International Journal of e-Collaboration
Volume 19 • Issue 3

Feature Extraction From Single-Channel EEG Using Tsfresh and Stacked Ensemble Approach for Sleep Stage Classification

Radhakrishnan B. L., Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

Kirubakaran Ezra, Department of Computer Science and Engineering, GRACE College of Engineering, Thoothukudi, India

Immanuel Johnraja Jebadurai, Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

ABSTRACT

The smart world under Industry 4.0 is witnessing a notable spurt in sleep disorders and sleep-related issues in patients. Artificial intelligence and IoT are taking a giant leap in connecting sleep patients remotely with healthcare providers. The contemporary single-channel-based monitoring devices play a tremendous role in predicting sleep quality and related issues. Handcrafted feature extraction is a time-consuming job in machine learning-based automatic sleep classification. The proposed single-channel work uses Tsfresh to extract features from both the EEG channels (Pz-oz and Fpz-Cz) of the SEDFEx database individually to realise a single-channel EEG. The adopted mRMR feature selection approach selected 55 features from the extracted 787 features. A stacking ensemble classifier achieved 95%, 94%, 91%, and 88% accuracy using stratified 5-fold validation in 2, 3, 4, and 5 class classification employing healthy subjects data. The outcome of the experiments indicates that Tsfresh is an excellent tool to extract standard features from EEG signals.

KEYWORDS

Automated Sleep Scoring, Machine Learning, Remote Sleep Monitoring, Single-Channel EEG, Sleep, Sleep Monitoring, Sleep Stage Classification, Stacking Classifier, Tsfresh

1. INTRODUCTION

Sleep stage classification, otherwise termed sleep scoring or identification, is crucial for treating sleep-related disorders (Wulff, Gatti, Wettstein, & Foster, 2010). Many people with sleep issues are at risk of developing other health problems such as obesity, diabetes, and neuropsychiatric disorders (Kammerer, Mehl, Ludwig, & Lincoln, 2021; Kim, Kang, Choe, & Yoon, 2021; Lu et al., 2021). Polysomnography (PSG) is a medically proven gold standard used to diagnose common sleep disorders (Thorpy, 2017). PSG acquires biosignals from the body like brain signals (electroencephalogram, EEG), movement of the eye (electrooculogram, EOG), heartbeat (electrocardiogram, ECG), and jaw movement, or limb muscular activity (electromyogram, EMG). The trained sleep scoring experts examine the recorded signals and assign a sleep stage to each 30-second PSG data termed as an

DOI: 10.4018/IJeC.316774

epoch. The sleep technicians either follow the guidelines of Rechtschaffen and Kales (Allan Hobson, 1969) or the American Academy of Sleep Medicine (AASM) (Richard B. Berry, MD; Rita Brooks, MEd, RST, RPSGT; Charlene E. Gamaldo & Susan M. Harding, MD; Robin M. Lloyd, 2016). In the manual approach, sleep experts have to visually evaluate epoch by epoch and label sleep stages to build a hypnogram indicating corresponding sleep stages. Sleep stages are defined as wake-state (W), rapid eye movement (REM), and non-REM (NREM). According to AASM, NREM comprises three stages N1, N2, and SWS (slow-wave sleep). Manual sleep scoring has drawbacks like time-consuming, labour-intensive, the requirement of highly trained sleep technicians, inter-rater variability and occasionally subjective (Cesari et al., 2021; Stepnowsky, Levendowski, Popovic, Ayappa, & Rapoport, 2013). Furthermore, overnight PSG studies have limitations such as expensive, big waiting lists in the clinics, unfamiliar sleeping environments, restricted privacy, skin irritation due to the adhesive from the electrode and multiple sensors connected to the person may obstruct sleep, lowering the recording accuracy (Zhang et al., 2021). Developing a robust and convenient automated sleep stage classification system could be highly beneficial to overcome the limitations mentioned above. The role of EEG signals is significant in identifying specific sleep stages among all the distinct signals recorded by PSG, either in manual or automated scoring methods (Kwon, Kim, & Yeo, 2021). Recent technological improvements play a significant role in designing and developing reliable in-home based automatic sleep stage classification (ASSC) systems (Kwon et al., 2021). This study proposes and uses a single-channel EEG to classify different sleep stages to develop a practical in-home sleep system, as prior studies have shown (Ghimatgar, Kazemi, Helfroush, & Aarabi, 2019a; Hassan & Bhuiyan, 2017a). Thus far, single-channel EEG based sleep classification approaches, particularly machine learning algorithms, have been widely researched.

Typical machine learning algorithms follow a traditional procedure of preprocessing, feature extraction, and selection of features before passing the data to the classifier. The major ASSC works in the literature rely on handcrafted features from the domain such as time, frequency and time-frequency (Boostani, Karimzadeh, & Nami, 2017a). In the time domain, major features obtained includes statistical features (mean, variance, standard deviation, skewness, kurtosis, etc.), Hjorth parameters, threshold and zero-crossing rate. In the frequency domain, frequently extracted features comprise spectral estimation, parametric methods (autoregressive, moving-average, autoregressive moving-average), non-parametric approaches (power spectral-density) and higher-order spectra. The most important features extracted from the time-frequency domain includes entropy and complexity based (Renyi's, Tsallis, permutation, Lempel–Ziv, multi-scale, approximate etc.), and fractal-based (correlation dimension, Lyapunov exponent, Hurst exponent, Petrosian, Higuchi etc.) (Boostani et al., 2017a), (Zhao, 2019).

Feature extraction from the EEG signal is an intensive process for people working in this domain. They have to employ multiple methods, algorithms and signal decomposition techniques to extract the most relevant features from signal domains such as time-domain (TD), frequency domain (FD) and time-frequency domain (TFD). Therefore, deciding the features to extract and select the useful features for the machine learning classifier is a cumbersome job. The authors attempt to test the effectiveness of automatic feature extraction python package Tsfresh (Christ, Braun, Neuffer, & Kempa-Liehr, 2018) in sleep stage classification; it is a popular feature extraction package for time-series data. This research strives to test the effectiveness of Tsfresh in extracting features for a machine learning-based automated sleep classification system employing a single-channel EEG. The following are the objectives of this research:

- Use Tsfresh to extract all possible features from a single-channel EEG.
- Test the significance of the feature selection method in Tsfresh and compare it with the standard method used in the EEG domain.
- Design a stacking classifier to classify the sleep stage using the selected features.

This system can be implemented in portable in-home sleep monitoring devices. The remainder of this article is organised as follows. Background section details the related works. The materials procedure section lists the details about the data, preprocessing, feature engineering and classification. The results and discussion section discusses the experiment results and compares them with the existing works. Finally, the conclusion section discusses the remarks about the work done and its advantages.

2. BACKGROUND

Monitoring sleep at home using single-channel EEG at user convenience is a hot research area. In home-based sleep monitoring, machine learning techniques play a significant role in automated sleep stage classification (Pan, Brulin, & Campo, 2020). Generally, EEG signals are non-stationary; feature extraction is the key step in any ASSC system that uses EEG signals. The performance of the AASC system that uses machine learning to classify sleep stages depends on the significant features extracted from the signals (Zhao, 2019). (Hassan, Bashar, & Bhuiyan, 2016) studied the application of spectral features in sleep stage classification. The authors have extracted flatness, spectral roll-off, spread, centroid, slope and degree features for every 30-second epoch. Then the extracted feature vector is utilised for training the boosted decision tree classifier, producing 82% accuracy in 5-class classification. Naturally, distinct frequency bands exist in the EEG signal. Signal decomposition filters or extracts the specific signal component from a composite signal. (Hassan & Bhuiyan, 2017a), decomposed the EEG signals utilising ensemble empirical mode decomposition (EEMD) and extracted moment-based features. The extracted features are passed to the RUboost classifier and attained the classification accuracy of 83.5% in a 5-class classification. A multimodel signal decomposition technique proposed by Jiang et al. decomposed the signals in the frequency domain as (delta (δ), alpha (α), theta (θ), beta (β), K-complex and spindle) and time-domain using (empirical mode decomposition (EMD)). Then, statistical and fractal features are extracted from the decomposed signals. In order to improve the classification outcome, a random forest (RF) algorithm is coupled with Hidden Markov (HM) method. This method achieved 86.9% accuracy in 5-class classification (JIANG, LU, MA, & WANG, 2019). ASSC work presented by (da Silveira, Kozakevicius, & Rodrigues, 2017) decomposed the EEG signals into five subbands utilising discrete wavelet transform (DWT). From every subband, statistical features are extracted, and using an RF classifier reached 91.5% accuracy in 5-class classification. In (Ilhan, 2017), the decomposed the Fpz-Cz EEG channel signals into sub-bands such as delta (δ), alpha (α), beta (β), theta (θ), saw-tooth, spindle, and K-complex. Energy features are extracted and provided to the ensemble classifier from these frequency bands. This method has achieved an accuracy of 92.81% for healthy people and 80.39% for people with sleep-related problems using a 6-class classification. Researchers employed the Tunable-Q factor wavelet-transform (TQWT) to process oscillatory signals, a data-adaptive flexible signal decomposition method. In (Hassan & Subasi, 2017), the author used TQWT to extract six sub-bands. From every subband, statistical moment characteristics are retrieved, fed into the Bagging classifier, which achieved an accuracy of 94.3% in 5-class classification. (Hassan & Bhuiyan, 2016) presented an ASSC system using an RF classifier. TQWT decomposes the EEG signal and extracts spectral features from every sub-band in their approach. Using the extracted features RF classifier produced a 91.50% accuracy in the 5-class classification. Decomposing the EEG signals as delta (δ), alpha (α), theta (θ), beta (β), gamma (γ) and sigma (σ) subbands and retrieving features from each subband is another approach used by the researchers (Memar & Faradji, 2018). In (Seifpour, Niknazar, Mikaeili, & Nasrabadi, 2018), the authors extracted local extrema features from the time domain. Additionally, statistical and pattern features are extracted from the six subbands (low and high delta, theta, alpha, sigma and beta). Using an SVM-based classifier, they achieved 88.6% and 90.2% accuracy in five-class and six-class sleep stage classification.

