IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 March 2024, accepted 28 April 2024, date of publication 4 June 2024, date of current version 1 July 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3409441*

RESEARCH ARTICLE

# Detection of Bruxism Using Inverse Discrete Wavelet Transformed Reconstructed Band Limited EEG Signals by Group Wise Feature Ranking

**AINUL ANAM SHAHJAMAL KHAN[1], (Graduate Student Member, IEEE), SHAIKH ANOWARUL FATTAH[2], (Senior Member, IEEE), MUHAMMAD QUAMRUZZAMAN[1], AND MOHAMMAD SAQUIB[3], (Senior Member, IEEE)**

[1]Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh
[2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
[3]Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA

Corresponding author: Shaikh Anowarul Fattah (fattah@eee.buet.ac.bd)

This work was supported by the Information and Communication Technology (ICT) Division, Government of the People's Republic of Bangladesh, under Grant 86 (fiscal year 2020-21) and Grant 61 (fiscal year 2021-22).

**ABSTRACT** Bruxism is a sleep disorder which is manifested by unintentional grinding and clenching of teeth during sleep. An automated sleep bruxism recognition system using single channel EEG data is proposed in this paper which is based on Inverse Discrete Wavelet Transformed Reconstructed Band Limited (IDWT-RBL) signals. These band limited EEG signals are used for extracting various features. Instead of using handcrafted features, feature reduction is done by ranking using statistical test scoring combined with classifier testing. This technique finds optimal features, reduces model complexity, lowers computational burden and increases model interpretability. Abundant features from time, frequency and statistical domains are used primarily so that no significant feature is ignored. Choosing a good subset of features from a larger set is a challenge for data with low sample size. To meet this challenge, a Group wise Feature Ranking (GFR) technique is introduced to reduce feature dimension. After statistical ranking and group wise averaging the scores, most significant groups of features are chosen. The proposed scheme is validated on a publicly available dataset. This process is examined for both unlabeled and labeled sleep stage. For segments with unlabeled sleep stage, cubic Support Vector Machine (SVM) performed best for F3C3 channel using 6 features with an accuracy of 97.83%. For segments with labeled sleep stage, F3C3 and REM sleep stage using 10 features performed best with 98.39% accuracy. The accuracy of proposed method is superior to most recent bruxism detection techniques. Finally, the GFR technique is applied to detect sleep disordered breathing (SDB). The outstanding performance to detect SDB clearly demonstrates the versatility of the proposed GFR technique to solve binary classification problem using EEG signals. Moreover, the introduced GFR technique enhances confidence and pellucidity of the system as it is more explainable.

**INDEX TERMS** Bruxism, classification, EEG signal, sleep stage, feature ranking, feature optimization.

## I. INTRODUCTION

The unintentional and unwanted grinding and clenching of teeth while one is asleep are the main clinical symptoms of sleep bruxism. The sleep bruxism possesses risk of various illnesses in the stomatognathic system which include tooth breakage, tooth wear etc. [1]. The dominance of bruxism is found to be 13.6% and it is comorbid with asthma, rhinitis, sinusitis, mental problem, back pain, allergy etc. [2]. Obstructive sleep apnea, gastroesophageal

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

pH/gastroesophageal reflux disease are also associated factor of bruxism [3]. Patients having previous implants may be fractured or damaged by the occlusal force [4]. Bruxism may be comorbid with acute neurological diseases such as encephalitis, intracerebral hemorrhage, traumatic brain injury, hypoxic-ischemic encephalopathy and acute ischemic stroke [5]. This may lead to recurrent aspiration pneumonia, endotracheal tube dislodgement etc. In its extreme case, a patient may face tongue amputation [5]. Recently, incidence of bruxism increased significantly during COVID-19 pandemic [6], [7]. Hence, detection and treatment of sleep bruxism is very important.

Bruxism can be diagnosed using clinical examination as well as self-reporting of the patient [8]. Polysomnography (PSG) is the standard method to detect bruxism. PSG has some pitfalls which include invasiveness to sleep in new environment, high cost, devoted infrastructure and lab [9], [10]. PSG requires continuous assistance by an expert person [11]. If number of patients increase, many expert persons need to be appointed for a long duration [9]. There is always a chance of human error in identifying the targeted instances. Hence, there is a demand for detecting bruxism automatically during sleep. Various studies are conducted to automatically detect bruxism.

In [12], EMG and ECG signals are used to detect bruxism episodes with the help of a neural network. In [13], convolutional neuronal network (CNN) is used to differentiate bruxism events by analyzing mandibular movements. Ultra-miniature EMG system is developed and used in [14] to detect bruxism. Force-based stress sensors are used for detecting bruxism [15]. In [16], EEG, EOG, EMG and ECG data are captured and a satisfactory detection performance is achieved.

The methods described in [12], [13], [14], [15], and [16] utilized multimodal signals to detect bruxism which resulted in complex systems. Some of them used device inserted into mouth or outer surface of mandible which creates discomfort to patients. Most of them used handcrafted features. Hence, to reduce the complexity of the system and ease the patient, a method with less modes and explainable features is required.

## II. RELEVANT WORKS

Some studies are done to use less number of channels without insertion of device in patients mouth. These studies also considered sleep stage as a biomarker for bruxism detection.

First study of bruxism detection based on labeled sleep stage annotations is reported in [17] where the rapid eye movement (REM) and wake stages are considered and EMG and/or ECG signals are utilized. They used normalized value of power spectral density as feature. 149 segments from bruxism patients and 95 segments from normal subjects are considered for their study. Each segment is of one minute duration. However, the study has some limitations. One of the major limitations of this study is the use of only two sleep stages without examining the other sleep stages. No justification is reported for using only two sleep stages. A very small number of segments are taken into consideration for the study. This number is very small compared to the whole night's data. Moreover, on which criteria or reason this small number of segments are chosen is not reported. In preprocessing stage, they used low pass filter with a cut-off frequency of 25Hz. The use of such low cut-off frequency might lose high frequency information. The reason for using such low cut-off frequency in preprocessing stage is not mentioned. Authors claimed higher accuracy but no comparison with previous study is presented. They reported that their method took much less time in comparison to the traditional systems without presenting any computational complexity comparison. Handcrafted feature of frequency domain is used in their study. Hence, time domain variations are not captured.

First study of bruxism detection using only EEG signal with sleep stage annotations is conducted in [18]. 140 segments from bruxism patients and 84 segments from normal subjects are considered for this study. Each segment is of one minute duration. Here, the power spectral density of a single-channel EEG data is utilized with the information of two sleep stages: S1 and REM. This study has some similar limitations as the study done in [17], such as the use of only two sleep stages without mentioning any justification. A very small number of segments are taken into consideration which are very few compared to the whole night's data. No reasoning or selection criteria is mentioned for choosing these small number of segments. Low pass filtering with a cut-off frequency of 25Hz is used in preprocessing which is prone to lose high frequency information. The reason behind this cut-off frequency selection is not reported. In their study, they used EEG data of two channels: C4-P4 and C4-A1. No explanation is given for choosing these channels and ignoring other channels in their study. In result comparison, they compared their result of bruxism detection with neruomascular disease detection using EMG, which is a rather irrelevant comparison. They also compared their result with general sleep disorder detection study not confined to bruxism detection. The only case where they compared their work with bruxism detection used EMG and ECG. This comparison is also not relevant because their work is solely based on bruxism detection using EEG. Some handcrafted features are used. They used only frequency domain features which is prone to not capturing temporal variations.

An approach very similar to [18] is reported in [19] using ten classifiers and a majority voting method. Their study has the same limitations as the study reported in [18]. Moreover, in comparison of their method with the prevailing approaches they have specified some irrelevant comparisons. They compared their bruxism detection technique to three sleep stage classification studies, two obstructive sleep apnea studies and one sleep apnea classification study. While comparing their proposed hybrid classifier with the present hybrid classifiers, they compared with the hybrid classifiers

which are used to detect cancer, heart disease, schizophrenia and Cushing's syndrome. These existing hybrid classifiers are not tested on the same data they used. Hence, this comparison does not indicate the true performance of their proposed classifier.

The aim of our study is to overcome the inadequacies of the methods described in [12], [13], [14], [15], [16], [17], [18], and [19]. Hence, we tried to use single EEG channel to detect bruxism which can result in simple system without the use of devices inserted into mouth or outer surface of mandible that can result comfort to the patients. Without using handcrafted features we selected important feature subset from a larger feature set so that no statistically important relevant feature is overlooked. Therefore, the study is not prone to arbitrariness of research methodology. Only frequency based analysis may not provide satisfactory performance because temporal variations are not properly captured in such case. Hence, without confining the study to only frequency domain features we extended it to incorporate statistical features, Hjorth parameters and entropy based features. Without using only two sleep stages we examined all other sleep stages so that most appropriate sleep stage to detect bruxism can be found. Without using a very small number of chosen segments we considered all segments from whole night's sleep data. This results in bias free data collection. Without limiting the signal to a cut-off frequency of 25Hz, we used frequency up to 128Hz so that high frequency information are kept. A Power Spectral Density (PSD) based analysis is presented in Section III to support our motivation of using very high frequency EEG data.

Most of the EEG based studies carried out on bruxism detection focused on bruxism events or symptoms detection from a given EEG segment. Event detection study is less challenging as there are visible changes in EEG during symptoms. During bruxism events heart rate increases as well as masseter muscle moves. These motor events generate from neural firing. Hence, simple statistical measures are sufficient to differentiate these events. Event independent bruxism detection is more challenging as there might not be dominant changes in all segments of EEG data collected from a bruxism patient, which is rarely investigated, e.g. in [17], [18], and [19]. The limitations of these studies are discussed earlier. Our study tries to overcome the limitations of these studies.

A great portion of human lifespan is consumed by sleep. Apparently inactive sleep has a great influence on wellbeing of a person. Sleep disorders put great negative impact on human health. Sleep epochs (each of 30 second duration) are labeled by sleep experts using standard rules [20]. According to American Academy of Sleep Medicine (AASM) rule sleep stages are labeled into 5 classes such as Wake, NI, N2, N3, and REM. Rechtschaffen and Kales (R & K) rule [21] labels sleep stages into 6 classes. They are Wake, NI, N2, N3, N4 and REM. Broadly, sleep stages are categorized as rapid eye movement (REM) sleep and non-REM (NREM) sleep [22]. Detection of sleep stage is a decisive way to assess sleep quality as well as diagnose sleep disorders [23]. Changes in sleep stage patterns and lengths are observed in various sleep disorders [21]. Hence, sleep staging is very crucial. The effectiveness of sleep stage classification can be greatly improved using automated sleep stage classification and has vital clinical significance. Some studies focused on this issue and had very good performance [20], [21], [23], [24]. Due to relation of sleep stage and sleep disorder, sleep stage based bruxism detection is done in our study, which is termed here as classification using labeled sleep stage.

The main unique contributions of our study are stated below:

- A novel Group wise Feature Ranking (GFR) technique is introduced to rank the features. To the best of our knowledge, this technique is completely a new one and such technique is never used before to rank features in any study which considered feature ranking. The technique uses a very simple algorithm to rank features in a group wise manner. This holds grouped structure in final selected features. The EEG segments are decomposed and then ranked in a group wise manner. This retains the explainable property of features.
- Data with low sample size are characteristic of many machine learning applications. Choosing subset of the features that can discriminate the classes very well, becomes the main importance for the user. In this study, a very simple algorithm is introduced and used to find the good subset of features which resulted in excellent performance.
- In this study, feature dimension is reduced from many less computationally expensive features. Highly computationally expensive features such as fractal dimension, approximate entropy etc. are excluded in initial feature extraction. This ensures overall reasonable computational complexity.

Other distinct contributions regarding bruxism detection in this study are:

- To the best of our knowledge, for the first time, average PSD for all segments from patients and normal subjects for a particular sleep stage are demonstrated in this study to show the distinguishing spectral characteristics between bruxism patients and normal subjects.
- In this study, we consider both very high frequency EEG data along with the low frequency EEG data unlike conventional methods where only traditional bands are used ignoring the very high frequency signals.
- Rather than using handcrafted/arbitrary features, this study first starts with some neurophysiological inspired features and then attempts to find a smaller useful feature subset from a larger set to classify bruxism and healthy subjects.
- This study is the first attempt to classify bruxism and healthy subjects considering both labeled and unlabeled sleep stages, unlike existing methods that only considers few labeled sleep stages.

