Sleep Stage Classification Using Signal Processing and Machine Learning

Bharath Kumar Swargam[a], Suyash Mishra[a]

[a]*Department of Electrical Engineering, Institute of Infrastructure, Technology, Research and Management (IITRAM), Ahmedabad, India*

Abstract

Sleep plays the most pivotal role in determining well-being and quality of life. Difficulties related to sleep affects human life in both mental and physical means. So, the diagnosis of sleep disorders should be comprehensive. A proper sleep stage classification is a must that helps in the efficient diagnosis of sleep-related disorders. The proposed research focuses on employing signal processing techniques and various machine learning algorithms to classify individuals' sleep stages. From National Sleep Research Resource (NSRR), the Wisconsin Sleep Cohort (WSC) dataset had utilized for this study. Two EEG, EOG, and EMG channels, namely C3_M2, O1_M2, E1, E2, chin, cchin_l with the sampling frequency of 100Hz and 200 Hz, are extracted from the polysomnogram data. These channels are employed for analysis since they can provide more information about sleeping patterns. In this study, the sleep stages are classified by using bi-orthogonal wavelet filter banks. The decomposition of the signals is carried out using five-level wavelet decomposition, giving six sub-bands. Later, data has fed into various supervised machine learning classifiers. A combination of channels is used and analyzed. The highest classification performance with an accuracy of 84.2% was achieved using EEG, EOG, and EMG channels combined, sampled at 100Hz, with an ensemble of bagged trees (EBT) classifier and 10-fold cross-validation. We also found that the classification accuracy was better when signals are combined. The model developed can be utilized in evaluating the quality of sleep and helps in diagnosing sleep disorders by using the sleep stage classification of an individual.

Keywords: EEG (electroencephalography), EOG(electrooculography), EMG(electromyography), PSG (polysomnographic), Machine Learning,

Classification, Sleep stages, Ensemble of Bagged Trees (EBT)

1. Introduction

Sleep is essential for maintaining well-being and optimal health. Getting a good amount of quality sleep is just as significant as regular exercise and eating a well-balanced diet. According to health professionals, Lower risk of (cancer, weight gain, and heart disease), better productivity and concentration, healthier calorie regulation, higher athletic performance, more emotional and social intelligence, preventing depression, and a healthy immune system are just a few of the many benefits associated with good sleep [1]. Sleep has links to many of the brain functions. So, the quality of sleep impacts our health and behaviour hugely. For good sleep, an individual should go through various sleep stages [2]. The human body functions several physio-chemical alterations during sleep which helps to differentiate these stages. These stages tend to repeat and follow a pattern for a good quality sleep. The breach of this pattern causes various sleep disorders leading to many higher risks of severe health issues. Hence, the diagnosis of these disorders is essential, which is possible using an accurate classification methodology. This classification involves tedious mathematical calculations, signal processing techniques. Hence, a computer-based approach is preferable. It involves the following steps: Acquisition of data, extraction of appropriate channels, pre-processing of the data, wavelet decomposition, evaluation of suitable features, and applying supervised machine learning classifiers.

1.1. Related Work done previously

Many other different methods have been done previously for classifying the sleep stages. Certain characteristics of the signals are retrieved to evaluate the recorded epochs, and a classification algorithm is utilized to determine the sleep stage in those techniques. Several EEG and EOG-based automatic sleep stage detection methods had published using various feature extraction techniques. Time-domain analysis, frequency-domain analysis, and time-frequency-domain analysis are examples of these. Also, they used non-linear parameters and complexity measures favorably. In some papers, feature selection and dimensional reduction were performed first and then the classification stage. The goal of this study is to lessen the quantity of features and produce low-dimensional features from the input features. For classification tasks, a variety of machine learning-based techniques have been suggested, including Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Decision Trees (DT).

A study titled "Automatic sleep stage classification with deep residual networks in a mixed-cohort setting" [3] used a different combination of cohorts. Using the Wisconsin Sleep Cohort dataset, they classified sleep stages. They obtained an accuracy of around 83-85 percent of sleep scoring data models.

Also, the Wisconsin dataset had used for other classification-related work. As, one of the papers titled "Obstructive Sleep Apnea during REM Sleep and Hypertension" [4] indicates that Obstructive Sleep Apnea (OSA) in the REM stage is linked to hypertension, which is clinically significant since OSA therapy is often restricted to the first half of the sleep cycle, leaving the majority of REM sleep untreated.

1.2. Sleep Stages

A conventional criterion, namely Rechtschaffen and Kales (R-K criterion), is introduced to classify sleep stages manually. According to R-K rules[5], sleep has mainly classified into two categories, REM (Rapid Eye Movement) sleep and NREM (Non-Rapid Eye Movement) [6]. NREM sleep accounts for 75-80% of overall sleep time, while REM sleep accounts for 20-25%. NREM is sub-divided into two groups, light and deep sleep. Light sleep consists of sleep stage 1(N1), and sleep stage 2(N2) whereas, deep sleep consists of N3. Wakefulness/stage W is not a part of REM and NREM sleep. It is not consistent, since it includes many rounds of the sleep cycle, each with its own set of stages. Each sleep stage is associated with various activities in the body.

The details of each sleep stage are as follows:

- **sleep stage 1 (N1)** - This stage involves the transition of wakefulness to sleep. It is simply the stage of "dozing off." Although the body and brain activity begin to decrease with periods of short movements during N1 sleep, the body has not completely rested. In this stage, there are minor changes in brain activity associated with falling asleep. If a person is undisturbed, within a few minutes, it moves to stage 2.

