Konya Mühendislik Bilimleri Dergisi, c. 11, s. 3, 783-800, 2023
Konya Journal of Engineering Sciences, v. 11, n. 3, 783-800, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1073932

DETERMINING THE MOST POWERFUL FEATURES IN THE DESIGN OF AN AUTOMATIC SLEEP STAGING SYSTEM

[1,*]Seral ÖZŞEN, [2]Yasin KOCA, [3]Gülay TEZEL, [4]Sena ÇEPER,
[5]Serkan KÜÇÜKTÜRK, [6]Hülya VATANSEV

[1, 2]Konya Technical University, Electrical and Electronics Engineering Department, Konya, TÜRKİYE
[3, 4]Konya Technical University, Computer Engineering Department, Konya, TÜRKİYE
[5, 6]Necmettin Erbakan University, Medicine Faculty, Sleep Laboratory, Konya, TÜRKİYE
[1] sozsen@ktun.edu.tr, [2]kocayasiin@gmail.com, [3]gtezel@ktun.edu.tr, [4]sceper@ktun.edu.tr,
[5]skuccukturk@kmu.edu.tr, [6]hulyavatansev@hotmail.com

Highlights

- The accuracies of automatic sleep staging systems using real sleep data are generally under 85% and studiying with real datasets is very hard because extracting useful features from a noisy environment is more complex. This study used real sleep dataset to reach high accuracy.
- The automatic sleep stage classifiers used in literature are generally designed for known state-of-art datasets and they are taken from healthy persons. A real dataset including healthy and Obstructive Sleep Apnea patients is used in this dataset.
- The main contribution of this study is feature engineering in that we scrutunized which features are more useful in a sleep staging system which uses real dataset with healthy and OSA patients.

***Corresponding Author:** Seral ÖZŞEN, sozsen@ktun.edu.tr

Konya Mühendislik Bilimleri Dergisi, c. 11, s. 3, 783-800, 2023
Konya Journal of Engineering Sciences, v. 11, n. 3, 783-800, 2023
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1073932

DETERMINING THE MOST POWERFUL FEATURES IN THE DESIGN OF AN AUTOMATIC SLEEP STAGING SYSTEM

[1,*] Seral ÖZŞEN, [2]Yasin KOCA, [3]Gülay TEZEL, [4] Sena ÇEPER, [5]Serkan KÜÇÜKTÜRK, [6]Hülya VATANSEV

[1, 2]Konya Technical University, Electrical and Electronics Engineering Department, Konya, TÜRKİYE
[3,4] Konya Technical University, Computer Engineering Department, Konya, TÜRKİYE
[5, 6] Necmettin Erbakan University, Medicine Faculty, Sleep Laboratory, Konya, TÜRKİYE
[1]sozsen@ktun.edu.tr, [2]kocayasiin@gmail.com, [3]gtezel@ktun.edu.tr, [4]sceper@ktun.edu.tr,
[5]skuccukturk@kmu.edu.tr, [6]hulyavatansev@hotmail.com

(Received: 15.02.2023; Accepted in Revised Form: 21.06.2023)

ABSTRACT: Spending too much time on manual sleep staging is tiring and challenging for sleep specialists. In addition, experience in sleep staging also creates different decisions for sleep experts. The search for finding an effective automatic sleep staging system has been accelerated in the last few years. There are many studies dealing with this problem but very few of them were conducted with real sleep data. Studies have been carried out on mostly processed and cleaned-ready data sets. In addition, there are few studies in which the data distribution in sleep stages is balanced (equal numbers of epochs from each stage are used), and it is seen that the performance of these studies is quite low compared to other studies. When the literature studies are examined, there is a wide range of studies in which many features are extracted, many feature selection methods are used, many classifiers are applied and various combinations of these are available. For this reason, to determine the best-performing features and the most powerful features, 168 features were extracted from the real EEG, EOG, and EMG signals of 124 patients. These features were selected with 7 different feature selection methods, and classification was carried out with 4 classifiers. In general, the ReliefF feature selection method has performed best, and the Bagged Tree classifier has reached the highest classification accuracy of 67.92% with the use of nonlinear features.

Keywords: *Automatic Sleep Staging, Frequency Analysis of EEG Signals, Sleep Signal Detection*

1. INTRODUCTION

Sleep staging is a process that is done for many reasons such as to detect sleep-related disorders, determine sleep quality,…etc. According to AASM standards, [27] sleep can be categorized into Wake, REM, and Non-REM stages. The Non-REM stages are further divided into Non-REM1, Non-REM2, and Non-REM3 stages. The staging is done by examining sleep-related signals which are named Polysomnography Signals (PSG) by a sleep expert. The most used ones in PSG signals are electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) signals. These signals are divided into 30 sec-long epochs and then the expert decides which stages should be selected as a label of the related epoch. In doing this, the experts use rules [27] and their experience. Whereas this seems to be a very straightforward process, making the correct decision would be cumbersome in many cases. First of all, the signals are mainly noisy, and recognizing some signal patterns which are used in making decisions is sometimes very hard in this noisy environment. Also, there isn't a clear distinction in some epochs between the stages. For example, an EEG signal belonging to a Wake stage can be very similar to one of the Non-REM1 stages. In such cases, the experience of the sleep expert can result in different staging labels. The manual staging process is also time-consuming and tiring because the expert should examine many signals in 30-second epochs in detail (the duration of the overall signals is 6-8 hours) and interpret

***Corresponding Author:** Seral ÖZŞEN, sozsen@ktun.edu.tr

the rules according to the overall sleep to make the decision. So, based on these reasons in place, the research for automatic sleep-staging systems has begun many years ago.

