**Faridpur Engineering College, Faridpur**



Department of Computer Science & Engineering

Course Name: Machine Learning (Sessional)

Course Code: CSE 816

A project report on:

**Sleep Stage Detection Using Machine Learning Algorithm**

Submitted to Submitted by

Md. Suman Reza Md Rakibul Islam Nahid

Lecturer ID: 2537 || Reg: 1797

Dept. of Computer Science & Engineering Rifa Tamanna Mitu

Faridpur Engineering College, Faridpur. ID: 2525 || Reg: 1780

Rifat Ara Nur Rekha

ID: 2534 || Reg: 1793

May 2023

**Index**

**Title Page No.**

**Abstract**  **2**

**Introduction 2**

**Data Preprocessing 3**

**Exploratory Data Analysis 4**

**Feature Engineering 5**

**Model Selection 5**

**Model Training 6**

**Model Evaluation 7**

**Result 7**

**Conclusion 8**

**Future Work 9**

**References 10**

**Abstract:** In this work, we present a machine learning-based method for identifying sleep stages using a variety of physiological data collected throughout various phases of sleep research. The goal of this research is to create an efficient and reliable algorithm that can automatically distinguish between different phases of sleep, such as waking, REM, and NREM. This is accomplished by analyzing a database of sleep-recorded physiological signals including electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG).

**Keywords:** Sleep stage detection, machine learning, physiological signals, electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), support vector machines (SVM), random forests, deep learning, sleep disorders, healthcare.

**Recommendations:**

* Incorporating Multi-modal Data,
* Long-Term Sleep Monitoring,
* Online Sleep Stage Detection,
* Fine-Grained Sleep Stage Classification,
* Transfer Learning and Generalization,
* Explainable AI and Interpretability,
* Robustness to Individual Differences,
* Validation and Clinical Trials.

The area of sleep stage identification using machine learning may be advanced by addressing these suggestions, leading to better sleep problem diagnosis, personalized sleep treatment, and a deeper knowledge of sleep-related events.

**Introduction:** When it comes to de-stressing, one of the most fundamental biological functions is sleep. The basic processes of the brain are essential for a person's capacity for learning, performance, and physical activity [1, 2]. The most pressing and intriguing question in the fields of neuroscience and sleep disorder diagnosis is how to better understand sleep quality. The current gold standard for studying sleep is sleep stage scoring [3]. Finding the critical phases of sleep for diagnosing and treating sleep disorders is the purpose of sleep stage scoring [4, 5].

For rating sleep stages, researchers employ polysomnographic (PSG) data, which are the continuous recording of several electrophysiological signals. The most generally used standard for sleep stage categorization is provided by the American Academy of Sleep Medicine (AASM). PSG signal recordings are categorized as either Non-Rapid Eye Movement (NREM) sleep, REM sleep, or waking (W) under this standard. Newer recommendations in this area may be found at [6], [7] from the American Academy of Sleep Medicine (AASM). Each of the five stages of sleep has its own unique wave pattern, and the AASM guidelines outline those waves in detail [8].

The nervous system and the rest of the organism undergo a number of functional changes during NREM and REM sleep [9]. Alterations in hormone levels also occur throughout these two phases [9]. Deep non-rapid eye movement (NREM) sleep is associated with reduced sympathetic nerve activity. During rapid eye movement (REM) sleep, breathing patterns shift and become more rapid and irregular ([10], [11]). The total metabolic rate and blood flow during REM are comparable to awake, but NREM sleep is associated with significant decreases in both [12].In NREM sleep, alpha, beta, and gamma waves were suppressed; in wakefulness, theta and delta waves increased; and in REM sleep, alpha and beta waves predominated. There are a number of abnormalities of the neurological system and other bodily functions that may be detected via the careful monitoring of NREM and REM sleep.

**Dataset:** ﻿We utilized the sleep recording of Haaglanden Medisch Centrum (HMC, The Hague, The Netherlands), available as an open-access public dataset in PhysioNet [12]. It was collected in 2018 and published very recently on 1 July 2021. The dataset includes a whole-night PSG sleep recording of 154 people (88 Male, 66 Female) with a mean age of 53.8 15.4 years. Patient recordings were chosen at random and represented a diverse group of people who were referred for PSG examinations in the context of various sleep disorders. All signals were captured at 256 Hz using AgAgCl electrodes on SOMNO- screen PSG, PSG+, and EEG 10–20 recorders (SOMNOmedics, Randersacker, Germany). Each recording consists of four-channel EEG (F4/M1, C4/M1, O2/M1, and C3/M2), two- channel EOG (E1/M2 and E2/M2), one-channel bipolar chin EMG, and one-channel ECG. The recordings also contain the sleep scoring, consisting of W, N1, N2, N3, and R for an epoch of 30 sec. The AASM guidelines were used to score sleep stages which were manually scored by well-trained sleep technicians [20]. We have decided to use three EEG channels (F4, C4, and O2) in this study according to the international 10–20 EEG system.