- **sleep stage 2 (N2)** - This stage is a period of light sleep. During N2 sleep, the body relaxes, breathing and heartbeat slow down, the temperature drops, movement of eye stops. Although brain wave activity declines, there are brief bursts of activity that help the body avoid being jolted awake by external stimuli. In the first sleep cycle, N2 lasts for around 10 to 25 minutes.The duration becomes longer as the cycle progresses. This stage covers around half of a person's sleep time.

- **sleep stage 3 (N3)** - This stage is a period of deep sleep that refreshes brain activity. During N3 sleep, brain activity, breathing, heartbeat slow down to their lowest levels. Muscles and eye movements relax even further down. This stage, according to experts, is important for restorative sleep, as it allows the body to repair and grow, as well as boost the immune system. An individual spends most time in this phase during the first sleep cycle. In the further sleep cycles, the duration gets shorter. It is harder to wake up someone during this phase.

- **REM** - Brain activity rises to levels similar to those seen while you're awake. This stage occurs around 90 minutes after falling asleep. Even though the eyelids are closed, the eyes dart swiftly in different directions. The body experiences atonia, temporary muscular paralysis, during REM sleep (The eyes and the muscles that control breathing being the exceptions). An individual gets dreams more often in this phase. It may also occur in NREM sleep stages, but less frequent and intense. REM sleep is crucial to cognitive functions. In the initial sleep cycle, this phase lasts just a few minutes, but it grows longer as the sleep cycle proceeds.

- **Wakefullness** - Maximum brain activity occurs during this period.

The technicians record the sleep data of the patient overnight. They obtain various signals like EEG (electroencephalogram), PSG (polysomnogram), ECG (electrocardiogram). These are examined and analyzed by the doctor. Later, they provide an appropriate diagnosis based on their observations. The sleep data from manual inspection of the patient contains many errors involved. The whole night recordings of sleep data are inaccurate, ineffective for proper sleep diagnosis. So, it becomes difficult for a doctor to interpret the sleep patterns of the patient accurately. If the sleep disorder is not diagnosed effectively, there may be severe abnormal consequences involved[7, 8]. So, researchers are trying to find the quick and best possible methods to identify sleep disorders. It is crucial to classify how much time an individual spends in each sleep stage. When known, we can say whether he is having a deep sleep or light sleep, or any sleep disorder, because an individual needs to get a good amount of deep sleep. Sleep stage classification helps in providing better information to a doctor. So, efficient, accurate, and reliable sleep stage classification is essential.

Currently, methods like convolutional neural networks, deep learning[3, 9], machine learning, and signal processing techniques are available for sleep stage classification. The medical technicians diagnose sleep using polysomnographic (PSG) recordings[10]. The polysomnogram includes various physiological signals like EEG (Electro- encephalography), EOG (Electrooculography), EMG (Electromyography), ECG (Electrocardiography), SpO2 (Pulse oximetry). While assessing sleep studies, EEG comes out to be more effective. It has proven to be the "gold standard" for the identification of sleep disorders. In our study, we included two EEG channels, i.e., C3_M2 O1_M2, for classification.

2. Data Collection

The dataset[11] used for this study was provided by the National Sleep Research Resource's (NSRR) Wisconsin Sleep Cohort (WSC). The NSRR provides de-identified physiological signals and clinical data elements from well-characterized research cohorts and clinical trials upon request. The WSC is one of the research projects in the Population Health Sciences department of the University of Wisconsin-Madison. The WSC collects data from overnight in-laboratory sleep studies from the Sleep Research Laboratory, a part of the Institute for Clinical and Translational Research (ICTR)'s

Clinical Translational Research Core (CTRC), University of Wisconsin. They are conducted with baseline samples of 1500 Wisconsin State employees and assessed at four-year intervals.

2.1. Data Structure

The WSC dataset consists of 2,570 records from 1,123 subjects aged 37-85, and time frame (2000-2015). The data consists of 11,387 files (with .eannot, .EDF, .log.txt, .scotxt, .stg.txt extensions) of 290 GB, 236 variables. Participants have up to five (5) records in the dataset. Each record is associated with a different research visit, distinguished by the wsc_id, wsc_vst (study visit number) variables. The record of the individual is collected using polysomnography (PSG), contains an EDF signal file. The study visit breakdown (no. of records per each visit) of 2,570 recordings is as follows:

- Visit 1 - 1,123 recordings
- Visit 2 - 758 recordings
- Visit 3 - 566 recordings
- Visit 4 - 121 recordings
- Visit 5 - 2 recordings .

The PSG signal is a combination of various channels. In the former half (2000-2009), the data was collected using the Grass Heritage system. It has 16 channels connected to the person to record various behaviours using EEG, EOG, EMG, and other sensors. In the latter half (2009-present), the collected data was from the Grass Comet Lab-Based system. It has 19 channels connected to the person similarly as before. The sampling rate for all these is 100 Hz for 2000-2009 and 200 Hz for 2009 on wards. The variables include Administrative, Anthropometry, Clinical Data, Demographics, General Health, Lifestyle and Behavioural Health, Medical History, Sleep Monitoring, Sleep Questionnaires, Sleep Treatment. It helps us to analyse using the required section from variables.

The WSC dataset also consists of sleep staging and respiratory event scoring annotations. These are available in the following formats:

- .eannot - One row per epoch with an indication of scored sleep stage (e.g. wake, N1, N2, N3, REM).
- .sco.txt - contains respiratory event scoring for Gamma studies.
- .stg.txt - contains staging format for Gamma studies.
- .log.txt - provides bio calibrations and light indicators for Gamma studies.
- .allscore.txt - gives staging and respiratory event scoring for Twin studies.