It can be found in a huge literature in the search for automatic sleep staging studies. In the study of Lee et. al. [33], a Detrended Fluctuation Analysis (DFA) was used for the classification of sleep data which is taken from the MIT/BIH sleep database [33]. Liu and Sun [36] conducted a three-class staging study on MIT/BIH database by using multifractal DFA [36]. In another study using DFA, Farag et. al. [17] reached an overall classification accuracy of 85.18% on their real recorded dataset with 22 patients [17]. Hassan et. al. [22] used a tunable Q-factor wavelet transform to classify sleep stages of Physionet and Dream datasets and obtained a classification accuracy of 91.50% [22]. Other methods were also applied to this dataset in similar studies [23, 24, 25] and performances between 90-94% were reached. Chlon et. al. [13] proposed a different method for the classification of sleep stages. They combined Hierarchical Driclet Process-Hidden Markov Model (HMM) with Wide-sense-stationary time series analysis and conducted the tests of the proposed method on simulated sleep data [13]. In the study of Jiang et. al. [29], a new method based on Empirical Mode Decomposition (EMD) was proposed and they reached a 92% classification accuracy on Physionet data [29]. They also applied a new rule-free refinement based on HMM to optimize classification results. In their study, Acharya et. al. extracted features by using nonlinear dynamic analysis, Higher Order Spectra (HOS), and Recurrence Quantification Analysis (RQA) [2].They analyzed the effectiveness of these features by ANOVA. Chaozsen et. al. used Hilbert Huang Transform (HHT) to obtain spectral features [10]. By adding a sample entropy feature to their feature set, they obtain 89.9% classification accuracy on the MIT database. In another study, Peker used Complex valued Neural Network to classify the Sleep-EDF database by extracting complex-valued nonlinear features and he obtain a classification accuracy of 91.57% [41]. Tian et. al. [44] used the aid of Multiscale entropy features of the Sleep-EDF database and they classified them by proportion-based SVM with a classification accuracy of 91.4% [44]. In their study, Liua et. al. [37] conducted a comprehensive feature selection application for automatic sleep staging [37]. They extracted 50 features from the EEG signals of the MIT/BIH database by using Multifractal DFA, visibility graph algorithm (VGA), frequency analysis, and nonlinear analysis. They utilized genetic algorithm (GA) as feature selector and used Least Squares SVM to compare the performances. In a study conducted with real data, Zhang et. al. [49] used frequency, time-domain, and nonlinear features and reached an average accuracy of 82.18% in three subjects [49]. Zhang et. al. [50] applied band-pass filtering, spectral feature extraction, gaussian parameters, and statistic-based feature selection [50]. They reached a classification accuracy of 85.5% in their dataset consisting of 39 subjects. In their study, Henri Korkalainen et. al. used overlapping different-size epochs (Korkalainen et. al., 2021). They reached 81.9% accuracy but their dataset includes many NonREM3 and REM epochs and they used deep learning. Ghimatgar et. al. [20] used HMM for staging with single-channel EEG [20]. They used four public EEG datasets and comparatively low accuracies for such public datasets (between 77.6%-97.4%). In their study, Arslan et. al used machine learning methods for sleep stage classification for data taken from 50 patients with sleep-related diseases (not from the OSA patients) [3]. They obtained comparatively high accuracies but 19 channels were used in that study.

Some literature reviews can be found in [1, 7, 18, 19, 26].

The vast majority of these studies use standard datasets such as the Sleep-EDF dataset, and MIT/BIH database. The accuracy ratios in these studies are generally over 85%. But when the studies using real data are analyzed, the accuracy values decrease under 85%. Dealing with real datasets is very hard because extracting useful features from a noisy environment is more complex. Besides, if the used dataset is taken from patients rather than healthy persons, the problem becomes more complicated. Besides this, when the studies obtaining high accuracy rates were searched, it can be seen that they used deep learning methods. In deep learning methods, it is not possible to see which feature gave the best result. Thus our study differs from theirs in that we also tried to see which features are more useful in the staging process. In our study, we used a dataset consisting of PSG signals of 124 persons (93 of them have Obstructive sleep apnea (OSA) and 31 of them are healthy). By utilizing time-domain analysis, frequency analysis, nonlinear analysis, and MDFA, 168 features were extracted from the EEG, EOG, and EMG signals. To detect best performing features, 7 feature selection methods were used in 5 applications. In performance comparison, classification was carried out with k-nearest neighbor (kNN), decision tree (DT), SVM, and Bagged Tree methods.

2. MATERIALS and METHOD

2.1. Used Data

During the study, PSG data were obtained from a total of 124 people, 31 of whom were healthy, and 93 patients, who were hospitalized in the sleep laboratory of the Meram Medical Faculty hospital. Whereas many signals are recorded in PSG, EEG, EOG, and EMG signals are the most commonly used ones in manual and automatic sleep staging studies. Patients with an AHI value above 5 were accepted. Using the EEG, EOG, and EMG signals from the obtained PSG data, these signals were divided into 30-sec epochs. PSG signals were sampled at 200 Hz. The signals obtained for each patient were normalized to avoid inter-patient amplitude differences. The EEG signals used were 0.3Hz-35Hz, EMG signals 1Hz-45Hz, and EOG signals were filtered between 0.3Hz-30Hz with a 6th-order Butterworth bandpass filter. The stage of these epochs was determined by the sleep specialist. The total number of epochs was noted as 67443. The distribution of these epochs according to the sleep stages is given in Table 1.

***Table 1**. Distributions of epochs after electrode disconnection deleting process*

	Patient	Healthy	Total
Wake	13600	3712	17312
NonREM-1	6031	1496	7527
NonREM-2	23066	7858	30924
NonREM-3	4765	1269	6034
REM	3625	2021	5646

As seen, there is a very high imbalance between the stages. This results in inappropriately trained classifiers because the classifiers arrange their parameters according to the most seen classes. To avoid this situation, almost equal numbers of epochs from each stage were selected and included in the study. 5000 epochs were selected from each stage and a data set consisting of 25000 epochs in total was prepared. When selecting 5000 epochs, a balanced distribution was taken into account between the sick and healthy individuals. In summary, the generated dataset consists of a total of Nx25000 epochs, 5000 epochs per stage, with N being the number of features. In this dataset, 1000 epochs from each stage are reserved for feature selection, 3000 epochs are reserved for training testing with 5-fold cross-validation, and 1000 epochs are reserved for validation. As a result, three sub-datasets were obtained: the dataset to be used for feature selection (Nx5000 size), the dataset to be used in cross-validation (Nx15000) and the dataset to be used in validation (Nx5000). Random epoch selection was carried out in these selections.

2.2. Feature Extraction

From the EEG, EOG (left eye and differential EOG), and EMG signals, time, frequency, nonlinear, and MDFA features were obtained.

Time Domain Features:

The time domain features used in the study were determined by examining the literature studies as Mean value, Standard deviation, Skewness, Kurtosis, Signal energy, Zero crossing rate, and maximum-minimum distance (MMD) [1]. These 7 features were extracted from the EEG, left-eye EOG, difference EOG (difference EOG signal is obtained by subtracting left eye EOG signal from the right-eye EOG signal), and EMG signal. Thus, 28 features were obtained as time features.

Frequency Domain Features:

The frequency content of the signals used for feature extraction in the frequency domain was obtained by the Welch method [40]. Many methods have been used to obtain the frequency content in sleep staging applications. In a recent study, many frequency analysis methods were compared in an application and it was concluded that the Welch method was more successful because it was less sensitive to noise than other methods [7]. In other similar studies, it has been seen that the Welch method has achieved successful results, so the Welch method has been preferred as a frequency analysis method. The frequency resolution was chosen as 0.05 Hz. The following frequency features were extracted from the

EEG signals after obtaining their Power Spectral Distributions (PSD) by the Welch method:

1. Relative power ratio of the alpha frequency band (8-12 Hz)
2. Relative power ratio of the Beta frequency band (12-16 Hz)
3. Relative power ratio of Theta frequency band (4-8 Hz)
4. Relative power ratio of Delta frequency band (0-4 Hz)
5. Difference of Alpha power between the current epoch and precious epoch
6. Difference of Beta power between current epoch and precious epoch
7. Difference of Theta power between the current epoch and precious epoch
8. Difference of Delta power between the current epoch and precious epoch
9. Relative power ratio of 12-14 Hz band (for sleep spindles)
10. Difference of 12-14Hz power between current epoch and precious epoch
11. Mean value of power spectra
12. Standard deviation of power spectra
13. Skewness of power spectra
14. Kurtosis of power spectra

Besides these 14 frequency features obtained from the EEG signal, 12 frequency features were taken from the EOG signals (6 features from the left-eye EOG, 6 features from the difference EOG). These are the Relative power of the 0.5 Hz-2 Hz frequency band, the Difference of 0.5-2Hz power between the current epoch and precious epoch, the Mean value of power spectra, the Standard deviation of power spectra, the Skewness of power spectra and Kurtosis of power spectra. Two more features were added to these frequency features which were extracted from the EMG signals: The total energy of EMG power spectra and the Difference of total power spectral energy between the current epoch and the previous epoch. Thus, 28 frequency features were obtained in total (14 from EEG, 12 from EOG, and 2 from EMG).