In automatic sleep stage classification (ASSC) systems, significant steps are signal preprocessing, extracting features, and classification. However, sleep stage classification accuracy depends on the precise analysis of EEG signals in ASSC systems. Extracting features from an EEG signal is a time-consuming process as well as non-trivial (Zhao, 2019). Automatic feature extraction tools for assisting in this job are crucial. Tsfresh (Christ et al., 2018) is a python based time series feature extraction library, tested for features extraction in EEG signals and any time-series data. In (Lazarevich, Prokin, & Gutkin, 2018), the authors demonstrated that automatic feature extraction using Tsfresh provides on par baseline performance of standard deep learning-based models' used in neural decoding problems. An EEG based epileptic seizures detection system proposed by (Wu, Zhou, & Li, 2020), used extracted features from the intrinsic-mode frequencies (IMFs) decomposed by EMD using Tsfresh and pyEntropy (a python based library for extracting entropy features). XGBoost classifier achieved enhanced performance in terms of accuracy using the above-extracted features. Another work done by (Grossi, Valbusa, & Buscema, 2021) used the Tsfresh package to extract the features to study the Autism signatures from the EEG signals for machine learning study. The extracted features using Tsfresh helps to achieve promising results in detecting Autism symptoms.

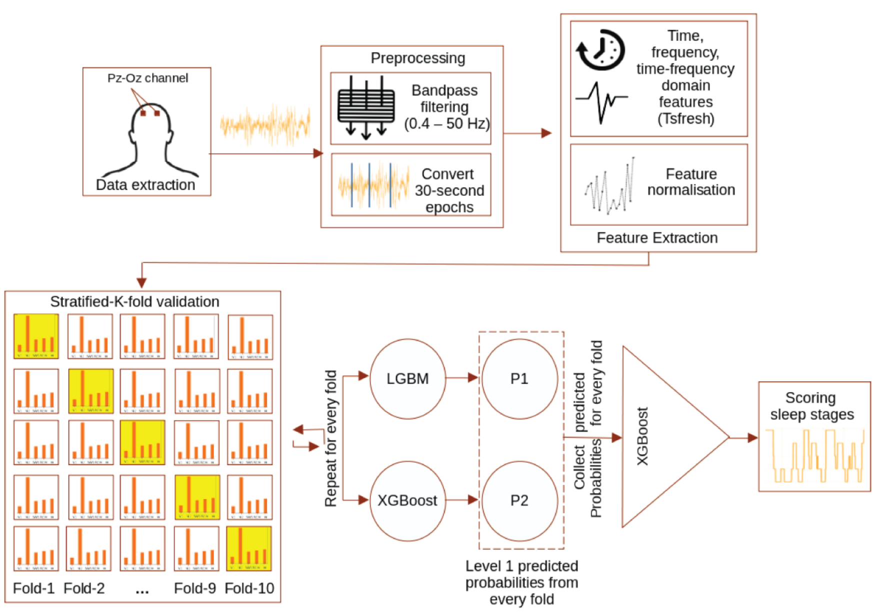
3. MATERIALS AND PROCEDURES

Using Tsfresh and the ensemble learning method, this research proposes a trustworthy automated sleep staging method for devices employing single-channel EEG for sleep monitoring. Figure1 depicts the proposed system architecture in schematic form. The proposed work is implemented using MNE, a special python library designed for handling and processing neurophysiological data (Gramfort et al., 2013).

3.1 Data Acquisition

The EEG data utilised in this study's experiments came from the public SEDFEx (Sleep-EDF extended) dataset. Two experiments, sleep cassette (SC) along with sleep telemetry (ST), recorded the whole night sleep of volunteers who participated in the sleep (SEDFEx) study. SC experiment recorded

Figure 1. Proposed stacked ensemble architecture for sleep stage classification



participants' sleep with no sleep disorders, yielding 153 recordings. ST experiment recorded the effects of Temazepam in sleep, yielding 44 recordings. From the above experiments, this database has 197 PSG (polySomnographic) recordings possessing data from the sensors such as EEG (from Pz-Oz and Fpz-Cz electrodes), EOG (horizontal), event marker and submental chin EMG. The EEG and EOG sensors sampling rate is fixed as 100 Hz, and the remaining sensors are fixed as 1 Hz. The sleep technicians manually scored whole sleep stages using the R&K method and stored them in the *Hypnogram file (Goldberger et al., 2000). The sleep stages stored in the hypnogram comprises sleep stages wake (w), rapid eye movement (REM), non-rapid eye movement (NREM) stages (N1, N2, N3, and N4), Movement (M), and ungraded epochs represented using a question mark (?). Based on AASM procedures, the NREM stages N3 and N4 fused as slow-wave sleep (SWS) in this work. This work used PSG recordings of ten healthy subjects' from the SC experiment and twenty PSG recordings with sleep difficulties. The participants who had sleep difficulties took a placebo or Temazepam during the sleep recording. The authors have categorised the ST recordings based on the placebo or temazepam intake and analysed them separately. The list of PSG recordings utilised for the experiments is presented in Table 1.

3.2 Data Pre-Processing

The proposed work utilised single EEG channel data from the SEDFEx database. Pz-oz has been more effective in producing good results in ML-based systems when compared with the Fpz-Cz channel. The proposed model is tested in both the channels individually. Many single-channel EEG works either uses Pz-Oz or Fpz-Cz channel from the SEDFEx database (Hassan & Bhuiyan, 2017a; da Silveira et al., 2017). In EEG signals, artefacts such as eye movement, the presence of gamma signals during waking and sleep are pretty common. Both substantially impact the identification of sleep stages. Hence, the EEG signals separated from all the chosen PSG recordings. Next, the signals are filtered using a bandpass filter with a hamming window using the lower and upper ranges of 0.5 and 49.5 Hz (Memar & Faradji, 2018). The Pz-oz and Fpz-Cz signal before and after filtering are presented in Figure 2. After filtering, the signals are converted into 30-s epochs without any overlapping. Then, the stages N3 and N4 have been merged as slow-wave sleep (SWS) based on AASM procedures for all the EEG recordings used in the experiment. The distribution of sleep stages from the used PSG recordings are presented in Table 2. The share of the N1 stage is quite low, and N2 accounts high when compared with the other sleep stages. As a consequence, it leads to severe data imbalance complications. Hence, adequate attention to be given to address the data imbalance complication while designing an ASSC system. The MNE-python package is used for the implementation of complete PSG data pre-processing (Gramfort et al., 2013).