2.2. Sampling rate information of the channels in the dataset

Channel name	EDF Label	Sampling Rate in (Hz)	Hardware Filters in (Hz)
Left EOG	E1	100	Low Pass 30
Right EOG	E2	100	Low Pass 30
Left Central EEG	C3_M2	100	Low Pass 30
Left Occipital EEG	O1_M2	100	Low Pass 30
Chin EMG	chin	100	Low Pass 30
Linked Left and Right Leg EMG	lleg_r	100	Low Pass 30
Snore	snore	100	Low Pass 30
ECG	ECG	100	Low Pass 30
Nasal Airflow	nasalflow	100	Low Pass 30
Oral Airflow	oralflow	100	Low Pass 30
Nasal Pressure	nas_pres	100	Low Pass 30
Thoracic IP	thorax	100	-
Abdominal IP	abdomen	100	-
IP Sum	sum	100	-
Position	position	100	-
SpO2	spo2	100	-

Table 1: Details of Channels in the PSG data of Grass Heritage System (2000-2009)

The PSG signal (2000-2009) consist of 16 channels.Each channel's signal were recorded at 100Hz and passed through the low pass filter of 30Hz. Table 1 lists the channels and their corresponding EDF label. EEG, EOG, EMG channels namely, C3_-M2, O1_M2, E1, E2, chin are extracted from this Table 1. While choosing EMG channel, we choose chin because, it is performing better in classifying sleep stages compared to very bad performance of other EMG channel (lleg_r). So, the inclusion of lleg_r degrades the classification performance.

Channel	EDF Label	Sampling Rate in (Hz)	Hardware Filters in (Hz)
Left EOG	E1	200	Low Pass 35
Right EOG	E2	200	Low Pass 35
Left Frontal EEG	F3_M2	200	Low Pass 35
Frontal EEG	Fz_M2	200	Low Pass 35
Central EEG	Cz_M2	200	Low Pass 35
Left Central EEG	C3_M2	200	Low Pass 35
Parietal EEG	Pz_M2	200	Low Pass 35
Left Occipital EEG	O1_M2	200	Low Pass 35
Linked Center and Left Chin EMG	cchin_l	200	Low Pass 70
Linked Left and Right Leg EMG	lleg_r	200	Low Pass 70
ECG	ECG	200	Low Pass 35
Snore	snore	200	Low Pass 70
Airflow	flow	200	Low Pass 15
Nasal Pressure	nas_pres	200	Low Pass 15
Thoracic IP	thorax	200	Low Pass 15
Abdominal IP	abdomen	200	Low Pass 15
IP Sum	sum	200	Low Pass 15
Position	position	200	-
SpO2	spo2	200	-

Table 2: Details of Channels in the PSG data of Grass Comet Lab Based system (2009-2015)

The PSG signal (2009-2015) consist of 19 channels. Each channel's signal were recorded at 200Hz and passed through the low pass filter of 70Hz, 35Hz. Table 2 lists the channels and their corresponding EDF label. EEG, EOG, EMG channels namely, C3_M2, O1_M2, E1, E2, chin are extracted from this Table 2. While choosing EMG channel, similar to channels sampled at 100Hz,we choose cchin_l because, it's performance in classification is better compared to other EMG channel(lleg_r).

After extraction of required channels, the number of epochs in each sleep stage are given below. The signals from Grass Heritage System (table1) have a total of 15,43,461 epochs. Whereas,the signals from Grass Comet Lab based system (table2) have a total of 6,46,535 epochs.

Sleep Stage	No.of Epochs
N1	1,29,132
N2	8,34,926
N2	81,430
REM	2,12,166
wake	2,85,807
Total	15,43,461

Table 3: Number of Epochs for Each sleep stage of channels sampled at 100Hz

Sleep Stage	No.of Epochs
N1	46,425
N2	3,39,747
N2	41,431
REM	80,068
wake	1,38,864
Total	6,46,535

Table 4: Number of Epochs for Each sleep stage of channels sampled at 200Hz

3. Methodology

We proposed a computer-assisted sleep stage classification system in this paper. It includes data acquisition, extraction of required channels from data, pre-processing of data, wavelet decomposition, feature extraction, and classification of sleep stages using different supervised machine learning classifiers. The classifiers used in this study are Ensemble Bagged Trees and Quadratic SVM.

