International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Revolutionizing Sleep Diagnostics: A Novel Deep Learning Approach for Real-Time Sleep Stage Analysis

Youcef Yahi[1,2,3], Yassine Benallou[1], Fatima Zohra Driss Khoudja[2], Manisha Jassal[3] , Giovanna Cincotti[4]

Submitted: 06/02/2024 **Revised**: 14/03/2024 **Accepted**: 20/03/2024

Abstract: Sleep stage scoring, traditionally done manually by specialists through the inspection of neurophysiological data from sleep studies, is a labor-intensive, monotonous, and time-consuming activity. This has led to an increased interest in the development of Automated Sleep Stage Classification (ASSC) technologies. Such systems are vital for assisting medical professionals in diagnosing and managing sleep-related disorders and neurological conditions, including Alzheimer's disease. This paper presents a cutting-edge classification technique that combines deep learning strategies, delivering outstanding outcomes. It also reviews progress and hurdles in current methods of sleep stage determination using Electroencephalogram (EEG) signals, covering steps like preprocessing, feature detection, and categorization. The paper's goal is to unveil a new classifier design that promises real-time, high-accuracy solutions recognized by the scientific community. This includes the classification of EEG signals into different patient sleep stages: Wake, N1, N2, N3, and REM. Employing a robust classifier system within the Electroencephalography Analysis System (EAS) based on a Brain-Computer Interface (BCI), this system utilizes hybrid classifiers beginning with feature extraction methods such as WDT and PCA, followed by a combination of BiLSTM and LightGBM classifiers. The process starts with training the BiLSTM model on raw EEG data to learn temporal patterns and feature extraction. Features from the BiLSTM outputs are then used as inputs for LightGBM, creating a potent classification system. Unlike previous approaches that often required multiple EEG channels and longer epochs, this research introduces an effective method for 10-second epochs from a single-channel EEG, incorporating novel statistical features and utilizing the PhysioNet Sleep Database, EDFx sleep DATA. The proposed method has shown an average classification sensitivity of 92.1%, specificity of 98.8%, and an overall accuracy of 97.42% using a decision tree classifier, outperforming previous studies in classification accuracy.

Keywords*: Automated Sleep Stage Classification (ASSC), Deep Learning, Electroencephalogram (EEG) Signals, Sleep Disorders and Brain Diseases, Feature Extraction ,WDT, PCA, Hybrid Classifiers, BiLSTM, LightGBM, Real-time Analysis*

I. Introduction

1.1. Background and Significance

Sleep, a fundamental brain function, significantly influences an individual's cognitive performance, learning, and physical abilities [1–9]. This reversible state renders an individual partially or completely unconscious, leading to reduced brain complexity [10–13]. With humans spending approximately one-third of their lives asleep, prevalent conditions like insomnia and Obstructive Sleep Apnea (OSA) can significantly impact physical health [14–16]. Globally, sleep disorders affect 12% of individuals in Algeria [17] and 50–70 million people in the United States [18,19], highlighting its international prevalence. Additionally, more than 90% of depressive disorder patients reportedly experience sleep-related issues [6,20].

[1]*Technology Laboratory of Communication, Dr. Tahar Moulay University of Saida, 20000 Saida, Algeria*
[2]*Laboratory of Physico-Chemical Studies, Dr. Tahar Moulay University of Saida, 20000 Saida, Algeria*
[3]*Sapienza University of Rome, 00185 Roma, Italy*
[4]*Department of Biosciences, University of Milan, Via Giovanni Celoria, 20133 Milan MI, Italy*
***email** : youcef.yahi@univ-saida.dz*

Estimates suggest that sleep apnea affects 2%–4% of adults and 1%–3% of children, while around 33% of the world's population exhibits symptoms of insomnia [15,21,22]. Sleep-related problems can lead to sleepiness, depression, and even fatalities [6,15]. Alarmingly, incidents of falling asleep while driving account for at least 100,000 automobile crashes annually in the United States [23–25]. Sleep-related factors contribute to a quarter of traffic accidents in Germany and 20% in England, with Australia spending over $1500 million on drowsiness-induced fatalities [26]. Police reports suggest that up to 3% of road traffic accidents and 4% of fatalities are due to sleep-related causes [27,28].

Given these alarming statistics, developing devices capable of automatically detecting and analyzing sleep patterns to identify conditions like fatigue, drowsiness, apnea, insomnia, or narcolepsy is imperative.

1.2. Importance of Sleep Stage Scoring

Sleep stage scoring, considered the gold standard in analyzing human sleep [17,29–36], aims to diagnose and treat sleep disorders effectively. This process relies on Polysomnographic (PSG) recordings obtained from patients during overnight sleep at hospitals [5,9,16,17]. Traditionally, experts visually score overnight PSG recordings, encompassing Electroencephalogram (EEG),

Electrooculogram (EOG), Electromyogram (EMG), and Electrocardiogram (ECG) data, using guidelines established by Rechtschaffen and Kales (R&K) in 1968 [37,38].

The PSG recordings are divided into 20- or 30-second epochs, classified as Wakefulness (W), Rapid Eye Movement (REM) sleep, and Non-Rapid Eye Movement (NREM) sleep. NREM sleep includes stages 1, 2, 3, and 4 (S1, S2, S3, S4), following the R&K guidelines or more recent American Academy of Sleep Medicine (AASM) guidelines from 2007 [40–44]. Notably, the AASM standards amalgamate NREM stages S3 and S4 into a single stage termed N3 or Slow Wave Sleep (SWS) [43,44].

2. Challenges in Sleep Stage Scoring and Need for Automation

Polysomnographic (PSG) recordings, examined visually and utilizing multiple channels, result in expensive, error-prone, tedious, and time-consuming sleep stage scoring [45,47]. Analyzing a full-night recording typically spans 2 to 4 hours, with expert agreement rates occasionally below 90% [8,32,38]. PSG-based sleep stage scoring usually occurs in hospital settings, necessitating subjects to be on waiting lists and spend uncomfortable nights in specially equipped sleep labs [7,12].

3. Importance of Automated Sleep Stage Classification

Automated techniques like Automatic Sleep Stage Classification (ASSC) would alleviate these challenges by reducing clinician time, enhancing analytical accuracy, and aiding sleep disorder diagnosis and treatment [8,12,38]. Despite the importance of the EEG signal in sleep staging, multiple EEG channels restrict subject movement, hindering device portability and wearability. Implementing a wearable single-channel EEG device garners interest among researchers and mitigates disturbances caused by PSG recording wires [7,11,12].

4. Approaches in Automatic Sleep Stage Classification

Numerous methods for automatic sleep stage classification utilize signal feature extraction and classification algorithms [43,47,50]. These include time, frequency, and time-frequency domain analyses, alongside successful use of nonlinear parameters and complexity measures [2,14,27,37]. Some systems perform feature selection and dimensionality reduction before classification, aiming to minimize features and generate low-dimensional inputs [9,45,47]. Machine learning-based classifiers such as Linear Discriminant Analysis (LDA), Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Decision Trees (DT) are widely employed for sleep stage classification [4,15,17].

A Bidirectional Long Short-Term Memory (BiLSTM) is a type of recurrent neural network (RNN) architecture that processes sequences bidirectionally. Introduced as an extension of the standard LSTM architecture, BiLSTMs incorporate information from both past and future contexts by utilizing two separate hidden states for each time step: one capturing information from past time steps (forward LSTM) and another capturing information from future time steps (backward LSTM) [74].

This architecture allows the model to understand context and dependencies in both directions within a sequence, which can be beneficial for tasks such as natural language processing (NLP), speech recognition, time series analysis, and more.

Long Short-Term Memory (LSTM): LSTM is a type of RNN designed to address the vanishing or exploding gradient problem that affects standard RNNs. It introduces gating mechanisms (input, forget, and output gates) that control the flow of information within the network, enabling it to learn and retain information over long sequences. Bidirectional LSTM A BiLSTM consists of two LSTM layers, one processing the input sequence in the forward direction and the other in the backward direction. The outputs of these two LSTM layers are concatenated or combined in some way to capture information from both past and future contexts simultaneously.

5. Challenges and Contributions

Existing ASSC methods display varying accuracies (70% to 94%), sensitivity, and specificity below 90%, presenting challenges in accuracy, sensitivity, and specificity [17,59]. Addressing efficient sleep feature extraction methods, improved classification algorithms, and portability for long-term, home-based monitoring remains partially unresolved. Many techniques involve complex methodologies, long computational times, and poor generalization, posing challenges for real-time hardware implementation like driver fatigue detection systems. Some studies necessitate multiple EEG channels or combinations with other PSG methods, making subjects uncomfortable [17]. Despite efforts with single-channel EEG usage, unresolved epoch ambiguity in distinguishing EEG signals between S1 and REM stages hampers classification performance [17,42].

6. Paper Overview and Proposed Methodology

This study provides a comprehensive new method in comprehensive with various approaches to Automated Sleep Stage Classification (ASSC) and the identification of sleep disorders through the use of EEG signals. It presents a detailed analysis framework for sleep stages that includes steps such as pre-processing, feature identification, selection, dimensionality reduction, and

categorization. Furthermore, the paper introduces an innovative automated technique that leverages a single-channel EEG for categorizing sleep stages. This method incorporates advanced filtering processes, novel features in the time domain, and a groundbreaking deep-learning classifier architecture. Such an architecture is designed to grasp both forward and backward context and dependencies within data sequences, proving advantageous for applications in natural language processing (NLP), speech recognition, time-series analysis, and beyond.

The paper discusses Long Short-Term Memory (LSTM) networks, a form of Recurrent Neural Network (RNN) created to solve the issues of vanishing or exploding gradients found in conventional RNNs. It achieves this through unique gating mechanisms (input, forget, and output gates) that manage data flow in the network, thus allowing it to capture and maintain information across extended sequences. The study also explores Bidirectional LSTM (BiLSTM), which employs two LSTM layers to process data both forwards and backwards, merging their outputs to effectively incorporate information from both prior and subsequent contexts. The features from BiLSTM are used like input of LightGBM classifiers, that could be the robust system used for the behavior states detection.

2. EEG (Electroencephalogram)

The human brain, an intricate and dynamic structure including numerous linked neurons that communicate via both dendrites and axons [61–63], is crucial for cognitive function. Figure 1 illustrates the overall architecture of a neuron.

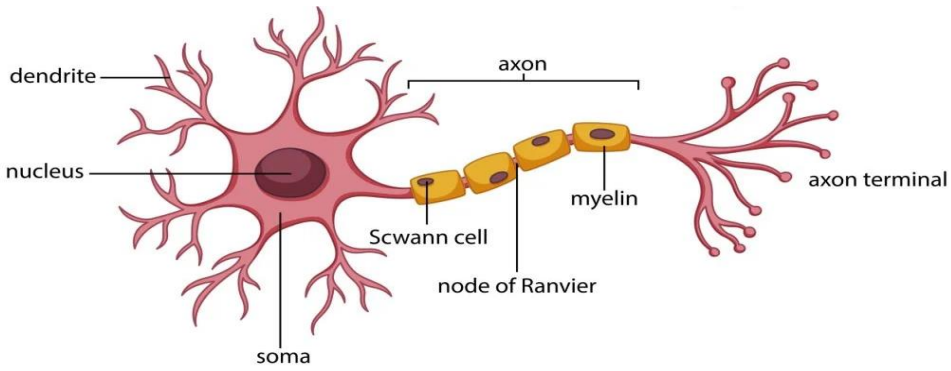


Fig 1: illustrates the anatomical arrangement of a neuron [65].