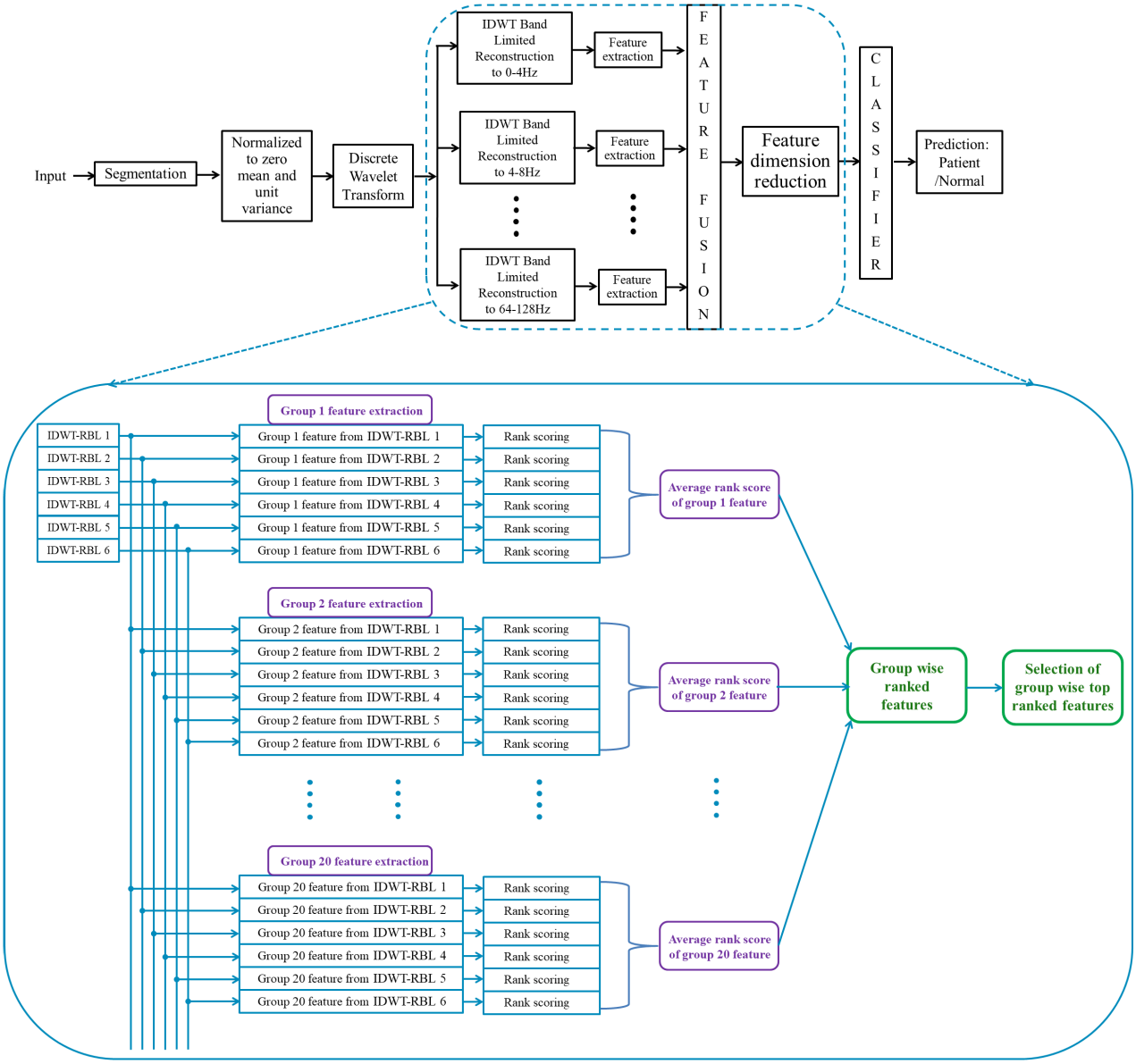


**FIGURE 1.** Block diagram of the proposed method.

- It is also the first attempt to use IDWT-RBL EEG signals to successfully classify bruxism and healthy subjects.
- All sleep stages are examined in this study to find the most significant sleep stage to classify bruxism and healthy subjects, unlike other studies where not all sleep stages are tested.
- It is also the first endeavor to investigate all available channels of EEG data in the dataset and channel selection is done experimentally which reduced data and made the system less complex to classify bruxism and healthy subjects.
- It examines all available data segments of each subject unlike existing methods where some segments are discarded.
- Features are ranked after decomposing the signal.
- The results of this study are superior to recently reported EEG based bruxism detection techniques

The comprehensive description of our scheme is presented next with proper rationalization and reasoning. The remaining writings are structured as follows. In Section III, a clear description of the main procedure is presented. In Section IV, the results obtained from labeled and unlabeled sleep stages are presented with logical justification. In Section V limitations of the study are discussed. Section VI concludes the paper with remarks and impending research directions.

## III. METHODS

Proposed method is illustrated through a block diagram in Fig. 1. Given EEG signal is first preprocessed and DWT

is applied. Band limited reconstructed signals are then obtained to extract different features. Features extracted from different bands are concatenated. Then feature dimension reduction is performed. Finally, supervised classifier is utilized.

The steps of the method are described in the following sections

### A. DATA COLLECTION

The EEG records corresponding to patients and normal subjects are collected from the cyclic alternating pattern (CAP) sleep database available in PhysioNet [25]. The data were logged at the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. The data include EEG channels, electroocoulogram, electromyogram (EMG) of the submentalis muscle, bilateral anterior tibial EMG, electrocardiogram signals etc..

Expert neurologists trained at the Sleep Center of the Ospedale Maggiore of Parma, Italy, provided the scoring of the sleep macrostructure, according to the Rechtschaffen Kales rules. The Cyclic Alternating Patterns (CAP) are also included in the dataset. CAP refers to episodic EEG activity that happens in NREM stages. Each CAP consists of activation and deactivation of cerebrum. These are termed as phase A and phase B. Phase A is separated into three subtypes: A1, A2 and A3.

In our study EEG recordings of all existing EEG channels in the dataset from six subjects (two with bruxism and four normal) are collected for processing. All subjects have 13 common EEG channels. Hence, experimentations on all channels could be done and it can be possible to find the best EEG channel for the classification task. Details of experimented data are specified in Table 1. Segments of one minute from each channel are considered for feature extraction. The segments are considered using and without using the labels of sleep stages during classification.

**TABLE 1. Details of experimented data.**

| **Data for Classification with Unlabeled Sleep Stage** | | | | |
|---|---|---|---|---|
| **Number of available segments** | **Segments from patients** | **Segments from normal subjects** | **Duration of available segments (Seconds)** | **Duration of available segments (Hours)** |
| **2424** | **603** | **1821** | **145440** | **40.40** |

| **Data for Classification with Labeled Sleep Stage** | | | | | |
|---|---|---|---|---|---|
| **Sleep stage** | **Number of available segments** | **Segments from patients** | **Segments from normal subjects** | **Duration of available segments (Seconds)** | **Duration of available segments (Hours)** |
| S1 | 95 | 51 | 44 | 5700 | 1.58 |
| S2 | 840 | 212 | 628 | 50400 | 14.00 |
| S3 | 236 | 70 | 166 | 14160 | 3.93 |
| S4 | 548 | 112 | 436 | 32880 | 9.13 |
| REM | 640 | 135 | 505 | 38400 | 10.67 |
| **Total** | **2359** | **580** | **1779** | **141540** | **39.32** |

### B. PREPROCESSING

At first, EEG data are downsampled to 256Hz to convert all the data to same sampling frequency. Then EEG data are segmented to one minute duration. Each EEG segment is preprocessed using z-score normalization to avoid bias due to acquisition process. Segments of data from patient and healthy person are collected serially. Except some faulty data, all segments are considered which reduces the chance of bias on result due to personal selection.

### C. BAND LIMITING BY DWT

Our objective is to find some distinguishable features which can easily differentiate two classes: normal and bruxism. It is required to observe or analyze the types of variations that are expected. The information content in EEG extends beyond the traditional maximum frequency band [26], [27]. Traditional consideration of the EEG up to 30Hz may not be sufficient. Hence, frequency bands higher than custom ceiling are considered in this study.

Only frequency based analysis may not provide satisfactory performance because temporal variations are not properly captured. Time resolution and frequency resolution, both need to be increased to capture time and frequency domain variations. Multi-resolution analysis is used here to extract band limited signals. Here, low frequency components of the EEG signals are investigated with less time resolution and higher frequency resolution. The high frequency components of the EEG signals are investigated with higher time resolution and lower frequency resolution.

A wavelet function of limited duration with zero mean can be expressed as in (1)

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{1}$$

where $a$ is known as dilation parameter and $b$ is known as translation parameter. Here, $\psi_{a,b}(t)$ is the mother wavelet

The continuous wavelet transform (CWT) coefficients of a function $f(t)$ is expressed as (2).

$$\begin{aligned} w_{a,b} &= \langle f(t), \psi_{a,b} \rangle \\ &= \int_{-\infty}^{\infty} \psi_{a,b} f(t)\, dt. \end{aligned} \tag{2}$$

To simplify the computation, parameters $a$ and $b$ are selected based on power of 2 i.e. dyadic.

$$a = a_0^m, \quad b = nb_0a_0^m \qquad m, n \in Z \tag{3}$$

If $a_0 = 2$ and $b_0 = 1$, the set of the wavelet becomes

$$\begin{aligned} \psi_{m,n}(t) &= a_0^{-m/2}\psi(a_0^{-m}t - nb_0) \\ &= 2^{-m/2}\psi(2^{-m}t - n). \end{aligned} \tag{4}$$

The DWT coefficients of a function $f(t)$ are now given as

$$\begin{aligned} w_{m,n} &= \langle f(t), \psi_{m,n}(t) \rangle \\ &= 2^{-m/2} \sum_{k=-\infty}^{\infty} f(t)\psi(2^{-m}t - n). \end{aligned} \tag{5}$$

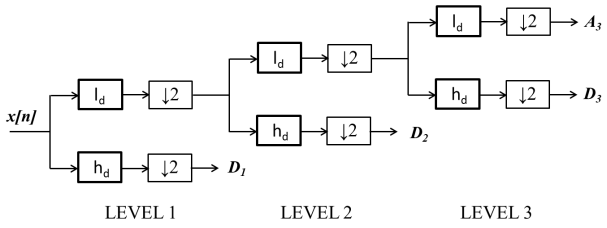


**FIGURE 2.** Analysis of a signal using DWT based on the filter banks.

The signal can be reconstructed at each level with the wavelet coefficients by using inverse DWT. Implementation of DWT based on multi resolution proposed by S.G. Mallat is accomplished using filter banks, [28] as shown in Fig. 2. For a three level decomposition $D_1$, $D_2$ and $D_3$ are detail coefficients of level 1, level 2 and level 3 respectively. $A_3$ denotes approximate coefficients of level 3. The signal at different levels can be reconstructed back again using inverse DWT based on the filter bank as shown in Fig. 3.

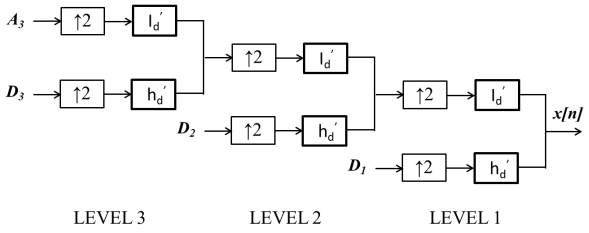


**FIGURE 3.** Reconstruction of a signal using inverse DWT based on the filter banks.

The EEG signal is decomposed into different levels using DWT and is reconstructed at each level again [28]. Five level decomposition is used. During reconstruction of band-limited signal of 0-4Hz, 4-8 Hz, 8-16Hz, 16-32 Hz, 32-64Hz and 64-128Hz are obtained as shown in Fig. 4.

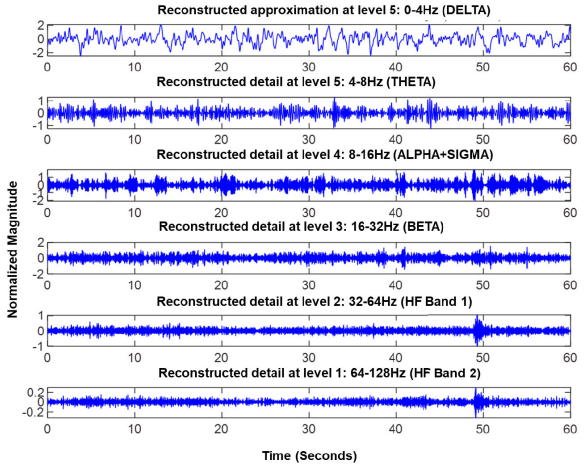


**FIGURE 4.** Reconstructed signals of one minute EEG of a subject.

Mostly, EEG signals are analyzed up to frequency range of 30Hz. In bruxism detection using EEG [18], [19], they used low pass filter in preprocessing stage to limit the EEG frequency up to 25Hz. As discussed earlier, EEG activity extends further than the customary frequency range. Hence high frequencies are kept in our analysis and 32-64Hz is considered as high frequency band 1 (HF Band 1) and 64-128Hz is considered as high frequency band 2 (HF Band 2). This is justified in Fig. 5 which demonstrates the average PSD by Welch's method for sleep stage 2 and 5 for F3C3 channel for all segments from patients and normal subjects. There is a clear distinction in high frequency components between normal subjects and patients.