Nonlinear Features:

While there are many nonlinear features used in the literature, the most commonly used ones were identified and included in the study. Accordingly, the nonlinear features extracted from EEG, EOG, and EMG signals are [2]: Approximate Entropy, Sample Entropy, Fuzzy Entropy, Renyi's Entropy, Permutation Entropy, Hurst Exponent, Lyapunov Exponent, Correlation Dimension, Kolmogorov Complexity, Lempel-Ziv Complexity, Higuchi Fractal Dimension, Hjorth mobility, Hjorth complexity. As a result, a total of 52 nonlinear features were obtained, 13 from EEG, 13 from left-eye-EOG, 13 from difference-EOG, and 13 from EMG.

MDFA features:

Peng et al. [42] proposed the Detrended Fluctuation Analysis (DFA) method to analyze the similarity and correlation in DNA sequences. Since then, this method has been frequently used to identify mono-fractal scaling features in non-stationary time series in many fields such as financial market analysis [14], analysis of biomedical time series, and detection of abnormal conditions [6,47], natural and social events [11]. is used. However, many time series, including biomedical signals, may not exhibit a mono-fractal structure [30]. In particular, biomedical signals may exhibit transient variations and erratic fluctuations. These variations and fluctuations cannot be explained by a single scale provided by the DFA method and are suitable for multiple fractal structures. Therefore, Kantelhardt et al. [30] removed the limitations of CFA by introducing the Multifractal Detrended Fluctuation Analysis (MDFA) method, which is the advanced version of the DFA method [8, 16, 38].

In the study, the following features were extracted from the time series obtained by the analysis of EEG, EOG, and EMG signals with MDFA [21]: maximum hurst exponent, minimum hurst exponent, generalized hurst exponent, maximum singularity exponent, minimum singularity exponent, Mean singularity exponent, Singularity exponent corresponding to the peak of the multifractal spectrum, Asymmetric index, Multifractal spectrum corrs. to max sing exp [8], multifractal spectrum corrs. to min sing exp [8], Vertical distance between f(amin)- f(amax) [8], Skewness of Multifractal spectrum, Kurtosis of Multifractal spectrum, the width of Multifractal spectrum, the height of Multifractal spectrum. Accordingly, a total of 60 MDFA features were obtained, 15 from the EEG signal, 15 from the left-eye-EOG signal, 15 from the differential EOG signal, and 15 from the EMG signal.

In summary, during the feature extraction process, features were extracted in 4 groups: time, frequency, nonlinear, and MDFA. 28 time, 28 frequency, 52 nonlinear, and 60 MDFA features were extracted from EEG, EOG, and EMG signals (See Table 2).

Table 2. *Summary of the extracted features*

	EEG	Left-eye-EOG	Differential EOG	EMG	Total
Time Features	7	7	7	7	28
Frequency Features	14	6	6	2	28
Nonlinear Features	13	13	13	13	52
MDFA Features	15	15	15	15	60

2.3. Feature Selection

As in other fields, the feature selection stage is as valuable as the feature extraction stage in biomedical classification problems. Studies using feature selection methods among sleep staging studies were examined in detail and it was concluded that these feature selection methods would be appropriate to use and compare in the study: Canonical Correlation Analysis (CCA) [35], 2016), Sequential Feature Selection (SFS) [43]. Feature selection with Fisher Score (FS) [34], Feature selection with chi-square test, ReliefF [45], Feature selection with Information Gain (IG) method [4], Feature selection with fast correlation-based filter method (FCBF) [31].

5 feature selection applications were conducted in the study. 4 of them were conducted to determine the best features in each feature sub-set (time-, frequency-, nonlinear- and MDFA-subsets) and the last one was done to determine the best-performing features among the whole feature set with 168 features. For each application, combinations of all the above-mentioned feature selection methods and all classification methods were run within the application. For example, in the 1st application done with the time-domain features, the number of features selected by the CCA feature selection method was made 2,3,4,5, … 27,28. Afterward, the same process was carried out by choosing SFS as the feature selection method. In this way, with each feature selection method, features are selected in all possible number of features and all classifiers are run for these feature combinations. Other applications (applications 2-4) with frequency properties, nonlinear properties, and MDFA properties were carried out in the same way. In application-2 made with frequency features, the number of features was from 2 to 28; In application-3 with nonlinear features, the number of features was from 2 to 52; In application-4 made with MDFA features, the number of features was changed from 2 to 60.

2.4. Classification

To compare the performances of the selected features, 4 different classification method was utilized. They are:

- K-nearest neighbor method-kNN [15]
- Support vector machines- SVM [46]
- Decision trees-DT [39]
- Bagged tree classifier-BT [9]

They have commonly used classifiers in automatic sleep staging literature. The following criteria were used while evaluating the classification performances [28]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$F1 - score = \frac{2TP}{2TP+FP+FN} \quad (5)$$

Where TP: true positive, TN: true negative, FP: false positive, FN: false negative.

These criteria were calculated separately for each stage using the complexity matrix obtained as a result of the classification, as well as the average values of all stages were calculated.

3. RESULTS

3.1. Results of Application-1 for Time Domain Features

In the feature selection application conducted with time features, the number of features was changed from 2 to 28 for each of the 7 feature selection methods, and classification was made with 4 different classifiers. In Figure 1, for the CCA feature selection method, the test classification accuracies of the classification methods concerning the number of features are seen. As can be seen, the highest classification accuracy was obtained with the Bagged tree classifier at 64.9% for 13 features.

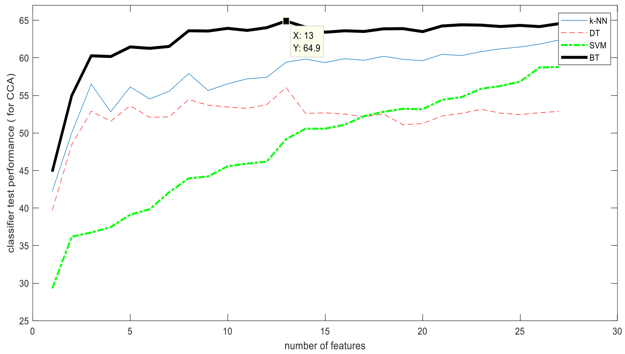

Figure 1. *Comparison of classifiers' test performances concerning the number of features for the CCA feature selection method*
(for frequency domain features).

When the same procedure was conducted for other feature selection methods, the best performances were obtained as given in Table 3.

