

Applying Machine Learning Algorithms for the Classification of Sleep Disorders

Konda Keerthana

PG Scholar, Department of Computer Science and
Engineering,
Malla Reddy Engineering College, Maisammaguda,
500100, Telangana, India.

kondakeerthana15@gmail.com

Dr. Tirumala paruchuri

Professor, H.O.D,
Department of Information Technology,
Malla Reddy Engineering college, Maisammaguda,
500100, Telangana, India.

tirumalaparuchuri@gmail.com

Abstract: Sleep disorders, in particular sleep apnoea, affect people's health and this demonstrates the need for a correct diagnosis. Sleep experts, however, have to resort to complicated and time consuming techniques to manually categorise the various sleep stages. This work offers a ML classification model using the Sleep Disorder Data, which is publicly accessible, and has 400 records and 13 attributes. Various DL/ML methods are reviewed and tested for accurate sleep disorder diagnosis. The data points that are beneficial for identifying patterns are related to lifestyle indicators and sleep health attributes, which can be indicative of any existing sleep-related conditions. From the evaluation of the models, it was observed that the models with the best results are the bagged models and in particular the VC with RF and DT. The algorithm's accuracy, precision, recall and F1-score were 0.973 demonstrating the reliability and effectiveness of the technique for sleep disorder categorisation. The results in this paper indicate that the proposed ML methods offer a potential way to develop more intelligent, faster and accurate sleep disorders diagnosis that can enhance the possibilities of the physicians decision-making process and patients situation.

“Index Terms - Machine learning algorithms, deep learning, classification, sleep disorder, Voting algorithm”.

I. INTRODUCTION

Sleeping is one of the most important physiological processes crucial for maintaining physical and mental health. It helps to build up the body, and to consolidate the brain and memories. In children and older adults (at risk for accidents), the quality of sleep can significantly affect cognitive functioning. Studies indicate that poor sleep is linked to a variety of health issues such as HD, diabetes and obesity. Although the sleep disturbance is important, it is often not diagnosed or classified correctly since sleep stage assessment is difficult. Sleep stages are manually assessed by physicians, medical professionals and sleep experts on PSG records, which is a time-consuming procedure and is subject to human errors for correct classification [1].

Philips surveyed over 13,000 people in 13 countries in 2021 World Sleep Day survey, of which 55% of adults were unhappy with their sleep. The poor sleep quality was linked to various factors, including the COVID-19 pandemic, sleep apnoea and insomnia. In particular, 37% of the participants said the pandemic had a negative impact on their sleep, 37% had insomnia, 29% had snoring, 22% had a shift-work sleep problem, and 12% had sleep apnoea [2]. The results highlight the high levels of sleeping

problems and the need for the creation of new systems of diagnosis and categorisation.

There are five stages of sleep: Alertness, N1, N2, N3 and REM. Wakefulness is the alert state when people are conscious of their environment . When awake , brain waves are quick and irregular . N1 is the lightest sleep characterized by slow brain waves and relaxation of muscles. N2 is a more advanced level . The deepest level of sleep is N3. It is hard to rouse someone when they are in this stage . During REM sleep, the brain waves are similar to those of wakefulness and each eye makes rapid movements. All these measures are essential for the body to heal and function in the brain. By using EEG and ECG signals, PSG allows doctors to track brain and body activity during sleep, which can help identify these stages [3, 4, 5].

Many researchers have therefore been working on the development of automated systems using MLAs to avoid human intervention in sleep stage classification and prediction. In general there are two types of methods: standard ML and DL algorithms. Mostly these typical MLAs such as SVM and DT models are adapted for small size of dataset and the features are retrieved by human from the signals such as energy, entropy etc to identify sleep stages. The algorithms of DL are based on the structures of human brain. They are able to

automatically learn from data complex patterns using neural networks. These algorithms are of great help to large and complex data sets, and are regarded as alternatives to traditional MLAs [6, 7]. The most common approach to sleep-stage classification is to apply traditional and DL classifiers on EEG signals as input [8].