According to [65], the brain can be structurally classified into three primary structures such as the cerebrum with the cerebellum and brainstem (shown in Figure 2a). Of them, the cerebrum, which is the largest part, consists of two hemispheres that contain the central nervous system on their outer layer. The cortex is divided into 4 lobes: the frontal, the parietal, the temporal, and the occipital, as shown in Figure 2b [62].

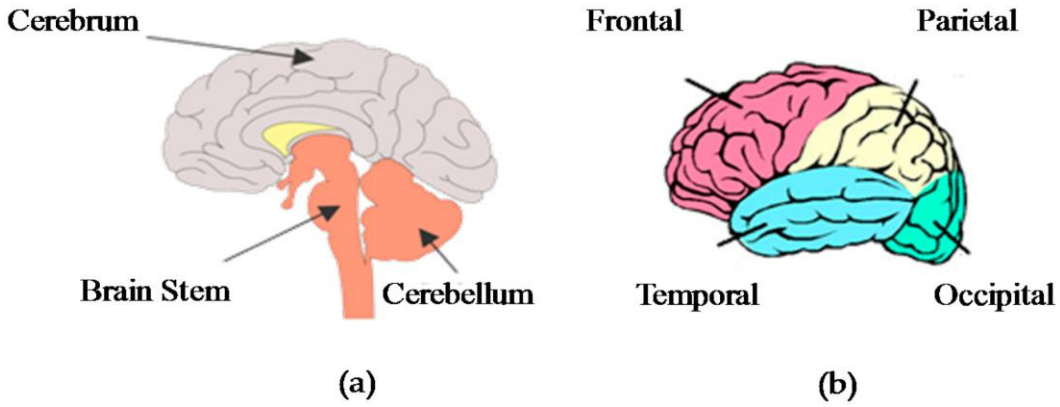


Fig 2: shows: (a) A depiction of the human brain.(b) A figure explicitly illustrating the anatomical organization of the brain [62].

Various techniques have been developed to evaluate the level of signal activity in the human brain, such as EEG, Magnetoencephalography (MEG), and functional Near-Infrared Spectroscopic (fNIRS), and Positron Emission Tomography (PET) [61,66,72]. Out of them, EEG is particularly notable as an important biological signal, with substantial practical importance in the neurology field [57]. This approach, which does not need any invasion of the body, quantifies the electrical activity of the cerebral cortex [73]. It has been widely used since it was

discovered by Berger in 1929 [70]. EEG incorporates a multitude of electrodes placed on the scalp according to the internationally recognized 10/20 placement technique [71]. With the progress in hardware technology, multi-channel EEG systems were developed. To accommodate extra electrodes, an expansion was made to the 10/10 system [69]. Figure 3 depicts the locations denoted by blue points are the 10/20 scheme, while the other points indicate the electrodes used in our studies.

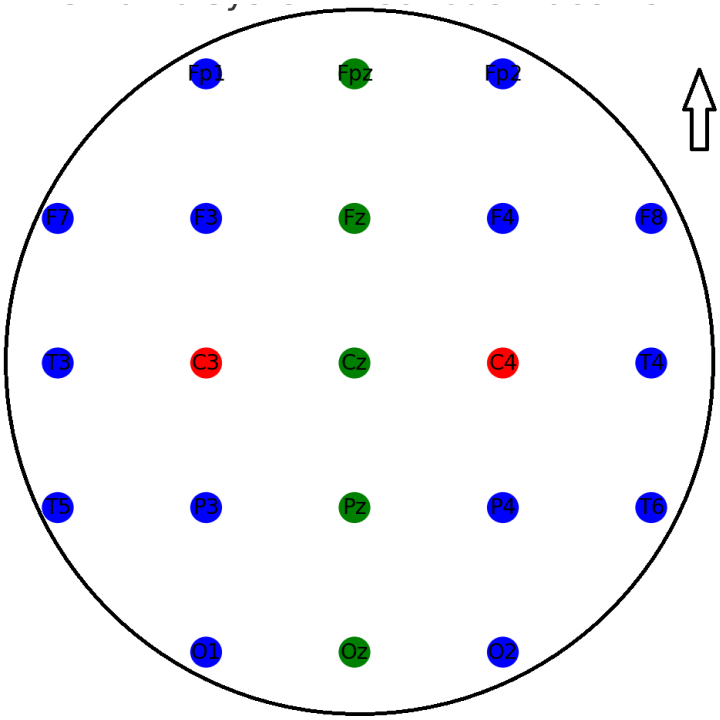


Fig 3: displays the EEG 10-20 system electrode placement as detailed in reference.

The EEG waveform may be categorized into five distinct frequency bands: delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ). These bands provide essential observations for the diagnosis, monitoring, and treatment of neurological characteristics and illnesses. Table 1 presents both the frequency and amplitude ranges that define the decomposition of EEG signals in different frequency bands, as shown by references [3,62].

Table 1 displays the frequency and amplitude ranges of the decomposed EEG signal with the δ, θ, α, β, and γ sub-bands in typical circumstances.

Frequency Band	Amplitude Range	Frequency Range
δ (Delta)	Low	0.5 - 4 Hz
θ (Theta)	Moderate	4 - 8 Hz
α (Alpha)	High	8 - 13 Hz
β (Beta)	Moderate	13 - 30 Hz
γ (Gamma)	Low	30 - 100 Hz

Various methods utilizing single or multi-channel EEG recordings have been intensively investigated in the field of automated sleep stage grading. Presented below are synopses of chosen investigations in this field:

1. Research conducted by [5]: By utilizing Fp1 and Fp2 EEG signals, quasi-stationary components were segmented, feature extraction was performed using Short Time Fast Fourier (STFT), dimension reduction was achieved with Fuzzy C-Means (FCM), and sleep stage classification was achieved with multiclass SVM. Attained a precision rate of 70.92%.

2. Mustfa et al. [10] employed six EEG signals and focused on several signal processing aspects, including time domain, frequency domain, and non-linear features. They used Random Forests (RF) and Support Vector Machines (SVM) as classifiers. Demonstrated peak efficiency using frontal EEG signals and spectrum linear characteristics in conjunction with RF.

3. In a study conducted by [11], the Bagging method was utilized with statistics and spectral characteristics extracted from a single EEG channel. This approach achieved accuracy rates ranging from 85.57% to 95.05% for various sleep state classifications.

4. The approach proposed by [12] uses Complete Ensemble Empirical Mode Decomposition (CEEMDAN) to score sleep stages based on single-channel EEG data. Bagging is employed for classification, resulting in accuracies ranging from 86.89% to 99.48% for different sleep phases.

5. Metrics based on entropy as proposed by Sotelo et al. [16]: Utilized entropy metrics, Q-algorithm for reducing dimensionality, and J-mean clustering for automated sleep stage scoring based on two-channel EEG data, resulting in an accuracy of up to 80%.

Study [28] employed Support Vector Machines (SVM) to differentiate between waking and sleepy states based on three EEG channels. The study achieved a remarkable accuracy of 98.01% and precision of 97.91%.

Using single-channel EEG and ANN, Fraiwan et al. [29] developed a methodology that achieved an 84% classification accuracy with WVD features by combining Wigner-Ville Distribution (WVD), Hilbert-Hough Spectrum (HHS), and Continuous Wavelet Transform (CWT).

8. The properties based on Renyi's entropy as described by [30]: Utilizing three time-frequency approaches, we were able to achieve an accuracy rate of 83% in identifying sleep stages. This was accomplished by employing an RF classifier and analyzing data from a single EEG channel.

9. The study utilized nine graph domain variables to extract the single-channel of the EEG signal to classify six sleep phases. This classification was achieved using a multiclass SVM algorithm, resulting in an accuracy of 87.5%.

10. The categorization of sleep stages according to [41]: investigated time- and frequency-domain characteristics from PSG data (two EEG channels, two EOG channels, and one EMG channel), using a Dendrogram-SVM (DSVM) to achieve 94% specificity, 82% sensitivity, and 92% accuracy.

11. Karkovská and Mezeiova [42] utilized quadratic discriminant analysis to extract 14 features from PSG data, including 6 EEG, 2 EOG, and 1 EMG channel. Their approach achieved an accuracy rate of 81%.

12. In a study conducted by [45], researchers extracted 39 variables from a single-channel EEG and used a binary SVM to categorize five sleep phases. The classification achieved an average sensitivity of 88.32%, specificity of 97.42%, and accuracy of 97.88%.

These papers demonstrate multiple methods for automatically scoring sleep stages using EEG data. Each study uses unique strategies and achieves high levels of accuracy in classifying different sleep states.

Table 2: Comparative results and contribution.

Study	Methodology	Results
[5]	EEG signals (Fp1, Fp2), STFT, FCM, multiclass SVM	Precision rate: 70.92%
Mustfa et al. [10]	Six EEG signals, time/frequency/non-linear features, RF & SVM classifiers	Peak efficiency with frontal EEG & RF
[11]	Single EEG channel, Bagging with statistics and spectral features	Accuracy: 85.57% to 95.05%
[12]	Single-channel EEG, CEEMDAN, Bagging	Accuracy: 86.89% to 99.48%
Sotelo et al. [16]	Two-channel EEG, entropy metrics, Q-algorithm, J-mean clustering	Accuracy up to 80%
Study [28]	Three EEG channels, SVM	Accuracy: 98.01%, Precision: 97.91%
Fraiwan et al. [29]	Single-channel EEG, ANN, WVD, HHS, CWT	Accuracy: 84%

[30]	Single EEG channel, RF, Renyi's entropy	Accuracy: 83%
[31]	Single-channel EEG, nine graph domain variables, multiclass SVM	Accuracy: 87.5%
[41]	PSG data, DSVM	Specificity: 94%, Sensitivity: 82%, Accuracy: 92%
Karkovská and Mezeiova [42]	PSG data (6 EEG, 2 EOG, 1 EMG), quadratic discriminant analysis	Accuracy: 81%
[45]	Single-channel EEG, 39 variables, binary SVM	Sensitivity: 88.32%, Specificity: 97.42%, Accuracy: 95.88%

4. Methodology for Sleep Stage Classification Proposal

This section presents a novel and efficient approach for categorizing different sleep phases. The methodology is specifically designed to be used in hardware systems, enabling real-time support in the diagnosis and treatment of sleep disorders. As shown in Figure 4, our method begins by applying band-pass filters to the EEG signal to filter and segment it into 5 EEG waves sub-bands. Afterwards, we get two new sets of statistical characteristics from each frequency range. In the last stage, a range of well-established machine learning classifiers are employed to choose the most effective one for categorizing W, REM, and NREM sleep phases. The suggested sleep stage categorization method is particularly noteworthy as it is completely automated and functions just utilizing a single EEG channel. The next subsections outline each stage with more precision.

4.1 Architecture of aimed system :

The Hybrid classifier system that combines BiLSTM and LightGBM for EEG data involves several steps, from preprocessing and feature extraction to model training and evaluation. Below is an outline of the process, along with a proposed approach for each step:

1. Data Preprocessing

Filtering: Apply band-pass filters to remove noise and focus on relevant EEG frequencies.