3.3 Feature Extraction and Selection

For modelling non-stationary and nonlinear EEG signals, feature extraction is critical. Previous research has shown that extracting characteristics from domains, such as time, frequency, and time-

Table 1. List of PSG recordings utilised for the experiments

Sleep Cassette Experiment	Sleep Telemetry Experiment	
	Placebo intake	Temazepam intake
SC4001E0, SC4051E0, SC4131E0, SC4002E0, SC4061E0, SC4182E0, SC4012E0, SC4011E0, SC4131E0, SC4031E0, SC4012E0	ST7011J0, ST7041J0, ST7061J0, ST7071J0, ST7101J0, ST7121J0, ST7132J0, ST7141J0, ST7151J0, ST7162J0	ST7021J0, ST7051J0, ST7081J0, ST7091J0, ST7111J0, ST7131J0, ST7142J0, ST7152J0, ST7161J0, ST7172J0

Figure 2. 10 seconds EEG signal (before and after filtering) using SC4012E0 PSG recording

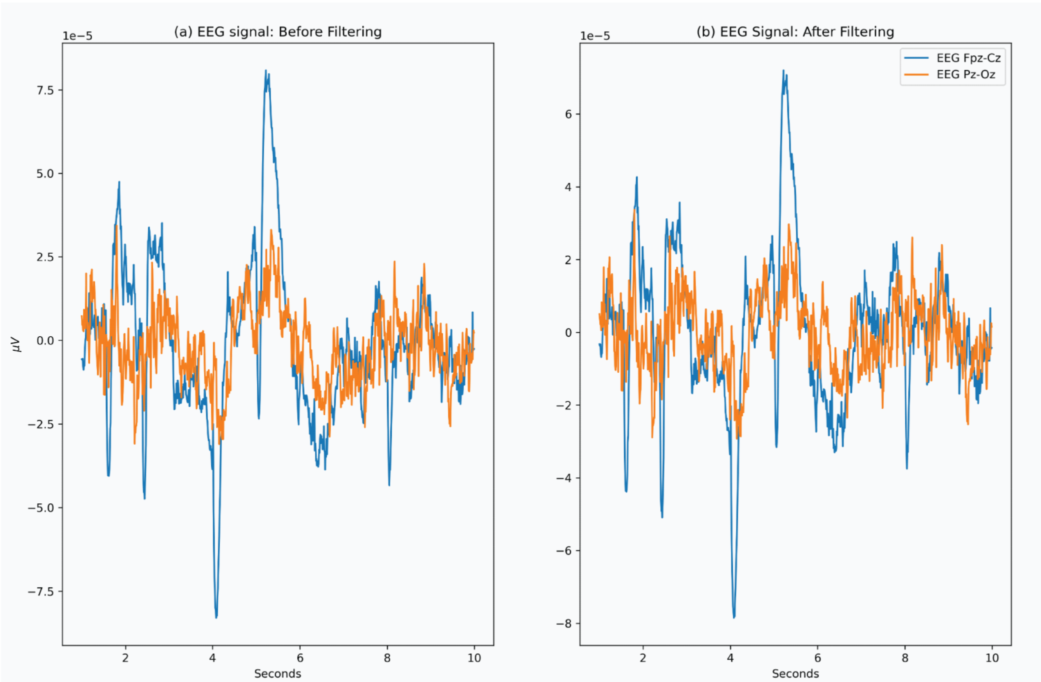


Table 2. Distribution of sleep stages from the PSG data used from the SEDFx database

Type of data	Wake	Light Sleep		Deep Sleep	REM Sleep
	W	N1	N2	SWS	R
Healthy	809	408	2195	755	760
Placebo (intake) *	579	401	2042	833	915
Temazepam (intake)*	409	399	2124	930	970

* Participants were facing a mild strenuous in falling asleep

frequency is effective for ASSC systems; that enhances the sleep stage detection potential of classifiers (Boostani et al., 2017a), (Zhao, 2019). Features are extracted from every 30-s epoch using Tsfresh (Christ et al., 2018)an open-source python based feature extraction package. A 30-second epoch with 3000 data points at a sampling rate of 100 Hz is denoted as $X\left(m\right)$. After epoching, the EEG signal, it becomes a sequence of epochs collected over a period of time.

The set of epochs $S = \left\{X_j\right\}_{j=1}^{M}$ is the input for the employed machine learning methods. Each epochs X_j mapped into features with a dimension K that is relevant to the problem $\overrightarrow{x_j} = \left(x_{j,1}, x_{j,2}, \ldots, x_{j,K}\right)$. In general, mapping epochs (sequence of signals) set S into feature matrix of rows (M) and columns (N) by taking K epochs from all X_j as an elements of $\overrightarrow{x_j}$. Hence, $\overrightarrow{x_j}$ is created by performing signal characterisation techniques $F_i : X_j \rightarrow x_{j,k}$ to a corresponding epoch from X_j that produces $\overrightarrow{x_j} = \left[F_1\left(x_j\right), F_2\left(x_j\right), \ldots, F_M\left(x_j\right)\right]$. Tsfresh is used to carry out the feature extraction process. The authors have employed the "MultiprocessingDistributor" method from Tsfresh to speed up the feature extraction process, yielding 787 features from time, frequency and time-frequency domains.

High dimensional data analysis such as EEG is a challenge since the overall performance of the machine learning algorithm employed in the ASSC system is strongly dependent on data dimensions. The feature selection process facilitates reducing the dimension of the data. The identification of all highly and weakly relevant attributes is the all-relevant problem of feature selection. This challenge is tough to tackle in applications like ASSC systems since each label, or target variable is linked to many time series features simultaneously. In this work, the authors have tested two approaches for feature selection; the fresh feature selection algorithm from Tsfresh and the minimum-redundancy and maximum-relevance (mRMR) feature selection algorithm (Peng, Long, & Ding, 2005).

The "fresh" algorithm is used by tsfresh to reduce the number of irrelevant features. It evaluates each feature vector individually and independently for its significance in predicting the target. The number of target classes used in this work is five (W, N1, N2, SWS, and REM) according to AASM guidance. Let us consider five stages of sleep as $y = \left(y_1, \ldots, y_5\right)$. When considering feature X, its relevancy to detect the correct sleep stage $y \in \left(y_1, \ldots, y_i\right)$ is estimated by employing a hypothesis testing. For each x_j Tsfresh employs a statistical test to examine the hypothesis. The outcome of every hypothesis test is calculating the p-value. A small p-value signifies the importance of x_j in predicting the y. Tsfresh returned a total of 280 selected features after feature significance tests p-values based on Benjamini-Yekutieli approach. The subset of extracted features are listed in Table 3.

The MRMR is a solution to the redundancy problem. It chooses a subset of features $x_s \subseteq \overrightarrow{x_j}$ from the list of features by maximising each feature's relevance to the target class and minimising redundancy between the features chosen. In our experiments, finally, MRMR is used in the implementation. It selected 55 features out of 787 features, the list of selected features is given in Table 4. For detailed information on MRMR refer to the original research (Peng et al., 2005).

3.4 Classification Model and Performance

In ASSC systems, classifiers play an indispensable role in classifying an epoch's evoked features and assigning a specific sleep stage to an epoch. The classifiers are trained to create a linear/non-linear

Table 3. A subset of features extracted from both Pz-oz and Fpz-Cz channels

Domain	Category	Feature
Time	Statistics	Maximum, median, standard deviation, number of peaks, quantiles, number of crossing, variance, skewness, root mean square
	Autocorrelation	Different lags, partial autocorrelation
	Linear trend intercept	Intercept, standard error, rvalue, slope
	Energy	Ratio, Absolute
	Auto regressive	Auto regressive coefficients
Frequency	Fourier Transform	Coefficient, aggregate
	Power	Welch density
Time-frequency	Complexity	Lempel ziv complexity
	Entropy	Sample entropy, approximate entropy, permutation entropy