Table 3. *Comparison of feature selection methods for time domain features*

Applied feature selection method	**Best performed classifier**	**The optimum number of features**	**Highest test classification accuracy (%)**
CCA	BT classifier	13	64.90
Chi-square	BT classifier	22	65.07
FS	BT classifier	24	64.89
FCBF	BT classifier	19	65.91
IG	BT classifier	18	65.57
ReliefF	BT classifier	**20**	**65.96**
SFS	BT classifier	25	65.02

When the results of Table 3 is analyzed, it was observed that the highest classification accuracy was obtained with ReliefF, although there was no significant difference between the feature selection methods. On the other hand, in all feature selection methods except the Canonical Correlation Method, the highest performances were obtained for the cases where the number of features was in the range of 18-25. Considering that only time features are used in practice, this can be interpreted as an expected result because the system needs as much information as possible to make an accurate classification.

When the performances of the classifiers are compared, kNN and BT classifiers have similar performance, while DT and SVM have relatively lower classification accuracies. When ranking in terms of performance in the classifications, it is striking that the classifiers are ranked from best to worst as BT, kNN, SVM, and DT, respectively. In addition, it was observed that while the DT performed better than SVM at low feature counts, SVM performance was better than the DT when the number of features increased.

3.2. Results of Application-2 for Frequency domain features

In this application, where only frequency features are used as in-time features, the number of features for each of the 7 feature selection methods is changed from 2 to 28 (because there are 28 frequency features), and classification with 4 different classifiers for each feature set. The change in test classification accuracy according to the feature number was recorded for each feature selection method and classifier.

This change is seen in Figure 2 for the ReliefF feature selection method. Again, as for the time domain features, the highest test classification accuracy was reached by the BT classifier as 61.99% with 11 features.

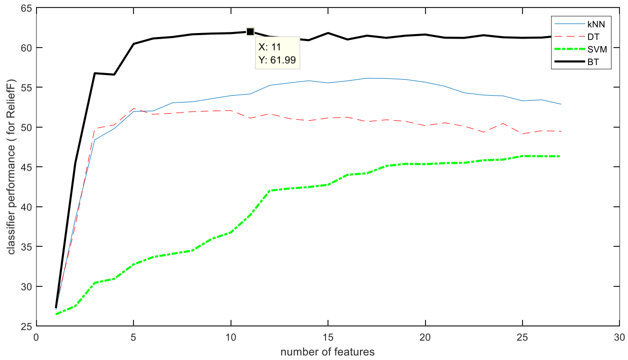

Figure 2. *Comparison of classifiers' test performances with regard to the number of features for the ReliefF feature selection method (for frequency domain features).*

The highest accuracy values obtained by the feature selection methods are summarized in Table 4.

Table 4. *Comparison of feature selection methods for frequency domain features*

Applied feature selection method	Best performed classifier	The optimum number of features	Highest test classification accuracy (%)
CCA	BT classifier	27	61.06
Chi-square	BT classifier	27	61.53
FS	BT classifier	27	61.11
FCBF	BT classifier	25	61.38
IG	BT classifier	24	61.55
ReliefF	BT classifier	**11**	**61.99**
SFS	BT classifier	15	61.94

When the results given in Table 4 are examined, the highest accuracy was reached with the ReliefF method, as in the time properties. The highest performances have been obtained when almost all features are used in feature selection methods except ReliefF and SFS methods. The fact that the features extracted during the study were determined to be those that are thought to give the best results based on the literature has a great effect on this. When the classifier performances were compared, it was observed that the BT method obtained the best performance, as in the time properties, followed by the kNN, DT, and SVM methods, respectively.

3.3. Results of Application-3 for Nonlinear Features

In this application, as stated before, a total of 52 features, 13 from each signal, were obtained and classification processes were carried out with feature selection methods. Thus, the feature number was changed from 1 to 52 in the feature selection applications. The same procedure was performed and the best feature number, best feature selection method, and best classifier were tried to be found. The change in test classification accuracy for the Chi-square method with regard to the feature number is given in Figure 3.

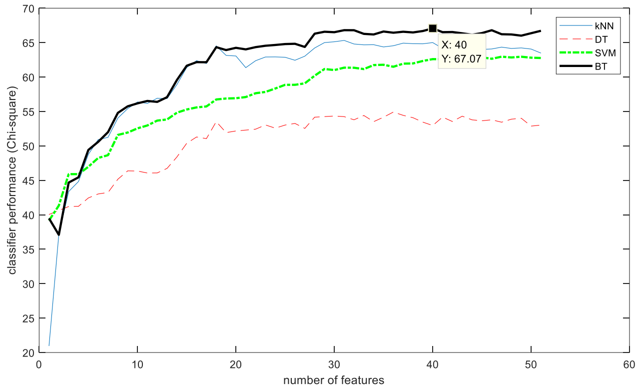

Figure 3. *Comparison of classifiers' test performances with regard to the number of features for the Chi-square feature selection method (for nonlinear features).*

The highest accuracy values obtained by the feature selection methods are summarized in Table 5.

Table 5. *Comparison of feature selection methods for nonlinear features*

Applied feature selection method	Best performed classifier	The optimum number of features	Highest test classification accuracy (%)
CCA	BT classifier	36	66.88
Chi-square	BT classifier	40	67.07
FS	BT classifier	49	66.97
FCBF	BT classifier	50	66.49
IG	BT classifier	50	66.25
ReliefF	BT classifier	**24**	**67.92**
SFS	BT classifier	17	67.26

When the results in Table 5 are examined, the highest classification accuracy was again obtained with ReliefF as a feature selection method. As a classifier, the BT method showed a more successful performance than in the previous applications. 24 features were selected in the ReliefF method, where the highest accuracy was obtained, and 17 features were selected in the sequential feature selection method. The performance is the best when almost all of the features are used in the other feature selection methods.

3.4. Results of Application-4 for MDFA features

As a result of MDFA applied to EEG, EOG, and EMG signals, a total of 60 features, 15 from each signal, were obtained as stated before, and classification processes were carried out with feature selection methods. Again, as with the other applications the feature number was changed, and the performances of feature selection methods were compared with the use of four classifiers. The results obtained for FCBF are shown in Figure 4.

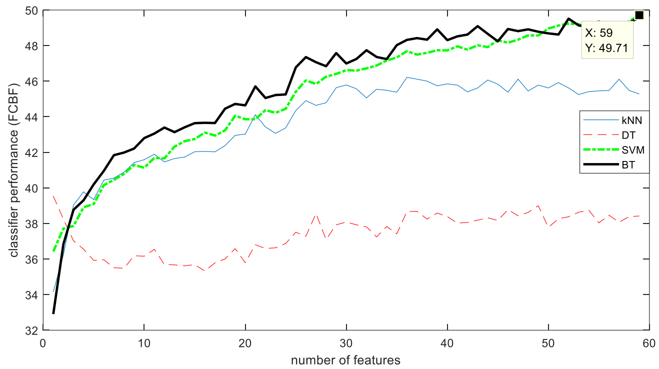

Figure 4. *Comparison of classifiers' test performances with regard to the number of features for the FCBF feature selection method (for MDFA features).*

Here, as seen from the figure, the performance of BT and SVM classifiers is close to each other. This was the case for other feature selection methods, too. Meanwhile, the performance of DT was very poor for all feature selection methods. The best accuracies reached for each feature selection method can be seen in Table 6.

