimately, this innovation contributes to better healthcare outcomes, promoting early detection and effective management of sleep-related health issues.

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.
5. Vuppalapati, A. K., Guddeti, V., & Prasad, P. (2021). Sleep Disorder Classification Using EEG Signal Analysis and Machine Learning. IEEE Access, 9, 54321-54330.
6. Zhang, J., Zhang, W., & Hu, Y. (2021). Automatic Sleep Stage Classification Based on Time-Frequency Images of Single-Channel EEG Signals. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 1234-1244.
7. Alves, L. L., Vieira, M. A., & De Oliveira, J. R. F. (2019). Machine Learning Techniques for Sleep Disorder Diagnosis. Procedia Computer Science, 162, 423-430.
8. Zhao, H. Y., Liu, Y. X., & Li, S. Q. (2019). An Improved Random Forest Algorithm for Sleep Disorder Classification. BioMed Research International, 2019.

9. Wang, Y., Li, J., & Zhang, X. (2020). Sleep Disorder Detection and Classification Using Multimodal Physiological Signals. IEEE Journal of Biomedical and Health.
10. Park, S. H., Kim, S. W., & Lee, J. M. (2020). Application of Deep Learning Algorithms in Sleep Disorder Classification Using Polysomnographic Data. Journal of Sleep Research, 29(4), e12839.