Table 4. Features selected by MRMR

#	Feature	#	Feature
1	EEG_ar_coefficient__coeff_2__k_10	29	EEG_agg_linear_trend__attr_"stderr"__chunk_len_10__f_agg_"max"
2	EEG_fft_coefficient__attr_"abs"__coeff_24	30	EEG_ar_coefficient__coeff_1__k_10
3	EEG_ar_coefficient__coeff_5__k_10	31	EEG_change_quantiles__f_agg_"mean"__isabs_True__qh_0.4__ql_0.2
4	EEG_quantile__q_0.4	32	EEG_standard_deviation
5	EEG_ar_coefficient__coeff_3__k_10	33	EEG_ar_coefficient__coeff_8__k_10
6	EEG_quantile__q_0.3	34	EEG_root_mean_square
7	EEG_permutation_entropy__dimension_3__tau_1	35	EEG_fft_aggregated__aggtype_"variance"
8	EEG_quantile__q_0.2	36	EEG_linear_trend__attr_"stderr"
9	EEG_partial_autocorrelation__lag_3	37	EEG_agg_autocorrelation__f_agg_"mean"__maxlag_40
10	EEG_change_quantiles__f_agg_"mean"__isabs_True__qh_0.6__ql_0.4	38	EEG_agg_linear_trend__attr_"stderr"__chunk_len_5__f_agg_"min"
11	EEG_partial_autocorrelation__lag_4	39	EEG_agg_linear_trend__attr_"stderr"__chunk_len_10__f_agg_"min"
12	EEG_permutation_entropy__dimension_4__tau_1	40	EEG_quantile__q_0.6
13	EEG_quantile__q_0.8	41	EEG_agg_linear_trend__attr_"intercept"__chunk_len_50__f_agg_"max"
14	EEG_ratio_beyond_r_sigma__r_2	42	EEG_fourier_entropy__bins_2
15	EEG_quantile__q_0.7	43	EEG_ratio_beyond_r_sigma__r_2.5
16	EEG_quantile__q_0.9	44	EEG_change_quantiles__f_agg_"var"__isabs_True__qh_0.6__ql_0.4
17	EEG_number_peaks__n_1	45	EEG_change_quantiles__f_agg_"var"__isabs_False__qh_0.6__ql_0.4
18	EEG_range_count__max_1__min_-1	46	EEG_number_peaks__n_10
19	EEG_partial_autocorrelation__lag_6	47	EEG_augmented_dickey_fuller__attr_"usedlag"__autolag_"AIC"
20	EEG_permutation_entropy__dimension_5__tau_1	48	EEG_large_standard_deviation__r_0.15000000000000002
21	EEG_agg_linear_trend__attr_"stderr"__chunk_len_10__f_agg_"mean"	49	EEG_agg_linear_trend__attr_"intercept"__chunk_len_50__f_agg_"min"
22	EEG_change_quantiles__f_agg_"mean"__isabs_True__qh_0.8__ql_0.6	50	EEG_change_quantiles__f_agg_"mean"__isabs_True__qh_0.8__ql_0.4
23	EEG_permutation_entropy__dimension_6__tau_1	51	EEG_agg_linear_trend__attr_"stderr"__chunk_len_50__f_agg_"mean"
24	EEG_agg_linear_trend__attr_"stderr"__chunk_len_5__f_agg_"max"	52	EEG_agg_autocorrelation__f_agg_"var"__maxlag_40
25	EEG_quantile__q_0.1	53	EEG_ratio_beyond_r_sigma__r_0.5
26	EEG_permutation_entropy__dimension_7__tau_1	54	EEG_change_quantiles__f_agg_"mean"__isabs_True__qh_0.6__ql_0.2
27	EEG_agg_linear_trend__attr_"stderr"__chunk_len_5__f_agg_"mean"	55	EEG_partial_autocorrelation__lag_2
28	EEG_agg_autocorrelation__f_agg_"median"__maxlag_40		

boundary among the feature vectors of different sleep stages. Ensemble learning is adopted across distinct domains for better prediction performance. The stacking technique learns and combines the predictions of good performing machine learning models to achieve improved results using a meta classifier (Freiberger et al., 2020). The idea behind the stacked model is to train distinct base classifiers and amalgamate them to train a meta-classifier. The meta-classifier predict the sleep stages by utilising the returned predictions of base classifiers.

As presented in Table 2, the N1 stage is relatively low, and N2 accounts for high compared with the other sleep stages. Indeed, it is an unbalanced classification issue in ASSC; this may misguide the classifier. Thus, using the python SMOTE package, this work employs a "minority oversampling" approach to promote the less repressed classes equal to the majority class. Oversampling the whole dataset may lead to overfitting the classifier. Hence, only the training data is oversampled before passing it to the classifier (Fernández, García, Herrera, & Chawla, 2018).

From the raw single-channel EEG, a significant number of features are extracted using Tsfresh. Classification of high-dimensional feature sets impacts the classifier's classification performance due to the vast amount of feature sets. Furthermore, the features set may contain some irrelevant characteristics, lowering classification performance and raising processing costs. In classifier training, using optimal features assists in achieving good classification results. Hence, using the mRMR algorithm, the redundant and less important features are removed from the feature set (Peng et al., 2005).

In the proposed stacking classifier architecture, all the base classifiers are first trained and evaluated in the dataset with default hyperparameters using a stratified K-fold-validation (k=5). Then, hyper-parameters of the stacking classifier's base and meta algorithms are tuned using a grid search technique. Finally, based on performance, the LGBM, XGBoost are picked as base classifiers, and XGBoost is chosen as meta-classifier. The meta-classifier takes the predicted probabilities of base classifiers as input parameters to classify the sleep stages. The authors have tested Multi-layer Perceptron, K-Nearest Neighbor, Support Vector Machines, Random Forest, XGBoost, Light Gradient Boosting Machine (LGBM) and Nu-Support Vector Classification for the stack and tried different stack combinations.

In this work, stratified k-fold-validation ($CV_{accuracy}$) is employed to evaluate the classification performance accurately and reduce potential bias (Boostani, Karimzadeh, & Nami, 2017b). The classification task is clearly an imbalanced data classification when referring to Table 1. This leads the classifier to end up classifying towards the of majority class data. Hence, the data need to be balanced. To avoid overfitting, the training data is only balanced during each fold. In general, the statistical measurements accuracy, precision, f1-score, recall and kappa are used to access the detection performance:

$$\mathrm{CV}_{\mathrm{accuracy}} = \frac{1}{K}\sum_{i=1}^{K} Ac_i$$

Here, K correspond to the count of k-folds, and Ac_i is the measured accuracy of each fold. The ratio of true positives to all positives is known as precision. Concerning sleep stage classification, precision is the percentage of a specific sleep stage correctly identified against other sleep stages. It is represented as:

$$CV_{precision} = \frac{Truepositive}{Actualclass}$$

Precision also provides the measure of pertinent data points. It is very essential that a person shouldn't be suspected as a patient who really doesn't have a sleep disorder. Recall provides the metric of how well our classifier detects true positives. As a result, recall tells us the count of accurately identified sleep stages. It is represented as:

$$CV_{recall} = \frac{Truepositive}{sum\ \ of\ \ actual\ \ positive}$$

Recall also provides the metric of how accurately the classifier identifies the relevant sleep stage. The F1-score brings the harmonic mean of a stacking classifier's precision as well as a recall to create a single statistic. It is mostly employed to compare the outcomes of two distinct classifiers. Assume that classifier A has a higher recall and precision than classifier B. The F1-scores of classifiers are utilised to identify superiority among them:

$$CV_{f1\text{-}score} = \frac{2x\left(CV_{precision}\ xCV_{recall}\right)}{CV_{precision} + CV_{recall}}$$

Metrics like accuracy or precision/recall don't provide us with a whole performance picture of the classifiers in a multi-class classification context. Cohen's kappa $\left(k\right)$ statistic is a great way to deal with multi-class and imbalanced-class problems (Boostani et al., 2017b):

$$k = \frac{p_o - p_e}{1 - p_e}$$

Here, p_o signifies the detected agreement, whereas p_e signifies the anticipated agreement. It basically tells you how well a classification algorithm performs in comparison to a classifier that guesses at random based on the frequency of each class. Kappa can never be more than or less than one. According to the Landis et al. technique, a score of 0 indicates no agreement, a score of 0.21–0.40 signifies fair, a score of 0.41–0.60 signifies moderate, a score of 0.61–0.80 signifies considerable, and a score of 0.81–1 signifies practically ideal agreement.

4. RESULTS AND DISCUSSIONS

The SEDFx database is utilised to assess the performance of the proposed stacked ensemble classifier and compare the results to the current literature. In order to realise a single-channel EEG setup, both EEG channel (Fpz-Cz/Pz-oz) data from PSG recordings have been experimented with separately for the experimental assessment. The configuration of the computer system used for the experiment is Dell R730 with Intel(R) Xeon CPU 2.60 GHz x 32GB of RAM. Different base and meta classifier combinations are evaluated; our proposed combination performed very well in sleep stage classification. The meta classifier takes the predicted probabilities of base classifiers as input parameters to classify the sleep stages. In this work, the authors conducted three different experiments using healthy, placebo intake, and temazepam intake EEG recordings listed in Table 1. All three experiments evaluated the performance of Pz-oz and Fpz-Cz EEG channels individually by employing the three category recordings detailed in Table 1. Each experiment tested the stacking classifier's performance on 5-class (wake, N1, N2, SWS and REM sleep), 4-class (wake, light-sleep, deep-sleep and REM sleep), 3-class (wake, light-sleep and deep-sleep) and 2-class (wake and sleep) classifications. Except for 2-class, multi-class classification is the objective of the classifier and 2-class classification is treated as binary classification.