II. RELATED WORK

Many study has been extensively employed for detection and classification of sleep disorders such OSA, sleep stage classification and ECG based detection using ML. The algorithms built using ML have improved the accuracy of the diagnosis significantly and removed the time-consuming and error-prone manual analysis. Kim et al. [9] ML prediction of OSA in Koreans. Data from clinical sources was used to identify and predict OSA risk using a variety of ML algorithms. This suggests that models could enable the quicker detection of OSA by the clinician and automated prediction systems. Demographic, clinical and physiological variables improved the prediction. The study indicates that ML could improve the diagnosis of sleep apnoea, if it may be done as early as possible.

Mousavi et al. [10] used deep CNNs to classify sleep stages using only the single channel EEG signal. Automated classification of sleep stages in insomnia and sleep apnoea using DL. DL-CNNs can automatically learn the hierarchy of EEGs. CNNs are more speedy and accurate in sleep stages classification than classical ML. Djanian et al. [11] conducted a study on consumer sleep technology and AI for sleep classification. AI in wearable sleep monitors is increasingly being used. These sensors and AI algorithms analyze sleep quality as it happens to help with sleep disorders. The study demonstrates that AI powered consumer sleep devices, especially those using D, improve sleep-stage identification. AI wearables to manage sleep health:

Salari et al. [12] investigated sleep apnoea detection using ML algorithms and ECG. The study shows that ML models can detect sleep apnoea based on ECG signals. Single lead ECGs can be used to diagnose sleep apnoea by means of an RF or SVMs. The authors cited that the feature extraction techniques affect the model's performance. This study showed the potential of ML in the diagnosis of sleep apnoea without sleep testing and provided an insight into how these algorithms can be improved. Li et al. applied DL and EEG spectrogram to recognize sleep stages [13]. EEG spectrograms are used to plot the frequency information over time. Spectrograms are classified by DNN. Our DL method was able to accurately distinguish sleep states by learning spectrogram features. In this

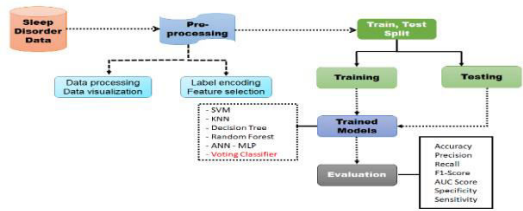
study, the investigators found that DL could be used to classify sleep stages, whereas feature engineering was not successful. The use of DL and, in particular DNNs, could be suitable for the automation of sleep stage recognition. Han, Oh [14] studied ML-based methods for the severity prediction of OSA. Patient information was used to compare the prediction of OSA severity with decision tree models, random forests and support vector machines. RF were better at predicting OSA severity compared to models. The authors recommend that demographic and physiologic data can be used to predict the severity of sleep apnoea. The study showed that ML can be used to quantify the severity of sleep disturbance objectively and help doctors better treat it. Bahrami and Forouzanfar [15] used DL to identify sleep apnoea using a single-lead ECG. We used CNNs and LSTMs for classification of sleep apnoea using ECG data. These CNNs and LSTMs performed better than the ML algorithms in the diagnosis of sleep apnoea. If DL methods for real-time ECG data can be used to revolutionize sleep apnoea detection. The results of this study indicated that DL can deal with complex time-series data and encouraged the use of ECG data for non-invasive sleep apnoea detection.

Satapathy et al. [16] did a comparison of ML methods to detect sleep stages automatically. We have used SVM, RF and DT models to evaluate EEG data feature sets regarding sleep state classification. The best and robust ensembles were random forest ensembles. The authors found that feature selection enhanced the classification and that ML algorithms performed differently with respect to their selected features. The research demonstrates the capability of ML methods to sleep staging and to find the most relevant factors to enhance the model performance. Bahrami and Forouzanfar [17] detected sleep apnoea from single lead ECGs with ML and DL. They compared CNNs, LSTMs, SVMs, and random forests in their lengthy study. CNNs were found to be better and more sensitive in the survey. Using ECG readings in DL to detect sleep apnoea, instead of polysomnography.