Table 6. *Comparison of feature selection methods for MDFA features*

Applied feature selection method	Best performed classifier	The optimum number of features	Highest test classification accuracy (%)
CCA	BT classifier	46	49.51
Chi-square	BT classifier	57	49.79
FS	SVM	54	49.77
FCBF	SVM	59	49.71
IG	BT classifier	59	49.52
ReliefF	BT classifier	**50**	**49.95**
SFS	BT classifier	58	49.59

When the results in Table 6 are examined, it is seen that MDFA features are not successful enough in this classification problem. It has been determined that the feature selection methods have obtained very close results (49. xx) and these results have been achieved with a very large number of features. On the other hand, what stands out for this feature set is the better performance of the SVM, unlike the situation with other feature sets.

3.5. Comparison of Results

In the above 4 applications, 5-fold cross-validation and validation processes were carried out for each number of features by applying 7 different feature selection methods and 4 different classifiers. The highest cross-validation results obtained for each application are summarized in Table 7.

Table 7. *Comparison of application results*

	Feature set	**Best feature selector**	**Best classifier**	**Total number of features**	**Number of selected features**	**Highest accuracy (%)-test in CV**	**Highest accuracy (%)-validation**
Application-1	Time	ReliefF	BT	28	20	65.96	64.94
Application-2	Frequency	ReliefF	BT	28	11	61.99	62.36
Application-3	Nonlinear	ReliefF	BT	52	24	**67.92**	**66.66**
Application-4	MDFA	ReliefF	BT	60	50	49.95	50.14

When Table 7 is examined, it has been observed that the best feature selection method in all applications is ReliefF, although there are no high differences between the feature selection methods. On the other hand, it has been determined that the best classification method is the BT classifier. When the number of selected features is evaluated, it is observed that most of the extracted features are used in applications where time and MDFA features are used. In time features, this can be attributed to the use of necessary features in the study based on literature and past studies. Considering the low accuracies obtained for MDFA, it can be concluded that MDFA features alone are insufficient in classification. On the other hand, a remarkable feature reduction was made in frequency and nonlinear features. 11 of 28 frequency features were selected in frequency domain analysis and 24 of 52 nonlinear features were selected in application-3 in which nonlinear features were used. This gives the impression that some (nearly half) of the frequency and nonlinear features used are unnecessary. If an evaluation is made between the feature sets, the highest accuracy was achieved with nonlinear features (67.92%), followed by time (65.96%), frequency (61.99%), and MDFA (49.95%), respectively.

When the obtained results were examined, it was observed that there were no high differences between the cross-validation results and the validation results. For example, Table 8 shows the cross-validation and validation results obtained with the Bagged tree method for the case where the ReliefF method is used in Application-3 (nonlinear properties). In Figure 5, the variation of cross-validation and validation accuracies according to the number of features is visualized for the situation in question. This situation was seen similarly for all other application results.

In Table 9, the complexity matrix obtained in cross-validation is given for the case where the highest classification accuracy is obtained. Almost similar results were obtained in other applications. Accordingly, while high accuracy was achieved in the classification of Non-REM 3 and REM stages, lower success was observed in the classification of Wake, Non-REM 1, and Non-REM 2 stages. When the previous studies are examined, the classification performances of the Wake and Non-REM 1 stages are generally poor. However, performances in the Non-REM 2 stage are generally higher. The reason for the low performance of the Non-REM 2 stage in this study can be interpreted as the low number of data used to create a balanced data set compared to other studies. In future studies, solutions will be explored at this point.

Table 8. *CV and Validation results of Application-3 (nonlinear features) for ReliefF feature selector and BT classifier.*

	CV results					Validation results				
Number of features	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1-score**	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**	**F1-score**
2	0.5159	0.5159	0.8790	0.5113	0.5133	0.5106	0.5106	0.8777	0.5060	0.5080
3	0.5567	0.5567	0.8892	0.5505	0.5529	0.5410	0.5410	0.8853	0.5344	0.5369
4	0.6140	0.6140	0.9035	0.6087	0.6107	0.5932	0.5932	0.8983	0.5884	0.5899
5	0.6299	0.6299	0.9075	0.6247	0.6267	0.6218	0.6218	0.9055	0.6176	0.6192
6	0.6267	0.6267	0.9067	0.6207	0.6229	0.6262	0.6262	0.9066	0.6231	0.6233
7	0.6399	0.6399	0.9100	0.6341	0.6363	0.6312	0.6312	0.9078	0.6251	0.6269
8	0.6580	0.6580	0.9145	0.6522	0.6545	0.6518	0.6518	0.9130	0.6471	0.6487
9	0.6529	0.6529	0.9132	0.6476	0.6497	0.6514	0.6514	0.9129	0.6460	0.6479
10	0.6550	0.6550	0.9138	0.6500	0.6520	0.6546	0.6546	0.9137	0.6512	0.6523
11	0.6619	0.6619	0.9155	0.6582	0.6597	0.6516	0.6516	0.9129	0.6492	0.6502
12	0.6568	0.6568	0.9142	0.6523	0.6541	0.6502	0.6502	0.9126	0.6465	0.6479
13	0.6588	0.6588	0.9147	0.6544	0.6561	0.6552	0.6552	0.9138	0.6503	0.6523
14	0.6581	0.6581	0.9145	0.6532	0.6551	0.6524	0.6524	0.9131	0.6494	0.6506
15	0.6627	0.6627	0.9157	0.6575	0.6595	0.6564	0.6564	0.9141	0.6543	0.6550
16	0.6755	0.6755	0.9189	0.6696	0.6717	0.6678	0.6678	0.9170	0.6631	0.6648
17	0.6685	0.6685	0.9171	0.6628	0.6649	0.6658	0.6658	0.9165	0.6598	0.6621
18	0.6703	0.6703	0.9176	0.6649	0.6670	0.6658	0.6658	0.9165	0.6606	0.6625
19	0.6733	0.6733	0.9183	0.6681	0.6702	0.6606	0.6606	0.9152	0.6563	0.6580
20	0.6713	0.6713	0.9178	0.6649	0.6675	0.6694	0.6694	0.9174	0.6631	0.6654
21	0.6765	0.6765	0.9191	0.6710	0.6732	0.6746	0.6746	0.9187	0.6704	0.6721
22	0.6759	0.6759	0.9190	0.6704	0.6726	0.6676	0.6676	0.9169	0.6636	0.6651
23	0.6704	0.6704	0.9176	0.6647	0.6670	0.6740	0.6740	0.9185	0.6701	0.6717
24	0.6717	0.6717	0.9179	0.6650	0.6676	0.6646	0.6646	0.9162	0.6581	0.6604
25	0.6792	0.6792	0.9198	0.6731	0.6754	0.6666	0.6666	0.9167	0.6608	0.6630
26	0.6716	0.6716	0.9179	0.6658	0.6681	0.6750	0.6750	0.9188	0.6703	0.6721
27	0.6751	0.6751	0.9188	0.6689	0.6713	0.6688	0.6688	0.9172	0.6632	0.6651
28	0.6715	0.6715	0.9179	0.6647	0.6672	0.6688	0.6688	0.9172	0.6641	0.6659
29	0.6737	0.6737	0.9184	0.6676	0.6700	0.6682	0.6682	0.9171	0.6631	0.6649
30	0.6771	0.6771	0.9193	0.6706	0.6730	0.6714	0.6714	0.9179	0.6654	0.6677
31	0.6681	0.6681	0.9170	0.6616	0.6641	0.6678	0.6678	0.9170	0.6616	0.6640
32	0.6709	0.6709	0.9177	0.6655	0.6676	0.6726	0.6726	0.9182	0.6678	0.6696
33	0.6766	0.6766	0.9192	0.6707	0.6730	0.6630	0.6630	0.9158	0.6566	0.6587
34	0.6780	0.6780	0.9195	0.6715	0.6740	0.6652	0.6652	0.9163	0.6607	0.6621
35	0.6701	0.6701	0.9175	0.6637	0.6662	0.6704	0.6704	0.9176	0.6665	0.6679
36	0.6687	0.6687	0.9172	0.6622	0.6647	0.6716	0.6716	0.9179	0.6661	0.6683
37	0.6760	0.6760	0.9190	0.6696	0.6722	0.6678	0.6678	0.9170	0.6630	0.6648
38	0.6727	0.6727	0.9182	0.6660	0.6686	0.6674	0.6674	0.9169	0.6614	0.6637
39	0.6739	0.6739	0.9185	0.6680	0.6703	0.6606	0.6606	0.9152	0.6566	0.6580
40	0.6729	0.6729	0.9182	0.6671	0.6694	0.6638	0.6638	0.9160	0.6584	0.6605
41	0.6723	0.6723	0.9181	0.6660	0.6685	0.6670	0.6670	0.9168	0.6620	0.6638
42	0.6748	0.6748	0.9187	0.6684	0.6709	0.6656	0.6656	0.9164	0.6594	0.6619
43	0.6769	0.6769	0.9192	0.6701	0.6726	0.6718	0.6718	0.9180	0.6666	0.6686
44	0.6696	0.6696	0.9174	0.6633	0.6657	0.6640	0.6640	0.9160	0.6586	0.6607
45	0.6717	0.6717	0.9179	0.6653	0.6677	0.6628	0.6628	0.9157	0.6564	0.6588
46	0.6666	0.6666	0.9167	0.6598	0.6624	0.6648	0.6648	0.9162	0.6597	0.6618
47	0.6718	0.6718	0.9180	0.6649	0.6674	0.6658	0.6658	0.9165	0.6596	0.6620
48	0.6715	0.6715	0.9179	0.6649	0.6674	0.6668	0.6668	0.9167	0.6606	0.6628
49	0.6727	0.6727	0.9182	0.6655	0.6681	0.6694	0.6694	0.9174	0.6629	0.6653
50	0.6738	0.6738	0.9185	0.6671	0.6696	0.6626	0.6626	0.9157	0.6574	0.6596
51	0.6673	0.6673	0.9168	0.6601	0.6629	0.6662	0.6662	0.9166	0.6597	0.6622
52	0.6673	0.6673	0.9168	0.6602	0.6629	0.6678	0.6678	0.9170	0.6627	0.6646