Artifact Removal: Use techniques like Independent Component Analysis (ICA) to remove artifacts (e.g., eye blinks, muscle movements).

Normalization: Normalize the data to have a standard scale, which is important for training neural networks effectively.

2. Feature Extraction

The Time-Domain Features Extract features like mean amplitude, variance, skewness, and kurtosis from the EEG signals. The Frequency-Domain Features is Compute power spectral density, band power in standard EEG bands (Delta, Theta, Alpha, Beta, Gamma).The Time-Frequency Features use wavelet transform or Short-Time Fourier Transform (STFT) for time-frequency analysis. Statistical Feature Calculate correlation, covariance, or other statistical measures between different EEG channels.

3. Sequence Modeling with BiLSTM

Input Preparation: Structure the EEG data into sequences suitable for input into the BiLSTM model. These could be segments of continuous EEG recordings.

-Model Architecture: Design a BiLSTM network. The network should have LSTM layers capable of processing sequences in both forward and backward directions.

Training: Train the BiLSTM model to learn representations of the EEG sequences. The output could be a feature vector representing each sequence.

4. Feature Engineering for LightGBM

Combine Features Combine the features extracted in step 2 with the learned representations from the BiLSTM model. This creates a rich feature set that includes both hand-engineered features and learned temporal patterns. Dimensionality Reduction (Optional) Apply methods like PCA (Principal Component Analysis) to reduce the dimensionality of the feature set, if necessary.

5. Classification with LightGBM

Model Configuration Configure the LightGBM classifier. LightGBM is effective with large datasets and can handle a variety of feature types. The training Train the LightGBM model on the combined feature set. This model will make the final predictions.

6. Model Evaluation

Cross-Validation Use techniques like k-fold cross-validation to evaluate the model performance. Metrics

Compute performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, depending on the specifics of your task.

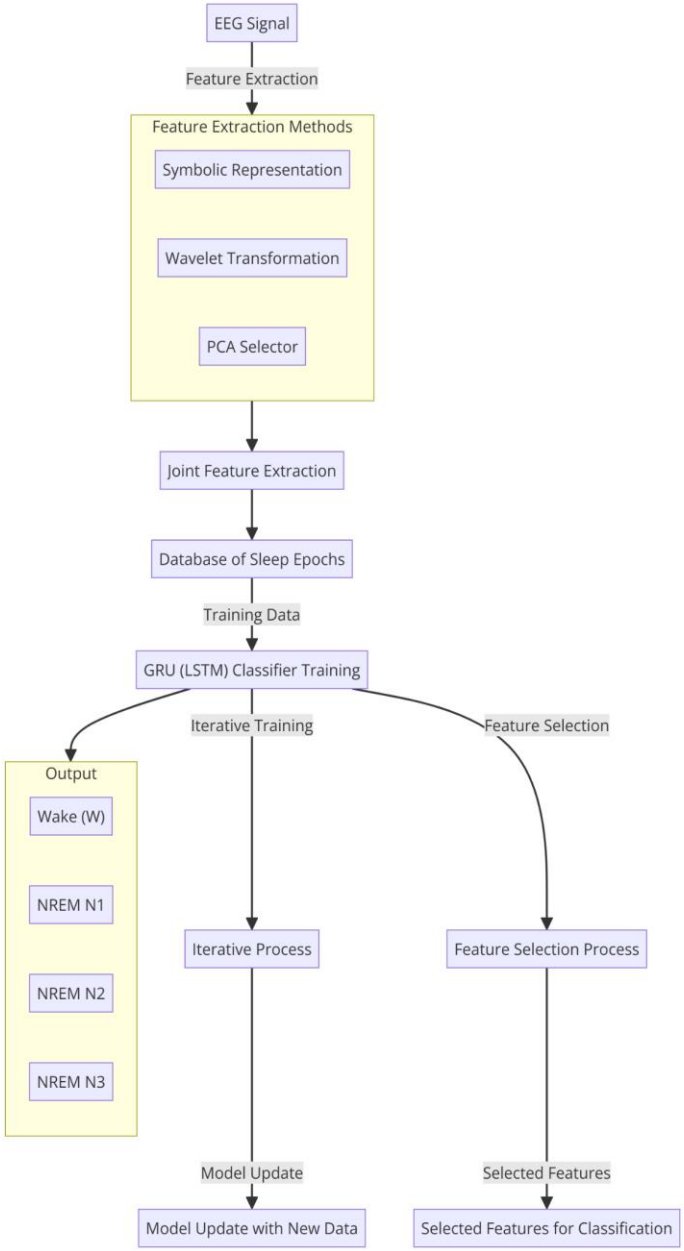
7. Model Interpretation and Analysis

Feature Importance Analyze which features are most important for predictions.

Error Analysis Examine cases where the model makes errors to understand its limitations.

Data and Computational Resources is an approach that can be computationally intensive and requires a significant amount of labeled EEG data.Customization Tailor each step to your specific dataset and classification task. Iterative Process Model development is iterative. You may need to go back and adjust earlier steps based on the results you get.

This hybrid approach leverages the strengths of both neural networks in capturing complex temporal patterns and gradient boosting in handling a wide range of features effectively. However, the success of this approach depends heavily on the quality of the data and the relevance of the features to the behavioral states you're trying to classify.



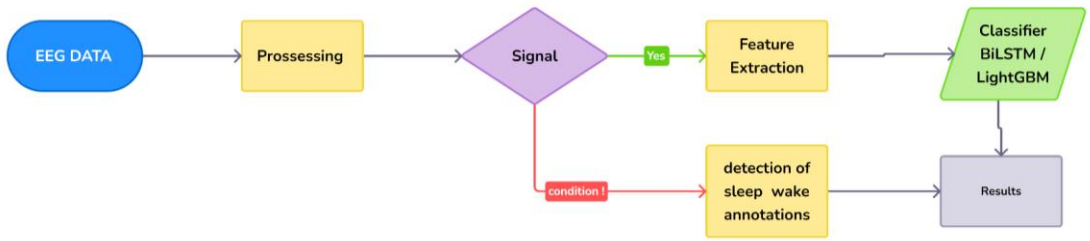


Fig 4: illustrates the process of the proposed Automatic Sleep Stage Classification (ASSC) System.

The Combining features extracted from the BiLSTM and CSP with hand-engineered features, and using LightGBM like a final classifier , the gradient boosting framework, to handle the classification task with five different classes give an accuracy developed after each time of classification the new data's.

4.2. Obtain EEG Signal

The EEG signal used in this study is derived from the Sleep-EDF database, which can be accessed through the Physionet website. This database contains 61 polysomnograms (PSGs) that were gathered from 1987 to 2002, with some records dating back to before 1991. The dataset consists of audio recordings obtained from 20 individuals who are in good health and fall within the age range of 25 to 34. The participants are evenly distributed between male and female. Each subject had two polysomnography (PSG) recordings, each lasting around 20 hours per night, spanning two consecutive days in their own natural home contexts. Regrettably, the recording cassette failure resulted in the unavailability of the second night for subject 13.

The recordings include EEG Fpz-Cz, EEG Pz-Oz, EOG horizontal, submental chin EMG, event markers, and other signals including oro-nasal breathing and rectal body temperature. The study only employed the EEG Fpz-Cz signal, which was collected at a frequency of 100 Hz, as the single-channel EEG. The hypnogram files provided contain sleep stage annotations, which include W, 1, 2, 3, REM, M (movement time), and ? (indicating unannotated phases). These annotations have been carefully assessed by qualified technicians, following the guidelines outlined in the R&K handbook.

In this study, every signal is analyzed in 10-second increments. Figure 5 demonstrates samples from five separate phases of EEG signals that were used as input for the developed filters. Furthermore, Table 5 provides a comprehensive breakdown of the number of waking, Stage 1, Stage 2, Stage 3, Stage 4, and REM epochs for all individuals in the database. The Sleep-EDF database has been extensively cited and employed in several research works in the literature [2].

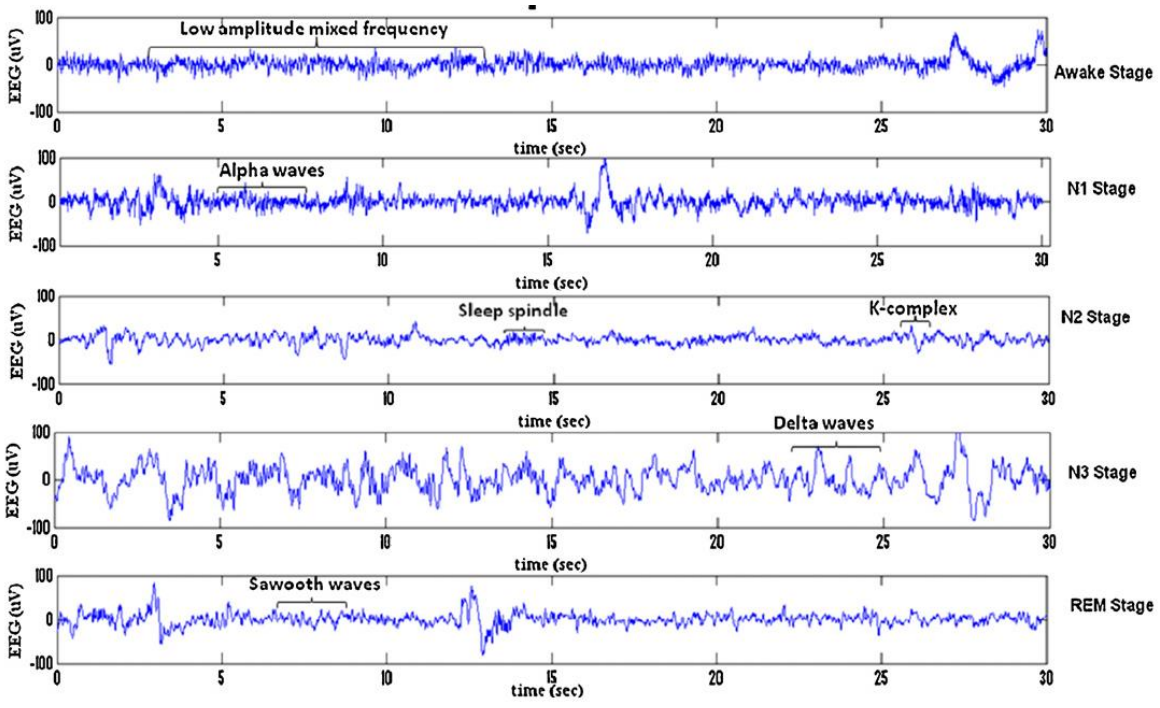


Fig 5 displays the EEG signals corresponding to various sleep stages.

Table 5. Number of epochs for different sleep stages in the dataset.

