Applying Machine Learning Algorithms For The Classification Of Sleep Disorders

SIVASANKAR CHITTOOR[1]
BOLIGORLA VIDYA[2], N.U.DHARSHAN[2], A.GEETHA[2], A.ROHITH[2], SHIVAM KUMAR[2]

[1] Assistant Professor, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India
[2] Research Scholar, Department of Computer Science & Information Technology, Siddharth Institute of Engineering & Technology, Andhra Pradesh, India

ABSTRACT
Sleep disorders significantly impact physical and mental health, necessitating accurate and accessible diagnostic methods. Traditional diagnostic techniques like Polysomnography (PSG) are often inconvenient, expensive, and limited in availability. This project aims to leverage machine learning algorithms to classify sleep disorders using health and lifestyle data from the Kaggle Sleep Health and Lifestyle Dataset. The existing system employs algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, and Artificial Neural Network (ANN), which have several limitations including computational expense and sensitivity to hyperparameters. To address these issues, the proposed system implements ensemble learning techniques, specifically Stacking Classifier and Voting Classifier, to enhance accuracy, robustness, and interpretability. By combining the strengths of multiple models, the project seeks to provide a more efficient, cost-effective, and accessible solution for diagnosing sleep disorders, ultimately improving patient outcomes and quality of life.

Keyword: - Sleep Disorders, Machine Learning, Stacking Classifier, Voting Classifier, Sleep Apnea, Insomnia.

1. INTRODUCTION

To address these limitations, this project aims to apply machine learning algorithms for the classification of sleep disorders using health and lifestyle data. By leveraging advanced machine learning techniques, we can develop a more accessible, cost-effective, and efficient system for diagnosing sleep disorders. The dataset used in this project is sourced from Kaggle and includes various health and lifestyle factors that influence sleep patterns, providing a comprehensive foundation for machine learning models.

The existing system employs traditional machine learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, and Artificial Neural Network (ANN). These algorithms have been widely used due to their simplicity and effectiveness in handling various classification tasks. However, they come with several disadvantages, including computational expense, sensitivity to hyperparameters, risk of overfitting, and lack of interpretability.

To overcome these challenges, the proposed system implements ensemble learning techniques, specifically the Stacking Classifier and Voting Classifier. Ensemble learning methods combine the strengths of multiple base models to enhance overall performance and robustness. The Stacking Classifier involves training several base models and then using another model to combine their predictions. This method leverages the individual strengths of different algorithms to achieve better predictive performance. The Voting Classifier aggregates the predictions of multiple models through majority voting or averaging, providing a more stable and accurate classification outcome.

To address these limitations, this project aims to apply machine learning algorithms for the classification of sleep disorders using health and lifestyle data. By leveraging advanced machine learning techniques, we can develop a more accessible, cost-effective, and efficient system for diagnosing sleep disorders. The dataset used in this project is sourced from Kaggle and includes various health and lifestyle factors that influence sleep patterns, providing a comprehensive foundation for machine learning models.

The existing system employs traditional machine learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, and Artificial Neural Network (ANN). These algorithms have been widely used due to their simplicity and effectiveness in handling various classification tasks. However, they come with several disadvantages, including computational expense, sensitivity to hyperparameters, risk of overfitting, and lack of interpretability.

To overcome these challenges, the proposed system implements ensemble learning techniques, specifically the Stacking Classifier and Voting Classifier. Ensemble learning methods combine the strengths of multiple base models to enhance overall performance and robustness. The Stacking Classifier involves training several base models and then using another model to combine their predictions. This method leverages the individual strengths of different algorithms to achieve better predictive performance. The Voting Classifier aggregates the predictions of multiple models through majority voting or averaging, providing a more stable and accurate classification outcome.

2. LITERATURE SURVEY

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.

3. METHODOLOGY

3.1EXISTING SYSTEM

The existing system for the classification of sleep disorders relies on traditional machine learning algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, and Artificial Neural Network (ANN). These algorithms are commonly used due to their simplicity and effectiveness in handling various classification tasks. Each algorithm has its own methodology for processing and classifying data based on different principles, such as distance measurement, hyperplane optimization, decision rules, ensemble learning, and neural network structures.