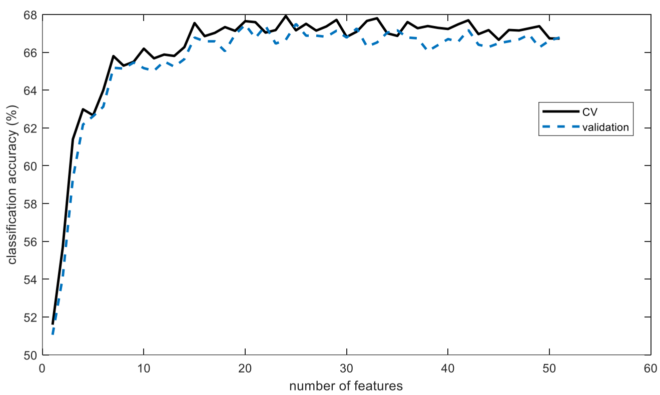

Figure 5. *The variation of cross-validation and validation accuracies according to the number of features.*

Table 9. *Complexity matrix of CV results where the highest accuracy was reached (67.92%)*

Predicted class / Real class	Wake	Non-REM 1	Non-REM 2	Non-REM 3	REM	Total	accuracy
Wake	1872	625	303	76	124	3000	0.6240
Non-REM 1	641	1565	430	98	266	3000	0.5216
Non-REM 2	324	472	1552	317	335	3000	0.5173
Non-REM 3	49	61	218	2630	42	3000	0.8766
REM	101	190	209	45	2455	3000	0.8183

3.6. Results of Application-5 for the Whole Features

In the applications made with the feature sets described above, the performance of the individual feature sets was observed and compared. In this application, in which all features are used, only the ReliefF method (because ReliefF gives the highest result in all applications) was changed from 2 to 168 and the classification results were obtained with the Bag Tree method (bag tree method shows the best performance). Accordingly, while the highest accuracy in cross-validation was 66.21% for 157 features, 66.44% classification accuracy was achieved for 84 features in validation. On the other hand, when the number of features is increased up to approximately 40, classification accuracy increases significantly, while increasing the number of features after 40 does not cause a significant increase in classification performance. Therefore, it can be interpreted that the first 40 features provide sufficient information for the classification process. These features are listed as follows (by selection order):

1. Kolmogorov complexity of EOG-diff signal (nonlinear)
2. Higuchi fractal dimension of EOG-left signal (nonlinear)
3. Fuzzy Entropy of EEG signal (nonlinear)
4. Kolmogorov complexity of EEG signal (nonlinear)
5. Lampel-Ziv complexity of EEG signal (nonlinear)
6. Hjorth mobility of EOG-diff signal (nonlinear)
7. Correlation dimension of EEG signal (nonlinear)
8. MMD-Maximum Minimum Distance of EEG signal (time)
9. Lyapunov exponent of EEG signal (nonlinear)
10. Sample entropy pf EOG-diff signal (nonlinear)
11. Total energy of EEG signal (time)
12. Approximate entropy of EEG signal (nonlinear)
13. Hurst Exponent of EEG signal (nonlinear)
14. Asymetric index of EEG signal (MDFA)
15. Renys' entropy of EOG-left signal (nonlinear)
16. Asymmetric index of EOG-diff signal (MDFA)
17. ZCR of EMG signal (time)
18. Sample entropy of EEG signal (nonlinear)
19. Total energy of EOG-left signal (time)
20. Lampel-Ziv complexity of EOG-left signal (nonlinear)
21. Higuchi Fractal Dimension of EOG-left signal (nonlinear)
22. Asymmetric index of EOG-left signal (MDFA)