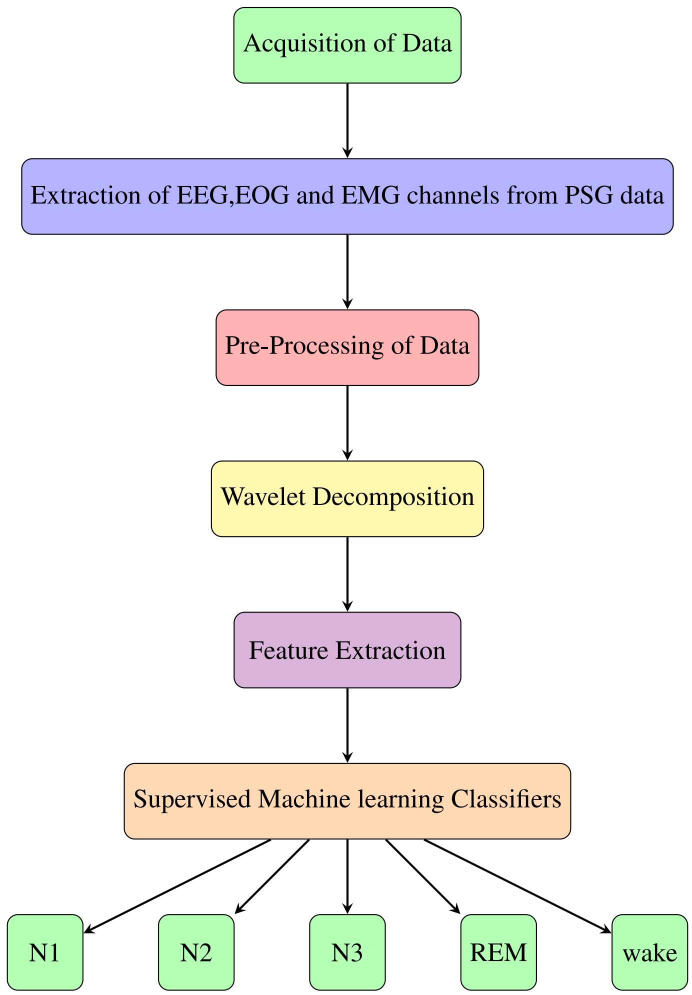
3.1. Overview

The outline of our study for the classification of sleep stage is reviewed in the flowchart below. We acquired our data (Wisconsin Sleep Cohort(WSC)) for this study from the National Sleep Research Resource(NSRR). It consists of polysomnographic (PSG) data comprising various physiological signals like EOG, EEG, EMG, and ECG. A MATLAB code was written and used to read the annotations of each subject's recording. We extracted EEG, EOG, EMG signals of specific channels from the data read. These channels, pre-processed to produce 30 seconds Epochs according to their sampling frequency(100Hz and 200Hz). Then, we applied filter bank followed by wavelet decomposition to the pre-processed data. We obtained feature extraction from this data and fed this into supervised machine learning classifiers. This process is carried out on the entire dataset.

3.1.1. Extraction of EEG, EOG, and EMG channels and Pre-Processing

We used a Matlab code 'edfread' to read the '.edf' file from the dataset. Then we extracted channels C3_M2, O1_M2, E1, E2, chin of sampling frequency 100Hz and C3_M2, O1_M2, E1, E2, cchin_l of sampling frequency 200Hz. We have chosen to use them because they are effective bio-medical signals and have high practical value in clinical diagnosis. The dataset provided consists of annotations for each sleep stage. These are scored according to R &K rules by the experts of the WSC team. We used a Matlab code 'importfile_eannot' to read the annotations. We arranged the read data as a matrix. This matrix contains each row that corresponds to an epoch of 30 seconds. We used Optimal Bi-orthogonal Wavelet Filter Bank to remove the noise obtained during the recording of the sleep data. This filter bank is efficient as it tends to preserve signal characteristics while de-noising.We used this filter bank in this study to perform wavelet decomposition.We utilized a 5-level wavelet decomposition on every epoch and generated 6 sub-bands/wavelets from 1 epoch. These sub-bands were applied in 200 the next step.

3.2. Flow chart



3.3. Feature Extraction

Feature extraction of EEG, EOG, and EMG data is a challenging problem. These signals have complexity,non-linearity and random in nature. They are considered stationary only in the small interval, i.e., "quasi-stationary.". Over prolonged periods, the signal characteristic nature is non-stationary. Entropy is a measure of the degree of uncertainty in a system. Entropy can be used to measure the level of chaos in a brain-computer interface system[12]. It's a non-linear metric for determining how complex a time series is. Entropy measures how effectively one can predict the response of the corresponding part of the trajectory from the other. The higher the entropy, the more chaotic the system is in nature, and hence the less predictable it is. The change in sleep stages causes a variation in the EOG, EMG, and EEG signals. As seen in many previous works, these changes are supposed to reflect in the entropy results[13]. The specific features computed in this study using the standard algorithms were: Shannon entropy and tsallis entropy.

In our work, we had extracted features from six sub-bands using the Tsallis entropy as the feature, yielded a total of 6 features for each channel. The best-performing model made was made using five channels (2 EOG channels,1 EMG channel, 2 EEG channels) with 30 features.

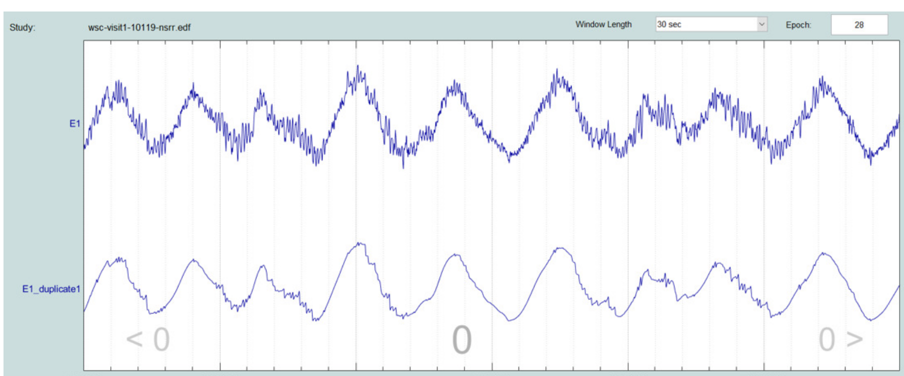


Figure 1: Signal representation of EOG channel i.e.,E1 before(above) and after(below) applying Wavelet De-noise filter(30-sec Epoch)

3.4. Machine Learning and classification

We applied machine learning methods to train a model in our study. A matrix for each combination of channels has formed. The rows of the matrix contain the epochs of sleep stages. The column represents features extracted using Tsallis entropy. The last column contains the label of the sleep stage attached to each Epoch. This matrix obtained is fed into various supervised machine learning classifiers for training the model. To prepare the machine learning model, we used the Classification Learner app in Matlab. To validate the model which is getting trained, we used the 10-fold cross-validation technique. It helps in preventing the over-fitting of the data. The model shows how well it can classify sleep stages (N1, N2, N3, REM, and wake). It also contains a confusion matrix that shows detailed classification, i.e., how many epochs are present in each sleep stage. For each combination of channels, a similar method followed in training the model. After observing the results of models using different classifiers, ' Ensemble Bagged Trees' gave the best sleep stage classification. The highest classification accuracy obtained is 84.2%. SVM Quadratic also gave classification accuracy around 84%, but this model consumed a significant amount of time for training.