3.1.1DISADVANTAGES OF EXISTING SYSTEM

- Extended Procedure: Polysomnography requires an overnight stay in a sleep laboratory, which can be inconvenient for patients and disrupt their daily routines.
- Preparation and Follow-Up: The process involves significant preparation and follow-up visits, adding to the overall time commitment for patients.
- Expensive Equipment and Personnel: PSG involves costly equipment and requires specialized healthcare personnel, making it an expensive diagnostic method.
- Long Waiting Times: Due to the limited number of sleep centers and high demand, patients often face long waiting times for appointments.
- Manual Analysis: The manual scoring and interpretation of sleep data by sleep technicians and physicians can introduce human error and inconsistency in diagnoses.
- Physical Discomfort: The numerous sensors and electrodes attached to the body during PSG can cause physical discomfort and interfere with sleep.

3.2 PROPOSED METHODOLOGY

The proposed system aims to improve the classification of sleep disorders by implementing ensemble learning techniques such as the Stacking Classifier and Voting Classifier. These methods combine the strengths of multiple base models to enhance overall performance and robustness. The Stacking Classifier involves training several base models and then using another model to combine their predictions. The Voting Classifier combines the predictions of multiple models through majority voting or averaging, providing a more stable and accurate classification outcome.

3.2.1 ADVANTAGES OF PROPOSED METHODOLOGY

- Automated Diagnosis: The machine learning models automate the classification of sleep disorders, reducing the need for manual analysis and interpretation by sleep specialists.
- Time-Saving: The system provides immediate predictions, significantly reducing the time required for diagnosis compared to traditional methods.
- Lower Costs: The proposed system eliminates the need for expensive equipment and specialized personnel associated with polysomnography, making it a cost-effective alternative.
- Accessibility: The web-based application can be accessed from anywhere, reducing the need for costly overnight stays in sleep laboratories.
- Enhanced Accuracy: By leveraging advanced ensemble learning techniques, the system achieves high predictive accuracy, improving the reliability of diagnoses.
- User-Friendly Interface: The web application provides an easy-to-use platform for users to input their health data and receive predictions, making it accessible to a broad audience.
- Remote Access: Patients can access the diagnostic tool from the comfort of their homes, making it especially beneficial for those in remote or underserved areas.

4. SYSTEM DESIGN

The Smart Door Locking System using Face Recognition is designed as a modular, intelligent, and secure solution for access control. The system consists of three core components: **Face Data Collection**, **Model Training**, and **Real-Time Recognition with Access Control**. During the data collection phase, images of authorized individuals are captured using a webcam and stored in a structured dataset. These images are then processed and used to train a facial recognition model using the **Local Binary Pattern Histogram (LBPH)** algorithm. The real-time recognition module continuously scans the camera feed, detects faces using **OpenCV**, and matches them against the trained model. If a match is found, the system activates an **electronic lock mechanism** through a microcontroller (such as Raspberry Pi or Arduino), granting access. A **Tkinter-based GUI** is implemented to manage user interactions, such as adding new users or viewing access logs. The system also includes logging functionality to record each access attempt, enabling monitoring and security auditing. This design ensures a touch-free, reliable, and user-friendly security system suitable for residential and institutional use.