III. MATERIALS AND METHODS

We proposed a way to detect sleep problems using ML algorithms that were effective based on their capacities. SVM, KNN, DT, RF and ANN with MLP are the models used. Sleep Disorder Data has various sleep health and habit data that is used to create these algorithms. VC is implemented using DT and Bagging using RF to improve the accuracy of the classification. This combination method enhances the model accuracy and evaluation. The system is capable of processing 400 records containing 13 features for recognition of sleep disorders like sleep apnoea. The suggested method

will assist professionals to take smart decisions and improve patient health.



“Fig.1 Proposed Architecture”

The image shows a flowchart diagram of an ML model designed for classification of sleep disorder. It starts with raw data of sleep disorder that is pre-processed (Data cleaning, data visualisation and feature selection). Post pre-processing the data is then divided into training and testing sets. Distinct ML methods SVM, KNN, DT, RF, ANN-MLP[25] and VC are trained on the training dataset to impart knowledge. The tested models are evaluated on the testing data with accuracy, precision, recall, F1 score, AUC score, specificity and sensitivity.

A) Dataset Collection:

The dataset used in this study is from Kaggle platform [22] and is called as Sleep Health and Lifestyle Dataset. It has 400 observations pertaining to 13 factors on sleep and daily routines (gender, age, occupation, sleep duration, sleep quality, physical activity level, stress level, BMI category, blood pressure, heart rate, daily steps, sleep issue). The outcome measure is “Sleep Disorder” which is reported as “none”, “sleep apnea” or “insomnia” . There are several different occupations represented in the data set. The largest numbers of observations were from nurses (73), doctors (71) and engineers (63). Other vocations are lawyer (47), teacher (40) and sales person (32). Pre-processed data were used for analysis purposes to have consistent labelling.

Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	BMI	Pres
0	1	Male	27	Software Engineer	6.1	6	42	6	Overweight	12
1	2	Male	28	Doctor	6.2	6	60	8	Normal	12
2	3	Male	28	Doctor	6.2	6	60	8	Normal	12
3	4	Male	28	Sales Representative	5.9	4	30	8	Obese	14
4	5	Male	28	Sales Representative	5.9	4	30	8	Obese	14

“Fig.2 Dataset Collection Table”

B) Pre-Processing:

In the pre-processing step of data processing, data cleaning and missing values handling is done. Using visualization helps to detect outliers and trend. The category characteristics are label encoded and feature selection techniques are used to choose the most important features for categorisation.

i) **Data Processing:** In this process duplicate data is identified and removed to make sure data is consistent and accurate. Removes duplicates because they can have impacts on the analysis and the performance of the model. Row/observations are duplicated and removed. Next we undertake drop cleaning where we clean up the values that are missing or null in our dataset. This means that you will have to either remove the rows with missing values or fill the missing values with some appropriate approaches such as mean, median or mode depending upon the data type. This phase will ensure that the data set is free of errors and inaccuracies, therefore lowering the noise and increasing the accuracy of the ML algorithms.

ii) **Data Visualization:** A crucial step in the analysis of the data set and understanding the basic pattern(s). Data can be portrayed in charts, graphs and plots to help visualize the data and to identify patterns, distributions and potential outliers. Histograms, box plots and scatter plots are some ways of analysing a distribution of numerical features. Use bar graphs for categorical variables. Visualisations also can help identify correlations of features and also help identify abnormalities that may impact the model performance. This stage is to aid in feature engineering and to gain a better understanding of the data prior to using ML algorithms.