To the best of our knowledge, for the first time, average PSDs for all segments from patients and normal subjects for a particular sleep stage are demonstrated in this study to show the distinguishing spectral characteristics between bruxism patients and normal subjects. As these PSDs are average of many segments they represent the characteristics of each group of subjects i.e. patient and normal.

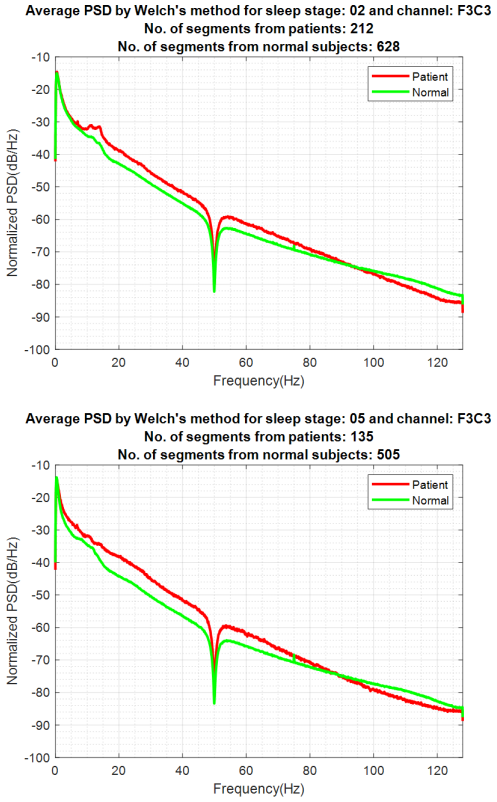


**FIGURE 5.** Average PSDs of patients and normal subjects for channel F3C3 for sleep stage 2 (top) and sleep stage 5 (bottom).

### D. NEUROPHYSIOLOGICAL INSPIRED FEATURES EXTRACTION

Bruxism patients have greater number of transient arousals during sleep, which is manifested by desynchronized EEG signals along with the shift in their frequencies [29]. It is well known that the cerebral triggering maintains a recurring structure (known as phase A) which is trailed by phases of deactivation (known as phase B), the overall phenomena is

**TABLE 2.** CAP summary of considered subjects.

| | patient a | patient b | normal a | normal b | normal c | normal d |
|---|---|---|---|---|---|---|
| **Total CAP time (sec)** | 8653 | 12236 | 7167 | 11601 | 5000 | 8554 |
| **CAP rate** | 0.39 | 0.58 | 0.35 | 0.50 | 0.29 | 0.46 |
| **No. of phase A1 events** | 207 | 228 | 141 | 462 | 164 | 235 |
| **No. of phase A2 events** | 35 | 47 | 106 | 24 | 34 | 80 |
| **No. of phase A3 events** | **145** | **207** | **108** | **60** | **61** | **50** |
| **Duration of phase A1 events (sec)** | 1186 | 1471 | 656 | 2863 | 1489 | 1724 |
| **Duration of phase A2 events (sec)** | 274 | 348 | 631 | 328 | 336 | 583 |
| **Duration of phase A3 events (sec)** | **2129** | **2335** | **1043** | **784** | **922** | **796** |

termed as the cyclic alternating pattern (CAP) [30]. Each Phase A is separated into three subtypes: A1, A2 and A3. CAP is the manifestation of sleep instability between sleep retaining phase A1 and stronger arousal phases A2 and A3. Bruxism patients have greater sleep instability (more phase A3) than that of healthy subjects [31].

For a better understanding of this fact, in Table 2, number and duration of different CAP phases (A1, A2 and A3) corresponding to some Bruxism patients and healthy subjects are presented. Moreover, for each subject, the CAP rate and total CAP time are also shown in the table. It can be noted from the table that the number and duration of Phase A3 are greater in case of Bruxism patients, as expected, which in turn will produce significant differences in spectral characteristics of the recorded EEG signals (between Bruxism patients and healthy subjects). In Fig. 6, in order to demonstrate the time-frequency characteristics in phases A1, A2 and A3, sample spectrogram plots for the Bruxism patient and healthy subject are shown. For a fair comparison, total duration of the EEG signal is kept similar for the two cases (Bruxism and healthy) and the starting point is selected where A1 phase is initiated. It is evidently perceived from the figure that the spectrogram of Bruxism patient exhibits dominancy of high frequency components (up to 16 Hz) while that of the healthy subject contains only low frequency components (mostly below 7 Hz). Above facts regarding the arousals (or instability) in sleep demonstrate that there is a strong chance of distinguished spectral variation in EEG signals during sleep for bruxism patients. Hence, some frequency domain features extracted from the EEG signal. Spectral entropy, spectral peak frequency and average spectral power (using Welch's method) are considered as potential frequency domain features in classifying bruxism patients and healthy persons.

Spectral entropy is actually use of the Shannon entropy perception to the power distribution of the Fourier-transformed signal. The spectral entropy portrays the information content

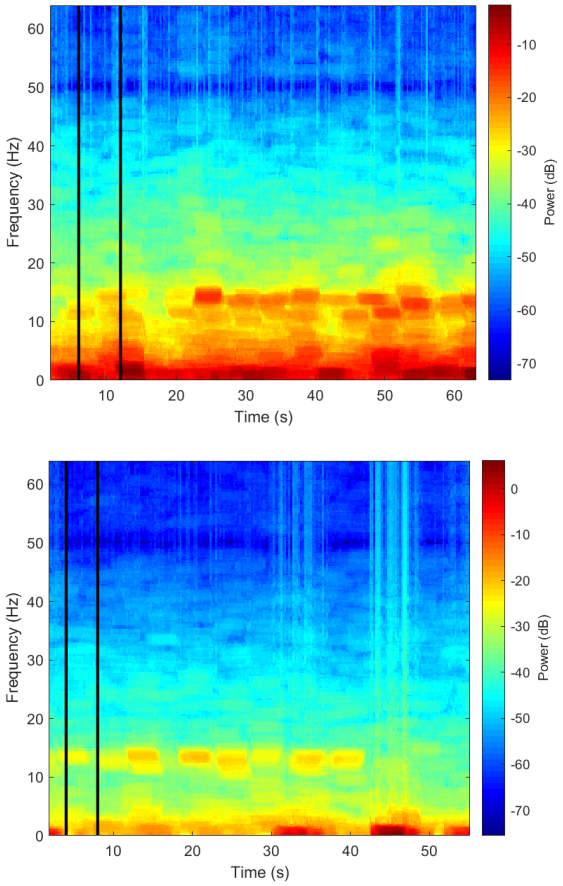


**FIGURE 6.** Spectrograms of consecutive phases A1, A2 and A3 of a patient (top) and a normal subject (bottom). The vertical black lines separate phase A1, A2 and A3. Channel considered: Fp2F4.

of the signal in spectral domain and thus it can be used proficiently to compare EEG signals obtained from Bruxism and healthy cases.

Welch's method is a popular technique to find the power spectral density of a signal. Here, the sub frames are overlapped and instead of using Periodograms, altered Periodograms (without using rectangular window) are averaged. It reduces noise in the estimated power spectra. Average spectral power obtained by using Welch's method can capture spectral variation in the EEG signals and are used as potential spectral feature in the proposed method. Moreover, spectral peak frequency is also considered as a distinguishing feature. In Fig. 6, spectra obtained in two cases (healthy and Bruxism) are shown where the dominance of the Bruxism case is clearly observed in the low frequency regions.

A negative-positive peak in the EEG signal with higher amplitude is regarded as a K-complex [32]. It has amplitude (measured peak to peak) which is twice as much compared to background activity in the preceding section. It has duration from 0.5 to 3 seconds [33]. This activity has also characteristics of biphasic or triphasic asymmetric slopes. Less number of K-complexes is observed in bruxism patients relative to healthy controls [34]. In [35], it is shown that

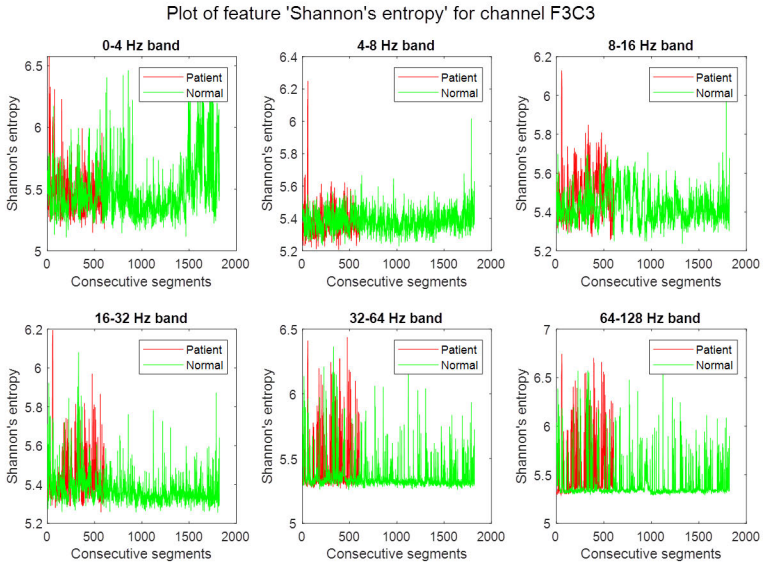


**FIGURE 7.** Plots of Shannon's entropy against consecutive segments for different bands for channel F3C3.

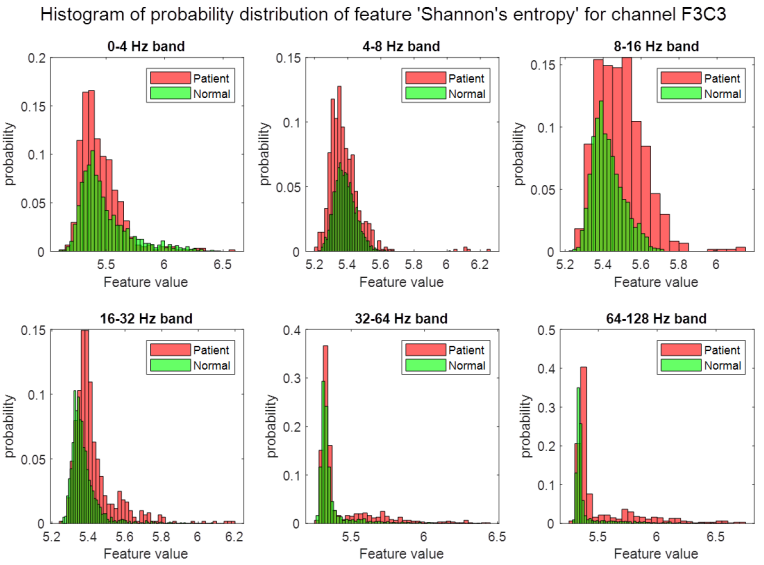


**FIGURE 8.** Plots of histograms of probability distribution of Shannon's entropy against feature values for different bands for channel F3C3.

the bruxism patients have distinct neurobehavioral pattern throughout the awake state. Greater EEG amplitude and smaller latency of visual evoked potential is observed in patients. A number of studies discovered that the behavioral pattern of the bruxism patient is different from that of the healthy subject and it is manifested by anxiety, depression, paranoid ideation etc. [36], [37]. These studies suggest alteration of neurobehavioral pattern in bruxism patients. If neurophysiology is changed due to disease this change should persist all the time. Therefore, magnitude variation of EEG may be present in patients throughout day and night. This motivated us to use statistical features namely, interquartile range (IQR), standard deviation, root mean square (RMS), mean, energy, mean absolute deviation, ranges of value, skewness, kurtosis, central sample moments of order 3, harmonic mean, sample mode and 25% trimmed mean in our study. Hjorth parameters (e.g. mobility and complexity) are established on the variation of the derivatives of the EEG signal. They have the ability to designate EEG signal characteristics simultaneously in the time and frequency domains and they are extensively used in analyzing nonstationary signals [38]. Hence, Hjorth parameters are considered as potential features for this study.