4.1 SYSTEM ARCHITECTURE

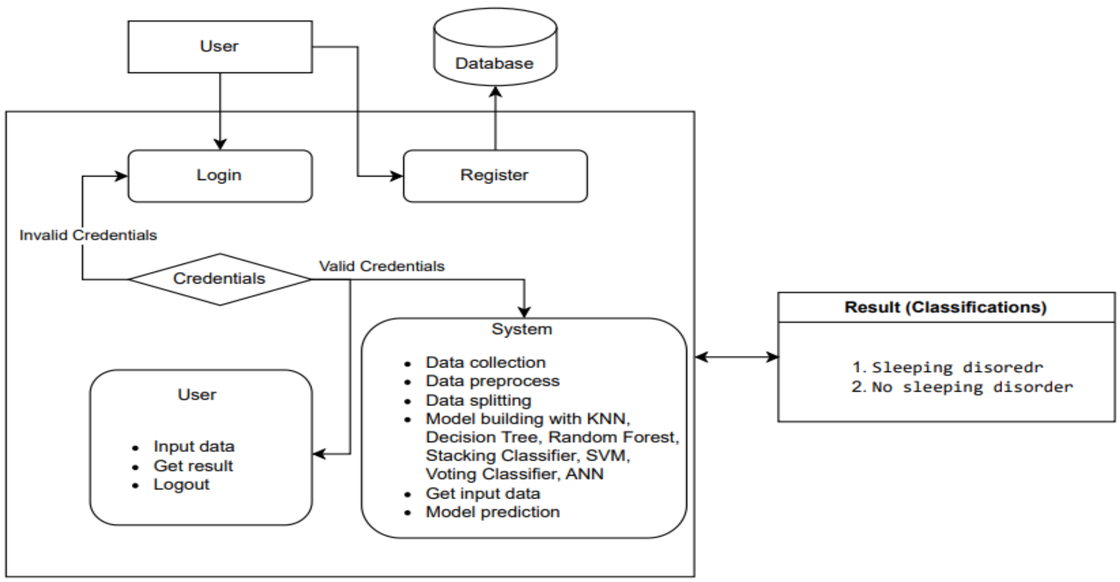


Fig. System Architecture

4.2 MODULES

1. SYSTEM:

1.1 Upload Data: Collect and upload a diverse dataset of health and lifestyle data relevant to sleep disorders. This dataset should include various features such as age, gender, occupation, sleep duration, quality of sleep, physical activity level, stress level, BMI category, blood pressure, heart rate, and daily steps.

1.2 Data Preprocessing: Once the data is loaded, undergo data cleaning and preprocessing procedures. This involves handling missing or corrupted data, encoding categorical variables, normalizing or standardizing numerical features, and applying data augmentation techniques if necessary to improve model generalization.

1.3 Model Building: Design and implement suitable ensemble learning architectures, such as Stacking and Voting Classifiers, for the classification task. Train the models using the preprocessed dataset, tuning hyperparameters to optimize performance. The base models for stacking might include XGBoost and AdaBoost, with Logistic Regression as the meta-model.

1.4 Model Prediction: Use the trained ensemble models to generate predictions on new, unseen health and lifestyle data. This involves preprocessing the new data similarly and using the models to predict the presence and type of sleep disorder.

1.5 Result: Present the results, including the predicted sleep disorder for each data instance along with confidence scores. Summarize performance metrics and use visual aids like confusion matrices and ROC curves to illustrate the models' performance.

2. User:

2.1 Register: Users should first register with their credentials to create an account in the system.

2.2 Login: Users can log in with their registered credentials to access the system.

2.3 Upload Data: Users can upload their health and lifestyle data, including various features such as age, gender, occupation, sleep duration, quality of sleep, physical activity level, stress level, BMI category, blood pressure, heart rate, and daily steps. This data should be in a structured format compatible with the system.

2.4 Viewing Results: After uploading their data, users can view the classification results provided by the model. The system will display the predicted sleep disorder along with confidence scores. Users can also view performance metrics of the model, such as accuracy, precision, recall, F1-score, and confusion matrices, to understand the reliability of the predictions.