3.4.1. Bagged Tree Algorithm

In ensemble methods, several decision trees combine to produce better predictive performance. It randomly creates several subsets of samples from training data. Each subset is used in training the decision trees. It leads to an ensemble of different models with prediction values. In comparison to a single decision tree, the average of all these predictions from numerous trees provides a more robust prediction. This algorithm helps in reducing the variance of a decision tree and improves the accuracy of the model.

3.4.2. Support Vector Machines(SVM)

The SVM (Support Vector Machine) is a supervised machine learning technique for classification and regression. In this algorithm, we use n-dimensional space (n is number of features) and plot each data at a point. The coordinates in this plane correspond to the value of each feature used. Then, a perfect hyperplane as a decision surface will form to achieve classification between the classes. It involves complex data transformations to sort the data according to the labels.

4. Results

The WSC dataset consists of signals sampled at 100Hz and 200Hz. We performed this entire procedure to both of them individually. First, we have taken the sampling rate as 100Hz and carried out the whole method on signals using the table1. Later, the sampling rate of 200Hz was applied and carried out the classification. The results are divided based on the combination of the signals and sampling frequency.

In our study, we utilized Accuracy, F1 Score, Cohen's kappa (k) and area under curve (AUC) of a Receiver Operating Characteristic (ROC) curve in evaluating the performance.

4.1. EEG Channels

4.1.1. Channels sampled at 100 Hz

True Class	**Predicted Class**					**F1 Score**
	N1	N2	N3	REM	wake	
N1	22%	35%	<1%	26%	17%	0.2893
N2	2%	92%	2%	3%	2%	0.8776
N3	<1%	44%	56%	<1%	<1%	0.6376
REM	8%	22%	<1%	64%	6%	0.6517
wake	4%	6%	<1%	3%	88%	0.8566
Cohen's kappa value ± kappa error = 0.6688 ± 0.0005						
Accuracy = 79.4%						

Table 5: Confusion Matrix of both EEG Channels

We began by classifying sleep stages using the EEG, EOG, and EMG channels separately. The model's performance was then improved by combining various channels. When they are combined, the number of features increases which helps in training the machine learning model.When both EEG channels (C3_M2 and O1_M2) of sampling rate 100Hz combined, the accuracy improved to 79.4%. Table 5 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In classification problems, accuracy is used as a metric to compare the performance. It is best only when each class has the same number of observations. From Tables 3 and 4, we can see each stage have an unequal number of Epochs.

Cohen's kappa coefficient (k) has become a popular metric for comparing performance in recent years. It is considered more robust than accuracy. The range of k determines whether the classification is good or not. If the value of k is between 0.75 and 1 then, it is considered an excellent classification. If the value of k is between 0.4 and 0.7 then, it is understood as fair to good classification. When the value of k is less than 0.4, the classification agreement is said to be poor. In this case, the value of k is 0.6688 ± 0005, which is fair to good classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.83	(0.03,0.22)
N2	0.94	(0.21,0.92)
N3	0.90	(0.01,0.56)
REM	0.92	(0.05,0.64)
wake	0.98	(0.04,0.88)

Table 6: ROC details when both EEG channels i.e.,C3_M2 and O1_M2 are used.Each sleep stage,here, is taken with respect to others

AUC(Area Under The Curve), ROC (Receiver Operating Characteristics) Curve is a graphical plot that helps in evaluating the performance. It is a probability curve in which AUC describes separability between classes. The greater the value of AUC, the more accurate the model. As seen from Table 6, the AUC is nearer to 1, which means the classification is appropriate.

4.1.2. *Channels sampled at 200 Hz*

True Class	**Predicted Class**					**F1 Score**
	N1	N2	N3	REM	wake	
N1	15%	58%	<1%	7%	20%	0.2348
N2	1%	92%	1%	2%	5%	0.8293
N3	<1%	44%	54%	1%	2%	0.6543
REM	1%	30%	<1%	65%	3%	0.7295
wake	2%	22%	<1%	1%	76%	0.7716
Cohen's kappa value ± kappa error = 0.6197 ± 0.0009						
Accuracy = 77.0%						

Table 7: Confusion Matrix of both EEG Channels

When both EEG channels (C3_M2 and O1_M2) of sampling rate 200Hz combined, the accuracy improved to 77.0%. Table 7 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In this case, the value of k is 0.6197 ± 0009, which is fair to good classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.79	(0.01,0.15)
N2	0.90	(0.32,0.92)
N3	0.89	(0.01,0.54)
REM	0.94	(0.02,0.65)
wake	0.95	(0.06,0.76)

Table 8: ROC details when both EEG channels i.e.,C3_M2 and O1_M2 are used.Each sleep stage,here, is taken with respect to others

As seen from Table 8, the AUC is approximately close to 1 (except for N1 Stage), which means the classification is appropriate.