**Data Preprocessing:** To guarantee that the data used for training models is of good quality and faithfully represents the actual world, machine learning (ML) relies heavily on data cleaning and preprocessing. Some typical data pretreatment and cleaning methods in ML are as follows:

* Data cleaning
* Feature scaling
* Feature encoding
* Dimensionality reduction
* Handling missing data
* Outlier detection and removal
* Data normalization
* Handling class imbalance

We only used Feature Encoding Data Preprocessing Technique.

**Missing Data:** There are several potential causes of missing data, including mistakes in data collecting or incomplete information. Since many machine learning algorithms fail when presented with missing values, dealing with missing data is an essential step. In ML, missing data may be dealt with in a number of ways.

1. Deletion: To delete is to remove all rows or columns that have null values. However, this approach sometimes leads to data loss.
2. Imputation: Imputation is a method for making a statistical guess to fill in gaps in data. Mean imputation, median imputation, and regression imputation are three common types of imputation.
3. Prediction: The missing values are predicted using machine learning methods and other aspects of the dataset.

**Outliers:** Extreme numbers that stand out from the rest of the data in a collection are called outliers. Data collecting mistakes may cause outliers, but sometimes unusual data points simply indicate exceptional circumstances. Since outliers may distort findings, they can significantly affect how well machine learning systems function. Several methods exist in ML for dealing with outliers, such as:

1. Removal: The outlier must be eliminated from the data set. However, this approach sometimes leads to data loss.
2. Transformation: Logarithmic transformation and Box-Cox transformation are two examples of transformation methods that may be used to make data more normal and mitigate the effect of outliers.
3. Winsorization: Winsorization is a method wherein outlying numbers are swapped for more moderate ones. For instance, the 95th percentile may be used to replace all numbers above it.

**Other Issues:** Problems such as multicollinearity and overfitting are also possible in machine learning. When there is an unequal number of samples between categories, we say that there is an imbalance in the data. When three or more variables in a dataset are strongly linked, this is known as multicollinearity. Overfitting happens when a model is too complicated and closely matches the training data, leading to subpar results when applied to novel data. Several methods exist for dealing with such problems, such as:

1. Data augmentation: To achieve statistical parity, data augmentation may be used to artificially inflate the proportion of minority group samples.
2. Feature selection: Features are picked: Using this method, only those independent variables that have a low correlation with the result variable are retained.
3. Regularization: Overfitting may be avoided by regularization, which entails including a penalty term in the objective function of the model. L1 regularization (Lasso) and L2 regularization (Ridge) are two well-known techniques for achieving this goal.

We had no missing data, outliers and other issues in our dataset.

**Exploratory Data Analysis:** To better comprehend the data, spot anomalies and trends, and choose the best model for the issue at hand, exploratory data analysis (EDA) is a crucial stage in the Machine Learning (ML) process. The features of the data are analyzed and summarized with the help of EDA methods. Some frequent EDA approaches in ML are as follows:

* Statistical descriptions
* Data Visualization
* Correlation analysis
* Missing value imputation
* Outlier detection
* Dimensionality Reduction

When it comes to laying the groundwork for a machine learning model, EDA methods are crucial. These methods aid in the discovery of meaningful connections within the data, which in turn facilitates the development of a more trustworthy model. Describe any associations or trends that were spotted in the data.

**Feature Extraction:** The electroencephalogram (EEG) may be characterized by its frequency and power in certain frequency ranges. The frequency range of the delta () band is 0.5 Hz to 4 Hz, that of the theta () band is 4 Hz to 8 Hz, that of the alpha () wave is 8 Hz to 13 Hz, that of the beta () band is 13 Hz to 30 Hz, and that of the gamma () wave is 30 Hz to 44 Hz [13,14]. To investigate the potential within EEG data, characteristics were retrieved from EEG signals using Fast Fourier transforms (FFT) and other techniques. Welch periodograms [15] were used to calculate power spectral density (PSD) for each time period. Average power, median frequency, mean frequency, spectral edge, and peak frequency were calculated from this PSD for each time period. A value of 30 seconds was chosen as the epoch width. Table 1 summarizes the EEG characteristics that were extracted for this investigation. The 75 EEG feature sets in this dataset represent a wide variety of brain activity.