2.5 Logout: Finally, users can log out of the system to secure their session and personal data.

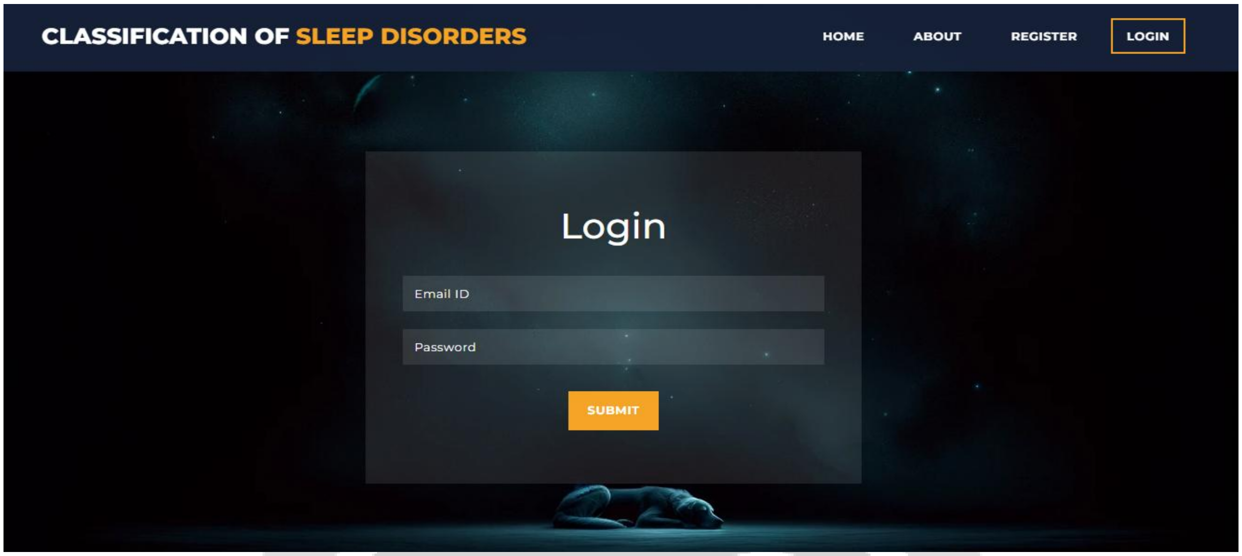
5. RESULTS AND PERFORMANCE

INDEX PAGE: This is the index page of our website

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

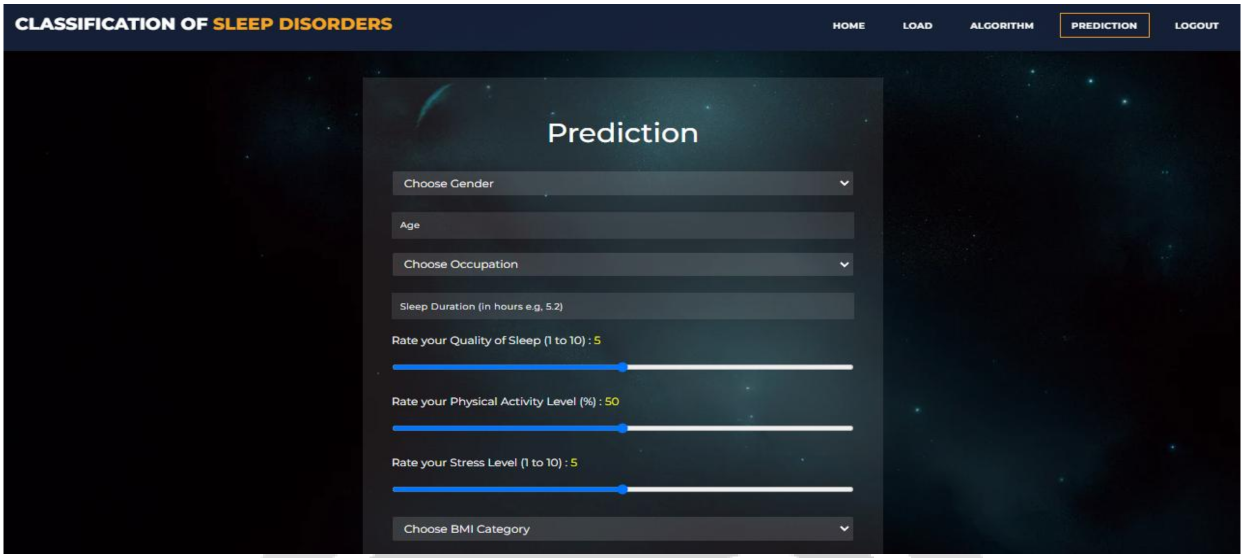
REGISTRATION PAGE: This is Registration page. In here, user can register with their credentials