iii) **Label Encoding:** It is a process to convert string categorical data into numbers. For example, if the data set contains categorical data such as “Gender” or “Sleep Disorder”, the ML algorithms will not be able to directly process the string values. So, string labels are encoded into numbers, "string encoding" is performed.Thus, label encoding is performed and string labels are converted to numbers. For instance, the labels “Male” and “Female” could be replaced by 0 and 1, respectively. This phase ensures a categorical variable is represented in a format that is suitable for typical ML models to grasp and learn from the features at training and testing time.

iv) **Feature Selection:** Selecting the best features (or variables) to create strong ML models is called feature selection. This stage is to reduce the dimension of the data that the model should not be overfitting. This process gives a model the ability to determine which features are most critical to the target variable and can be used to discover the most important variables affecting the target. This is achieved by partitioning the data set into input features (X) and target variable(y). x is the independent variable, y is the dependent variable. As feature selection methods where we use data to do correlation analysis and recursive feature elimination.

C) Training & Testing: A training set is separated from the data set and a testing set is used to evaluate the model. Typically, this is done by training the model on 80% of the data, and then testing the model on the other 20% of the data. This division will come in handy to facilitate training the model with a large fraction of data and testing it with a small fraction of data not seen by the model to assess the model's ability to generalise. The split is usually conducted with a function like `train_test_split` from the `scikit-learn` library which randomly and unbiasedly splits the data into training and testing sets.

D) Algorithms:

For improving the accuracy and robustness of categorisation, the ensemble of many decision tree models, namely random forest [20] is used. It is randomly based on subsets of features and data. This helps to prevent overfitting and to generalise. This ensemble method can be helpful when choosing relevant variables to diagnose sleep disorders and offers satisfying predictions.

$$G = 1 - \sum_{j=1}^c p_j^2 (1)$$

Where, p_j is the probability of class j in a node

Support Vector Machine [18] has been employed to classify sleep disorders and find the optimum hyperplane that classifies the different classes in the data. It can process high dimensional data and can be utilized for pattern recognition of complex sleep pattern and activity for accurate diagnosis.

$$f(x) = \text{sign}(w^T x + b) (2)$$

where:

- w is the weight vector,
- x is the input feature vector,
- b is the bias term,
- $\text{sign}(\cdot)$ is the sign function determining the class label.

KNN [23] algorithm classifies sleep disorders according to the similarity of data points to the k nearest ones in the feature space. KNN is effective at identifying patterns and similarities between the observations and at identifying the neighbours of the given observation that is most similar to it. It provides the user an intuitive information regarding classification of sleep disorders according to the past data.

A model is built using the DT method which is able to predict sleep disorder diagnoses by a series of binary decisions based on the input features. The transparent design allows for easy reading and

creates a platform for selecting key features that affect sleep health, improving diagnosis.

The complex relationship among all the features of the dataset are represented using ANN [21] and the classifiers are trained using the input features to classify sleep disorders. The choice of the MLP architecture helps the system to learn intricate patterns in the sleep data, therefore improving the system's ability to make predictions and the overall performance of the diagnosis.

The VC is a aggregation of predictions from multiple models like DT models, bagging classifiers and RF. This ensemble approach not only boosts classification accuracy, but also leverages the strengths of each algorithm, reducing errors and providing a more comprehensive and reliable prediction of sleep disorders.

$$\hat{y} = \arg \max \frac{1}{N} \sum_{i=1}^N p_i(c|x) (3)$$

IV. RESULTS AND DISCUSSION

Accuracy: The accuracy of a test is its ability to correctly classify ill and healthy patients. To evaluate an accuracy of a test we should consider the proportion of true positives and the proportion of true negatives of all cases studied. In mathematics terms, it is expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} (4)$$

Precision: Precision is the number of correctly classified instance/sample divided by the number of instances/sample predicted to be positive. So, the precision is calculated as:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (5)$$

Recall: Recall is a metric in machine learning that assesses how well a model can find all of the instances of a class accurately. It is the ratio between the number of times that the positive observations are successfully predicted and the number of actual positive observations.