Table 1. Features extracted from the EEG signal. The Global channel is averaged over F4, C4, and O2 electrodes.

| EEG Channel | EEG Spectral Waves | EEG Feature | Number of Feature |
| --- | --- | --- | --- |
| F4, C4 and O2 | δ,θ,𝛼,β and 𝛾 | Median Frequency | 15 |
| F4, C4 and O2 | δ,θ,𝛼,β and 𝛾 | Mean Power | 15 |
| F4, C4 and O2 | δ,θ,𝛼,β and 𝛾 | Spectral Edge | 15 |
| F4, C4 and O2 | δ,θ,𝛼,β and 𝛾 | Mean Frequency | 15 |
| F4, C4 and O2 | δ,θ,𝛼,β and 𝛾 | Peak Frequency | 15 |

Selecting, manipulating, and synthesizing new features from the raw data is what feature engineering is all about. There are a number of ways in which feature engineering may boost a model's effectiveness:

* Better Feature Representation.
* Improved Model Interpretability.
* Reduced Overfitting.
* Improved Robustness.

**Model Selection:** In machine learning, Random Forest is widely used for both classification and regression. It's a kind of ensemble learning in which many decision trees are constructed during training and the median or average prediction (depending on the task at hand) is returned.

Using random feature subsets and bootstrap aggregation (bagging) to increase the trees' variety, Random Forest generates a collection of decision trees on various subsets of the training data. Each decision tree is built with a different selection of attributes, and the final forecast is the mean of all the trees' estimates. This method helps to increase the model's generalization abilities while decreasing the likelihood of overfitting.

Due to its superior performance in terms of criteria like accuracy, precision, recall, and F1-score, Random Forest was ultimately chosen as the best method for a given job. The excellent accuracy, scalability to huge datasets with high-dimensional features, and provision of a measure of feature relevance that Random Forest offers led to its adoption. Random Forest is a popular option for many machine learning applications because of its low learning curve, as well as the fact that it is well-implemented in many different machine learning frameworks.

**Model Training:** In machine learning, "training" a model is running it against a dataset to identify patterns and correlations that will help it correctly predict or classify incoming data.

Several hyperparameters allow for fine-tuning of the machine learning method Random Forest. The maximum depth of each tree (max\_depth) and the total number of trees (n\_estimators) are two crucial hyperparameters.

The forest's tree count is determined by the n\_estimators hyperparameter. There is a tradeoff between improved performance and longer training times and computational complexity when increasing the number of trees used in an algorithm. How many estimators (n\_estimators) you should use to get the best results is data- and resource-specific. In practice, n\_estimators are often varied within a range, and the best performing value is chosen through validation or cross-validation.

In a decision forest, the maximum depth of each tree is set by the max\_depth hyperparameter. In general, the performance of the individual trees improves with increasing the maximum depth, but this comes with a higher danger of overfitting. If you want to strike a good balance between model complexity and generalization performance, the best value for max\_depth will change depending on the size and complexity of the dataset. In practice, it is customary to experiment with several max\_depth settings before settling on the one that yields the best results in validation or cross-validation.

TABLE I: OPTIMUM PARAMETER VALUES

| Parameters | Random Forest |
| --- | --- |
| n\_estimator | 240 |
| max\_depth | 75 |

Training machine learning (ML) models is not without its difficulties, some of which include:

1. Quantity and quality of the data: Obtaining high-quality and adequate data for training a model is a significant difficulty in ML. The information must be fair and accurate in its portrayal of the demographic of interest.
2. Engineering of features: Selecting and modifying the most useful characteristics from the given data is known as feature engineering. This is an essential part of the ML process since it determines how well the model will function in practice.
3. Excessive or insufficient snugness: When a model becomes too complicated and learns to fit the training data too well, it is said to have overfit, and it will perform poorly when applied to fresh data.
4. Tuning hyperparameters is the process of adjusting a model's initial settings and parameters before it is trained.
5. Lack of computing resources, such as a strong central processing unit or graphics processing unit, is a typical issue. This may slow down the training process and restrict the models' potential complexity and scale.
6. To get around this problem, you may leverage cloud-based services like Google Collab that provide you access to powerful computers. Even if these tools are accessible, there may not be enough memory or computing power to fully complete the training.

**Model Evaluation:** Metrics are used to evaluate a machine learning model's accuracy and how effectively it can learn from data it has never seen before. Commonly used assessment measures for sleep stage identification include:

* Accuracy: The proportion of samples in the dataset whose labels are accurate.
* Precision: The percentage of correct model predictions relative to all correct forecasts.
* Recall: The percentage of positive samples that are really representative of the dataset.
* F1 score: Harmonic mean of accuracy and recall with weights.