Entropy quantifies the information in a certain process. It is anticipated that there are differences in information content in EEG of patient and healthy control. This anticipation

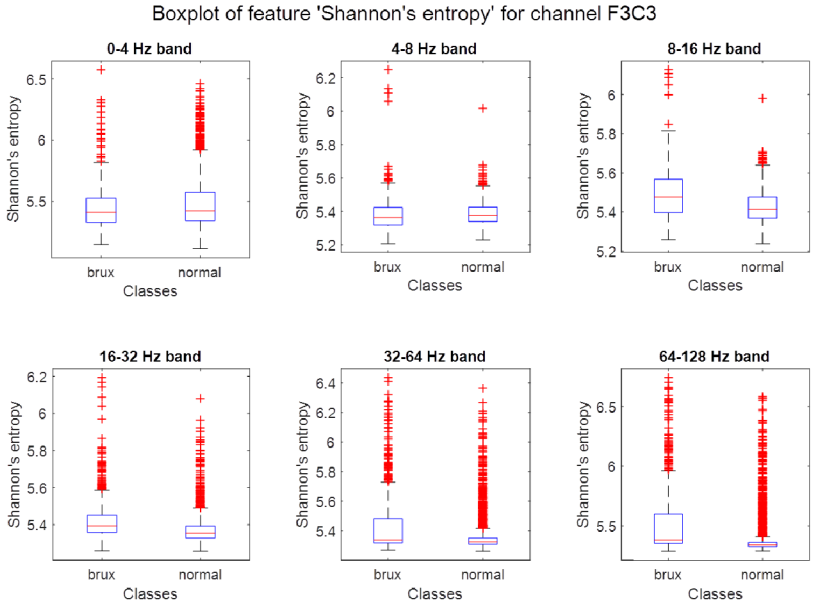


**FIGURE 9.** Boxplots of Shannon's entropy for different bands for channel F3C3.

is vindicated by using general feature plot, probability distribution and boxplots as shown in Fig. 7, Fig. 8 and Fig. 9. Two entropy based features, Shannon's entropy and log energy entropy are also included.

From Fig. 7, Fig. 8 and Fig. 9, the justification of using IDWT-RBL signal is eminent. The features are well separated in some of the bands. This is due to fact that multi resolution analysis has the ability to capture details of high frequencies with finer time resolution and that of low frequencies with wider time resolution.

Hence, the feature vector in this study comprises some spectral features, some statistical features, Hjorth parameters and entropy based features to find the characteristics of 6 IDWT-RBL signals for each 60 second segment.

### *E. PROPOSED GROUP WISE FEATURE RANKING (GFR) TECHNIQUE AND DIMENSIONALITY REDUCTION USING IT*

As 6 IDWT-RBL signals are obtained for each 60 second segment, particular feature from these six signals inherently exhibit group structures. It is more interpretable and interesting to retain the group structure of the features during feature selection. Moreover, as the features of same group come from different frequency bands, this may retain good frequency resolution too. This is our motivation for introducing GFR technique. Features of same group may be selected or not selected based on some criteria. In our proposed GFR technique, group selection is performed using statistical rank scoring to find the discriminative properties of the features. The strength of the proposed GFR technique lies in its simplicity while retaining excellent performance.

The GFR technique is introduced in this study. To the best of our knowledge, this technique is completely new and novel one and such technique is never used before to rank features. This technique and its convenience are described next. The robust performance of this technique is elaborated through experimentations in Section IV.

Firstly, 6 IDWT-RBL signals are constructed for each 60 second segment. Same features are obtained from 6 IDWT-RBL signals. For example, 6 standard deviation values are obtained from 6 IDWT-RBL signals. These 6 standard deviation values are termed as standard deviation feature group. 20 feature groups i.e. total 120 features (=6 bands$\times$20 features) are collected from 6 bands. All 120 features are ranked using Wilcoxon rank-sum test. Then average ranking of each feature group is calculated with respect to 6 bands as shown in Fig. 1. One feature group means 6 same features from 6 IDWT-RBL signals of each segment. 20 ranks are obtained for 20 feature groups. Then top ranked feature groups are chosen serially based on the average score of each group of features for dimensionality reduction. Result of this ranking procedure and its worth is presented in Section IV. Feature ranking and selection process for a fixed channel is depicted in Algorithm 1 during classification with unlabeled sleep stage.

Since features of same group comes from different band limited signals there might be less chance of correlation among these features. As aforementioned 20 features come from same IDWT-RBL signals similar spectrum content may cause more correlation among them. Hence top ranked features from different feature groups may not have good distinguishing influence. There is a more chance of class separation supremacy of classifier if we use group wise features.

For ranking the features Wilcoxon rank-sum test is utilized. The reason behind choosing this test is that it does not assume any hypothesis about the shape of the distribution of the

**Algorithm 1** Algorithm for Feature Ranking by GFR

**Input:** Feature data $\boldsymbol{D}$ with $n$ groups of features and $m$ bands, number of patients $p$, number of healthy controls $c$. $\boldsymbol{D}[{}_1X_1^1, {}_1X_2^1, \ldots, {}_1X_m^1, \ldots, {}_nX_m^1;$
$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
${}_1X_1^p, {}_1X_2^p, \ldots, {}_1X_m^p, \ldots, {}_nX_m^p;$
$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
${}_1X_1^{p+c}, {}_1X_2^{p+c}, \ldots, {}_1X_m^{p+c}, \ldots, {}_nX_m^{p+c}]$
**Output:** Ranks assigned to feature group 1 to $n$ by GFR, ${}_1R, {}_2R, {}_3R, \ldots, {}_nR$
**START**
**for** *i=1 to n* **do**
  **for** *j=1 to m* **do**
    **for** *k=1 to (p+c)* **do**
      Sort from smallest to largest, ${}_iX_j^k$
      Assign ranks, ${}_ir_j^k$
    **end**
    Initialize ranksum score, ${}_iW_j^{\text{pat}} \leftarrow 0$
    **for** *s=1 to p* **do**
      Summing the ranks, ${}_iW_j^{\text{pat}} \leftarrow {}_iW_j^{\text{pat}} + {}_ir_j^s$
    **end**
    Initialize ranksum score, ${}_iW_j^{\text{con}} \leftarrow 0$
    **for** *t=(p+1) to (p+c)* **do**
      Summing the ranks, ${}_iW_j^{\text{con}} \leftarrow {}_iW_j^{\text{con}} + {}_ir_j^t$
    **end**
    **if** *p<n* **then**
      ${}_iz_j \leftarrow {}_iW_j^{\text{pat}}$
    **else**
      ${}_iz_j \leftarrow {}_iW_j^{\text{con}}$
    **end**
  **end**
**end**
Z-score normalization of calculated ranks, ${}_iz_j \leftarrow |\text{zscore}({}_iz_j)|$
Averaging the rank score of each group,
**for** *i=1 to n* **do**
  ${}_iz_{\text{sum}} \leftarrow 0$
  **for** *j=1 to m* **do**
    Summing the ranks of each group, ${}_iz_{\text{sum}} \leftarrow {}_iz_{\text{sum}} + {}_iz_j$
  **end**
  ${}_ir_{\text{avg}} \leftarrow {}_iz_{\text{sum}}/m$
**end**
Sorting average ranks, ${}_ir_{\text{avg}}$
Assign ranks, ${}_iR$
**Return** ${}_iR$ ;
**END**

data [39]. Other tests assume certain shape of the distribution of the data. Hence, chance of error or bias due to distribution assumption in ranking is reduced. Moreover, as Wilcoxon rank-sum test does not start with any assumption it is a more comprehensive test. The test is a statistical test to decide if two sets of data come from same population. It checks for any difference in sum of ranks that exists between two groups. Standardized test-statistic is used to find the ranks [40], [41].

As there are 13 EEG channels available in the dataset we sought after finding out which channel is best suited to classify bruxism. We also wanted to find the classifier which can distinguish the classes remarkably well for this data. Hence, channel selection and classifier selection using extensive experimentation is also included in our study.

Instead of using a fixed channel averaged ranks of features for different channels could be used and then average of all channels may be considered into account. It should be noted that different channels have different spatial location. Some spatial location might have insignificant neurological role. Channels from these locations can lower the rank score of a significant feature. Hence, instead of using all the channels for feature ranking one best performing channel in terms of accuracy is selected.

At first all 20 feature groups are utilized using a single classifier on all 13 EEG channels. All 20 feature groups are extracted and ranked for best performing channel. Top ranked 5 feature groups are tested on 13 channels using 10 classifiers. Best performed classifier is used for next step. Top ranked different number of feature groups (4, 5, 6 etc.) are tested on 13 channels using 1 selected classifier. In this tuning manner optimized features and best channel is found.

Feature ranking and selection process for a fixed channel and fixed sleep stage is same as in Algorithm 1 during classification with labeled sleep stage. 20 feature groups are utilized using a single classifier on all 13 EEG channels and 5 sleep stages. Best sleep stage according to classification performance is selected. All 20 feature groups are extracted and ranked for data of best sleep stage and best performing channel. Top ranked 5 feature groups are tested on 13 channels and 1 selected sleep stage using 10 classifiers. Top performed classifier is chosen for next phase. Top ranked different numbers of feature groups are tested on top performed channel using 1 selected classifier. Then, optimized features and best channel is obtained.

### F. CUBIC SVM CLASSIFIER

In previous subsection it is shown that the classes are distributed to some extent in an overlapping manner. The two-dimensional space cannot separate the features as they are linearly non distinguishable. These features can be linearly distinguishable in the nonlinear feature space which is characterized indirectly by nonlinear kernel function. This type of kernel function is used in support vector machine (SVM) [42]. Hence, cubic SVM is used in this study to get the maximum class separability. It uses polynomial kernel which utilizes exact optimization rather than approximation. The core strength of SVM is its kernel trick. A kernel function provides techniques to avoid complicated calculations. Using it, data can be converted to higher dimensions and a hyperplane is obtained in this multidimensional space to

separate the classes in an efficient manner. Polynomial kernel of order 3 is used in this study. Cubic SVM's robustness for this dataset is tested by comparing with nine other types of classifiers namely: Decision Tree (DT), Medium DT, Liner Discriminant Analysis (LDA), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Cosine KNN, Bagged Trees, Subspace KNN and Boosted Trees.

## IV. RESULT AND DISCUSSION

There are two classification investigations: classification with unlabeled sleep stage and classification with labeled sleep stage. In the subsections to follow, the assessment technique of the proposed method is described with detailed steps with relevant explanation and discussion.

### A. PERFORMANCE MEASUREMENT MATRICES

Commonly used performance measurement matrices such as sensitivity ($S_{\text{sen}}$ ), specificity ($S_{\text{spec}}$) and accuracy ($A_{\text{acc}}$) are used to measure the performance of the method.

$$S_{\text{sen}} = \left(\frac{\text{TP}}{\text{FN} + \text{TP}}\right) \times 100 \tag{6}$$

$$S_{\text{spec}} = \left(\frac{\text{TN}}{\text{FP} + \text{TN}}\right) \times 100 \tag{7}$$

$$A_{\text{acc}} = \left(\frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}}\right) \times 100 \tag{8}$$

where, FN refers to number of false negative detections, TN refers to number of true negative detections, FP means number of false positive detections and TP means number of true positive detections.

5-fold cross validation is used to evaluate the performance of the classifiers. Features from all of the one minute segments from all subjects (both patients and normal subjects) are randomly distributed into 5 equal subgroups for this purpose. In each fold, four subgroups are used for training and residual one subgroup is utilized for testing. This procedure is reiterated 5 times.

Classification with unlabeled sleep stage and classification with labeled sleep stage for all channels and all sleep stages for all data from six subjects (two with bruxism and four normal) are tested. The results are summarized in next sections followed by discussion. Each accuracy is calculated by averaging the accuracies of at least 10 independent trials.

The term overall average accuracy is defined here as average of accuracies of all channels in case of unlabeled sleep stage. It is defined as average of accuracies of all channels for all sleep stages in case of labeled sleep stage.

### B. CLASSIFICATION WITH UNLABELED SLEEP STAGE

During classification with unlabeled sleep stage 20 features from each band are collected initially to test the classification performance. Initially Decision Tree (DT) classifier is chosen to test the performance with 5-fold cross validation and 100 independent trials for each classification result.

Experimental and investigative procedure is followed to find a good subset of features which would result in a reasonable classification performance. Twenty less computationally expensive features are chosen to reduce the computational complexity of the system. Computationally expensive features such as approximate entropy, fractal dimension etc. are not included in our study. DT classifier is chosen to test initial performance because of its simplicity and faster performance. 5-fold cross validation is chosen as it is a standard practice as well as the recent reported methods with which our results are compared also used it. In some random executions of 5-fold cross validations, the results are exceptionally outstanding. These outstanding results may not be produced in most of the executions of the system. Hence, reporting such results represents an excellent performance of the system. These true results give a false impression of the system. To avoid this false representation and become more confident on the proposed method, each 5-fold cross validation is repeated 100 times and average results are reported.