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

KNN

Class	Precision	Recall	F1-Score	Support
0	0.91	0.93	0.92	46
1	0.93	0.90	0.92	42
Accuracy			0.92	88
Macro Avg	0.92	0.92	0.92	88
Weighted Avg	0.92	0.92	0.92	88

Fig. Classification Report

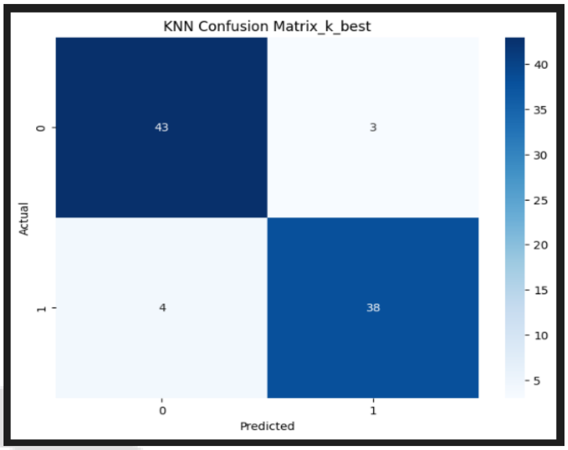


Fig. Confusion matrix

SVM

Class	Precision	Recall	F1-Score	Support
0	0.92	0.98	0.95	46
1	0.97	0.90	0.94	42
Accuracy			0.94	88
Macro Avg	0.95	0.94	0.94	88
Weighted Avg	0.95	0.94	0.94	88

Fig. Classification Report

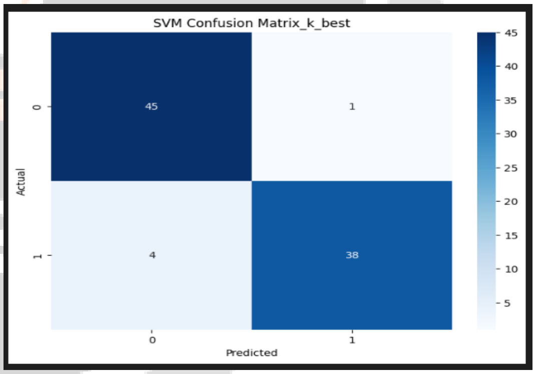


Fig. Confusion matrix

DECISION TREE

Class	Precision	Recall	F1-Score	Support
0	0.87	0.98	0.92	46
1	0.97	0.83	0.90	42
Accuracy	-	-	-	**0.91**
Macro Avg	0.92	0.91	0.91	88
Weighted Avg	0.92	0.91	0.91	88

Fig. Classification Report

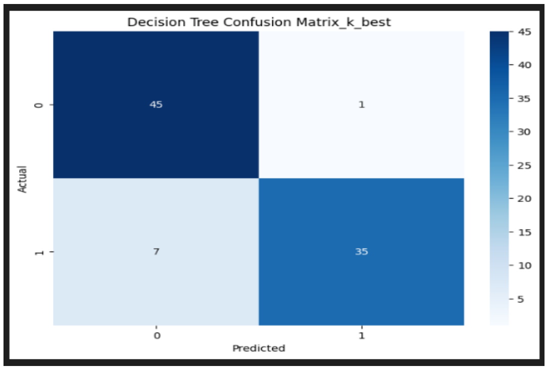


Fig. Confusion matrix

RANDOM FOREST

Class	Precision	Recall	F1-Score	Support
0	0.92	0.98	0.95	46
1	0.97	0.90	0.94	42
Accuracy			0.94	88
Macro Avg	0.95	0.94	0.94	88
Weighted Avg	0.95	0.94	0.94	88