Stage	W	S1	S2	S3	S4	REM
Total	5961	3552	7175	4321	1900	897

4.3. Pre-Processing

The first phase is pre-processing the input EEG data, with the goal of removing undesirable background signals and separating it into five separate frequency bands (δ, θ, α, β, and γ). We did this by using effective Butterworth band-pass filters with infinite impulse response (IIR). Due to their simple transfer functions, these filters are useful and may be implemented with ease in digital hardware engines, embedded systems, and digital signal processors [71]. The frequency response of the Butterworth filter guarantees little passband ripple and a uniformly flat passband. The minimal order of the filter, which is crucial for accurate and efficient design, is found using the specified equations (1) and (2). In these equations, Gp represents the gain in the passband, Gs represents the gain in the stop-band, ωp represents the frequency at the corner of the passband, and ωs represents the frequency at the corner of the stop-band.

$$N = 12 \times \frac{\ln(G_p/G_s)}{\ln(\omega_p/\omega_s)}$$

$$\omega_c = \omega_s \times (G_s)^{\frac{1}{2N}}$$

4.4. the aim Feature Extraction

Following pre-processing, the filtered EEG signals undergo feature extraction to capture specific characteristics within each 10-second EEG epoch. Time-domain statistical features are widely acknowledged for differentiating diverse EEG classes, offering insights into the data's underlying statistics. In this study, two novel statistical features are introduced. The first, Maximum-Minimum Distance (MMD), stems from segmenting the non-stationary EEG signal into sub-windows in the time domain. In addition, EnergySis (Esis) quantifies the energy and velocity of the EEG signal. Our methodology entails dividing the signal into smaller sections, known as sub-windows. The length of each sub-window, or the total number of samples it contains, is a multiple of 10, starting at 100. The length of the EEG waveform is equivalent to its wavelength. Figure 6 depicts the wavelength assumptions for 10 seconds.

This technology seeks to improve the technique of classification by extracting unique characteristics from the EEG data, enabling precise identification of different sleep phases.

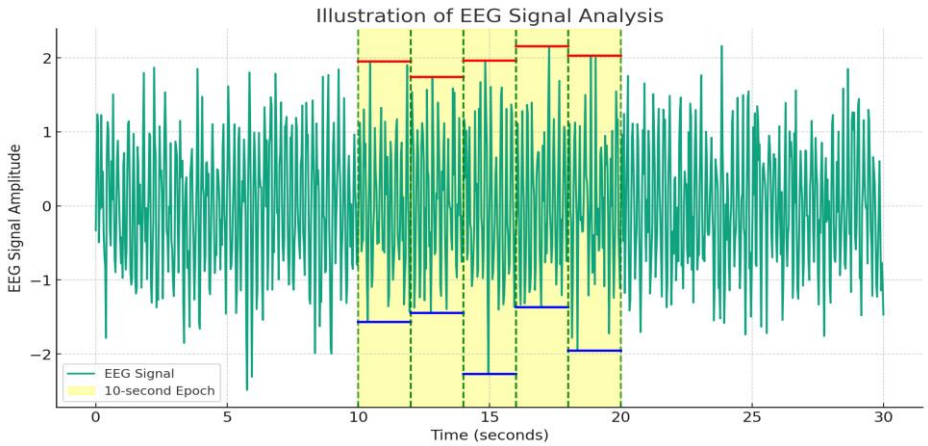


Fig 06: Size selection of samples EEG

Let λ denote the size of a sliding window, which is quantified in the number of samples and is also referred to as the wavelength of the EEG signal. Our method is based on the number of samples in each epoch: λ is set to 100 when there are less than 10,000 samples; it is set to 1000 when there are from 10,000 to 100,000 samples, and so on. This guideline is consistently applied throughout all time periods. Equation (3) demonstrates the procedure for determining the quantity of samples in a sliding window, often known as the wavelength.

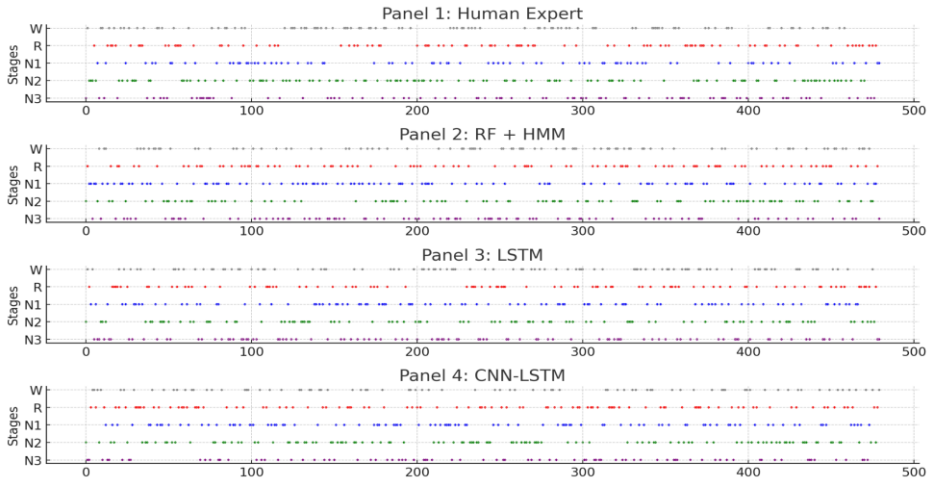


Fig 07 : classification of EEG behavior state

Figure 07 presents the automatic sleep stage classification using a dataset of healthy individuals (dataset 1; taken from the validation set). The first panel displays a hypnogram (indicating wakefulness, REM sleep, and NREM sleep stages N1–N3) scored by an expert. The second panel shows a hypnogram derived from Random Forest (RF) classification utilizing specific features and further refined through Hidden Markov Model (HMM) temporal smoothing. The third panel illustrates a hypnogram obtained through the application of a 3-layer bidirectional LSTM network, featuring 8 LSTM neurons per layer and based on certain features, with a sequence length of eight epochs (equivalent to 160 seconds). The fourth panel presents a hypnogram generated by a combined LightGBM-LSTM network, incorporating 11 convolutional layers and a 2-layer bidirectional LSTM, each layer having 32 LSTM neurons, using raw input data (1 EEG and 2 EOG signals) along with EMG power (one value per epoch). The final panel is a spectrogram showcasing the power density spectra of EEG signals from the C3A2 derivation over 20-second epochs, represented in a color-coded logarithmic scale (ranging from -10 dB to 20 dB, with 0 dB set to 1 µV2/Hz). Additional information on the naming conventions used for the algorithms can be found in the Supplementary Material.

using the applying of method x which presents an individual classifier we can see the amplitude range between wake and N3 figure 08.

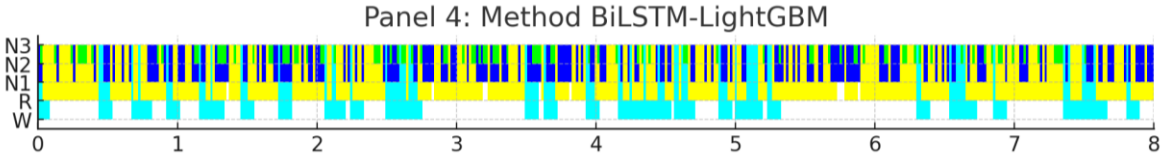


Fig 08: applying the method X on the first epoch.

Following stage 1, stage 2 (N2) represents a deeper level of sleep, marked by the appearance of sleep spindles and K-complexes, with the muscle tone remaining at an intermediate level.

This stage is the precursor to the deeper stages of sleep, namely stages 3 and 4 (also known as SWS or N3), which are characterized by slow oscillations (less than 1 Hz) and delta waves (1–4 Hz) in the brain's electrical activity, making up at least 20% of the time period measured. During these stages, muscle tone is significantly reduced.

Rapid eye movement (REM) sleep occurs at intervals throughout the night, distinguished by quick eye movements, a pattern of brain activity similar to that of being awake but with low amplitude, and very relaxed muscles (atonia).

The transition between these sleep stages is not arbitrary but follows a systematic pattern of non-REM and REM sleep phases that repeat roughly every 90 minutes, contributing to a typical structure of 3–5 such cycles per night for a restful sleep[75].

5. Our Contribution

We explored various machine learning approaches, including random forests (RF), feature-based networks (BiLSTM networks), and networks that process raw data (BiLSTM-LightGBM networks), applying them to both healthy individuals and patients. We evaluated the performance of these algorithms by reporting the Cohen's kappa values for different sleep stages.

Our findings showed that all algorithms performed well on data from healthy subjects, indicated by high Cohen's kappa values. When applied to patient data, the performance dropped, although artificial neural networks (ANNs) exhibited a smaller decline. By incorporating some patient data into our training set, we noticed an enhancement in performance on patient data. This outcome hints at the necessity for larger and more varied datasets to develop an algorithm that is robust and reliable in real-world scenarios. Notably, our deep neural networks (DNNs) demonstrated impressive results even when utilizing just a single EEG channel, marking a significant insight from our research.

II. Materials and Methods

2.1 Polysomnographic (PSG) MNE Data

We conducted training and testing of automatic sleep stage classification algorithms using the open datasets from Physionet [76]

The initial dataset contained 54 full-night recordings of sleep from healthy individuals. The second set included 22 full-night sleep recordings and 21 instances of a Multiple Sleep Latency Test (MSLT) involving patients. The MSLT is commonly employed to assess a patient's level of daytime sleepiness. In this test, participants are given four or five opportunities to nap for 20 minutes each, spaced by intervals lasting 1.5 hours. Figure 09 illustrates a typical MSLT hypnogram. Whereas typically only the naps are recorded, our dataset uniquely captures continuous recording over roughly 9 hours, during which additional sleep episodes were noted beyond the planned naps—a scenario that conventional setups might overlook. For the purposes of analysis and algorithmic classification, data was collected using the EEG channel C3A2, alongside one myographic and two oculographic channels.

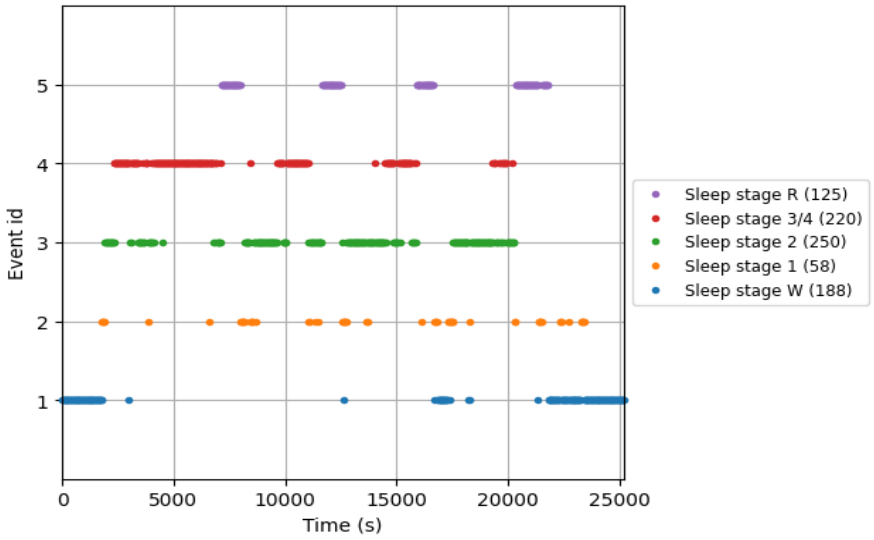


Fig 09: training it on a combined dataset consisting of both healthy participants and patient data (datasets 1 and 2).

In a research project exploring the impacts of vestibular stimulation on sleep, polysomnographic (PSG) data were collected from 18 healthy young men, aged between 20 and 28 years, with an average age of 23.7 years. Each participant underwent three sleep recordings, each lasting 8 hours. These included two nights where motion was introduced (the bed was rocked until sleep onset or for the first 2 hours after turning off the lights) and one control night without any movement [77].