4.1 Experiment 1: Healthy PSG Recordings

In this experiment, ten healthy recordings mentioned in Table 1 are used. First, the features are extracted separately from the Pz-Oz channel and Fpz-Cz channel. Next, using the mRMR feature selection method, 55 features are selected from the list of extracted features and listed in Table 4. The selected features become the input to the stacking classifier and are tested for distinct classes of classification. In a 5-class classification, Table 5 presents the classification results, Figure-3 depicts the mean confusion matrix, and Figure 4 exhibits the learning curve of the stacking classifier. Table 6 summarises the stacking classifier's performance as mean accuracy, f1-score, precision, recall and kappa in 4, 3, and 2 class classification.

4.2 Experiment 2: Placebo Intake PSG Recordings

In experiment 2, ten placebo intake recordings mentioned in Table 1 are used. First, the features are extracted separately from the Pz-Oz channel and Fpz-Cz channel. Next, using the mRMR feature selection method, 55 features are selected from the list of extracted features and listed in Table 4. The selected features become the input to the stacking classifier and are tested for distinct classes of classification. In a 5-class classification, Table 7 presents the classification results, Figure 5 depicts the mean confusion matrix, and Figure 6 exhibits the learning curve of the stacking classifier. Table 8 presents the stacking classifier's performance as mean accuracy, f1-score, precision, recall and kappa in 4, 3, and 2 class classification.

4.3 Experiment 3: Temazepam Intake PSG Recordings

In experiment 3, ten temazepam intake recordings mentioned in Table 1 are used. First, the features are extracted separately from the Pz-Oz channel and Fpz-Cz channel. Next, using the mRMR feature selection method, 55 features are selected from the list of extracted features and listed in Table 4. The selected features become the input to the stacking classifier and are tested for distinct classes of classification. In a 5-class classification, Table 9 presents the classification results, Figure 7 depicts the mean confusion matrix, and Figure 8 portrays the learning curve of the stacking classifier. Table

Table 5. 5-Class classification performance of stacking classifier using stratified-5-Fold-CV

Data	Channel	Fold	Accuracy	F1-score	Precision	Recall	Kappa
Sleep-EDFx (healthy recordings)	Pz-Oz	0	0.88	0.83	0.84	0.83	0.83
		1	0.88	0.84	0.84	0.83	0.83
		2	0.85	0.81	0.81	0.81	0.80
		3	0.88	0.83	0.84	0.83	0.83
		4	0.87	0.83	0.82	0.83	0.82
		Mean	0.87	0.83	0.83	0.83	0.82
	Fpz-Cz	0	0.90	0.85	0.86	0.85	0.86
		1	0.86	0.81	0.81	0.81	0.81
		2	0.88	0.83	0.83	0.83	0.83
		3	0.88	0.83	0.84	0.83	0.83
		4	0.90	0.86	0.87	0.86	0.86
		Mean	0.88	0.84	0.84	0.84	0.84

Figure 3. Stacking classifier mean confusion matrix using stratified-5-Fold CV

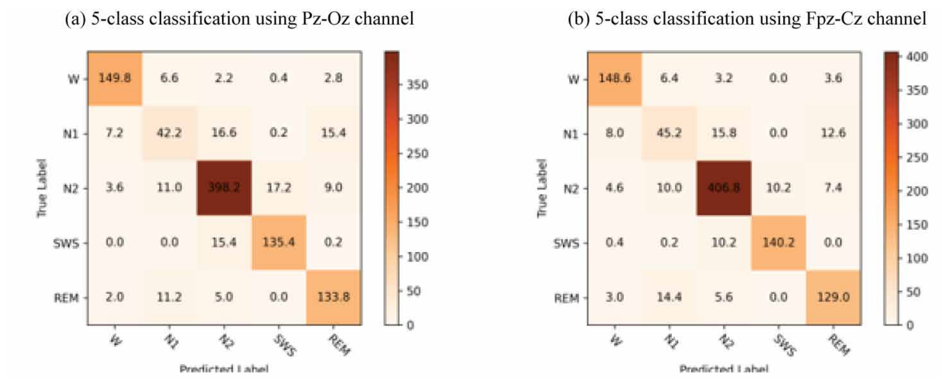


Figure 4. Learning curve of stacking classifier

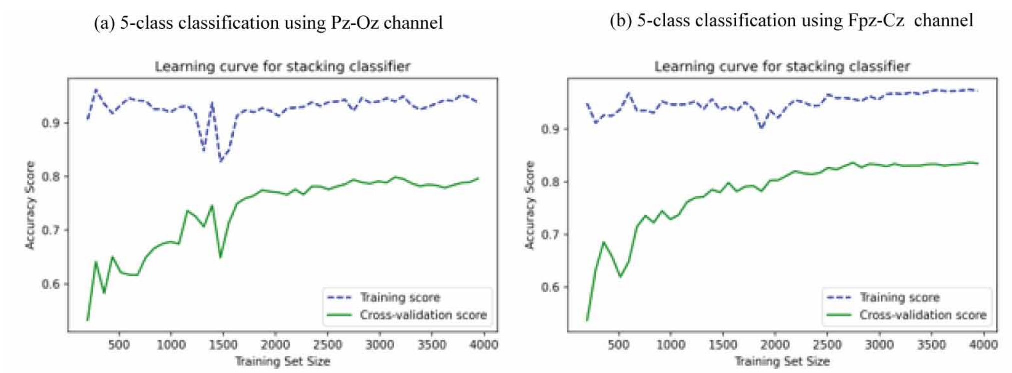


Table 6. Comparison of Stacking classifier performance on different target classes using stratified-5-Fold-CV

Data	Channel	Classifiers	Accuracy*	F1-score*	Precision*	Recall*	Kappa*
Sleep-EDFx (healthy recordings)	Pz-Oz	4-Class	0.90	0.89	0.89	0.90	0.85
		3-Class	0.90	0.89	0.89	0.89	0.84
		2-Class	0.91	0.90	0.90	0.90	0.86
	Fpz-Cz	4-Class	0.91	0.90	0.90	0.90	0.86
		3-Class	0.94	0.94	0.94	0.94	0.93
		2-Class	0.95	0.95	0.95	0.95	0.94

* Mean score

10 presents the stacking classifier's performance as mean accuracy, f1-score, precision, recall and kappa in 4, 3, and 2 class classification.

In experiment 1, as indicated in Table 5, for a 5-class classification, the proposed stacked ensemble classifier produced the mean, minimum and maximum accuracy of 87%, 85% and 88% employing the Pz-Oz channel. Similarly, using the Fpz-Cz channel achieved a mean, minimum, and maximum accuracy of 88%, 86%, and 90%. Both the channels produced improved results when compared with the existing ones (Ghimatgar et al., 2019a; Hassan, 2015; Hassan & Bhuiyan, 2017b). Interestingly,

Table 7. 5-Class classification performance of stacking classifier using stratified-5-Fold-CV

Data	Channel	Fold	Accuracy	F1-score	Precision	Recall	Kappa
Sleep-EDFx (placebo intake recordings)	Pz-Oz	0	0.79	0.74	0.74	0.74	0.71
		1	0.80	0.74	0.75	0.75	0.72
		2	0.80	0.74	0.76	0.73	0.72
		3	0.79	0.75	0.75	0.74	0.71
		4	0.80	0.74	0.75	0.74	0.72
		Mean	0.79	0.74	0.74	0.73	0.71
	Fpz-Cz	0	0.81	0.76	0.77	0.76	0.74
		1	0.83	0.77	0.79	0.76	0.76
		2	0.82	0.74	0.75	0.74	0.74
		3	0.82	0.77	0.78	0.76	0.75
		4	0.82	0.77	0.78	0.76	0.76
		Mean	0.82	0.76	0.77	0.76	0.75

Figure 5. Stacking classifier mean confusion matrix using stratified-5-Fold CV

(a) 5-class classification using Pz-Oz channel (b) 5-class classification using Fpz-Cz channel