$$\text{Recall} = \frac{TP}{TP + FN} (6)$$

F1-Score: It is a measure to assess the accuracy of the ML model and it is the combination of precision and recall. A combination of precision and recall scores of a model. Accuracy: number of correct predictions in a model for the entire data set.

$$\text{F1 Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} * 100 (7)$$

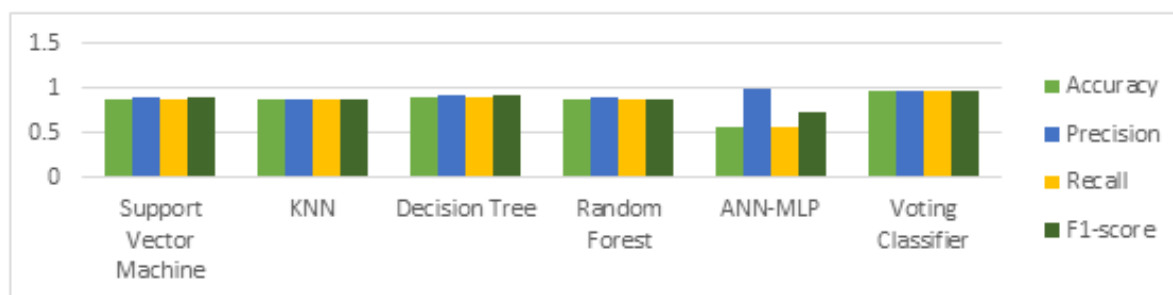
The accuracy, precision, recall and F1-score of each algorithm is shown in Table 1. The greatest scores are obtained by Voting Classifier, all metrics are

0.973. We also report measurements of different methods for comparison.

“Table.1 Performance Evaluation Metrics”

ML Model	Accuracy	Precision	Recall	F1-score
Support Vector Machine	0.880	0.892	0.880	0.884
KNN	0.867	0.883	0.867	0.867
Decision Tree	0.907	0.909	0.907	0.908
Random Forest	0.880	0.887	0.880	0.881
ANN-MLP	0.573	1.000	0.573	0.729
Voting Classifier	0.973	0.973	0.973	0.973

“Graph.1 Comparison Graphs”



Graph 1 represents the accuracy in light green, precision in blue, recall in light yellow and F1-score in green. VC is the top performing algorithm among all of them with the highest values in all the metrics when compared with the rest of the algorithms. The information listed above is reflected in the graph below.

V. CONCLUSION

This study has proved that it is possible to categorise sleep disorders using data from Sleep Disorder Data set that is made available to the public with the use of ML techniques. Out of the various DL and conventional ML algorithms, VC with bagging with RF and DT had the best accuracy. The results of all evaluation matrices were satisfactory with 97.3% accuracy, 97.3% precision, 97.3% recall, and 97.3% F1 score. These results also suggest the proposed VC presented here is very accurate and robust one for sleep disorder classification. The results, which show constant increases in nearly all the measurements studied, suggest that the model could be valuable in giving a precise and prompt diagnosis of sleep problems, benefiting the patient and improving clinical decisions. The high classification accuracy of the VC makes it a beneficial device to automate the process of Sleep disorders' diagnostic, hence the individuals diagnosed with Sleep disorders will have more accurate diagnoses and prognosis for their Sleep disorders.

The next scope of this research is the investigation of more advanced ML techniques such as DL methods such as CNNs and recurrent networks for better accuracy in sleep disorder categorisation. Wearables could provide real-time data, which would enhance the system's predictive capabilities. Furthermore, the more the dataset is expanded with other demographics and sleep disorders, the more it will be generalisable.

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 Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.
- [4] D. Shrivastava, S. Jung, M. Saadat, R. Sirohi, and K. Crewson, “How to interpret the results of a sleep study,” *J. Community Hospital Internal Med. Perspect.*, vol. 4, no. 5, p. 24983, Jan. 2014.