In addition to these measures, the confusion matrix is often used to see how well a model does on various phases of sleep. For each sleep stage, the confusion matrix details how many samples were properly identified and how many were misclassified.

TABLE II: PERFORMANCE METRIC RESULTS FOR MODELS

| Evaluation Metric | Random Forest | |
| --- | --- | --- |
| NREM | REM |
| Precision | 96% | 77% |
| Recall | 94% | 84% |
| f1-Score | 95% | 80% |

**Results:**

TABLE III: FINAL PERFORMANCE METRIC RESULTS FOR MODELS

| Evaluation Metric | Random Forest |
| --- | --- |
| Accuracy | 93% |

The bulk of studies employ EEG data collected overnight to categorize different types of sleep [15]. These papers suggest using a number of different feature reduction procedures to zero in on the most important aspects. The stages of sleep have been categorized in a number of different ways. However, none of these methods limits sleep to NREM or REM states. Using the ISRUCSleep [18] dataset, Satapathy et al. [16] suggested a technique for detecting two phases of sleep, such as awake and sleep. Overall, accuracy in the waking state was 91.67 percent, while in the sleep state, it was 93.8 percent. With an accuracy of 81.65%, Shen et al. [17] categorized sleep states into five phases including waking, NREM, and REM using a machine learning model they suggested based on the essence of characteristics applied to the ISRUC [19] dataset. Because these parameters tend to decline during NREM sleep and rise during REM sleep, biomarkers were used to analyze them.

Several practical applications exist for the sleep stage detection model, all of which highlight the significance of sleep stage monitoring and analysis. Some instances are as follows:

* Sleep disorder diagnosis: Sleep apnea, insomnia, and narcolepsy are all diagnosable by looking at a person's sleep phases. Sleep problems may be diagnosed and treated more effectively with the help of this model's automated detection and classification of sleep phases.
* Sleep quality monitoring: The model can track sleep conditions and pinpoint causes of poor slumber. If the model finds that a person isn't getting enough deep sleep, for instance, it's a sign that they might need some lifestyle adjustments or medical help.
* Performance optimization: Cognitive function, emotions, and general well-being are all profoundly affected by sleep's quality and quantity. The model may be used to enhance sleep quality and performance by pinpointing limiting variables and suggesting changes in routine.
* Sports performance: Athletic performance and recuperation are greatly aided by adequate sleep. Athletes may use the model to track their sleep cycles and develop better routines to boost their performance.

**Conclusion:** In this study, we created a machine learning model to analyze EEG data and identify different stages of sleep. To train a random forest classifier, we first retrieved numerous characteristics from the epochs.

Accuracy, precision, recall, and the F1 score were only a few of the criteria used to assess the model's efficacy. The model's excellent accuracy and F1 score on the test data show that it successfully detects and labels different phases of sleep using just EEG data.

The following goals were met by the project:

1. A machine learning model for identifying sleep stages from EEG data has been built and validated.
2. Used a number of criteria to assess the model's efficacy; test results showed good accuracy and an F1 score.
3. Several potential real-world applications of the model were identified and discussed. These included the diagnosis of sleep disorders, monitoring of sleep quality, improvement of performance, and enhanced athletic ability.

**Limitation:**

It's important to recognise the following constraints on this work:

* Weak generalization due to small sample size; additional data from other populations would help.

**Future Work:**

1. Gathering additional data from a wider range of people will improve the model's ability to generalize.
2. To enhance the model's precision and performance, we will investigate alternative feature extraction methods and machine learning algorithms.
3. Creating a tool that can be used by both medical professionals and patients to monitor and analyze sleep patterns in real time.

**Reference:**

[1] Estrada, E and Nava, Patricia and Nazeran, Homayoun and Behbehani, Khosrow and Burk, J and Lucas, Itakura distance: A useful similarity measure between EEG and EOG signals in computer-aided classification of sleep. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2006 , pp. 1189–1192.

[2] Li, Yi and Yingle, Fan and Gu, Li and Qinye, Tong, “Sleep stage classification based on EEG Hilbert-Huang transform,” in IEEE, 2009 4th IEEE Conference on Industrial Electronics and Applications, 1963, pp. 3676–3681.

[3] Ebrahimi, Farideh and Mikaili, Mohammad and Estrada, Edson and Nazeran, Homayoun, “Assessment of itakura distance as a valuable feature for computer-aided classification of sleep stages,” 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp.3300–3303.