The collected data encompassed 12 EEG channels set up according to the 10–20 system, two EOG channels, one submental EMG channel, one ECG channel, and respiration signals from the chest and abdomen. The recordings were made using a polygraphic amplifier from Artisan, Micromed, in Mogliano Veneto, Italy, with a sampling rate of 256 Hz, using Rembrandt DataLab software. Additionally, filters were applied to the analog signals to ensure data quality: a high pass filter for EEG signals at 0.16 Hz, EMG at 10 Hz, ECG at 1 Hz, and an anti-aliasing filter at 67.4 Hz. EEG signals were also adjusted to reference the contra-lateral mastoids (A1, A2). Sleep stages were identified in 20-second epochs, adhering to the AASM guidelines[78].

2.2 Classifiers:

Machine learning is a subset of computer science focused on enabling computers to learn from data characteristics and solve issues without explicitly programmed decision-making rules. It primarily revolves around two methodologies: supervised learning, where the algorithm learns from labeled data to make predictions or classifications, and unsupervised learning, which deals with unlabeled data, aiming to find hidden patterns or structures within. In our research, we adopted the supervised learning strategy to address a classification

challenge. This entails using algorithms to categorize data into predefined labels, leveraging a dataset where each instance is already associated with a label to teach the algorithm about the data features and the relevant labels.[79]

Our study approached the classification problem by employing supervised learning techniques, specifically focusing on two methods: (1) feature-based classification using Random Forests (RF) and a complex deeplearning LightGBM, and (2) raw data-based classification utilizing ANNs.

2.2.1 Classification Based features

Feature-based classification leverages the intricate yet informative patterns present in polysomnographic signals to identify different sleep stages. Notable patterns, such as sleep spindles (12–14 Hz), slow waves (0.5–4 Hz), alpha waves (8–12 Hz), and theta oscillations (4–8 Hz), play a crucial role in sleep stage differentiation as observed by experts. These patterns can be quantified effectively within the frequency domain. To accomplish this, we utilized classical spectral analysis methods, although a multi-taper method could also be effective, especially when employing spectrograms as features. In addition to these, other significant indicators of sleep stages, including rapid and slow eye movements, eye blinks, and muscle tone, can be measured. This process of identifying and measuring such indicators is known as feature engineering. Employing well-crafted, domain-specific features in machine learning not only minimizes the amount of training data required but also enhances the speed of the analysis and the interpretability of the results. In contrast, another method involves using deep learning to analyze raw data, which we discuss in a subsequent section [80].

Preprocessing and feature extraction

Initially, we opted for spectrograms of EEG signals over raw signals to capitalize on the well-established fact that spectra embody the key characteristics of sleep EEG, thus allowing for a substantial reduction in data dimensionality. Power density spectra for 20-second epochs (extended to 30 seconds for patient data) were computed using the Welch method in MATLAB, which involved FFT analysis with an average derived from either four or six 5-second windows, employing Hanning windows, without overlap, and achieving a frequency resolution of 0.2 Hz. The resulting spectra were visualized and color-coded on a logarithmic scale, limited to a frequency range of 0.8–40 Hz to decrease the data matrix size.

For the task of classification, we identified a suite of 20 engineered features, including but not limited to, power across various frequency bands and their ratios, eye movement indicators, and measures of muscle tone. The details of these features are elaborated in the Supplementary Material. We chose not to eliminate any epochs from our analysis, even those with artifacts, aiming for a system that necessitates minimal manual preprocessing. This decision is backed by the observation that artifacts often carry pertinent information, such as the association of wakefulness with movement artifacts and the likelihood of transitioning to stage 1 sleep following a movement. However, for quantitative analyses like average power density spectra computation, artifact exclusion becomes necessary, which can be efficiently handled using straightforward algorithms[81].

Our feature-based classification utilized two distinct methods: Random Forests (RF) and Artificial Neural Networks (ANNs).

2.3 Random forest (RF)

Decision trees are a foundational approach for tackling classification challenges, with each tree node representing a feature alongside a threshold value. To classify a data point, the process involves navigating the tree by comparing the feature of the data point to the node's threshold, moving left or right according to the comparison's outcome. This navigation continues until reaching a leaf, which assigns the data point to a specific class.

However, decision trees are susceptible to issues like overfitting, where the model becomes too tailored to the training data, impairing its ability to generalize to new data. To mitigate such limitations, an ensemble of trees strategy is employed. This involves constructing numerous trees, each based on a random segment of the training dataset. Classification of a data point is then achieved by aggregating the outcomes from all trees, with the class probability determined by the proportion of trees that designate the data point to that class. Techniques like Random Forest (RF) classifiers and other recent tree-based methods have shown excellent performance across diverse problems by adopting this ensemble approach [82].

For our sleep stage classification, we utilized the RF approach, analyzing feature vectors consisting of 20 elements. We calculated probability vectors for each epoch, whether 20 or 30 seconds long. Additionally, to account for the inherent temporal structure of sleep, we integrated time course learning through the use of a Hidden Markov Model (HMM) and applied a median filter (MF) with a window spanning three epochs (each being either 20 or 30 seconds) to refine the data.

2.4 Deep learning with raw data:

Deep learning leverages Deep Neural Networks (DNNs), a subset of Artificial Neural Networks (ANNs), capable

of modeling complex relationships within data. One of the key advantages of DNNs is their ability to autonomously identify features from raw data, eliminating the need for manual feature engineering. This feature learning can be effectively conducted through the use of LightGBM, among other techniques. Generally, DNNs outperform traditional feature-based classification methods in terms of accuracy, albeit with higher computational costs and a greater need for extensive training datasets. Despite these demands, DNNs require significantly fewer manual adjustments compared to feature-based approaches, making them more straightforward to develop and manage [83].

2.5 Learning Time Dependencies

Traditional machine learning models typically treat each data sample as independent from others, which applies to Random Forest (RF) classification and most Artificial Neural Networks (ANNs). However, when it comes to sleep scoring, experts often consider information from preceding epochs, suggesting the value of incorporating some degree of temporal information into the classification algorithm.

Given that sleep exhibits both local and overarching structures, such as sleep cycles, integrating this temporal dimension into models is essential. Nevertheless, it's crucial to not overly rely on global structures for scoring, as these may vary significantly in cases of sleep disorders or during short naps, potentially misleading the algorithm. Hence, our models are designed with limited temporal memory to avoid bias from long sequence patterns that may not apply universally, particularly in scenarios like Multiple Sleep Latency Test (MSLT) recordings or instances of disrupted sleep.

To incorporate the temporal aspect of sleep data, we adopted two strategies. Initially, we utilized a Hidden Markov Model (HMM) to refine the outcomes of RF classification, supplemented by a median filter (MF) with a three-epoch window. This simple yet effective method helps smooth the data, offering a more consistent analysis over short time spans.

Furthermore, we explored the use of Recurrent Neural Networks (RNNs), which inherently account for data's temporal structure by feeding back the output from a previous step as an input along with new data. Among RNNs, Bidirectional Long Short-Term Memory (LSTM) networks stand out for their efficiency in preventing gradient vanishing issues and their ability to handle long-term dependencies. Bidirectional RNNs, which consider both past and future data, offer an even more comprehensive view of temporal patterns.

To avoid biases from overly long sequences that might not be representative of all sleep patterns, we limited our models to learning from sequences no longer than 8, 32, and 128 epochs, equating to approximately 2.8 to 64 minutes. We adopted a dynamic batching approach for training, where the start of each sequence was randomly selected, allowing for overlapping sequences and a more versatile training dataset. Further details on sequence batching and processing can be found in the Supplementary Material.

III. Results

Throughout the training phase of Artificial Neural Networks (ANNs), we typically witness an improvement in classification accuracy. To determine when a network has been adequately trained and further training would not yield significant improvements, we evaluated both the cross-entropy loss and accuracy (the ratio of correctly classified instances; refer to the “Materials and Methods” section for more details). These metrics often display an exponential approach to a saturation point as training progresses. Once either accuracy or the loss function stabilizes at a plateau, it indicates that the network has reached convergence. These phenomena are captured in what are known as learning curves.

Our LSTM networks designed for feature-based analysis demonstrated solid convergence on data from healthy individuals as well as on a combined dataset from both groups of participants, indicative of effective training.

For ANNs that process raw data, the learning curves reflect varying degrees of convergence. Most networks exhibited consistent progress, with loss decreasing steadily and accuracy approaching a saturation point. However, certain networks experienced notable fluctuations in loss and accuracy, particularly those processing limited inputs such as a single EEG channel, or those handling inputs from EEG and EOG channels with sequence lengths of eight epochs, and networks incorporating EEG, EOG, and EMG over 128 epochs. The most erratic learning curves were observed in networks with residual connections, characterized by a larger parameter set, suggesting a need for more data and training iterations to achieve smooth convergence. Enhancing the dataset is anticipated to improve the performance of such complex networks.

3.1 Classification Performance

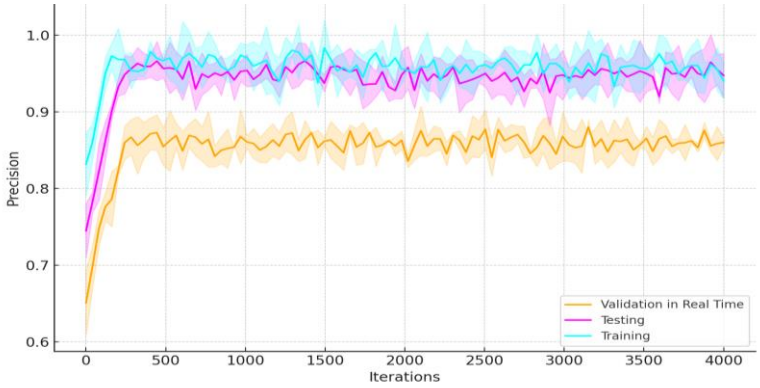


Fig 10: accuracy of the global system.

Figure 10 presents the Accuracy serves as a broad indicator of a model's correctness, computed by dividing the sum of correct predictions (true positives and true negatives) by the total number of instances in the dataset. Precision reflects the model's reliability in predicting positive outcomes, with a higher precision score indicating fewer false positives, thereby showing that the model is more precise in recognizing positive instances. Recall, in contrast, assesses the model's capability to identify all positive instances, with a higher recall score suggesting fewer false negatives, thus indicating that the model effectively captures positive instances. The F1-score, being the harmonic mean of precision and recall, strikes a balance between the two, accounting for both the accuracy and the comprehensiveness of the predictions. An F1-score of 1 is ideal, highlighting that a model evenly balances precision and recall without heavily favoring one over the other.

In our initial analysis, we focus solely on a LightGBM model, employing the same architecture outlined for the LightGBM component. This LightGBM model undergoes compilation and training with the same optimizer, loss function, and batch sizes as those used for the hybrid ANN–LightGBM model. The training and validation performance metrics for the LightGBM model, after an identical number of training epochs, are depicted in Figure 11.