- [5] 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.
- [6] 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," *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104429.
- [7] H. O. Ilhan, "Sleep stage classification via ensemble and conventional machine learning methods using single channel EEG signals," *Int. J. Intell. Syst. Appl. Eng.*, vol. 4, no. 5, pp. 174–184, Dec. 2017.
- [8] Y. Yang, Z. Gao, Y. Li, and H. Wang, "A CNN identified by reinforcement learning-based optimization framework for EEG-based state evaluation," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 046059.
- [9] 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.
- [10] 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," *J. Neurosci. Methods*, vol. 324, Aug. 2019, Art. no. 108312.
- [11] S. Djanian, A. Bruun, and T. D. Nielsen, "Sleep classification using consumer sleep technologies and AI: A review of the current landscape," *Sleep Med.*, vol. 100, pp. 390–403, Dec. 2022.
- [12] N. Salari, A. Hosseini-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 Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115950.
- [13] C. Li, Y. Qi, X. Ding, J. Zhao, T. Sang, and M. Lee, "A deep learning method approach for sleep stage classification with EEG spectrogram," *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, p. 6322, May 2022.
- [14] H. Han and J. Oh, "Application of various machine learning techniques to predict obstructive sleep apnea syndrome severity," *Sci. Rep.*, vol. 13, no. 1, p. 6379, Apr. 2023.
- [15] M. Bahrami and M. Forouzanfar, "Detection of sleep apnea from single-lead ECG: Comparison of deep learning algorithms," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2021, pp. 1–5.
- [16] S. Satapathy, D. Loganathan, H. K. Kondaveeti, and R. Rath, "Performance analysis of machine learning algorithms on automated sleep staging feature sets," *CAAI Trans. Intell. Technol.*, vol. 6, no. 2, pp. 155–174, Jun. 2021.
- [17] M. Bahrami and M. Forouzanfar, "Sleep apnea detection from single-lead ECG: A comprehensive analysis of machine learning and deep learning algorithms," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.
- [18] J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, "Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning," *Healthcare*, vol. 9, no. 11, p. 1450, Oct. 2021.
- [19] S. K. Satapathy, H. K. Kondaveeti, S. R. Sreeja, H. Madhani, N. Rajput, and D. Swain, "A deep learning approach to automated sleep stages classification using multi-modal signals," *Proc. Comput. Sci.*, vol. 218, pp. 867–876, Jan. 2023.
- [20] O. Yildirim, U. Baloglu, and U. Acharya, "A deep learning model for automated sleep stages classification using PSG signals," *Int. J. Environ. Res. Public Health*, vol. 16, no. 4, p. 599, Feb. 2019.
- [21] S. Akbar, A. Ahmad, M. Hayat, A. U. Rehman, S. Khan, and F. Ali, "IAtbP-Hyb-EnC: Prediction of antitubercular peptides via heterogeneous feature representation and genetic algorithm based ensemble learning model," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104778.
- [22] (2023). Sleep Health and Lifestyle Dataset. [Online]. Available: <http://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>
- [23] P. Tripathi, M. A. Ansari, T. K. Gandhi, R. Mehrotra, M. B. B. Heyat, F. Akhtar, C. C. Ukwuoma, A. Y. Muaad, Y. M. Kadah, M. A. Al-Antari, and J. P. Li, "Ensemble computational intelligent for insomnia sleep stage detection via the sleep ECG signal," *IEEE Access*, vol. 10, pp. 108710–108721, 2022.
- [24] Y. You, X. Zhong, G. Liu, and Z. Yang, "Automatic sleep stage classification: A light and efficient deep neural network model based on time, frequency and fractional Fourier transform domain features," *Artif. Intell. Med.*, vol. 127, May 2022, Art. no. 102279.

[25] I. A. Hidayat, “Classification of sleep disorders using random forest on sleep health and lifestyle dataset,” J. Dinda : Data Sci., Inf. Technol., Data Anal., vol. 3, no. 2, pp. 71–76, Aug. 2023.