[4] Mora, Antonio Miguel and Fernandes, Carlos M and Herrera, Luis Javier and Castillo, Pedro A and Merelo, Juan Juli´an and Rojas, Fernando and Rosa, Agostinho C, “TSleeping with ants, SVMs, multilayer perceptrons and SOMs,”2010 10th International Conference on Intelligent Systems Design and Applications, 2010, pp.126–131.

[5] Vatankhah, Maryam and Akbarzadeh-T, Mohammad-R and Moghimi, Ali, “An intelligent system for diagnosing sleep stages using wavelet coefficients,” The 2010 International Joint Conference on Neural Networks (IJCNN), pp.1–5, 2010.

[6] Shuyuan, Xiao and Bei, Wang and Jian, Zhang and Qunfeng, Zhang and Junzhong, Zou and Nakamura, Masatoshi, “Notice of Removal: An improved K-means clustering algorithm for sleep stages classification,” 2015 54th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 2015, pp.1222–1227.

[7] Lajnef, Tarek and Chaibi, Sahbi and Ruby, Perrine and Aguera, Pierre- Emmanuel and Eichenlaub, Jean-Baptiste and Samet, Mounir and Kachouri, Abdennaceur and Jerbi, Karim, “Learning machines and sleeping brains: automatic sleep stage classification using decision-tree multiclass support vector machines,”Journal of neuroscience methods, 2015, pp.94–105.

[8] Berry, Richard B and Brooks, Rita and Gamaldo, Charlene E and Harding, Susan M and Marcus, C and Vaughn, Bradley V and others, “The AASM manual for the scoring of sleep and associated events,”Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine, 2012.

[9] Institute of Medicine (US) Committee on Sleep Medicine and Research; Colten HR, Altevogt BM, editors. Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. Washington (DC): National Academies Press (US); 2006. 2, Sleep Physiology. Available from: https://www.ncbi.nlm.nih.gov/books/NBK19956/.

[10] Krieger J. Respiratory physiology: Breathing in normal subjects. In: Kryger M, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine. 4th ed. Philadelphia: Elsevier Saunders; 2000. pp. 229–241.

[11] Simon PM, Landry SH, Leifer JC. Respiratory control during sleep. In: Lee-Chiong TK, Sateia MJ, Carskadon MA, editors. Sleep Medicine. Philadelphia: Hanley and Belfus; 2002. pp. 41–51.

[12] Madsen PL, Schmidt JF, Wildschiodtz G, Friberg L, Holm S, Vorstrup S, Lassen NA. Cerebral O2 metabolism and cerebral blood flow in humans during deep and rapid-eye-movement sleep. Journal of Applied Physiology. 1991b;70(6):2597–2601.

[13] Ellis CA, Zhang R, Carbajal DA, Miller RL, Calhoun VD, Wang MD. Explainable Sleep Stage Classification with Multimodal Electrophysiology Time-series. Annu Int Conf IEEE Eng Med Biol Soc. 2021 Nov;2021:2363-2366. doi: 10.1109/EMBC46164.2021.9630506. PMID: 34891757.

[14] Santaji, S., Santaji, S., Desai, V. Automatic sleep stage classification with reduced epoch of EEG. Evol. Intel. 15, 2239–2246 (2022)

[15] Santosh Kumar Satapathy, Akash Kumar Bhoi, D. Loganathan, Bidita Khandelwal, Paolo Barsocchi, Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal, Biomedical Signal Processing and Control, Volume 69, 2021, 102898, ISSN 1746-8094,  
https://doi.org/10.1016/j.bspc.2021.102898

[16] Satapathy, S.K., Loganathan, D. A Study of Human Sleep Stage Classification Based on Dual ChannelSatapathy, S.K., Loganathan, D. A Study of Human Sleep Stage Classification Based on Dual Channels of EEG Signal Using Machine Learning Techniques. SN COMPUT. SCI. 2, 157 (2021).

[17] Shen H, Ran F, Xu M, Guez A, Li A, Guo A. An Automatic Sleep Stage Classification Algorithm Using Improved Model Based Essence Features. Sensors (Basel). 2020 Aug 19;20(17):4677. doi: 10.3390/s20174677. PMID: 32825024; PMCID: PMC7506989.

[18] Khalighi Sirvan, Teresa Sousa, Jos´e Moutinho Santos, and Urbano Nunes. “ISRUC-Sleep: A comprehensive public dataset for sleep researchers.“ Computer methods and programs in biomedicine 124 (2016): 180-192.

[19] B Kemp, AH Zwinderman, B Tuk, HAC Kamphuisen, JJL Obery´e. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE-BME 47(9):1185-1194 (2000).

[20] Scott M. Lundberg, Su-In Lee, “A Unified Approach to Interpreting Model Predictions,” 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.