4.2. EOG channels

4.2.1. Channels sampled at 100 Hz

True Class	**Predicted Class**					**F1 Score**
	N1	N2	N3	REM	wake	
N1	18%	45%	<1%	15%	22%	0.2439
N2	2%	91%	2%	3%	3%	0.8461
N3	<1%	59%	36%	3%	1%	0.4659
REM	7%	29%	1%	57%	6%	0.6228
wake	4%	10%	<1%	3%	83%	0.8061
Cohen's kappa value $\pm$ kappa error = 0.6012 $\pm$ 0.0006						
Accuracy = 75.7%						

Table 9: Confusion Matrix of both EOG Channels

When both EOG channels (E1 and E2) of sampling rate 100Hz combined, the accuracy improved to 75.7%. Table 9 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In this case, the value of k is 0.6012 $\pm$ 0006, which is fair to good classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.79	(0.03,0.18)
N2	0.90	(0.28,0.91)
N3	0.84	(0.01,0.36)
REM	0.90	(0.04,0.57)
wake	0.96	(0.05,0.83)

Table 10: ROC details when both EOG channels i.e.,E1,E2 are used.Each sleep stage,here, is taken with respect to others

As seen from Table 10, the AUC is approximately close to 1 (except for N1 and N2 Stages), which means the classification is appropriate.

4.2.2. Channels sampled at 200 Hz

True Class	**Predicted Class**					**F1 Score**
	N1	N2	N3	REM	wake	
N1	16%	48%	<1%	10%	26%	0.2318
N2	1%	91%	2%	2%	4%	0.8526
N3	<1%	37%	60%	2%	1%	0.6736
REM	3%	32%	2%	59%	4%	0.6677
wake	2%	10%	<1%	1%	86%	0.8322
Cohen's kappa value $\pm$ kappa error = 0.6568 $\pm$ 0.0008						
Accuracy = 78.8%						

Table 11: Confusion Matrix showing True Positive Rates and False Negative Rates of both EOG Channels

When both EOG channels (E1 and E2) of sampling rate 200Hz combined, the accuracy improved to 78.8%. Table 11 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In this case, the value of k is 0.6568 $\pm$ 0008, which is fair to good classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.80	(0.01,0.16)
N2	0.91	(0.25,0.91)
N3	0.91	(0.01,0.60)
REM	0.90	(0.03,0.59)
wake	0.96	(0.06,0.86)

Table 12: ROC details when both EOG channels i.e.,E1,E2 are used.Each sleep stage,here, is taken with respect to others

As seen from Table 12, the AUC is approximately close to 1 (except N1), which means the classification is appropriate.

4.3. EEG and EOG channels combined

4.3.1. Channels sampled at 100 Hz

True Class	Predicted Class					F1 Score
	N1	N2	N3	REM	wake	
N1	29%	37%	<1%	18%	16%	0.3565
N2	2%	94%	2%	2%	1%	0.8981
N3	<1%	42%	57%	<1%	<1%	0.6628
REM	7%	13%	<1%	74%	5%	0.7628
wake	4%	5%	<1%	2%	89%	0.8685
Cohen's kappa value ± kappa error = 0.7223 ± 0.0005						
Accuracy = 82.7%						

Table 13: Confusion Matrix of both EEG and EOG Channels

When both EEG, and EOG channels (C3_M2, O1_M2, E1, and E2) of sampling rate 100Hz combined, the accuracy improved to 82.7%. Table 13 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In this case, the value of k is 0.7223 ± 0005, which is fair to good classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.86	(0.03,0.29)
N2	0.95	(0.18,0.94)
N3	0.92	(0.01,0.57)
REM	0.96	(0.03,0.74)
wake	0.98	(0.04,0.89)

Table 14: ROC details when both EEG,EOG channels i.e.,C3_M2, O1_M2, E1, E2 are used.Each sleep stage,here, is taken with respect to others

As seen from Table 14, the AUC is approximately nearer to 1 , indicating the appropriate classification.

4.3.2. Channels sampled at 200 Hz

True Class	Predicted Class					F1 Score
	N1	N2	N3	REM	wake	
N1	23%	44%	<1%	9%	24%	0.3254
N2	1%	93%	1%	1%	3%	0.8808
N3	<1%	32%	67%	1%	1%	0.7424
REM	3%	22%	1%	71%	4%	0.7750
wake	2%	7%	<1%	1%	89%	0.3812
Cohen's kappa value ± kappa error = 0.6015 ± 0.0011						
Accuracy = 82.8%						

Table 15: Confusion Matrix of both EOG and EEG Channels

When both EEG, and EOG channels (C3_M2, O1_M2, E1, and E2) of sampling rate 200Hz combined, the accuracy improved to 82.8%. Table 15 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In this case, the value of k is 6015 ± 0011, which is fair to good classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.85	(0.01,0.23)
N2	0.94	(0.20,0.93)
N3	0.94	(0.01,0.67)
REM	0.95	(0.02,0.71)
wake	0.98	(0.05,0.89)

Table 16: ROC details when both EEG,EOG channels i.e.,C3_M2, O1_M2, E1, E2 are used.Each sleep stage,here, is taken with respect to others

As seen from Table 16, the AUC is approximately nearer to 1 , indicating the appropriate classification.