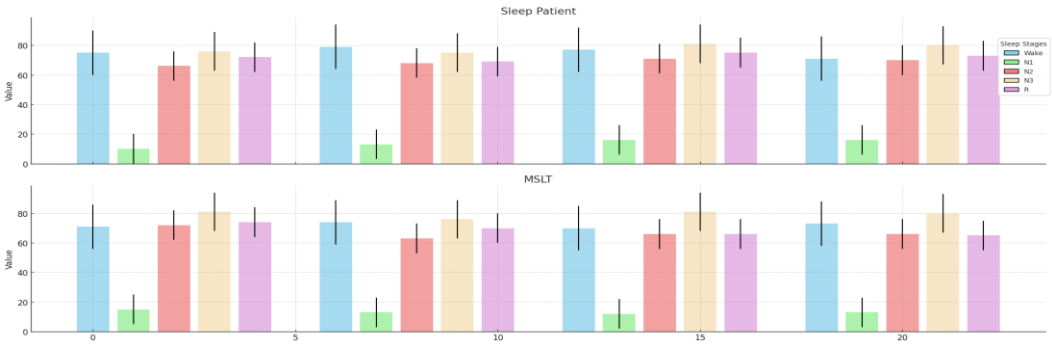


Figure 11: final classification using Kohen Cappa parameters ,trained on a mixture of data of healthy participants and patients data.

Figure 11 shows that all four methods demonstrated strong performance across all sleep stages, with the exception of stage 1 (N1), where Cohen's kappa was approximately 0.4. Despite this, such a result is still regarded positively, especially given its similarity to the relatively low agreement rates among human scorers for this particular stage.

Cohen's kappa for all the methods, as applied to the validation segment of the first dataset, are illustrated in supplementary materials. The majority of networks showed commendable performance on the validation dataset. Networks relying solely on a single EEG channel for input exhibited a marginally reduced effectiveness, likely due to the absence of data on eye movements and muscle tone in either the EEG spectrogram or the raw EEG signal. This limitation was not consistent across all recordings; in some instances, the performance was notably high. Interestingly, these networks performed significantly better on the test set, suggesting that the validation set might have included recordings that were challenging to score with just a single EEG channel.

The network analyzing EEG, EOG, and EMG data over 128 epochs displayed lower performance on both validation and test datasets, attributed to significant

accuracy fluctuations in the final training phase. This suggests a potential improvement could have been achieved by either halting training sooner or extending the training period.

Networks with either 16 or 32 units per layer showed slightly less accuracy in scoring stage 1 compared to a network with 8 units, possibly due to overfitting. However, this discrepancy was minimal, and a better outcome might be attainable with more extensive datasets. The unidirectional network was slightly less adept at predicting REM sleep than its bidirectional counterparts, although it has the advantage of being applicable in real-time analyses. Surprisingly, methods using Random Forests smoothed with either a median filter or a Hidden Markov Model performed nearly as well as those employing artificial neural networks for feature and raw data classification.

3.2 Generalization to the Patient Data

Our validation efforts extended to dataset 2, which comprises patient data, to assess the efficacy of our methodologies. Cohen's kappa values for a curated selection of methods are displayed in Figure 6, highlighting four specific approaches. For a comprehensive overview of the performance across all algorithms applied to patient data, refer to the supplementary materials, including Figures 12, as well as Supplementary Tables S3 and S4. It is important to note that none of the data from the patient dataset were utilized during the training phase of these algorithms.

The classifiers' performance on the patient sleep data was generally lower compared to the results from healthy participants, and this trend was even more pronounced for the Multiple Sleep Latency Test (MSLT) data, where kappa values varied widely. Notably, the Random Forest (RF) classification method had the poorest performance for stage 1 sleep classification within this dataset. Methods relying solely on a single EEG input, whether it be a spectrogram or raw EEG signal, were less effective on patient data.

Significantly low kappa scores were observed in several instances, particularly for stages 2, 3, and REM sleep among patient recordings when the training dataset did not include patient data. This led to frequent confusion between stages 2 and 1, likely due to the distinct characteristics of patient sleep, which tended to be more fragmented and disturbed. Consequently, algorithms lacking exposure to patient sleep patterns during training often misclassified these stages. The kappa values for stage 3 were notably low, primarily because of the rare occurrence or complete absence of deep sleep in patients, meaning that even minor errors significantly impacted kappa scores. Additionally, differences in the characteristics of REM sleep between patients and healthy subjects led to misclassifications; for example, low muscle tone during wakefulness in patients sometimes resulted in the erroneous identification of REM sleep, although some instances of what was initially considered false REM sleep upon further inspection turned out to be genuine episodes missed by experts.

Algorithms that analyzed only EEG data tended to make the most errors. Incorporating ocular channels into the model input improved accuracy, and the inclusion of muscle tone data yielded the best performance outcomes.

Performance enhancements were observed when the algorithms were trained on datasets that included patient data, underscoring the value of incorporating a diverse range of sleep patterns into the training process.

3.3 Networks Trained on the Data From Both Datasets

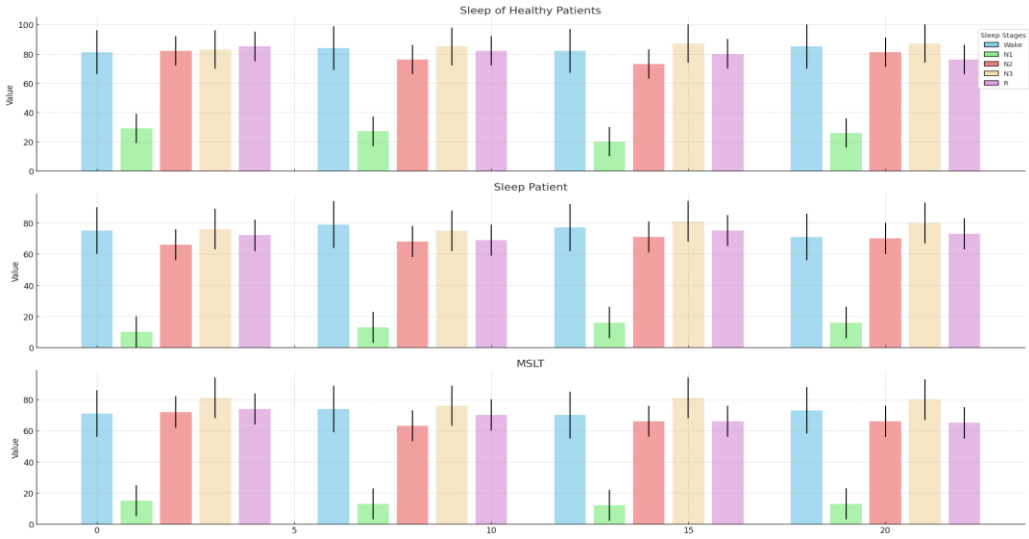


Fig 12: Cohen's kappa for the methods illustrated

The improved performance observed after training on datasets encompassing both healthy participants and patients aligns with expectations. Given the distinct nature of patient sleep patterns specifically in cases of narcolepsy

and hypersomnia compared to those of healthy individuals, this outcome is logical. Algorithms that utilized EEG, EOG, and EMG inputs generally achieved satisfactory kappa values across most recordings, with exceptions primarily occurring in situations where certain sleep stages were either absent or minimally represented. This scenario was common for stage 3 sleep, particularly in Multiple Sleep Latency Test (MSLT) recordings and some patient sleep records.

Discrepancies in classification often arose during transitions between stages or states, a challenging area where even human experts might not always concur. Achieving consensus on a "ground truth" for these transitional periods would likely necessitate multiple expert evaluations of the same recording. Additionally, instances of EEG signals contaminated with ECG artifacts were noted to detract from classification accuracy. Therefore, preprocessing steps to eliminate such ECG artifacts could potentially enhance the overall performance of the sleep stage classification algorithms.

IV. Discussion

Our methods achieved high Cohen's kappa values (around 0.8) for all sleep stages when the training and validation were performed on data from the same group, except for stage 1 (N1), where the kappa was less than 0.5. Stage 1 is generally acknowledged as challenging to score accurately.

The kappa values we obtained were on par with those reported for human experts, emphasizing that stage 1 was the most difficult for both manual and automated scoring, reflecting a common trend of low agreement among scorers.

The performance of our BiLSTM networks was in line with previous studies that applied LightGBM to EEG features and those that utilized LightGBMfor analyzing spectral features of EEG, EOG, and EMG signals. Our LightGBM-LSTM networks also showed comparable results to recent studies employing LightGBM for sleep scoring from single EEG channels and from multiple EEG, EOG, and EMG channels. One study reported a Cohen's kappa of 0.81 across all classes, which is close to our findings, though direct comparisons are cautious as we assessed each sleep stage separately, acknowledging the differing contributions of each to overall sleep architecture.

Although we did not use residual sequence learning specifically, we implemented residual connections and processed different signals as independent inputs, which were then combined for the BiLSTM portion of the network, potentially enhancing performance.

Despite the high accuracy achieved by automatic scoring algorithms, the sleep research community has not yet reached a consensus on their adequacy to fully replace human scoring.

Our study indicates that sleep data can be accurately scored using just a single EEG channel, though slightly improved results were achieved with a combination of 1 EEG, 2 EOG, and 1 EMG channels. The choice of the best method remains inconclusive due to minor performance differences. The addition of channels provides more information but also increases the likelihood of noise interference. Poor EMG signal quality, in particular, was found to negatively impact algorithm performance, echoing findings from other research that explored alternatives to EMG signals to boost scoring accuracy. It was notably remarkable that neural networks could successfully classify sleep stages, especially REM sleep, using only EEG data, a task traditionally challenging without eye movement and muscle tone indicators. This suggests neural networks may be capable of identifying specific EEG patterns indicative of REM sleep.

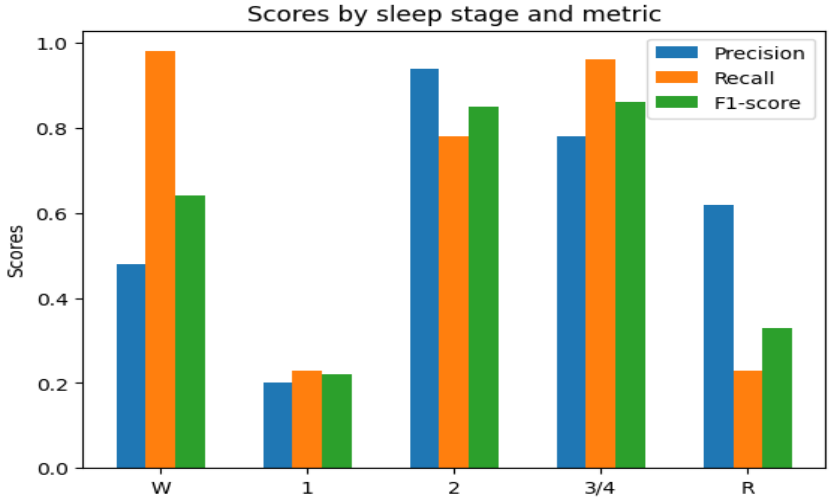


Fig 13: Final identification of the brain behavior state using the proposed system.

Figure 13 visualizes the precision, recall, and F1-score for different sleep stages: W (Wake), 1, 2, 3/4, and R (REM sleep). Here's an interpretation of the data presented:

- Sleep stage W (Wake):

  - Precision is low (around 0.2), indicating that when the model predicts this stage, it is correct about 20% of the time.

  - Recall is very high (nearly 1.0), suggesting that the model is excellent at identifying all the actual instances of this stage.

  - The F1-score, which balances precision and recall, is moderately high (around 0.6), due to the high recall rate.

- Sleep stage 1:

  - Both precision and recall are low (precision slightly below 0.5, recall around 0.2), indicating the model struggles to predict this stage accurately.