The results are summarized in Table 3. The plot of average accuracies of different channels is shown in Fig. 10. On average the F3C3 channel shows highest average accuracy of 92.99%. Overall average accuracy for all channels is 91.28±1.33%

Sensitivity and specificity are also highest for F3C3 channel. Sensitivity values are lower than specificity values. This is due to the unbalanced nature of the dataset. There are less disease data than normal data. The results are biased towards the normal class. Lower sensitivity indicates some wrong prediction in positive class i.e. disease class. This results in Type I error. During disease classification Type I error should be minimized as much as possible. In ideal case all disease segments should be classified as disease even at the cost of some wrong prediction of normal classes as disease classes. Hence, no patient segment remains undiagnosed. Our feature selection algorithm not only increased accuracy but also increased sensitivity which is described in later section. There is an increasing trend of accuracies from pre-frontal region towards parietal region for both hemispheres. 20 features from each 6 band limited signals for each 1 minute segment are used. It is observed that channel F3C3 has the highest average accuracy. Best performing channel i.e. F3C3 is considered for further feature optimization process. A feature vector for F3C3 is compiled using 20 features from each band. Those 20 features are ranked using absolute values of the standardized test-statistic of a two-sample unpaired Wilcoxon test. Fig. 11 shows the average score of each group of features. Then performance is tested with top five features namely sample mode, Shannon's entropy, complexity, spectral entropy and standard deviation.

Ten different classifiers including cubic SVM are used to check their performances. 5-fold cross validation and 10 independent trials are used for each classification. The results are summarized in Table 4, Table 5 and Table 6 and accuracies are shown in Fig. 12. It is evident that among all classifiers cubic SVM performs best with less deviation for

**TABLE 3.** Summary of classification performances with unlabeled sleep stage using 20 features.

| Classifier: Decision Tree | Fp2F4 | F4C4 | C4P4 | P4O2 | F8T4 | T4T6 | Fp1F3 | F3C3 | C3P3 | P3O1 | F7T3 | T3T5 | C4A1 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy (%) | 89.94 | 91.48 | 91.28 | 92.19 | 90.94 | 89.93 | 90.95 | **92.99** | 92.75 | 92.17 | 90.82 | 92.78 | 88.43 | **91.28±1.33** |
| Sensitivity (%) | 80.12 | 83.33 | 82.48 | 84.69 | 81.57 | 80.27 | 81.75 | **85.65** | 85.10 | 84.41 | 81.99 | 85.22 | 77.67 | **82.63±2.39** |
| Specificity (%) | 93.19 | 94.18 | 94.19 | 94.68 | 94.04 | 93.12 | 94.00 | **95.42** | 95.29 | 94.74 | 93.74 | 95.29 | 91.99 | **94.14±0.99** |

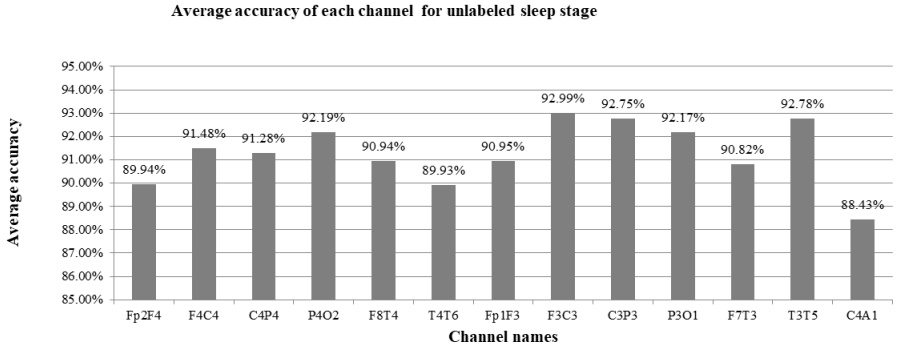


**FIGURE 10.** Average accuracy of each channel for classification with unlabeled sleep stage using 20 features.

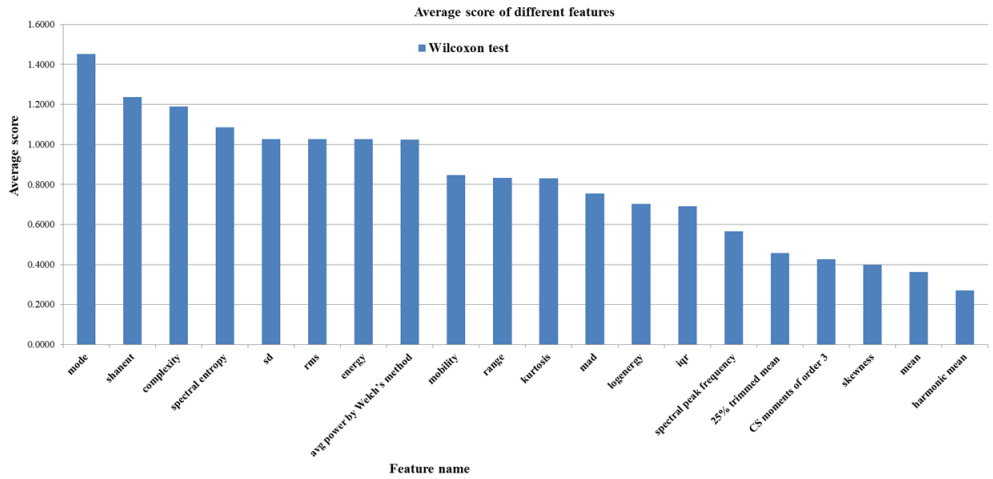


**FIGURE 11.** Average score of each group of features used for classification with unlabeled sleep stage.

all channels. Channel F3C3 showed best performance with 97.67% accuracy.

The individual channel wise and average sensitivity values are much higher for cubic SVM as expected. This reflects the ability of cubic SVM to detect the minority class in an unbalanced dataset. All other classifiers' average sensitivity values are below 90%. Only four of them crossed 80% mark. Hence, cubic SVM clearly outperforms other classifiers in terms of sensitivity. This is due to the fact that, in used dataset classes are distributed in an overlapping style. The two dimensional space cannot separate the features as they are linearly non distinguishable. These features are distinguished using nonlinear feature space which is characterized indirectly by nonlinear kernel function used in cubic SVM. High sensitivity also reduced Type I error. This is much desired for disease classification.

Average specificity is highest for Bagged Trees but it has low average sensitivity. SVM has very good average specificity but it has very poor sensitivity. Boosted Trees has also very good average specificity but it has low average sensitivity. Cubic SVM has very good average specificity along with highest average sensitivity. As our aim is mainly to reduce Type I error, our choice of using cubic SVM as classifier, seems justified. This is also reflected in overall average accuracy and individual channel accuracy values. Channel F3C3 continues to show better performance.

Hence, channel F3C3 and cubic SVM is selected for further feature optimization process.

**TABLE 4.** Summary of classification accuracies (%) with unlabeled sleep stage using 5 features and 10 classifiers.

| Classifier | Fp2F4 | F4C4 | C4P4 | P4O2 | F8T4 | T4T6 | Fp1F3 | F3C3 | C3P3 | P3O1 | F7T3 | T3T5 | C4A1 | Average accuracy |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DT** | 90.99 | 92.74 | 92.02 | 92.71 | 91.19 | 90.67 | 91.75 | 93.72 | 93.57 | 93.35 | 91.34 | 93.16 | 89.34 | **92.04±1.31** |
| **Medium DT** | 90.65 | 91.98 | 92.01 | 92.88 | 90.98 | 91.49 | 91.58 | 93.70 | 93.96 | 92.97 | 91.05 | 93.23 | 87.58 | **91.85±1.67** |
| **LDA** | 90.73 | 93.08 | 91.36 | 92.48 | 86.42 | 90.37 | 92.31 | 92.50 | 92.74 | 92.44 | 91.36 | 91.17 | 89.86 | **91.29±1.77** |
| **SVM** | 84.03 | 83.55 | 80.42 | 81.19 | 79.46 | 82.89 | 82.10 | 79.24 | 80.15 | 84.05 | 82.40 | 83.00 | 78.16 | **81.59±1.96** |
| **Cubic SVM** | **96.29** | **96.99** | **96.91** | **96.54** | **95.12** | **95.49** | **96.60** | **97.67** | **96.87** | 95.97 | **95.97** | **96.39** | **96.32** | **96.39±0.67** |
| **KNN** | 76.58 | 77.48 | 72.69 | 74.92 | 74.29 | 75.60 | 77.11 | 75.73 | 73.40 | 75.90 | 74.62 | 74.79 | 76.02 | **75.32±1.39** |
| **Cosine KNN** | 94.63 | 96.23 | 95.50 | 95.10 | 93.04 | 94.14 | 95.43 | 96.19 | 95.61 | 95.61 | 93.57 | 95.20 | 94.60 | **94.99±0.96** |
| **Bagged Trees** | 95.19 | 95.91 | 96.31 | 95.92 | 95.10 | 95.00 | 96.25 | 96.90 | 96.84 | **96.30** | 94.70 | 95.59 | 94.74 | **95.75±0.76** |
| **Subspace KNN** | 76.17 | 77.77 | 72.47 | 74.67 | 74.67 | 75.85 | 76.98 | 75.40 | 73.51 | 75.93 | 74.62 | 74.84 | 76.19 | **75.31±1.41** |
| **Boosted Trees** | 95.17 | 95.72 | 95.86 | 95.94 | 94.51 | 93.97 | 95.75 | 96.79 | 96.53 | 96.03 | 94.81 | 96.00 | 93.24 | **95.41±1.02** |

**TABLE 5.** Summary of classification sensitivities (%) with unlabeled sleep stage using 5 features and 10 classifiers.

| Classifier | Fp2F4 | F4C4 | C4P4 | P4O2 | F8T4 | T4T6 | Fp1F3 | F3C3 | C3P3 | P3O1 | F7T3 | T3T5 | C4A1 | Average sensitivity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DT** | 82.79 | 85.54 | 84.08 | 85.87 | 81.86 | 81.77 | 83.80 | 87.08 | 86.82 | 86.75 | 82.47 | 86.17 | 78.36 | **84.1±2.59** |
| **Medium DT** | 74.06 | 77.41 | 76.38 | 82.01 | 76.25 | 72.95 | 77.06 | 82.52 | 83.35 | 81.99 | 76.93 | 81.00 | 66.88 | **77.6±4.66** |
| **LDA** | 70.60 | 77.15 | 69.93 | 73.65 | 55.85 | 70.76 | 75.62 | 76.37 | 73.71 | 74.38 | 71.16 | 70.15 | 70.38 | **71.52±5.34** |
| **SVM** | 38.41 | 43.37 | 24.79 | 27.61 | 20.23 | 34.53 | 31.26 | 20.33 | 22.79 | 38.79 | 31.92 | 35.64 | 17.99 | **29.82±8.18** |
| **Cubic SVM** | **92.57** | **93.38** | **92.90** | **92.74** | **89.20** | **89.37** | **92.82** | **95.01** | **93.75** | **91.58** | **90.75** | **92.17** | **92.24** | **92.19±1.65** |
| **KNN** | 52.40 | 55.66 | 47.43 | 50.68 | 50.25 | 46.58 | 53.80 | 49.93 | 48.92 | 49.67 | 48.44 | 48.72 | 52.59 | **50.39±2.6** |
| **Cosine KNN** | 85.64 | 90.15 | 87.00 | 84.98 | 79.27 | 82.80 | 89.22 | 90.68 | 89.00 | 89.14 | 82.02 | 88.08 | 86.85 | **86.53±3.46** |
| **Bagged Trees** | 83.68 | 86.37 | 87.21 | 86.32 | 82.22 | 83.17 | 87.01 | 90.15 | 88.91 | 88.18 | 81.91 | 85.51 | 83.40 | **85.7±2.64** |
| **Subspace KNN** | 51.87 | 56.48 | 46.63 | 50.68 | 50.73 | 46.60 | 52.94 | 49.92 | 48.34 | 49.60 | 48.16 | 48.91 | 52.84 | **50.28±2.78** |
| **Boosted Trees** | 84.71 | 86.48 | 85.70 | 89.30 | 81.61 | 79.82 | 86.45 | 90.50 | 88.49 | 88.51 | 83.23 | 87.35 | 80.45 | **85.58±3.43** |

Then top 6,7 and 4 features from Wilcoxon test are also tested for all channels. For obtaining each classification result cubic SVM with 5-fold cross validation and 10 independent trials are used. The results are shown in Fig. 13. Using top 6 features highest average accuracy for all channels and highest accuracy using single channel (F3C3) is obtained.

Classification performance depends not only on number of features but also on feature quality. It is observed that the classification accuracy increases with the increase of number of top ranked features up to a certain limit. This increase in classification accuracy is expected as we are incorporating features from the top ranks. After certain numbers, features from the relatively lower rank are incorporated and thus the classification accuracy does not increase and it eventually decreases. It is expected that if the quality of a new feature to be incorporated is not relatively good with respect to the existing (already chosen) features, it may not enhance the classification performance and even it may reduce the existing classification performance.