Fig. Classification Report

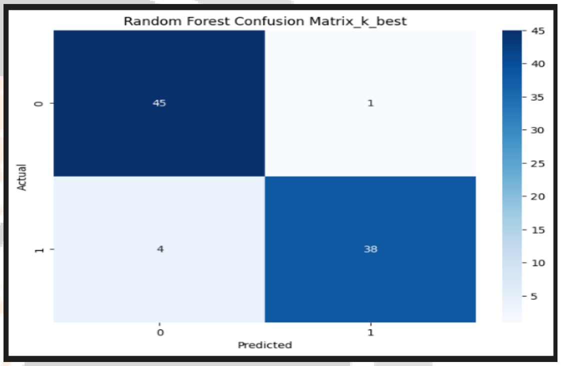


Fig. Confusion matrix

ANN

Class	Precision	Recall	F1-Score	Support
0	0.92	0.98	0.95	46
1	0.97	0.90	0.94	42
Accuracy			**0.94**	**88**
Macro Avg	0.95	0.94	0.94	88
Weighted Avg	0.95	0.94	0.94	88

Fig. Classification Report

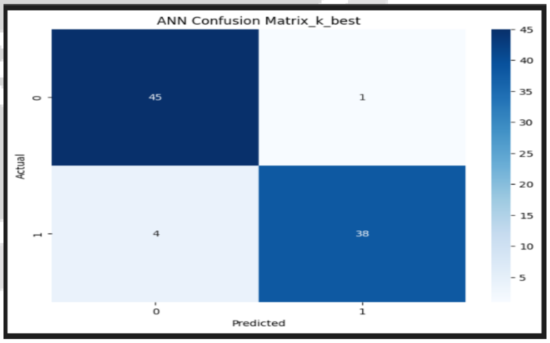


Fig. Confusion matrix

STACKING CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	0.92	0.98	0.95	46
1	0.97	0.90	0.94	42
Accuracy			0.94	88
Macro Avg	0.95	0.94	0.94	88
Weighted Avg	0.95	0.94	0.94	88

Fig. Classification Report

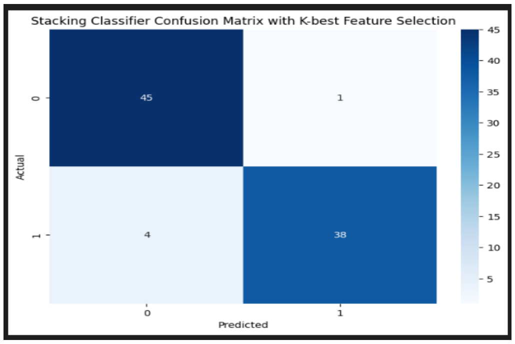


Fig. Confusion matrix

VOTING CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	0.92	0.98	0.95	46
1	0.97	0.90	0.94	42
Accuracy			0.94	88
Macro Avg	0.95	0.94	0.94	88
Weighted Avg	0.95	0.94	0.94	88