4.4. EEG, EOG, EMG channels combined

4.4.1. Channels sampled at 100 Hz

True Class	Predicted Class					F1 Score
	N1	N2	N3	REM	wake	
N1	32%	37%	<1%	15%	16%	0.3988
N2	2%	94%	1%	1%	1%	0.9060
N3	<1%	39%	60%	<1%	<1%	0.6961
REM	6%	12%	<1%	77%	5%	0.7993
wake	4%	5%	<1%	2%	90%	0.8761

Cohen's kappa value ± kappa error = 0.7463 ± 0.0005

Accuracy = 84.2%

Table 17: Confusion Matrix of EEG,EOG,EMG Channels Combined

When both EEG, EOG, and EMG channels (C3_M2, O1_M2, E1, E2, and chin) of sampling rate 100Hz combined, the accuracy improved to 84.2%. Table 17 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In this case, the value of k is 0.7463 ± 0005, which is close to 0.75 indicating excellent classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.88	(0.03,0.32)
N2	0.96	(0.17,0.94)
N3	0.93	(0.01,0.60)
REM	0.97	(0.03,0.77)
wake	0.98	(0.03,0.90)

Table 18: ROC details when EEG,EOG,EMG channels combined i.e.,C3_M2, O1_M2, E1, E2, and chin are used.Each sleep stage,here, is taken with respect to others

As seen from Table 18, the AUC is approximately nearer to 1 , indicating excellent classification of sleep stages.

4.4.2. Channels sampled at 200 Hz

True Class	**Predicted Class**					**F1 Score**
	N1	N2	N3	REM	wake	
N1	24%	44%	<1%	8%	23%	0.3462
N2	1%	94%	1%	1%	3%	0.8876
N3	<1%	31%	68%	1%	1%	0.7539
REM	2%	21%	1%	73%	3%	0.7904
wake	2%	7%	<1%	1%	91%	0.8718
Cohen's kappa value $\pm$ kappa error = 0.7411 $\pm$ 0.0007						
Accuracy = 83.8%						

Table 19: Confusion Matrix of EEG,EOG,EMG Channels Combined

When both EEG, EOG, and EMG channels (C3_M2, O1_M2, E1, E2, and cchin_l) of sampling rate 200Hz combined, the accuracy improved to 83.8%. Table 19 shows the confusion matrix obtained after classifying sleep stages using the Ensemble Bagged Trees classifier. In this case, the value of k is 0.7411 $\pm$ 0007, which is close to 0.75 indicating excellent classification.

Sleep Stage	Area Under Curve (AUC)	Point
N1	0.87	(0.01,0.24)
N2	0.95	(0.19,0.94)
N3	0.94	(0.01,0.68)
REM	0.96	(0.02,0.73)
wake	0.98	(0.05,0.91)

Table 20: ROC details when EEG,EOG,EMG channels combined i.e.,C3_M2, O1_M2, E1, E2, and cchin_l are used.Each sleep stage,here, is taken with respect to others

As seen from Table 20, the AUC is approximately nearer to 1 , indicating the appropriate classification.

5. Discussion

In this work, we are trying to classify sleep stages using PSG data. We mainly used EEG, EOG, and EMG signals for our study. First, we extracted the required signals from the PSG data and then made sleep stage classification. We divided the sleep stages into five categories, i.e., N1, N2, N3, REM, and wake. We applied Tsallis Entropy as feature extraction for classification. In contrast, other studies used convolutional neural networks(CNN), deep learning approaches[14, 15, 9]. At first, the study was performed using individual channels. But the results are not satisfactory. To improve the performance, we combined the signals (to increase features). It increased the performance by a significant amount. The highest classification accuracy (best model) obtained was using the combination of three signals (EMG, EOG, and EEG). When compared, the best accuracy obtained is equivalent to that of [3].

We particularly focused on 5 channels i.e one EMG channel, two EOG channels, and two EEG channels. In this paper, we found out that Tsallis entropy outperforms the existing feature selection of the EEG and EOG signal. It is also noteworthy that if big data set is being used for the classification, the accuracy increases. We have used around 1123 subject's PSG data record-sized over more than 290GB of data. Each subject's signal was segmented into epochs of 30 sec over the recording period. Then, carefully dividing the signal into the total no of Epochs and making the matrix out of it. We then applied a wavelet decomposition filter and sent the resultant matrix for features extraction like Tsallis entropy. For training of the data, to avoid over-fitting of the machine learning model, we found out that 10-folds cross-validation performs best. Also, we tried to use the hold-out validation of 25%. It also gives results near to 10 folds cross-validation. But it was way faster than cross-validation.

Table 21 shows the classification accuracy's obtained for various channels and their combinations. From the below table, we could see that as the number of channels increases in the training dataset, the accuracy of the result also increases. Also, both Ensemble Bagged Trees and SVM Quadratic have nearer classification accuracy. But the training of the model using SVM Quadratic was time-consuming. And the Ensemble Bagged Trees performed a bit better by utilizing less time.

Channels	Feature Used	Classifier	Performance	
			100Hz	200Hz
E1	Tsallis Entropy	Bagged Trees	71.4	74.6
		SVM quadratic	70.8	72.3
E2	Tsallis Entropy	Bagged Trees	70.8	74.0
		SVM quadratic	70.1	73.1
C3_M2	Tsallis Entropy	Bagged Trees	75.7	66.5
		SVM quadratic	75.4	64.0
O1_M2	Tsallis Entropy	Bagged Trees	73.5	72.0
		SVM quadratic	73.2	71.8
EMG	Tsallis Entropy	Bagged Trees	64.7	61.5
		SVM quadratic	63.8	61.2
E1 and E2	Tsallis Entropy	Bagged Trees	75.7	78.8
		SVM quadratic	74.3	77.9
C3_M2 and O1_M2	Tsallis Entropy	Bagged Trees	79.3	77.0
		SVM quadratic	78.6	76.5
EEG and EOG combined	Tsallis Entropy	Bagged Trees	82.7	82.8
		SVM quadratic	81.8	82.0
EEG,EOG and EMG combined	Tsallis Entropy	Bagged Trees	84.2	83.8
		SVM quadratic	83.5	81.9

Table 21: Classification accuracy's obtained for various channels

The benefits concerning our process are as follows:

- The first advantage of our study is that we have used EEG, EOG, and EMG signals from the PSG data. PSG data has multiple channels involved in it, making the process of recording sleep data more complex. It involves more work and effort to get sleep data. The patient may feel stressed during this, which results in inaccurate sleep recordings. So, fewer channels involve less complexity and make the patient comfortable.