Finally, top 6 features (sample mode, Shannon's entropy, complexity, spectral entropy, standard deviation and RMS),

**TABLE 6.** Summary of classification specificities (%) with unlabeled sleep stage using 5 features and 10 classifiers.

| Classifier | Fp2F4 | F4C4 | C4P4 | P4O2 | F8T4 | T4T6 | Fp1F3 | F3C3 | C3P3 | P3O1 | F7T3 | T3T5 | C4A1 | Average specificity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DT | 93.71 | 95.12 | 94.65 | 94.97 | 94.28 | 93.62 | 94.39 | 95.91 | 95.80 | 95.54 | 94.27 | 95.48 | 92.98 | **94.67±0.9** |
| Medium DT | 96.14 | 96.81 | 97.19 | 96.47 | 95.85 | 97.62 | 96.39 | 97.41 | 97.47 | 96.61 | 95.73 | 97.28 | 94.44 | **96.57±0.89** |
| LDA | 97.40 | 98.35 | 98.45 | 98.72 | 96.54 | 96.86 | 97.83 | 97.84 | 99.04 | 98.42 | 98.05 | 98.13 | 96.32 | **97.84±0.84** |
| SVM | **99.13** | 96.85 | 98.84 | 98.93 | 99.07 | 98.91 | 98.93 | 98.74 | 99.14 | **99.03** | **99.11** | 98.68 | 98.09 | **98.73±0.63** |
| Cubic SVM | 97.52 | 98.19 | 98.24 | 97.80 | 97.07 | 97.52 | 97.85 | 98.54 | 97.90 | 97.42 | 97.69 | 97.79 | 97.67 | **97.78±0.38** |
| KNN | 84.59 | 84.70 | 81.05 | 82.95 | 82.26 | 85.21 | 84.83 | 84.27 | 81.51 | 84.59 | 83.28 | 83.42 | 83.78 | **83.57±1.32** |
| Cosine KNN | 97.61 | 98.25 | 98.31 | 98.45 | 97.60 | 97.89 | 97.48 | 98.02 | 97.79 | 97.75 | 97.40 | 97.56 | 97.17 | **97.79±0.38** |
| Bagged Trees | 99.00 | **99.07** | **99.32** | **99.10** | **99.36** | **98.92** | **99.31** | **99.14** | **99.47** | 98.99 | 98.94 | **98.93** | **98.49** | **99.08±0.25** |
| Subspace KNN | 84.22 | 84.82 | 81.03 | 82.61 | 82.59 | 85.54 | 84.94 | 83.84 | 81.84 | 84.65 | 83.38 | 83.43 | 83.93 | **83.6±1.31** |
| Boosted Trees | 98.63 | 98.78 | 99.23 | 98.13 | 98.79 | 98.66 | 98.84 | 98.87 | 99.19 | 98.52 | 98.64 | 98.87 | 97.48 | **98.66±0.45** |

**TABLE 7.** Summary of classification performances with unlabeled sleep stage using 6 features.

| Classifier: Cubic SVM | Fp2F4 | F4C4 | C4P4 | P4O2 | F8T4 | T4T6 | Fp1F3 | F3C3 | C3P3 | P3O1 | F7T3 | T3T5 | C4A1 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy (%) | 96.49 | 96.87 | 96.84 | 96.51 | 95.33 | 95.41 | 96.61 | **97.83** | 96.91 | 96.06 | 96.02 | 96.61 | 96.25 | **96.44±0.66** |
| Sensitivity (%) | 92.79 | 93.12 | 92.90 | 93.05 | 89.57 | 89.07 | 92.72 | **95.14** | 94.08 | 91.99 | 90.85 | 92.16 | 92.09 | **92.27±1.67** |
| Specificity (%) | 97.72 | 98.12 | 98.15 | 97.66 | 97.23 | 97.51 | 97.90 | **98.71** | 97.85 | 97.41 | 97.73 | 98.09 | 97.63 | **97.82±0.38** |

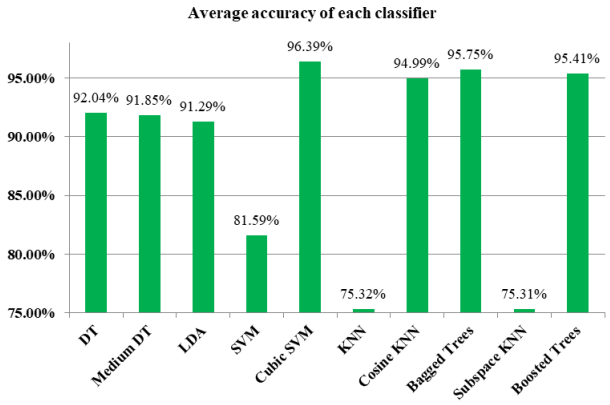


**FIGURE 12.** Average accuracy of each classifier with unlabeled sleep stage using 5 features.

channel F3C3 and cubic SVM are selected for detection bruxism with unlabeled sleep stage. The results are summarized in Table 7 and shown in Fig. 14. Channel F3C3 showed best performance using cubic SVM with 6 features. The accuracy obtained is 97.83%. It is worth noting that the pattern of classification accuracy is similar for both using 20 features and using top 6 features as shown in Fig. 10 and Fig. 14. But for the latter case the accuracy levels are increased.

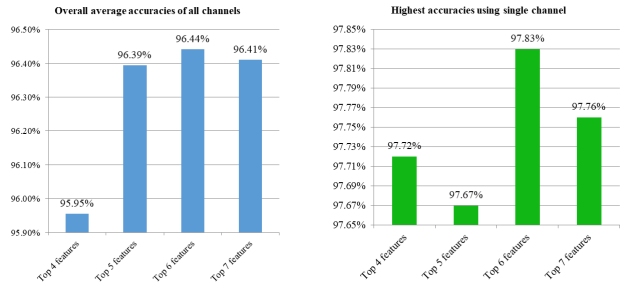


**FIGURE 13.** Overall average accuracies of all channels and highest accuracies using single channel for different feature subset.

### C. CLASSIFICATION WITH LABELED SLEEP STAGE

As our method performed well with data of unlabeled sleep stage and it is a sequential method, it is anticipated to converge for optimum features with good result for labeled sleep stage too.

Initially cubic SVM classifier is chosen as it performed better during classification with unlabeled sleep stage. To test the performance 5-fold cross validation and 10 independent trials are used for each classification result.

The results are summarized in Table 8. The plots of average accuracy of each channel for all sleep stages and average accuracy of each sleep stage for all channels are shown in Fig. 15. On average the F3C3 channel for REM sleep stage

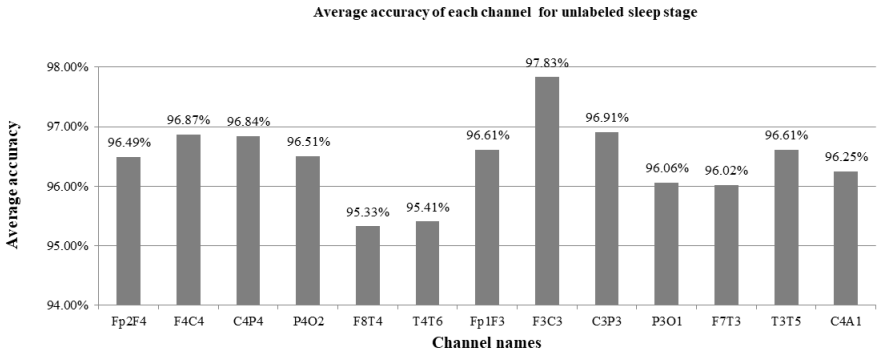


**FIGURE 14.** Average accuracy of each channel for classification with unlabeled sleep stage using 6 features.

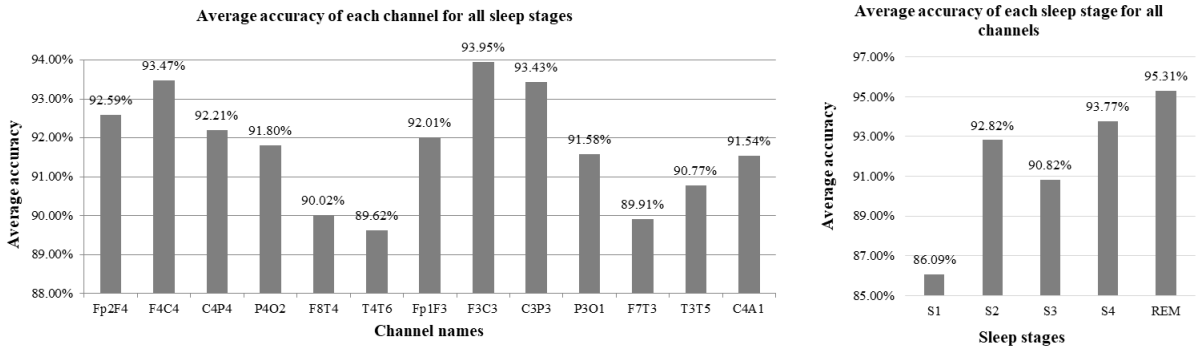


**FIGURE 15.** Average accuracy of each channel and average accuracy of each sleep stage for classification with labeled sleep stage using 20 features.

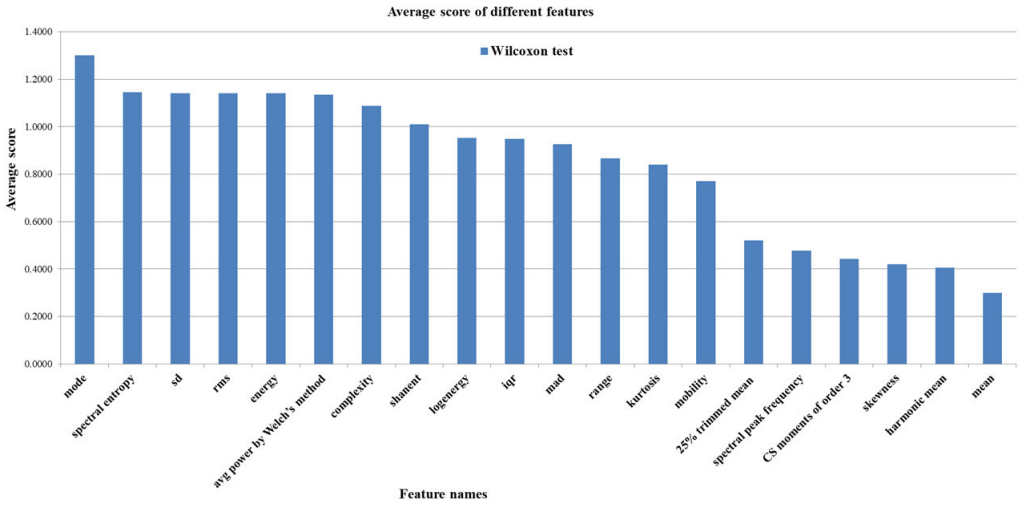


**FIGURE 16.** Average score of each group of features used for classification with labeled sleep stage.

shows highest average accuracy of 97.44%. Overall average accuracy for all channels and all sleep stages is 91.76%.

It is observed that using 20 features for 6 band limited signals for each 1 minute segment, combination of channel F3C3 and REM sleep stage has the highest average accuracy. This channel-sleep stage combination also has best performance in terms of sensitivity resulting in reduced Type I error. This combination has also specificity performance very close to best. Best performing channel i.e. F3C3 and best performed sleep stage i.e. REM is considered for further feature reduction process for getting maximum accuracy. A feature vector for the segments from F3C3-REM is compiled using 20 features from each band. Those 20 features are groupwise ranked using absolute values of the standardized test-statistic