23. Hjorth complexity of EMG signal (nonlinear)
24. Correlation dimension of EOG-diff signal (nonlinear)
25. Mean value of time domain EEG signal (time)
26. The difference of 0.5-2 Hz between the current and previous epoch in the EOG-diff signal (frequency)
27. Hjorth mobility of EMG signal (nonlinear)
28. Permutation entropy of EMG signal (nonlinear)
29. Fuzzy entropy of EOG-left signal (nonlinear)
30. Permutation entropy of EEG signal (nonlinear)
31. Mean value of time domain EOG-left signal (time)
32. Kolmogorov complexity of EOG-left signal (nonlinear)
33. Mean value of time domain EOG-diff signal (time)
34. Renys' entropy of EEG signal (nonlinear)
35. MMD-Maximum Minimum Distance of EOG-left signal (time)
36. Multifractal spectrum corrs. to max sing exp [16] of EEG signal (MDFA)
37. MMD-Maximum Minimum Distance of EOG-diff signal (time)
38. Hjorth mobility of EEG signal (nonlinear)
39. Approximate entropy of EOG-diff signal (nonlinear)
40. Renys' entropy of EOG-diff signal (nonlinear)

As can be seen, the majority of these features are nonlinear features (26 nonlinear features). Then time (9 features), MDFA (4 features), and frequency (1 feature) features are seen. As can be understood from this, it has been observed that the effect of nonlinear features on the classification performance is quite high. On the other hand, when evaluated based on the signals used, it is seen that 16 features are used for EEG, 20 features are from EOG (12 for difference EOG and 8 for left-eye EOG), and 4 features for EMG. Again, it can be interpreted that EOG and EEG signals are more decisive.

When the obtained results were examined, it should be noted that the obtained accuracies are not high enough. The use of real noisy data is the major reason for this. Besides we recorded the data mostly from the patients. Also, we used balanced data in classifications to increase the accuracy of all stages and when the studies in literature conducted with balanced data are examined, it can be seen that the accuracies are not very high as the unbalanced ones. In the study [5], done with reals data, the authors have reached a classification accuracy of 76.30% in their 19-subject data.

4. CONCLUSIONS

For the search for an effective automatic sleep staging system that can be used in real data, we conducted a comparative sleep staging study in which feature types, feature selection methods, and classifiers were compared. PSG data (EEG, EOG, and EMG signals) were obtained from a total of 124 people, 31 of whom were healthy and 93 patients. By using the epochs obtained from this dataset, 28 time-domain, 28 frequency-domain, 52 nonlinear, and 60 MDFA features were extracted. For selecting appropriate features; CCA, SFS, Fisher Score, Chi-square, ReliefF, IG, and FCBF feature selection methods were used. 5 feature selection applications were conducted in the study. 4 of them were conducted to determine the best features in each feature sub-set (time-, frequency-, nonlinear- and MDFA-subsets) and the last one was done to determine the best-performing features among the whole feature set with 168 features. Also, k-NN, SVM, DT, and BT classifiers were used for the classification of stages. For each application, combinations of all the above-mentioned feature selection methods and classification methods were run within the application. ReliefF feature selection method and BT classifier have given the best performance in all applications. In the first application in which Time domain features were used, 20 features reached the highest accuracy as 67.92%. When frequency features were used, the maximum accuracy was obtained as 61.99% with 11 features. In the case of nonlinear features, 24 features were selected over 52 features giving the highest accuracy of 67.92%. Lastly, MDFA features resulted in a poor accuracy level of 49.95% with the best 50 features. As seen from these results, nonlinear features give the best performance. This situation was seen in Application-5, too. In application-5 in which all of the 168 features were selected by ReliefF feature selector, 40 features were selected to give better performance. There are 26, 9, 4, and 1 feature from the nonlinear, time, MDFA, and frequency set respectively. The highest accuracy was reached as 66.21% in this application. The obtained performance is yet not sufficient to use practically in real life but after some further improvements, it will hopefully be possible to use automatic sleep scoring systems. Also maybe the other PSG recordings would be used by the other signals.

Acknowledgement

This study is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) with project number: 119E127.

REFERENCES

[1] Aboalayon, K.A.I., Faezipour, M., Almuhammadi, W.S. & Moslehpour, S., (2016), Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation, Entropy, 18, 272; doi:10.3390/e18090272.

[2] Acharya, U.R., Bhat, S., Faust, O., Adeli, H., Chua, E.C., Lim, W.J., & Koh, J.E. (2015). Nonlinear Dynamics Measures for Automated EEG-Based Sleep Stage Detection. European Neurology, 74, 268 - 287.

[3] Arslan, R. S., Ulutaş, H., Köksal, A.S., Bakır, M., Çiftçi, B. (2022), "Automated sleep scoring system using multi-channel data and machine learning", Computers in Biology and Medicine 146, 105653.

[4] Azhagusundari, B., & Thanamani, A.S. (2013). Feature Selection based on Information Gain. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075.

[5] Balci, M., Tasdemir, S., Ozmen, G., Golcuk A., (2022), Machine Learning-Based Detection of Sleep-Disordered Breathing Type Using Time and Time-Frequency Features, Biomedical Signal Processing and Control, 73, 103402.

[6] Barbi, M., Chillemi, S., Garbo, A. D., Balocchi, R., Carpeggiani, C., Emdin, M., Michelassi, C. & Santarcangelo, E., (1998), Predictability and nonlinearity of the heart rhythm, Chaos, Solitons & Fractals, 9 (3), 507-515.

[7] Boostani, R., Karimzadeh, F., & Nami, M. (2017). A comparative review on sleep stage classification methods in patients and healthy individuals. Computer methods and programs in biomedicine, 140, 77-91.

[8] Bose, R., Pratiher, S., & Chatterjee, S. (2019). Detection of epileptic seizure employing a novel set of features extracted from multifractal spectrum of electroencephalogram signals. IET Signal Process., 13, 157-164.

[9] Breiman, L. (2004). Bagging predictors. Machine Learning, 24, 123-140.

[10] Chaozhen, G., Siyang, L., Fulan, L. & Weichao, X., (2015), Sleep EEG staging based on Hilbert-Huang transform and sample entropy, 2015 International Conference on Computational Intelligence and Communication Networks

[11] Chattopadhyay, G. & Chattopadhyay, S., (2014), Study on statistical aspects of monthly sunspot number time series and its long-range correlation through detrended fluctuation analysis, Indian Journal of Physics, 88 (11), 1135-1140.

[12] Chi-Square. Available online: https://towardsdatascience.com/chi-square-test-for-feature-selection-in machine-learning-206b1f0b8223

[13] Chlon, L., Song, A.H., Subramanian, S., Soulat, H., Tauber, C., Ba, D. & Prerau, M., Multitaper Spectral Estimation HDP-HMMs for EEG Sleep Inference, arXiv:1805.07300v1 [stat.ML] 18 May 2018, (last visited on 7 October 2021)

[14] Di Matteo, T., Aste, T. & Dacorogna, M. M., (2003), Scaling behaviors in differently developed markets, Physica A: Statistical Mechanics and its Applications, 324 (1-2), 183-188.

[15] Fix, E. & Hodges, J.L., (1951), An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation, International Statistical Review / Revue Internationale de Statistique Vol. 57, No. 3 (Dec. 1989), pp. 233-238

[16] Erdoğan, N. K., (2017), BİST100 Endeksinin Çokfraktallı Eğimden Arındırılmış Dalgalanma Analizi, Journal of Current Researches on Business and Economics, 7 (2), 555-564.