- The next advantage is that our method involves less complexity than convolutional neural networks(CNN) and deep learning approaches. We have used only Tsallis Entropy as our feature for the classification.The computational time is less due to the simplicity involved. To avoid over-fitting the model, we employed 10-fold cross-validation during training.

The limitations of our study are:

- We have used EEG signals in our method, the collection of which requires the placement of electrodes on the human skull, making the patient difficult.
- The less accuracy of N1, N3 sleep stages are due to the fewer number of Epochs associated with them.

Today, convolution neural networks (CNN) are getting much more attention than machine learning for classification due to their better performance. But complexity is more in CNN methods. In the future, due to increased technology, there might be more computational power and more algorithms with less complexity for sleep stage classification.

6. Conclusion

Sleep is an essential key, which has many benefits associated with it. So, diagnosis of sleep disorders is very crucial. Approach by Medical practitioners to detect sleep disorders from sleep data manually takes time. The error present in the data makes it intricate for the doctor to identify sleep disorders. So, in this study, we presented a simple and effective machine learning sleep stage classification system using EMG, EEG, and EOG signals. We used the WSC dataset from NSRR for this study. We pre-processed the data, applied wavelet decomposition, and then feature extraction. Later, we classified this using supervised machine learning techniques to five sleep stages (N1, N2, N3, REM, and wake). We processed the whole data set consisting of 1123 subjects with 2570 recordings. Our proposed method obtained a maximum classification accuracy of 84.2%, cohen's kappa coefficient of 0.7463±0.0005 using five channels (C3_M2, O1_M2, E1, E2, and chin) of 100 Hz sampling frequency with

Ensemble Bagged Trees classifier and 10-fold cross-validation. Such computer-aided techniques can impact the medical industry in reducing the tedious work by medical practitioners. Our method is simple, robust, and produced good results. In the future, new researchers may develop simple, accurate, and better results in the classification of sleep stages.

References

[1] J. Fletcher, Why sleep is essential for health (May 2019). URL https://www.medicalnewstoday.com/articles/325353

[2] E. Suni, Stages of sleep (2020 Aug). URL https://bit.ly/3whexL1

[3] Automatic sleep stage classification with deep residual networks in a mixed-cohort setting, Sleep 44 (2020 Aug). doi:10.1093/sleep/zsaa161.

[4] Obstructive sleep apnea during rem sleep and hypertension. results of the wisconsin sleep cohort, American Journal of Respiratory and Critical Care Medicine 190 (10) (2014 Nov). doi:10.1164/rccm.201406-1136OC. URL https://pubmed.ncbi.nlm.nih.gov/25295854/

[5] A. R. Anthony Kales, A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects: Allan Rechtschaffen and Anthony Kales, Editors, NIH publication, U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968.

[6] Understanding sleep (2019 Aug). URL https://bit.ly/3dJT3zY

[7] E. Mignot, Why we sleep: the temporal organization of recovery, PLoS Biol 6 (4) (2008). doi:10.1371/journal.pbio.0060106.

[8] M. J. Fowler, M. J. Sullivan, B. R. Ekstrand, Sleep and memory, Science 179 (4070) (1973) 302–304. doi:10.2307/j.ctv6hp38g.2.

[9] O. Yildirim, U. B. Baloglu, U. R. Acharya, A deep learning model for automated sleep stages classification using psg signals, International Journal of Environmental Research and Public Health 16 (4) (2019). doi:10.3390/ijerph16040599. URL https://www.mdpi.com/1660-4601/16/4/599

[10] C. O'Reilly, T. Nielsen, Assessing eeg sleep spindle propagation. part 1: theory and proposed methodology, Journal of neuroscience methods 221 (2014) 202–214. doi:10.1016/j.jneumeth.2013.08.013.

[11] Wisconsin sleep cohort (2020).
URL https://sleepdata.org/datasets/wsc

[12] D. Zhang, X. Jia, H. Ding, D. Ye, N. V. Thakor, Application of tsallis entropy to eeg: Quantifying the presence of burst suppression after asphyxial cardiac arrest in rats, IEEE Transactions on Biomedical Engineering 57 (4) (2010) 867–874. doi:10.1109/TBME.2009.2029082.

[13] K. A. I. Aboalayon, M. Faezipour, W. S. Almuhammadi, S. Moslehpour, Sleep stage classification using eeg signal analysis: A comprehensive survey and new investigation, Entropy 18 (9) (2016). doi:10.3390/e18090272.
URL https://www.mdpi.com/1099-4300/18/9/272

[14] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, Y. Guo, Mixed neural network approach for temporal sleep stage classification, IEEE Transactions on Neural Systems and Rehabilitation Engineering 26 (2) (2018) 324–333. doi:10.1109/TNSRE.2017.2733220.

[15] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, M. De Vos, Joint classification and prediction cnn framework for automatic sleep stage classification, IEEE Transactions on Biomedical Engineering 66 (5) (2019) 1285–1296. doi:10.1109/TBME.2018.2872652.