Fig. Classification Report

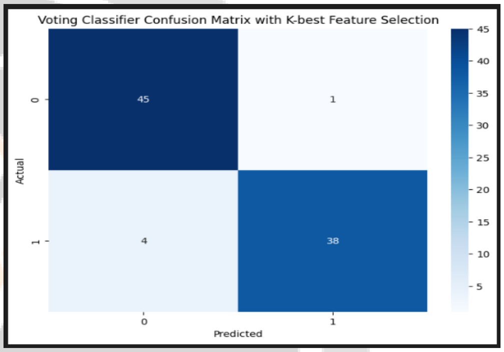


Fig. Confusion matrix

COMPARISION RESULTS

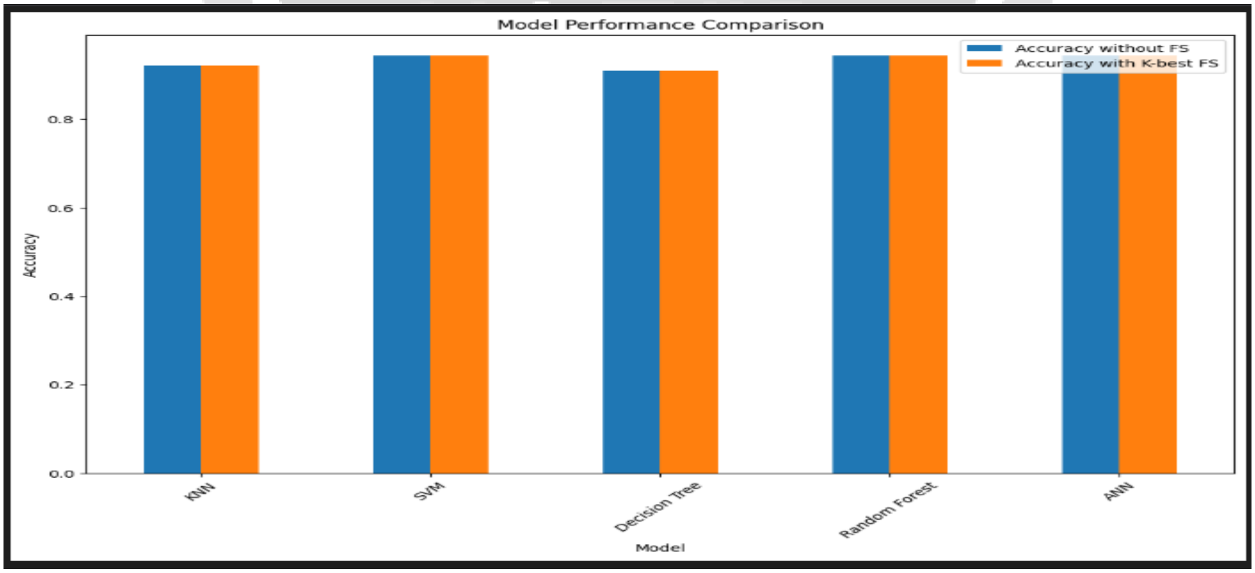


Fig. Accuracy comparison graph

Overall, the ensemble methods—Stacking Classifier and Voting Classifier—demonstrated competitive performance and improved accuracy compared to individual classifiers. The ensemble approaches successfully combined the strengths of multiple models to enhance classification performance and robustness, with the Stacking Classifier slightly outperforming the Voting Classifier. These results underscore the value of ensemble techniques in achieving more reliable and accurate predictions in complex classification tasks.

6. CONCLUSION

This project demonstrates the application of machine learning algorithms to classify sleep disorders using health and lifestyle data. By utilizing ensemble learning techniques such as Stacking and Voting Classifiers, the system aims to enhance the accuracy and robustness of sleep disorder classification compared to traditional methods. The use of diverse algorithms addresses the limitations of individual models, providing a more reliable and efficient diagnostic tool. The project successfully preprocesses and formats data for effective model training and prediction, resulting in actionable insights for identifying sleep disorders. This approach not only improves diagnostic accessibility and cost-effectiveness but also offers a scalable solution that can be adapted to various data sources and real-world applications. Future work may explore integrating additional data features and advanced models to further refine and enhance classification performance, ultimately contributing to better patient outcomes and management of sleep disorders.

7. REFERENCE

[1] 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.

[2] Zhao, H. Y., Liu, Y. X., & Li, S. Q. (2019). An Improved Random Forest Algorithm for Sleep Disorder Classification. BioMed Research International, 2019.

[3] Wang, Y., Li, J., & Zhang, X. (2020). Sleep Disorder Detection and Classification Using Multimodal Physiological Signals. IEEE Journal of Biomedical and Health.

[4] 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.

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

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

[7] 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.

[8] 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.

[9] Vuppalapati, A. K., Guddeti, V., & Prasad, P. (2021). Sleep Disorder Classification Using EEG Signal Analysis

and Machine Learning. IEEE Access, 9, 54321-54330.

[10] 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