  - The F1-score is also low (approximately 0.2), reflecting the poor performance on this stage.

- Sleep stage 2:

  - Precision is high (over 0.9), suggesting that predictions of this stage are very reliable.

  - Recall is moderately high (around 0.8), meaning the model identifies most of the actual instances of this stage.

  - The F1-score is high (around 0.85), showing good model performance on this stage.

- Sleep stage 3/4:

  - Precision is moderately high (around 0.8), indicating reliable predictions for this combined stage.

  - Recall is very high (nearly 1.0), suggesting that the model is excellent at detecting almost all instances of this stage.

  - The F1-score is high (also around 0.85), showing strong performance for this stage.

- Sleep stage R (REM sleep):

  - Precision is moderate (around 0.6), which means when the model predicts REM sleep, it's correct about 60% of the time.

  - Recall is low (around 0.2), indicating the model misses many actual instances of REM sleep.

  - The F1-score is low (around 0.3), showing that the model has difficulty with accurate predictions of REM sleep.

Overall, the model excels in identifying Wake and Sleep stage 3/4, does well with Sleep stage 2, but has significant room for improvement in Sleep stages 1 and REM. The overall accuracy of the model is very high (0.97), which might suggest that the most common sleep stages (likely stage 2 in this case) are predicted very well, bolstering the overall accuracy despite weaker performance in other stages. The macro and weighted averages are approximations and not derived from the individual stage scores, suggesting a generally high performance across all stages which does not align completely with the individual scores shown in the chart. This discrepancy indicates that the macro and weighted averages are not weighted by support (the number of true instances for each class) in this interpretation.

V. **Conclusion**

The exploration into automatic sleep stage classification using machine learning techniques, particularly deep learning models like BiLSTM and LightGBM-LSTM networks, has demonstrated promising results, closely mirroring the accuracy of human experts in many instances. These technologies have shown a strong ability to navigate the complex landscape of sleep data, achieving high Cohen's kappa values across various sleep stages, with the notable exception of stage 1 sleep which remains challenging due to its inherent ambiguity and the lower agreement rates among human scorers.

The inclusion of data from both healthy individuals and patients with conditions like narcolepsy and hypersomnia has enriched the training sets, enhancing the algorithms' robustness and their ability to generalize across different sleep patterns. This approach has particularly improved the performance metrics on patient data, underscoring the importance of diverse training datasets.

Furthermore, the study has underscored the potential of utilizing minimal channel configurations for sleep stage classification, demonstrating that accurate scoring can be achieved even with a single EEG channel, though the incorporation of additional EOG and EMG channels can slightly enhance performance. This finding suggests a scalable flexibility in the deployment of sleep stage classification systems, balancing between minimal hardware requirements and the desire for optimal accuracy.

Despite these advances, the research highlights a few areas for further refinement, such as the need for improved preprocessing to remove artifacts like ECG interference and the exploration of more complex models or training strategies to better capture the nuances of sleep stage transitions.

Moreover, while the automated systems have achieved impressive accuracies, the field has not yet reached a consensus on the sufficiency of these technologies to fully replace traditional human scoring. This ongoing debate suggests a need for continued development and validation of these systems, potentially focusing on hybrid models

that combine the strengths of human expertise with the scalability and efficiency of automated systems.

The study marks a significant step forward in the field of sleep research, offering robust, efficient tools for sleep stage classification that edge closer to the reliability and nuanced understanding of human experts. However, the journey towards fully automated sleep scoring systems that can operate across a wide range of conditions and populations with the same level of trust as human scorers continues, calling for further innovation and interdisciplinary collaboration.

References:

[1] Estrada, E.; Nazeran, H.; Nava, P.; Behbehani, K.; Burk, J.; Lucas, E. Itakura distance: A useful similarity measure between EEG and EOG signals in computer-aided classification of sleep stages. In Proceedings of the 27th IEEE Annual International Conference of Engineering in Medicine and Biology Society, Shanghai, China, 1–4 September 2005; pp. 1189–1192.

[2] Li, Y.; Yingle, F.; Gu, L.; Qinye, T. Sleep stage classification based on EEG Hilbert–Huang transform. In Proceedings of the 4th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 25–27 May 2009; pp. 3676–3681.

[3] Aboalayon, K.A.; Faezipour, M. Multi-class SVM based on sleep stage identification using EEG signal. In Proceedings of the IEEE Healthcare Innovation Conference (HIC), Seattle, WA, USA, 8–10 October 2014; pp. 181–184.

[4] Huang, C.-S.; Lin, C.-L.; Ko, L.-W.; Liu, S.-Y.; Sua, T.-P.; Lin, C.-T. A hierarchical classification system for sleep stage scoring via forehead EEG signals. In Proceedings of the IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), Singapore, 16–19 April 2013; pp. 1–5.

[5] Huang, C.-S.; Lin, C.-L.; Yang, W.-Y.; Ko, L.-W.; Liu, S.-Y.; Lin, C.-T. Applying the fuzzy c-means based dimension reduction to improve the sleep classification system. In Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ), Hyderabad, India, 7–10 July 2013; pp. 1–5.

[6] Lee, Y.-H.; Chen, Y.-S.; Chen, L.-F. Automated sleep staging using single EEG channel for REM sleep deprivation. In Proceedings of the Ninth IEEE International Conference on Bioinformatics and BioEngineering, Taichung, Taiwan, 22–24 June 2009; pp. 439–442.

[7] Hassan, A.R.; Bhuiyan, M.I.H. Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. *Biocybern. Biomed. Eng.* **2016**, *36*, 248–255. [**Google Scholar**] [**CrossRef**]

[8] Khalighi, S.; Sousa, T.; Pires, G.; Nunes, U. Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels. *Expert Syst. Appl.* **2013**, *40*, 7046–7059. [**Google Scholar**] [**CrossRef**][**Green Version**]

[9] Şen, B.; Peker, M.; Çavuşoğlu, A.; Çelebi, F.V. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *J. Med. Syst.* **2014**, *38*, 1–21. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[10] Radha, M.; Garcia-Molina, G.; Poel, M.; Tononi, G. Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal. In Proceedings of the 36th IEEE Annual International Conference of Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 1876–1880.

[11] Hassan, A.R.; Bashar, S.K.; Bhuiyan, M.I.H. On the classification of sleep states by means of statistical and spectral features from single channel electroencephalogram. In Proceedings of the IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India, 10–13 August 2015; pp. 2238–2243.

[12] Hassan, A.R.; Bhuiyan, M.I.H. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. *Biomed. Signal Process. Control* **2016**, *24*, 1–10. [**Google Scholar**] [**CrossRef**]

[13] Ge, J.; Zhou, P.; Zhao, X.; Wang, M. Sample entropy analysis of sleep EEG under different stages. In Proceedings of the IEEE/ICME International Conference on Complex Medical Engineering, Beijing, China, 23–27 May 2007.

[14] Kuo, C.-E.; Liang, S.-F. Automatic stage scoring of single-channel sleep EEG based on multiscale permutation entropy. In Proceedings of the IEEE Conference on Biomedical Circuits and Systems (BioCAS), San Diego, CA, USA, 10–12 November 2011; pp. 448–451.

[15] Liang, S.-F.; Kuo, C.-E.; Hu, Y.-H.; Cheng, Y.-S. A rule-based automatic sleep staging method. In

Proceedings of the 33rd IEEE EMBS Annual International Conference of the Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 6067–6070.

[16] Rodríguez-Sotelo, J.L.; Osorio-Forero, A.; Jiménez-Rodríguez, A.; Cuesta-Frau, D.; Cirugeda-Roldán, E.; Peluffo, D. Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques. *Entropy* **2014**, *16*, 6573–6589. [**Google Scholar**] [**CrossRef**]

[17] Lan, K.-C.; Chang, D.-W.; Kuo, C.-E.; Wei, M.-Z.; Li, Y.-H.; Shaw, F.-Z.; Liang, S.-F. Using off-the-shelf lossy compression for wireless home sleep staging. *J. Neurosci. Methods* **2015**, *246*, 142–152. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[18] Estrada, E.; Nazeran, H.; Ebrahimi, F.; Mikaeili, M. EEG signal features for computer-aided sleep stage detection. In Proceedings of the 4th International IEEE/EMBS Conference on Neural Engineering, Antalya, Turkey, 29 April–2 May 2009; pp. 669–672.

[19] Estrada, E.; Nazeran, H. EEG and HRV signal features for automatic sleep staging and apnea detection. In Proceedings of the 20th IEEE International Conference on Electronics, Communications and Computer (CONIELECOMP), Cholula, Mexico, 22–24 February 2010; pp. 142–147.

[20] Leistedt, S.; Dumont, M.; Lanquart, J.-P.; Jurysta, F.; Linkowski, P. Characterization of the sleep EEG in acutely depressed men using detrended fluctuation analysis. *Neurophysiol. Clin.* **2007**, *118*, 940–950. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[21] Redmond, S.J.; Heneghan, C. Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 485–496. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[22] Alvarez, D.; Hornero, R.; Marcos, J.V.; del Campo, F.; Lopez, M. Spectral analysis of electroencephalogram and oximetric signals in obstructive sleep apnea diagnosis. In Proceedings of the 31st IEEE Annual International Conference of the Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 400–403.

[23] Correa, A.G.; Leber, E.L. An automatic detector of drowsiness based on spectral analysis and wavelet decomposition of EEG records. In Proceedings of the 32nd IEEE EMBS Annual International Conference of the Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 1405–1408.

[24] Gurudath, N.; Riley, H.B. Drowsy driving detection by EEG analysis using wavelet transform and k-means clustering. *Procedia Comput. Sci.* **2014**, *34*, 400–409. [**Google Scholar**] [**CrossRef**]

[25] Correa, A.G.; Orosco, L.; Laciar, E. Automatic detection of drowsiness in EEG records based on multimodal analysis. *Med. Eng. Phys.* **2014**, *36*, 244–249. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[26] Kumari, K. Review on drowsy driving: Becoming dangerous problem. *Int. J. Sci. Res.* **2014**, *3*, 49–51. [**Google Scholar**]

[27] Tsai, P.-Y.; Hu, W.; Kuo, T.B.; Shyu, L.-Y. A portable device for real time drowsiness detection using novel active dry electrode system. In Proceedings of the 31st IEEE EMBS Annual International Conference of the Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 3775–3778.

[28] Yu, S.; Li, P.; Lin, H.; Rohani, E.; Choi, G.; Shao, B.; Wang, Q. Support vector machine based detection of drowsiness using minimum EEG features. In Proceedings of the IEEE International Conference on Social Computing (SocialCom), Alexandria, VA, USA, 8–14 September 2013; pp. 827–835.

[29] Fraiwan, L.; Lweesy, K.; Khasawneh, N.; Fraiwan, M.; Wenz, H.; Dickhaus, H. Time frequency analysis for automated sleep stage identification in fullterm and preterm neonates. *J. Med. Syst.* **2011**, *35*, 693–702. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[30] Fraiwan, L.; Lweesy, K.; Khasawneh, N.; Wenz, H.; Dickhaus, H. Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. *Comput. Methods Progr. Biomed.* **2012**, *108*, 10–19. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[31] Ebrahimi, F.; Mikaili, M.; Estrada, E.; Nazeran, H. Assessment of itakura distance as a valuable feature for computer-aided classification of sleep stages. In Proceedings of the 29th IEEE EMBS Annual International Conference of the Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007; pp. 3300–3303.