**TABLE 8.** Summary of classification performances with labeled sleep stage using 20 features.

| Classification Accuracies (%) | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Sleep stage** | **Fp2F4** | **F4C4** | **C4P4** | **P4O2** | **F8T4** | **T4T6** | **Fp1F3** | **F3C3** | **C3P3** | **P3O1** | **F7T3** | **T3T5** | **C4A1** | **Average** |
| **S1** | 88.53 | 87.26 | 87.58 | 85.05 | 86.00 | 82.84 | 83.89 | 88.32 | 88.84 | 86.63 | 83.26 | 84.42 | 86.53 | **86.09±2.04** |
| **S2** | 92.45 | 93.60 | 93.45 | 94.18 | 90.61 | 90.50 | 93.30 | 93.77 | 93.92 | 92.37 | 91.37 | 92.69 | 94.48 | **92.82±1.31** |
| **S3** | 91.91 | 94.28 | 91.61 | 90.42 | 89.36 | 87.71 | 92.63 | 93.69 | 92.58 | 90.30 | 87.63 | 88.73 | 89.83 | **90.82±2.16** |
| **S4** | 94.34 | 95.68 | 93.03 | 93.59 | 90.42 | 91.72 | 94.96 | 96.51 | 94.87 | 93.34 | 92.88 | 94.34 | 93.28 | **93.77±1.62** |
| **REM** | 95.72 | 96.52 | 95.36 | 95.78 | 93.70 | 95.34 | 95.27 | **97.44** | 96.94 | 95.28 | 94.42 | 93.67 | 93.59 | **95.31±1.23** |
| **Average** | **92.59±2.73** | **93.47±3.65** | **92.21±2.91** | **91.8±4.25** | **90.02±2.77** | **89.62±4.68** | **92.01±4.67** | **93.95±3.55** | **93.43±3.02** | **91.58±3.3** | **89.91±4.49** | **90.77±4.16** | **91.54±3.31** | |
| **Overall average accuracy for all channels and all sleep stages** | | | | | | | | | | | | | | **91.76** |
| **Classification Sensitivities (%)** | | | | | | | | | | | | | | |
| **Sleep stage** | **Fp2F4** | **F4C4** | **C4P4** | **P4O2** | **F8T4** | **T4T6** | **Fp1F3** | **F3C3** | **C3P3** | **P3O1** | **F7T3** | **T3T5** | **C4A1** | **Average** |
| **S1** | 90.00 | 86.67 | 89.80 | 87.25 | 90.00 | 86.27 | 86.86 | 87.06 | 88.63 | 89.61 | 88.63 | 86.67 | 89.22 | **88.21±1.44** |
| **S2** | 84.81 | 84.95 | 84.91 | 88.82 | 80.14 | 76.56 | 83.82 | 86.23 | 84.58 | 81.98 | 80.33 | 83.68 | 87.55 | **83.72±3.3** |
| **S3** | 83.57 | 88.14 | 85.00 | 86.00 | 84.14 | 79.86 | 87.43 | 87.43 | 85.71 | 80.86 | 81.14 | 79.57 | 84.29 | **84.09±2.94** |
| **S4** | 85.98 | 86.16 | 81.43 | 81.25 | 78.30 | 80.63 | 85.00 | 91.52 | 85.63 | 81.34 | 81.88 | 89.73 | 83.84 | **84.05±3.78** |
| **REM** | 85.48 | 89.70 | 86.00 | 86.30 | 84.89 | 88.59 | 85.93 | **92.07** | 89.11 | 85.63 | 82.89 | 81.78 | 83.41 | **86.29±2.91** |
| **Average** | **85.97±2.43** | **87.12±1.84** | **85.43±2.99** | **85.92±2.83** | **83.49±4.55** | **82.38±4.92** | **85.81±1.45** | **88.86±2.72** | **86.73±2.01** | **83.88±3.71** | **82.97±3.3** | **84.29±4.01** | **85.66±2.57** | |
| **Overall average sensitivity for all channels and all sleep stages** | | | | | | | | | | | | | | **85.27** |
| **Classification Specificities (%)** | | | | | | | | | | | | | | |
| **Sleep stage** | **Fp2F4** | **F4C4** | **C4P4** | **P4O2** | **F8T4** | **T4T6** | **Fp1F3** | **F3C3** | **C3P3** | **P3O1** | **F7T3** | **T3T5** | **C4A1** | **Average** |
| **S1** | 86.82 | 87.95 | 85.00 | 82.50 | 81.36 | 78.86 | 80.45 | 89.77 | 89.09 | 83.18 | 77.05 | 81.82 | 83.41 | **83.64±3.92** |
| **S2** | 95.03 | 96.51 | 96.34 | 95.99 | 94.14 | 95.21 | 96.50 | 96.32 | 97.07 | 95.88 | 95.10 | 95.73 | 96.82 | **95.9±0.83** |
| **S3** | 95.42 | 96.87 | 94.40 | 92.29 | 91.57 | 91.02 | 94.82 | 96.33 | 95.48 | 94.28 | 90.36 | 92.59 | 92.17 | **93.66±2.11** |
| **S4** | 96.49 | 98.12 | 96.01 | 96.77 | 93.53 | 94.56 | 97.52 | 97.80 | 97.25 | 96.42 | 95.71 | 95.53 | 95.71 | **96.26±1.3** |
| **REM** | 98.46 | 98.34 | 97.86 | 98.32 | 96.06 | 97.15 | 97.76 | 98.87 | **99.03** | 97.86 | 97.50 | 96.85 | 96.32 | **97.72±0.92** |
| **Average** | **94.44±4.46** | **95.56±4.32** | **93.92±5.14** | **93.17±6.37** | **91.33±5.8** | **91.36±7.33** | **93.41±7.34** | **95.82±3.55** | **95.58±3.84** | **93.52±5.92** | **91.14±8.31** | **92.5±6.18** | **92.89±5.6** | |
| **Overall average specificity for all channels and all sleep stages** | | | | | | | | | | | | | | **93.44** |

**TABLE 9.** Summary of classification accuracies (%) with labeled sleep stage using 5 features and 10 classifiers.

| **Classifier** | **Fp2F4-REM** | **F4C4-REM** | **C4P4-REM** | **P4O2-REM** | **F8T4-REM** | **T4T6-REM** | **Fp1F3-REM** | **F3C3-REM** | **C3P3-REM** | **P3O1-REM** | **F7T3-REM** | **T3T5-REM** | **C4A1-REM** | **Average** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DT** | 91.56 | 92.05 | 91.25 | 94.39 | 90.41 | 91.69 | 91.61 | 92.91 | 92.94 | 92.70 | 92.89 | 91.59 | 89.58 | **91.97±1.23** |
| **Medium DT** | 91.30 | 91.59 | 90.91 | 94.92 | 90.81 | 92.03 | 91.84 | 92.67 | 92.88 | 93.05 | 92.53 | 91.83 | 89.75 | **92.01±1.28** |
| **LDA** | 92.58 | 93.75 | 92.30 | 93.73 | 91.75 | 92.66 | 93.05 | 92.98 | 92.95 | 93.11 | 93.09 | 91.97 | 90.73 | **92.67±0.83** |
| **SVM** | 56.75 | 60.48 | 51.05 | 58.50 | 47.61 | 57.81 | 59.95 | 50.56 | 55.63 | 55.45 | 55.77 | 48.56 | 55.98 | **54.93±4.2** |
| **Cubic SVM** | **94.98** | **96.20** | **95.94** | 95.30 | 93.70 | 93.89 | **95.13** | **97.52** | **96.56** | 94.55 | **94.23** | **94.20** | **93.83** | **95.08±1.19** |
| **KNN** | 86.42 | 84.53 | 81.36 | 81.94 | 81.05 | 83.22 | 82.44 | 81.00 | 81.31 | 81.86 | 81.34 | 82.16 | 85.11 | **82.6±1.73** |
| **Cosine KNN** | 92.67 | 93.27 | 91.77 | 91.52 | 92.02 | 91.98 | 91.69 | 93.39 | 94.63 | 92.89 | 91.94 | 92.45 | 89.22 | **92.26±1.26** |
| **Bagged Trees** | 93.61 | 94.63 | 94.20 | **95.91** | **94.63** | **94.98** | 94.08 | 95.72 | 95.72 | **94.56** | 94.03 | 93.97 | 92.55 | **94.51±0.94** |
| **Subspace KNN** | 86.63 | 84.61 | 81.20 | 82.16 | 81.48 | 83.02 | 82.50 | 80.98 | 80.81 | 81.63 | 80.84 | 82.36 | 85.17 | **82.57±1.83** |
| **Boosted Trees** | 93.56 | 93.13 | 92.52 | 88.22 | 93.59 | 93.42 | 93.63 | 89.84 | 83.58 | 90.66 | 93.78 | 93.50 | 92.73 | **91.7±3** |

**TABLE 10.** Summary of classification sensitivities (%) with labeled sleep stage using 5 features and 10 classifiers.

| Classifier | Fp2F4-REM | F4C4-REM | C4P4-REM | P4O2-REM | F8T4-REM | T4T6-REM | Fp1F3-REM | F3C3-REM | C3P3-REM | P3O1-REM | F7T3-REM | T3T5-REM | C4A1-REM | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DT** | 78.89 | 81.19 | 79.63 | 85.33 | 79.11 | 79.93 | 81.19 | 82.00 | 82.67 | 80.89 | 82.00 | 79.63 | 75.41 | **80.61±2.34** |
| **Medium DT** | 78.81 | 78.07 | 78.30 | 86.00 | 79.48 | 80.74 | 80.52 | 82.44 | 83.04 | 83.63 | 79.70 | 79.78 | 73.26 | **80.29±3.14** |
| **LDA** | 73.85 | 75.33 | 70.07 | 73.41 | 75.19 | 76.15 | 71.78 | 74.37 | 72.81 | 71.04 | 73.19 | 71.04 | 72.30 | **73.12±1.85** |
| **SVM** | 65.19 | 61.19 | 54.37 | 50.89 | 61.41 | 57.04 | 61.41 | 64.81 | 58.22 | 55.93 | 60.00 | 55.11 | 57.78 | **58.72±4.14** |
| **Cubic SVM** | **85.78** | **91.48** | **92.30** | **89.33** | **85.26** | **86.00** | **88.44** | **94.30** | **92.07** | **86.00** | **84.59** | **83.26** | **84.07** | **87.91±3.64** |
| **KNN** | 69.26 | 60.67 | 56.67 | 58.15 | 53.63 | 58.30 | 57.33 | 51.85 | 50.37 | 53.04 | 51.63 | 55.56 | 61.11 | **56.74±5.09** |
| **Cosine KNN** | 75.04 | 81.26 | 75.11 | 75.48 | 72.74 | 73.41 | 77.26 | 82.89 | 81.33 | 75.41 | 72.96 | 73.33 | 76.22 | **76.34±3.41** |
| **Bagged Trees** | 73.78 | 78.37 | 77.85 | 85.56 | 78.52 | 83.26 | 77.78 | 84.15 | 82.67 | 82.30 | 76.22 | 78.00 | 72.22 | **79.28±4.05** |
| **Subspace KNN** | 68.67 | 60.67 | 56.37 | 58.59 | 55.26 | 57.85 | 57.26 | 51.93 | 49.70 | 52.37 | 50.59 | 56.22 | 61.41 | **56.68±5.14** |
| **Boosted Trees** | 82.22 | 80.07 | 80.37 | 50.89 | 79.78 | 80.96 | 82.67 | 65.56 | 27.19 | 69.85 | 81.63 | 80.07 | 75.41 | **72.05±16.24** |

**TABLE 11.** Summary of classification specificities (%) with labeled sleep stage using 5 features and 10 classifiers.

| Classifier | Fp2F4-REM | F4C4-REM | C4P4-REM | P4O2-REM | F8T4-REM | T4T6-REM | Fp1F3-REM | F3C3-REM | C3P3-REM | P3O1-REM | F7T3-REM | T3T5-REM | C4A1-REM | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DT** | 94.95 | 94.95 | 94.36 | 96.81 | 93.43 | 94.83 | 94.40 | 95.82 | 95.68 | 95.86 | 95.80 | 94.79 | 93.37 | **95±0.99** |
| **Medium DT** | 94.63 | 95.21 | 94.28 | 97.31 | 93.84 | 95.05 | 94.87 | 95.41 | 95.50 | 95.56 | 95.96 | 95.05 | 94.16 | **95.14±0.89** |
| **LDA** | 97.58 | 98.67 | 98.24 | **99.17** | 96.18 | 97.07 | **98.73** | 97.96 | 98.34 | **99.01** | 98.42 | 97.56 | 95.66 | **97.89±1.06** |
| **SVM** | 54.50 | 60.30 | 50.16 | 60.53 | 43.92 | 58.02 | 59.56 | 46.75 | 54.93 | 55.33 | 54.63 | 46.81 | 55.50 | **53.92±5.44** |
| **Cubic SVM** | 97.45 | 97.47 | 96.91 | 96.89 | 95.96 | 96.00 | 96.91 | 98.38 | 97.76 | 96.83 | 96.81 | 97.13 | 96.44 | **97±0.67** |
| **KNN** | 91.01 | 90.91 | 87.96 | 88.30 | 88.38 | 89.88 | 89.15 | 88.79 | 89.58 | 89.56 | 89.29 | 89.27 | 91.52 | **89.51±1.09** |
| **Cosine KNN** | 97.39 | 96.48 | 96.22 | 95.80 | 97.17 | 96.95 | 95.54 | 96.20 | 98.18 | 97.56 | 97.01 | 97.56 | 92.69 | **96.52±1.38** |
| **Bagged Trees** | **98.91** | **98.97** | **98.57** | 98.67 | **98.93** | **98.12** | 98.44 | **98.81** | **99.21** | 97.84 | **98.79** | **98.24** | **97.98** | **98.58±0.42** |
| **Subspace KNN** | 91.43 | 91.01 | 87.84 | 88.46 | 88.50 | 89.74 | 89.25 | 88.75 | 89.13 | 89.45 | 88.93 | 89.35 | 91.52 | **89.49±1.16** |
| **Boosted Trees** | 96.59 | 96.61 | 95.76 | 98.20 | 97.29 | 96.75 | 96.55 | 96.34 | 98.65 | 96.22 | 97.03 | 97.09 | 97.37 | **96.96±0.79** |

of a two-sample unpaired Wilcoxon test. Fig. 16 shows the average score of each group of features. Then performance is tested with top five features namely: sample mode, spectral entropy, standard deviation, RMS and energy.