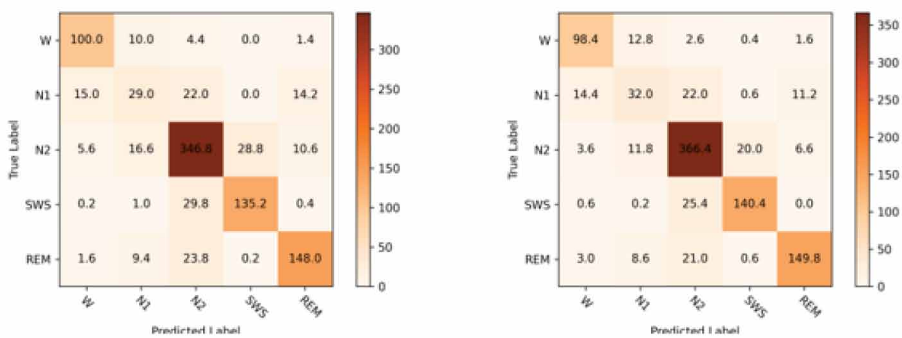


Figure 6. Learning curve of stacking classifier

(a) 5-class classification using Pz-Oz channel (b) 5-class classification using Fpz-Cz channel

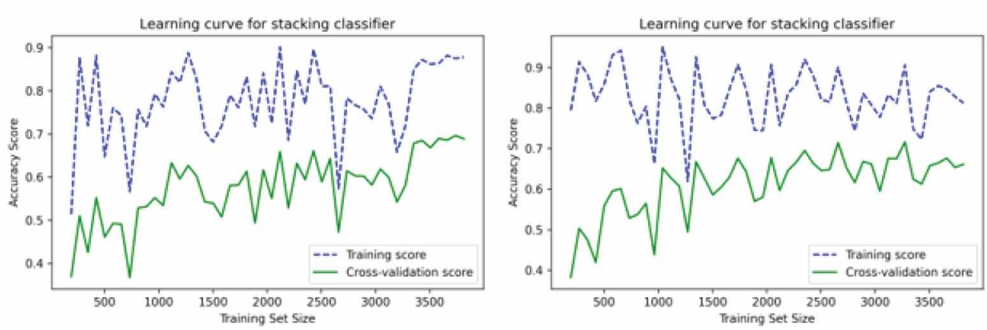


Table 8. Comparison of Stacking classifier performance on different target classes using stratified-5-Fold-CV

Data	Channel	Classifiers	Accuracy*	F1-score*	Precision*	Recall*	Kappa*
Sleep-EDFx (placebo intake recordings)	Pz-Oz	4-Class	0.82	0.82	0.83	0.81	0.72
		3-Class	0.82	0.81	0.82	0.81	0.72
		2-Class	0.82	0.81	0.83	0.80	0.72
	Fpz-Cz	4-Class	0.85	0.85	0.87	0.83	0.77
		3-Class	0.85	0.84	0.87	0.82	0.76
		2-Class	0.85	0.84	0.87	0.83	0.77

* Mean score

Table 9. 5-Class classification performance of stacking classifier using stratified-5-Fold-CV

Data	Channel	Fold	Accuracy	F1-score	Precision	Recall	Kappa
Sleep-EDFx (Temazepam intake recordings)	Pz-Oz	0	0.76	0.72	0.73	0.71	0.66
		1	0.80	0.74	0.74	0.74	0.72
		2	0.80	0.77	0.76	0.78	0.73
		3	0.79	0.74	0.74	0.74	0.71
		4	0.79	0.72	0.73	0.72	0.70
		Mean	0.79	0.74	0.74	0.74	0.70
	Fpz-Cz	0	0.83	0.78	0.78	0.77	0.76
		1	0.81	0.75	0.75	0.75	0.74
		2	0.83	0.79	0.79	0.80	0.77
		3	0.83	0.78	0.77	0.78	0.76
		4	0.83	0.78	0.78	0.78	0.75
		Mean	0.83	0.78	0.78	0.78	0.76

Figure 7. Stacking classifier mean confusion matrix using stratified-5-Fold CV

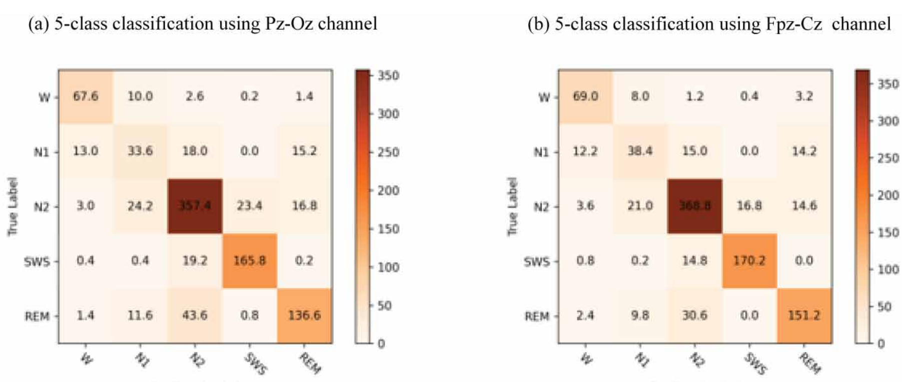


Figure 8. Learning curve of stacking classifier

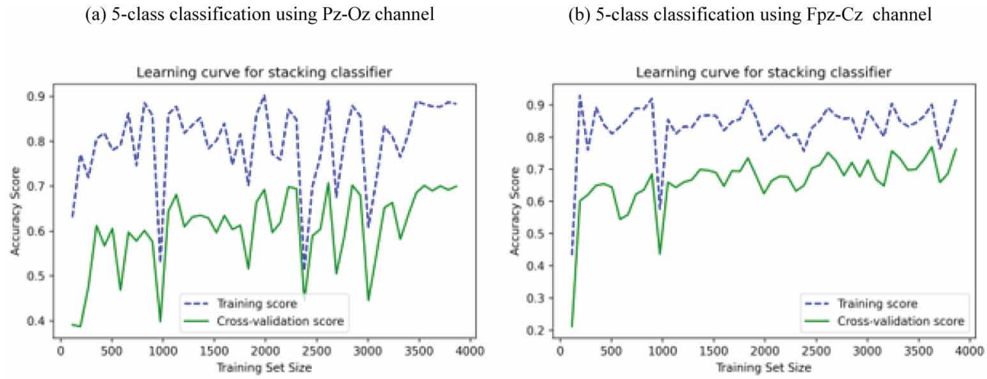


Table 10. Comparison of Stacking classifier performance on different target classes using stratified-5-Fold-CV

Data	Channel	Classifiers	Accuracy*	F1-score*	Precision*	Recall*	Kappa*
Sleep-EDFx (Temazepam intake recordings)	Pz-Oz	4-Class	0.83	0.82	0.83	0.82	0.73
		3-Class	0.83	0.82	0.83	0.82	0.73
		2-Class	0.83	0.82	0.83	0.82	0.73
	Fpz-Cz	4-Class	0.85	0.84	0.84	0.84	0.77
		3-Class	0.86	0.85	0.85	0.85	0.78
		2-Class	0.85	0.84	0.84	0.84	0.77

* Mean score

the proposed model produced on par results using the Pz-Oz channel and the Fpz-Cz channel showed a 1% improvement in accuracy compared with the deep learning approach (Sors, Bonnet, Mirek, Vercueil, & Payen, 2018b). The mean confusion matrix from stratified 5-fold-validation given in Figure 3 shows good classification accuracy. However, many N1 classes are misclassified as N2; due to the poor discrimination between N1 and N2. Besides, N1 epochs are misclassified as REM since there is a subtle inter-class dissimilarity. The classifier achieved 90%, 90%, and 91% accuracy in 4, 3, and 2 class classification, respectively, using the Pz-Oz channel. In the Fpz-Cz channel, the classifier attained 91%, 94%, and 95% accuracy in 4, 3, and 2 class classification. In healthy subjects' data, the Fpz-Cz channel has produced better results when compared with Pz-Oz.