[32] Mora, A.M.; Fernandes, C.M.; Herrera, L.J.; Castillo, P.A.; Merelo, J.; Rojas, F.; Rosa, A.C. Sleeping with ants, SVMs, multilayer perceptrons and SOMs. In Proceedings of the 10th International Conference on Intelligent Systems Design and Applications, Cairo, Egypt, 29 November–1 December 2010; pp. 126–131.

[33] Vatankhah, M.; Akbarzadeh-T, M.R.; Moghimi, A. An intelligent system for diagnosing sleep stages using wavelet coefficients. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–5.

[34] Brignol, A.; Al-Ani, T.; Drouot, X. EEG-based automatic sleep-wake classification in humans using short and standard epoch lengths. In Proceedings of the 12th IEEE International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11–13 November 2012; pp. 276–281.

[35] Jo, H.G.; Park, J.Y.; Lee, C.K.; An, S.K.; Yoo, S.K. Genetic fuzzy classifier for sleep stage identification. *Comput. Biol. Med.* **2010**, *40*, 629–634. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[36] Zhu, G.; Li, Y.; Wen, P.P. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1813–1821. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[37] Gudmundsson, S.; Runarsson, T.P.; Sigurdsson, S. Automatic sleep staging using support vector machines with posterior probability estimates. In Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC), Vienna, Austria, 28–30 November 2005; pp. 366–372.

[38] Hsu, Y.-L.; Yang, Y.-T.; Wang, J.-S.; Hsu, C.-Y. Automatic sleep stage recurrent neural classifier using energy features of EEG signals. *Neurocomputing* **2013**, *104*, 105–114. [**Google Scholar**] [**CrossRef**]

[39] Le, Q.K.; Truong, Q.D.K.; Vo, V.T. A tool for analysis and classification of sleep stages. In Proceedings of the IEEE International Conference on Advanced Technologies for Communications (ATC), Da Nang, Vietnam, 2–4 August 2011; pp. 307–310.

[40] Shuyuan, X.; Bei, W.; Jian, Z.; Qunfeng, Z.; Junzhong, Z.; Nakamura, M. An improved k-means clustering algorithm for sleep stages classification. In Proceedings of the 54th IEEE Annual Conference on Society of Instrument and Control Engineers of Japan (SICE), Hangzhou, China, 28–30 July 2015; pp. 1222–1227.

[41] Lajnef, T.; Chaibi, S.; Ruby, P.; Aguera, P.-E.; Eichenlaub, J.-B.; Samet, M.; Kachouri, A.; Jerbi, K. Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines. *J. Neurosci. Methods* **2015**, *250*, 94–105. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[42] Krakovská, A.; Mezeiová, K. Automatic sleep scoring: A search for an optimal combination of measures. *Artif. Intell. Med.* **2011**, *53*, 25–33. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[43] Güneş, S.; Polat, K.; Yosunkaya, Ş. Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting. *Expert Syst. Appl.* **2010**, *37*, 7922–7928. [**Google Scholar**] [**CrossRef**]

[44] Charbonnier, S.; Zoubek, L.; Lesecq, S.; Chapotot, F. Self-evaluated automatic classifier as a decision-support tool for sleep/wake staging. *Comput. Biol. Med.* **2011**, *41*, 380–389. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[45] Koley, B.; Dey, D. An ensemble system for automatic sleep stage classification using single channel EEG signal. *Comput. Biol. Med.* **2012**, *42*, 1186–1195. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[46] Ronzhina, M.; Janoušek, O.; Kolářová, J.; Nováková, M.; Honzík, P.; Provazník, I. Sleep scoring using artificial neural networks. *Sleep Med. Rev.* **2012**, *16*, 251–263. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[47] Zoubek, L.; Charbonnier, S.; Lesecq, S.; Buguet, A.; Chapotot, F. Feature selection for sleep/wake stages classification using data driven methods. *Biomed. Signal Process. Control* **2007**, *2*, 171–179. [**Google Scholar**] [**CrossRef**]

[48] Phan, H.; Do, Q.; Do, T.-L.; Vu, D.-L. Metric learning for automatic sleep stage classification. In Proceedings of the 35th IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 5025–5028.

[49] Kayikcioglu, T.; Maleki, M.; Eroglu, K. Fast and accurate PLS-based classification of EEG sleep using single channel data. *Expert. Syst. Appl.* **2015**, *42*, 7825–7830. [**Google Scholar**] [**CrossRef**]

[50] Sousa, T.; Cruz, A.; Khalighi, S.; Pires, G.; Nunes, U. A two-step automatic sleep stage classification method with dubious range detection. *Comput. Biol. Med.* **2015**, *59*, 42–53. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[51] Khalighi, S.; Sousa, T.; Oliveira, D.; Pires, G.; Nunes, U. Efficient feature selection for sleep staging based on maximal overlap discrete wavelet transform and SVM. In Proceedings of the 33rd IEEE EMBS Annual International Conference of the Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; pp. 3306–3309.

[52] Zhovna, I.; Shallom, I.D. Automatic detection and classification of sleep stages by multichannel EEG signal modeling. In Proceedings of the 30th IEEE EMBS 2008 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 2665–2668.

[53] Yu, S.; Chen, X.; Wang, B.; Wang, X. Automatic sleep stage classification based on ECG and EEG features for day time short nap evaluation. In Proceedings of the 10th World Congress on Intelligent Control and Automation (WCICA), Beijing, China, 6–8 July 2012; pp. 4974–4977.

[54] Ebrahimi, F.; Mikaeili, M.; Estrada, E.; Nazeran, H. Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. In Proceedings of the 30th IEEE EMBS Annual International Conference of the Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 1151–1154.

[55] Fell, J.; Röschke, J.; Mann, K.; Schäffner, C. Discrimination of sleep stages: A comparison between spectral and nonlinear EEG measures. *Electroencephalogr. Clin. Neurophysiol.* **1996**, *98*, 401–410. [**Google Scholar**] [**CrossRef**]

[56] Weiss, B.; Clemens, Z.; Bódizs, R.; Halász, P. Comparison of fractal and power spectral EEG features: Effects of topography and sleep stages. *Brain Res. Bull.* **2011**, *84*, 359–375. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[57] Liu, Y.; Yan, L.; Zeng, B.; Wang, W. Automatic sleep stage scoring using Hilbert–Huang transform with BP neural network. In Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE), Chengdu, China, 18–20 June 2010; pp. 1–4.

[58] Dursun, M.; Gunes, S.; Ozsen, S.; Yosunkaya, S. Comparison of artificial immune clustering with fuzzy c-means clustering in the sleep stage classification problem. In Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications (INISTA), Trabzon, Turkey, 2–4 July 2012; pp. 1–4.

[59] Estrada, E.; Nazeran, H.; Nava, P.; Behbehani, K.; Burk, J.; Lucas, E. EEG feature extraction for classification of sleep stages. In Proceedings of the 26th IEEE EMBS Annual International Conference of the engineering in Medicine and Biology Society, San Francisco, CA, USA, 1–5 September 2004; pp. 196–199.

[60] Obayya, M.; Abou-Chadi, F. Automatic classification of sleep stages using EEG records based on fuzzy c-means (fcm) algorithm. In Proceedings of the 31st National Radio Science Conference (NRSC), Cairo, Egypt, 28–30 April 2014; pp. 265–272.

[61] Ahmad, R.F.; Malik, A.S.; Kamel, N.; Amin, H.; Zafar, R.; Qayyum, A.; Reza, F. Discriminating the different human brain states with EEG signals using fractal dimension: A nonlinear approach. In Proceedings of the IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Kuala Lumpur, Malaysia, 25–25 November 2014; pp. 1–5.

[62] Borzenko, A. Neuron mechanism of human languages. In Proceedings of the International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009; pp. 375–382.

[63] Myelin. Available online: **https://en.wikipedia.org/wiki/Myelin** (accessed on 22 August 2016).

[64] Khalifa, W.; Salem, A.; Roushdy, M.; Revett, K. A survey of EEG based user authentication schemes. In Proceedings of the 8th International Conference on Informatics and Systems (INFOS), Cairo, Egypt, 14–16 May 2012.

[65] Hu, X.-S.; Hong, K.-S.; Ge, S.S. Fnirs-based online deception decoding. *J. Neural Eng.* **2012**, *9*, 26012–26019. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[66] Aqil, M.; Hong, K.-S.; Jeong, M.-Y.; Ge, S.S. Detection of event-related hemodynamic response to neuroactivation by dynamic modeling of brain activity. *Neuroimage* **2012**, *63*, 553–568. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[67] Khan, M.J.; Hong, K.-S. Passive BCI based on drowsiness detection: An fNIRS study. *Biomed. Opt. Express* **2015**, *6*, 4063–4078. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[68] Naseer, N.; Hong, K.-S. FNIRS-based brain-computer interfaces: A review. *Front. Hum. Neurosci.* **2015**, *9*, 3. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[69] Bhutta, M.R.; Hong, M.J.; Kim, Y.-H.; Hong, K.-S. Single-trial lie detection using a combined fNIRS-polygraph system. *Front. Psychol.* **2015**, *6*, 709. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[70] Hong, K.-S.; Naseer, N. Reduction of delay in detecting initial dips from functional near-infrared spectroscopy signals using vector-based phase analysis. *Int. J. Neural. Syst.* **2016**, *26*, 1650012. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[71] Hong, K.-S.; Santosa, H. Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. *Hear. Res.* **2016**, *333*, 157–166. [**Google Scholar**] [**CrossRef**] [**PubMed**]

[72] Rao, R.; Derakhshani, R. A comparison of EEG preprocessing methods using time delay neural networks. In Porceedings of the 2nd IEEE EMBS International Conference on Neural Engineering, Arlington, VA, USA, 16–19 March 2005; pp. 262–264.

[73] Mantri, S.; Patil, V.; Mitkar, R. EEG based emotional distress analysis–a survey. *Int. J. Eng. Dev.* **2012**, *4*, 24–28. [**Google Scholar**]

[74] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[75] Achermann, P., and Tarokh, L. (2014). Human sleep and its regulation. *Kosmos* 2, 173–180.

[76] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[77] Omlin, X., Crivelli, F., Näf, M., Heinicke, L., Skorucak, J., Malafeev, A., et al. (2018). The effect of a slowly rocking bed on sleep. Sci. Rep. 8:2156. doi: 10.1038/s41598-018-19880-3

[78] Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. F. (2007). The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Westchester, IL: American Academy of Sleep Medicine.

[79] Bishop, C. (2016). Pattern Recognition and Machine Learning. New York, NY: Springer-Verlag.

[80] Babadi, B., and Brown, E. N. (2014). A review of multitaper spectral analysis. IEEE Trans. Biomed. Eng. 61, 1555–1564. doi: 10.1109/TBME.2014.2311996

[81] Malafeev, A., Omlin, X., Wierzbicka, A., Wichniak, A., Jernajczyk, W., Riener, R., et al. (2018). Automatic artefact detection in single-channel sleep EEG recordings. J. Sleep Res. doi: 10.1111/jsr.12679 [Epub ahead of print].

[82] Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32. doi: 10.1023/A:1010933404324

[83] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., and Lang, K. J. (1989). Phoneme recognition using time-delay neural networks. IEEE Trans. Acoust. Speech Signal Proc. 37, 328–339. doi: 10.1109/29.21701