[17] Farag, A.F., El-Metwally, S.M., & Morsy, A.A. (2014). A Sleep Scoring System Using EEG Combined Spectral and Detrended Fluctuation Analysis Features. Journal of Biomedical Science and Engineering, 07, 584-592.

[18] Faust, O., Razaghi, H., Barika, R., Ciaccio, E. J., & Acharya, U. R. (2019). A review of automated sleep stage scoring based on physiological signals for the new millennia. Computer methods and programs in biomedicine, 176, 81–91. https://doi.org/10.1016/j.cmpb.2019.04.032

[19] Fiorillo, L., Puiatti, A., Papandrea, M., Ratti, P. L., Favaro, P., Roth, C., Bargiotas, P., Bassetti, C. L., & Faraci, F. D. (2019). Automated sleep scoring: A review of the latest approaches. Sleep medicine reviews, 48, 101204. https://doi.org/10.1016/j.smrv.2019.07.007

[20] Ghimatgar H., Kazemi K., Helfroush M. S., Aarabi A. (2019) "An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov Model", Journal of Neuroscience Methods, 324, 1082320

[21] Göğüş, F. Z., 2020, Osas Hastaları için CPAP Cihazlarındaki Optimum Basıncın Yapay Zeka ile Tahmini, PhD thesis, Konya-Türkiye

[22] Hassan, A. R., & Bhuiyan, M.I.H., (2016a). 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. https://doi.org/10.1016/j.jneumeth.2016.07.012

[23] Hassan, A.R., & Bhuiyan, M.I.H., (2016b), Automatic sleep scoring using statistical features in the EMD domain and ensemble methods, Biocybernetics and Biomedical Engineering 36, 248-255

[24] Hassan, A.R., & Subasi, A. (2016). Automatic identification of epileptic seizures from EEG signals using linear programming boosting. Computer methods and programs in biomedicine, 136, 65-77.

[25] Hassan, A.R., & Bhuiyan, M.I.H., (2017), An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting, Neurocomputing219, 76–87

[26] Huang, W., Guo, B., Shen, Y., Tang, X., Zhang, T., Li, D., & Jiang, Z. (2020). Sleep staging algorithm based on multichannel data adding and multifeature screening. Computer methods and programs in biomedicine, 187, 105253.

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

[28] Japkowicz, N., Shah, M., (2014), Evaluating Learning Algorithms: A Classification Perspective, Cambridge University Press.

[29] Jiang, D., LU, Y., 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

[30] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A. & Stanley, H. E., (2002), Multifractal detrended fluctuation analysis of nonstationary time series, Physica A: Statistical Mechanics and its Applications, 316, 87-114.

[31] Khourdifi, Y., & Bahaj, M. (2018). Feature Selection with Fast Correlation-Based Filter for Breast Cancer Prediction and Classification Using Machine Learning Algorithms. 2018 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), 1-6.

[32] Korkailen, H., Leppanen, T., Duce, B., Kainulainen, S., et. al. (2021), "Detailed Assessment of Sleep Architecture With Deep Learning and Shorter Epoch-to-Epoch Duration Reveals Sleep Fragmentation of Patients With Obstructive Sleep Apnea", IEEE Journals of Niomedical and Health Informatics, 25(7)

[33] Lee, J. M., Kim, D. J., Kim, I. Y., Suk Park, K., & Kim, S. I. (2004). Nonlinear analysis of human sleep EEG using detrended fluctuation analysis. Medical Engineering & Physics, 26(9), 773–776. https://doi.org/10.1016/j.medengphy.2004.07.002

[34] Li, C., & Xu, J. (2019). Feature selection with the Fisher score followed by the Maximal Clique Centrality algorithm can accurately identify the hub genes of hepatocellular carcinoma. Scientific reports, 9(1), 17283. https://doi.org/10.1038/s41598-019-53471-0

[35] Liu, L., Wang, Q., Adeli, E., Zhang, L., Zhang, H., & Shen, D. (2016). Feature Selection Based on Iterative Canonical Correlation Analysis for Automatic Diagnosis of Parkinson's Disease. Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 9901, 1–8. https://doi.org/10.1007/978-3-319-46723-8_1

[36] Liu, Z., & Sun, J. (2015). Sleep Staging from the EEG Signal Using Multifractal Detrended Fluctuation Analysis. 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), 63-68.

[37] Liua, Z., Suna, J., Zhanga, Y. & Rolfea, P., (2016), Sleep staging from the EEG signal using multi-domain feature extraction, Biomedical Signal Processing and Control 30, 86–97

[38] Márton, L. F., Brassai, S. T., Bakó, L. & Losonczi, L., (2014), Detrended Fluctuation Analysis of EEG Signals, Procedia Technology, 12, 125-132.

[39] Myles, A.J., Feudale, R.N., Liu, Y., Woody, N., & Brown, S.D. (2004). An introduction to decision tree modeling. Journal of Chemometrics, 18.

[40] Otis M. Solomon, Jr, PSD Computations Using Welch's Method, SANDIA REPORT, S A N D 91-1533 • U C -7 0 6, Unlimited Release, Printed December 1991, USA

[41] Peker, M. (2016). An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms. Neurocomputing, 207, 165-177.

[42] Peng, C., Havlin, S., Stanley, H.E., & Goldberger, A.L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos, 5 1, 82-7.

[43] Rückstieß, Thomas., Osendorfer, Christian., & van der Smagt, Patrick. (2011). Sequential Feature Selection for Classification. 10.1007/978-3-642-25832-9_14.

[44] Tian, P., Hua, J., Qi, J., Ye, X., Che, D., Ding, Y. & Peng, Y., (2017), A hierarchical classification method for automatic sleep scoring using multiscale entropy features and proportion information of sleep architecture, Biocybernetics and Biomedical Engineering 37, 263-271

[45] Urbanowicz, R.J., Meeker, M., Cava, W.L., Olson, R.S., & Moore, J. (2018). Relief-Based Feature Selection: Introduction and Review. Journal of biomedical informatics, 85, 189-203.

[46] Vapnik, V. N., (2000), The Nature of Statistical Learning Theory., New York: Springer

[47] Vaquerizo-Villar, F., Alvarez, D., Kheirandish-Goza, L., Gutiérrez-Tobal, G. C., Barroso-García, V., Crespo, A., Campo, F. d., Gozal, D. & Hornero, R., (2018), Detrended fluctuation analysis of the oximetry signal to assist in pediatric sleep apnoea-hypopnoea syndrome diagnosis, Physiological Measurement, 39 (11).

[48] Zero Crossing Rate[online], https://www.sciencedirect.com/topics/engineering/zero-crossing-rate

[49] Zhang, Y., Wang, B., Jing, J., Zhang, J., Zou, J., & Nakamura, M. (2017). A Comparison Study on Multidomain EEG Features for Sleep Stage Classification. Computational Intelligence and Neuroscience, 2017.

[50] Zhang, Z. & Guan, C., (2017), An Accurate Sleep Staging System with Novel Feature Generation and Auto-Mapping, IEEE International Conference on Orange Technologies (ICOT)