In experiment 2, as shown in Table 7, in a 5-class classification, the proposed classifier accomplished mean, minimum and maximum accuracy of 79%, 79% and 80% employing the Pz-Oz channel. Also, using the Fpz-Cz channel achieved a mean, minimum, and maximum accuracy of 82%, 81%, and 83%. The mean confusion matrix from stratified 5-fold-validation shown in Figure 3 demonstrates good classification accuracy. Nevertheless, N1, N2 and REM misclassification is very high due to the high variance in placebo data. The classifier achieved 82% accuracy in 4, 3, and 2 class classifications using the Pz-Oz channel. Similarly, employing the Fpz-Cz channel, the classifier attained 85% accuracy in 4, 3, and 2 class classifications. In placebo intake subjects' data, the Fpz-Cz channel has produced better results when compared with Pz-Oz. Generally, the signal patterns differ across healthy subjects and subjects having sleep disorders. As a result, there is a distinct difference in recognising sleep stages between healthy subjects and subjects having sleep disorders. Notably, the Fpz-Cz channel provided a 1% to 2% increase in overall classification results compared to previous investigations (Ghimatgar, Kazemi, Helfroush, & Aarabi, 2019b; Hassan, Bashar, & Bhuiyan, 2015).

In experiment 3, as exhibited in Table 9, in a 5-class classification, the proposed classifier accomplished mean, minimum and maximum accuracy of 79%, 76% and 80% employing the Pz-Oz channel. Again, using the Fpz-Cz channel achieved a mean, minimum, and maximum accuracy of 83%, 81%, and 83%. The mean confusion matrix from stratified 5-fold-validation presented in Figure 3 shows good classification accuracy. Nonetheless, N1, N2 and REM misclassification is low compared to placebo data and high compared to healthy subjects' data. The classifier achieved 83% accuracy in 4, 3, and 2 class classification, respectively, using the Pz-Oz channel. In the Fpz-Cz channel, the classifier attained 85%, 86% and 85% accuracy in 4, 3, and 2 class classification. In temazepam intake subjects' data, the Fpz-Cz channel has produced better results when compared with Pz-Oz. Notably, the Fpz-Cz channel provided a 1% to 3% increase in overall classification results compared to previous investigations (Ghimatgar et al., 2019b; Hassan et al., 2015; Hassan & Bhuiyan, 2017b).

The performance achieved using the proposed classifier depends on the representative features extracted. The representative features extracted from the Fpz-Cz EEG channel by Tsfresh achieved a decent mean accuracy across 5, 4, 3, and 2 class classifications. However, this is insufficient to correctly classify the sleep stages with a reasonable accuracy rate. Class imbalance is another typical issue faced by sleep stage classification systems. The less represented class, such as N1, accounts for lesser classification accuracy Some existing studies have balanced the less represented classes to improve the classification accuracy (Zhao, 2019). Table 1 clearly shows that the dataset used in this experiment is imbalanced. The proposed study balanced the less represented classes equal to the highly repressed classes of training data to avoid overfitting in classification. After checking the learning curves in Figure 4, Figure 6 and Figure 8, it is evident that the class balancing approach assists the classifier to some extent improve the classification concerning specific sleep stages. According to the 2 class classification results presented in Table 6, Tsfresh can be a good fit for extracting features for an application requiring sleep and wake classification alone (Li, Lee, & Chung, 2015).

The pattern of sleep is a successive event, and there is a subtle difference between classes. Hence, it is hard to distinguish the sleep stages that are in a transition state. That is evident from the confusion matrix in Figure 3, Figure 5 and Figure7. Generally, there is a subtle or narrow difference between sleep stages such as N1 and N2 as well as N2 and SWS. This kind of subtle difference can be represented well using fractal-based features (Memar & Faradji, 2018). Moreover, during our experiments, depending on the data size, the feature extraction took a minimum of 38.57 seconds to a maximum of 121.78 seconds. For a real-time application, time-consuming feature extraction is not a viable option because it produces time delays in processing the sleep stages. Tsfresh package provides an option to extract selected features and permits the users to add custom features. Adding custom features enables the extraction of the required representative features. The time needed to extract the features can be reduced by extracting only the potential features using the selected feature option of Tsfresh. Table 11 presents the comparison of the proposed work with the existing work. The proposed sleep stage classification approach outperformed well against the listed existing works.

Table 11. Comparison with existing work

Database	Author, year	Classifier	5-Class %	Kappa
Sleep-EDFx	(Hassan, 2015)	AdaBoosting	80	78
	(Hassan & Bhuiyan, 2017b)	RUBoost	83.4	84
	(Sors, Bonnet, Mirek, Vercueil, & Payen, 2018a)	Deep convolution neural network	87	-
	(Ghimatgar et al., 2019a)	Random Forest	81.8	74
	Proposed		88	84

5. CONCLUSION

Using single-channel EEG, the proposed study developed a stacked ensemble classifier for identifying different sleep states. The proposed work employs Tsfresh to overcome the drawbacks of handcrafted feature extraction. The experiment results indicate that Tsfrsh can be used to extract the generic features used by the sleep stage classification works. Next, the experiment results also reveal that automatic feature extraction using TSfresh produces decent classification results. The proposed system achieved a decent mean accuracy of 88% with a kappa score of 0.84, using healthy subjects' sleep data. Similarly, it performed well in 4, 3, and 2 class classifications with a mean accuracy of 91% ~ 95% and a kappa score of 0.86 ~ 0.94. The developed system performs very well for applications requiring sleep and wake classification. In the present work, Tsfresh extracted features from the traditional 30-second data. Applying Tsfresh in subbands extracted from the sleep signal is left for future work.

ACKNOWLEDGMENT

The authors thank the SUSE Center of Excellence for Cloud Computing, Karunya Institute of Technology and Sciences for providing the computing facility to implement this project.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

FUNDING AGENCY

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY

The data used in this study is available at Physionet database: https://www.physionet.org/content/sleep-edfx/1.0.0/

REFERENCES

Allan Hobson, J. (1969). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Electroencephalography and Clinical Neurophysiology*, *26*(6), 644. doi:10.1016/0013-4694(69)90021-2

Berry, Brooks, Gamaldo, Harding, & Lloyd. (2016). American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications, Version 2.2. *American Academy of Sleep, 28*(3), 391–397. www.aasmnet.org

Boostani, R., Karimzadeh, F., & Nami, M. (2017a, March 1). A comparative review on sleep stage classification methods in patients and healthy individuals. In *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd.

Boostani, R., Karimzadeh, F., & Nami, M. (2017b). A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer Methods and Programs in Biomedicine, 140*, 77–91. 10.1016/j.cmpb.2016.12.004

Cesari, M., Stefani, A., Penzel, T., Ibrahim, A., Hackner, H., Heidbreder, A., Szentkiralyi, A., ... (2021). Interrater sleep stage scoring reliability between manual scoring from two European sleep centers and automatic scoring performed by the artificial intelligence-based Stanford-STAGES algorithm. *Journal of Clinical Sleep Medicine, 17*(6), 1237–1247. https://jcsm.aasm.org/doi/abs/10.5664/jcsm.9174

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing*, *307*, 72–77. doi:10.1016/j.neucom.2018.03.067

da Silveira, T. L. T., Kozakevicius, A. J., & Rodrigues, C. R. (2017). Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Medical and Biological Engineering and Computing, 55*(2), 343–352. https://link.springer.com/article/10.1007/s11517-016-1519-4

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. doi:10.1613/jair.1.11192

Freiberger, M., Sackesyn, S., Ma, C., Katumba, A., Bienstman, P., & Dambre, J. (2020). Improving time series recognition and prediction with networks and ensembles of passive photonic reservoirs. *IEEE Journal of Selected Topics in Quantum Electronics, 26*(1).

Ghimatgar, H., Kazemi, K., Helfroush, M. S., & Aarabi, A. (2019a). An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov Model. *Journal of Neuroscience Methods*, *324*, 108320. doi:10.1016/j.jneumeth.2019.108320 PMID:31228517

Ghimatgar, H., Kazemi, K., Helfroush, M. S., & Aarabi, A. (2019b). An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov Model. *Journal of Neuroscience Methods, 324*, 108320. 10.1016/j.jneumeth.2019.108320

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., & Mietus, J. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation, 101*(23). https://www.ahajournals.org/doi/abs/10.1161/01.CIR.101.23.e215

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., & Goj, R. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience,* (7), 267.

Grossi, E., Valbusa, G., & Buscema, M. (2021). Detection of an Autism EEG Signature From Only Two EEG Channels Through Features Extraction and Advanced Machine Learning Analysis. *Clinical EEG and Neuroscience, 52*(5), 330–337. https://journals.sagepub.com/doi/abs/10.1177/1550059420982424

Hassan, A. R. (2015). *A comparative study of various classifiers for automated sleep apnea screening based on single-lead electrocardiogram. In 2015 International Conference on Electrical & Electronic Engineering (ICEEE)*. IEEE. Retrieved from https://ieeexplore.ieee.org/document/7428288/

Hassan, A. R., Bashar, S. K., & Bhuiyan, M. I. H. (2015). On the classification of sleep states by means of statistical and spectral features from single channel Electroencephalogram. *2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015*, 2238–2243. doi:10.1109/ICACCI.2015.7275950

Hassan, A. R., Bashar, S. K., & Bhuiyan, M. I. H. (2016). Automatic classification of sleep stages from single-channel electroencephalogram. In *12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015*. Institute of Electrical and Electronics Engineers Inc.

Hassan, A. R., & Bhuiyan, M. I. H. (2016). A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *Journal of Neuroscience Methods*, *271*, 107–118. doi:10.1016/j.jneumeth.2016.07.012 PMID:27456762

Hassan, A. R., & Bhuiyan, M. I. H. (2017a). Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Computer Methods and Programs in Biomedicine*, *140*, 201–210. doi:10.1016/j.cmpb.2016.12.015 PMID:28254077

Hassan, A. R., & Bhuiyan, M. I. H. (2017b). Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting. *Computer Methods and Programs in Biomedicine, 140*, 201–210. 10.1016/j.cmpb.2016.12.015

Hassan, A. R., & Subasi, A. (2017). A decision support system for automated identification of sleep stages from single-channel EEG signals. *Knowledge-Based Systems*, *128*, 115–124. doi:10.1016/j.knosys.2017.05.005

Ilhan, H. O. (2017). Sleep stage classification via ensemble and conventional machine learning methods using single channel EEG signals. *International Journal of Intelligent Systems and Applications in Engineering*, *4*(5), 174–184. doi:10.18201/ijisae.2017533859

Jiang, D., Lu, Y., Nan, Ma, Y., & Wang, Y. (2019). Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement. Expert Systems with Applications, 121, 188–203.

Kammerer, M. K., Mehl, S., Ludwig, L., & Lincoln, T. M. (2021). Sleep and circadian rhythm disruption predict persecutory symptom severity in day-to-day life: A combined actigraphy and experience sampling study. *Journal of Abnormal Psychology, 130*(1), 78–88.

Kim, H. O., Kang, I., Choe, W., & Yoon, K. S. (2021). Sleep duration and risk of obesity: A genome and epidemiological study. *World Academy of Sciences Journal, 3*(2), 1.Retrieved November 17, 2021, from https://www.spandidos-publications.com/10.3892/wasj.2021.91/abstract

Kwon, S., Kim, H., & Yeo, W. H. (2021, May 21). Recent advances in wearable sensors and portable electronics for sleep monitoring. *iScience*.

Lazarevich, I., Prokin, I., & Gutkin, B. (2018). *Neural activity classification with machine learning models trained on interspike interval series data*. https://www.biorxiv.org/content/10.1101/2021.03.24.436765v1

Li, G., Lee, B. L., & Chung, W. Y. (2015). Smartwatch-Based Wearable EEG System for Driver Drowsiness Detection. *IEEE Sensors Journal*, *15*(12), 7169–7180. doi:10.1109/JSEN.2015.2473679

Lu, H., Yang, Q., Tian, F., Lyu, Y., He, H., Xin, X., & Zheng, X. (2021). A meta-analysis of a cohort study on the association between sleep duration and type 2 diabetes mellitus. *Journal of Diabetes Research*.

Memar, P., & Faradji, F. (2018). A Novel Multi-Class EEG-Based Sleep Stage Classification System. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(1), 84–95. doi:10.1109/TNSRE.2017.2776149 PMID:29324406

Pan, Q., Brulin, D., & Campo, E. (2020). Current Status and Future Challenges of Sleep Monitoring Systems: Systematic Review. *JMIR Biomedical Engineering, 5*(1), e20921. https://biomedeng.jmir.org/2020/1/e20921

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(8), 1226–1238. doi:10.1109/TPAMI.2005.159 PMID:16119262

Seifpour, S., Niknazar, H., Mikaeili, M., & Nasrabadi, A. M. (2018). A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal. *Expert Systems with Applications*, *104*, 277–293. doi:10.1016/j.eswa.2018.03.020

Sors, A., Bonnet, S., Mirek, S., Vercueil, L., & Payen, J.-F. (2018a). A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomedical Signal Processing and Control*, *42*, 107–114. doi:10.1016/j.bspc.2017.12.001

Sors, A., Bonnet, S., Mirek, S., Vercueil, L., & Payen, J. F. (2018b). A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomedical Signal Processing and Control*, *42*, 107–114. doi:10.1016/j.bspc.2017.12.001

Stepnowsky, C., Levendowski, D., Popovic, D., Ayappa, I., & Rapoport, D. M. (2013). Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters. *Sleep Medicine*, *14*(11), 1199–1207. doi:10.1016/j.sleep.2013.04.022 PMID:24047533

Thorpy, M. (2017). International classification of sleep disorders. In *Sleep Disorders Medicine: Basic Science, Technical Considerations and Clinical Aspects: Fourth Edition* (pp. 475–484). Springer. Retrieved November 17, 2021, from https://link.springer.com/chapter/10.1007/978-1-4939-6578-6_27

Wu, J., Zhou, T., & Li, T. (2020). Detecting epileptic seizures in EEG signals with complementary ensemble empirical mode decomposition and extreme gradient boosting. *Entropy, 22*(2), 140. https://www.mdpi.com/1099-4300/22/2/140/htm

Wulff, K., Gatti, S., Wettstein, J. G., & Foster, R. G. (2010, July 14). Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nature Reviews Neuroscience*. https://www.nature.com/articles/nrn2868

Zhang, J., Tang, Z., Gao, J., Lin, L., Liu, Z., Wu, H., Liu, F., & Yao, R. (2021). Automatic detection of obstructive sleep apnea events using a deep CNN-LSTM model. *Computational Intelligence and Neuroscience*, *2021*, 2021. doi:10.1155/2021/5594733 PMID:33859679

Zhao, D. (2019, July 1). Comparative analysis of different characteristics of automatic sleep stages. In *Computer Methods and Programs in Biomedicine*. Elsevier.

B. L. Radhakrishnan is currently pursuing his Ph.D. in Computer Science and Engineering, at KITS. Received his M.Tech in Computer Science and Engineering from Dr. MGR University, India in 2008, and did his B.E in Computer Science and Engineering from Anna University, India in 2006. He holds Industry Certifications such as Certified OpenStack Administrator, SUSE Certified Engineer in Enterprise Linux and Google Certified Associate Cloud Engineer. His research interests include remote sleep monitoring, cloud computing and Blockchain.

E. Kirubakaran is presently a Professor in the Department of Computer Science and Engineering and Director at Grace College of Engineering, Thoothukudi, India. He obtained his B.E. (Honours) degree in Mechanical Engineering in 1978, M.E. in Computer Science in 1984 and Ph.D. degree in Computer Science in 1999 from the National Institute of Technology (the then Regional Engineering College), Tiruchirappalli, Trichy, India. In 2005, he received his M.B.A. degree from IGNOU. He had served Bharat Heavy Electricals Ltd. Trichy, India for 38 years and retired as Additional General Manager. During his tenure at BHEL, he worked in the areas of designing, developing and maintenance of Computer Systems. After his BHEL service he worked at Karunya (Deemed)University, Coimbatore as Professor of Computer Science and Engineering for 3 years and 6 months . He had been a visiting faculty to many institutions in and around Trichy. He has also guided a number of Research Scholars for Ph.D. He is a Lifetime Member of the Computer Society of India, Member of IEEE and Fellow of the Institution of Engineers (India). He has held the post of Secretary, Vice Chairman and Chairman in the Computer Society of India, Trichy Chapter. Dr. E. Kirubakaran is a Certified Lead Auditor for Information Security Management Systems.

Immanuel Johnraja Jebadurai received his Bachelor's degree in Engineering during 2003 from M.S. University, India and his Master's degree in Engineering during 2005 from Anna University, India. He received his Doctor of Philosophy degree from Karunya Institute of Technology and Sciences, India during 2017. Currently, he is the Professor and Head of Computer Science and Engineering, Karunya Institute of Technology and Sciences. His areas of research include Mobile Ad Hoc network security, routing Software Defined Networkjng and Vehicular Ad Hoc Networking. He has been the instructor and the course coordinator of Cisco Certified Network Associate (CCNA) course in Cisco Networking Academy, Karunya University since 2007. He was one of the CCNP instructors recognized by Cisco Systems, Bangalore.