Previously mentioned 10 different classifiers are used to check their performances. 5-fold cross validation and 10 independent trials are used for each classification. The results are summarized in Table 9, 10 and 11 and average accuracy of each classifier is shown in Fig. 17.

Again, F3C3-REM combination has shown highest sensitivity. Average specificity is highest for Bagged Trees but it has low average sensitivity. LDA and boosted trees both have good specificity but they have poor sensitivity. Cubic SVM has very good average specificity along with highest average sensitivity. Hence, cubic SVM shows balanced performance in terms of both sensitivity and specificity, which in turns results in reduced Type I error and good accuracy.

It is evident that among all classifiers cubic SVM performs best for all channels. Channel F3C3 and REM stage showed best performance with 97.52% accuracy. Hence, channel F3C3, REM stage and cubic SVM is selected for further feature optimization process.

During feature optimization, top ranked features are stacked serially and accuracy is observed. With top 10 features F3C3 channel and REM sleep stage obtained highest accuracy. These results are summarized in Table 12 and shown in Fig. 18.

It is anticipated that training a model with less data may require more features to give desired performance which

**TABLE 12.** Summary of performance metrics for F3C3-REM with labeled sleep stage using top ranked features.

| F3C3-REM | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| **Top 5 features** | 97.52 | 94.30 | 98.38 |
| **Top 6 features** | 97.72 | 95.04 | 98.44 |
| **Top 7 features** | 97.86 | 94.81 | 98.67 |
| **Top 8 features** | 98.19 | 95.33 | 98.95 |
| **Top 9 features** | 98.31 | 95.04 | 99.19 |
| **Top 10 features** | **98.39** | 95.04 | 99.29 |
| **Top 11 features** | 98.36 | 94.59 | 99.37 |
| **Top 12 features** | 98.16 | 94.44 | 99.15 |
| **Top 13 features** | 98.09 | 94.22 | 99.13 |

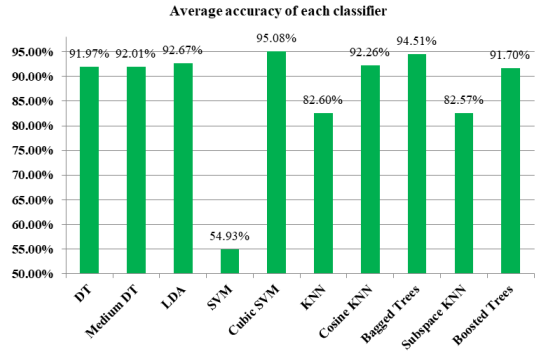


**FIGURE 17.** Average accuracy of each classifier with labeled sleep stage using 5 features.

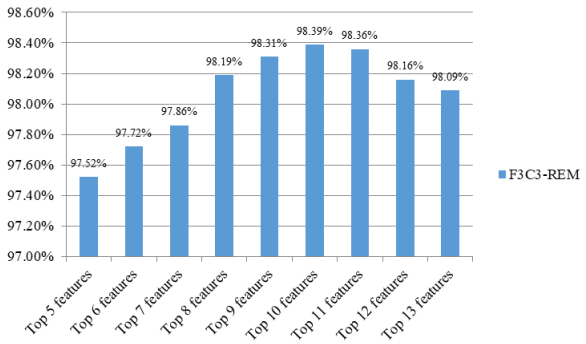


**FIGURE 18.** Average accuracies of F3C3 channel for REM stage for different number of features.

is reflected in the results. Top 10 features (sample mode, spectral entropy, standard deviation, RMS, energy, average power by Welch's method, complexity, Shannon's entropy, log energy entropy and IQR), channel F3C3, REM sleep stage and cubic SVM are selected for detection bruxism with labeled sleep stage. Combination of channel-sleep stage F3C3-REM showed best performance with 10 features using cubic SVM with 98.39% accuracy.

In summary, our study concluded that accuracy of channel F3C3 is best for classification with unlabeled sleep stage and channel-sleep stage combination F3C3-REM is best for labeled sleep stage. This agrees with the functional Magnetic Resonance Imaging (f-MRI) study [43] of bruxism patients. In their study, the right inferior parietal lobe demonstrated lower stimulation in bruxism patients. Hence, active regions in left and right portions of brain are not balanced in bruxism patients. Blood oxygenation level-dependent signal difference images revealed increased activation during teeth clenching in the localized confined region of frontal and parietal region boundaries of both lobe of brain. There are more active areas in left lobe. There are also increased activations in localized regions of frontal lobe. This fact indicates more activation in left brain region and hence left brain region can be more distinguishable to detect bruxism. F3C3, which is located in left region, is found to have more class separation ability. This finding is concordant with clinical finding.

Our study is not event based, rather our objective is to distinguish every one minute EEG segment (irrespective of events/symptoms occurring or not). Our study reveals REM sleep stage as a potential biomarker to distinguish bruxism without checking all other sleep stages. REM sleep stage is easily identifiable by eye movements and all other stages need not be checked. Hence computational cost can be reduced. Another study [44] suggests that patients with bruxism have more REM sleep than healthy individuals. This also confirms the potential credibility of using REM sleep stage to detect bruxism.

In both DT and medium DT, split criterion used is Gini's diversity index. In medium DT, maximum number of splits is set to 20 to make the classification tree less complex or deep. Using medium DT performance is equivalent to DT. In medium DT, performance is almost same both for unlabeled sleep stage and labeled sleep stage. Performance is not high as it uses simple if-else structure which is not sufficient to classify overlapping classes.

In LDA, regularized LDA is used where discriminant type specified as linear. In linear LDA computation is less expensive. Compared to DT, using LDA performance is slightly decreased for unlabeled sleep stage and slightly improved for labeled sleep stage. Linear LDA assumes distributions of all classes have the same shape. This is not true for the data used in this study. Unsatisfactory performance is due to oversimplified algorithm of LDA to separate complicated overlapping classes.

In SVM linear kernel is used and in cubic SVM polynomial kernel of order 3 is used. Both for unlabeled sleep stage and labeled sleep stage SVM performs poorly but cubic SVM shows the best performance among all classifiers. The classes in used data are distributed in an overlapping way. Hence, the two-dimensional space is unable to separate the features as they are linearly indistinguishable. They can be linearly distinguishable in the nonlinear feature space which is categorized by nonlinear kernel function. This nonlinear kernel function is used in cubic SVM.

In KNN, number of nearest neighbors is only 1 in predictors for classifying every instance while predicting. The distance metric used is Euclidean. In cosine KNN, number of nearest neighbors is 10 and cosine distance metric

**TABLE 13.** Summary of performances of all channels for classification with unlabeled sleep stage utilizing proposed method using 6, 4 and 2 bands.

| | Fp2F4 | F4C4 | C4P4 | P4O2 | F8T4 | T4T6 | Fp1F3 | F3C3 | C3P3 | P3O1 | F7T3 | T3T5 | C4A1 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Accuracy (%) (6 bands)** | 96.49 | 96.87 | 96.84 | 96.51 | 95.33 | 95.41 | 96.61 | 97.83 | 96.91 | 96.06 | 96.02 | 96.61 | 96.25 | **96.44±0.66** |
| **Accuracy (%) (4 bands)** | 92.15 | 91.78 | 90.63 | 89.72 | 89.12 | 88.94 | 92.03 | 91.38 | 90.46 | 89.91 | 86.98 | 92.27 | 92.54 | **90.61±1.64** |
| **Accuracy (%) (2 bands)** | 84.89 | 84.20 | 82.85 | 82.26 | 79.61 | 80.92 | 82.89 | 83.77 | 84.08 | 83.11 | 80.50 | 83.91 | 84.16 | **82.86±1.62** |
| **Sensitivity (%) (6 bands)** | 92.79 | 93.12 | 92.90 | 93.05 | 89.57 | 89.07 | 92.72 | 95.14 | 94.08 | 91.99 | 90.85 | 92.16 | 92.09 | **92.27±1.67** |
| **Sensitivity (%) (4 bands)** | 84.16 | 84.81 | 80.22 | 79.62 | 78.56 | 76.65 | 84.53 | 81.49 | 79.60 | 79.27 | 71.24 | 83.45 | 85.39 | **80.69±3.97** |
| **Sensitivity (%) (2 bands)** | 66.87 | 66.80 | 62.50 | 57.48 | 53.18 | 51.77 | 59.04 | 63.35 | 60.03 | 58.18 | 49.07 | 62.60 | 61.06 | **59.38±5.46** |
| **Specificity (%) (6 bands)** | 97.72 | 98.12 | 98.15 | 97.66 | 97.23 | 97.51 | 97.90 | 98.71 | 97.85 | 97.41 | 97.73 | 98.09 | 97.63 | **97.82±0.38** |
| **Specificity (%) (4 bands)** | 94.79 | 94.09 | 94.07 | 93.06 | 92.62 | 93.01 | 94.52 | 94.66 | 94.05 | 93.43 | 92.19 | 95.19 | 94.91 | **93.89±0.95** |
| **Specificity (%) (2 bands)** | 90.86 | 89.96 | 89.59 | 90.46 | 88.36 | 90.57 | 90.79 | 90.54 | 92.04 | 91.36 | 90.91 | 90.97 | 91.81 | **90.63±0.95** |

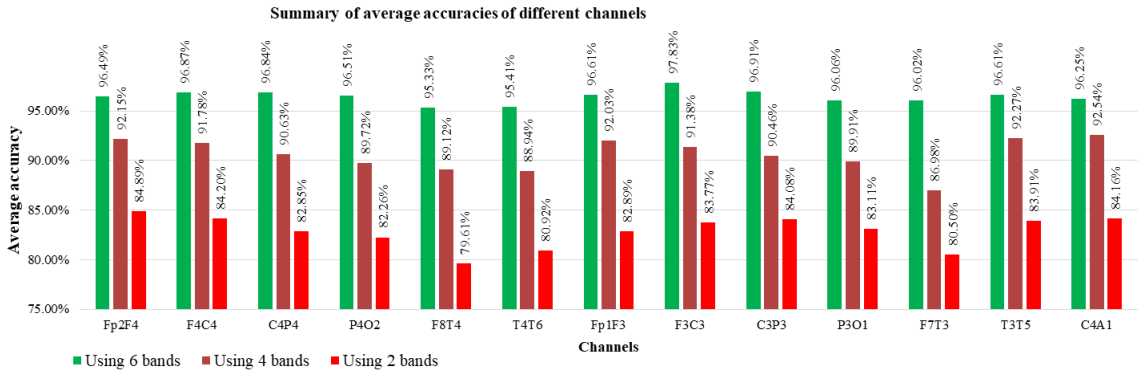


**FIGURE 19.** Accuracies of different channels for classification with unlabeled sleep stage utilizing proposed method using 6,4 and 2 bands.

is used. Additionally it centers and scales each column of the predictor data by the column mean and standard deviation, respectively. KNN performs poorly for unlabeled sleep stage and moderately for labeled sleep stage. Cosine KNN performs better in both cases. Increase in number of neighbour from 1 to 10 resulted in improved performance because majority voting from 10 neighbours is more reliable than considering only 1 neighbour for prediction.

Bagged trees is an ensemble learner which is composed of several weak learners i.e. decision trees to improve performance. Bootstrap aggregating is used as ensemble-aggregation method and 30 DT are used. As expected, performance of bagged trees is better than both DT and medium DT in case of unlabeled and labeled sleep stage.

Subspace KNN is an ensemble learner. For each learner feature subspace is chosen at random with replacement. 30 learners are used. Its performance is similar to KNN both for unlabeled and labeled sleep stage. The reason may be number of nearest neighbors is only 1 both for KNN and subspace KNN. This low number of nearest neighbors degrades predictive power.

Boosted trees is also an ensemble learner. Here adaptive boosting is used with 30 trees. Maximum number of splits is set to 20. Its performance is superior to DT and medium DT for unlabeled sleep stage. But for labeled sleep stage its performance is similar to DT and medium DT. The reason might be the less number of data used for labeled sleep stage, which is insufficient for growing adaptive trees.

To test the justification of considering non traditional high frequency bands, the proposed method is also tested by using 4 bands which have maximum 32Hz frequency components. These 4 bands consist of IDWT-RBL signals of 0-4Hz, 4-8Hz, 8-16Hz and 16-32Hz respectively. It is also tested with 2 bands consisting of 0-4Hz, 4-8Hz bands. Results are summarized in Table 13 and shown in Fig. 19. It clearly indicates that considering two high frequency bands (32-64Hz and 64-128Hz) in analysis proved their worth.

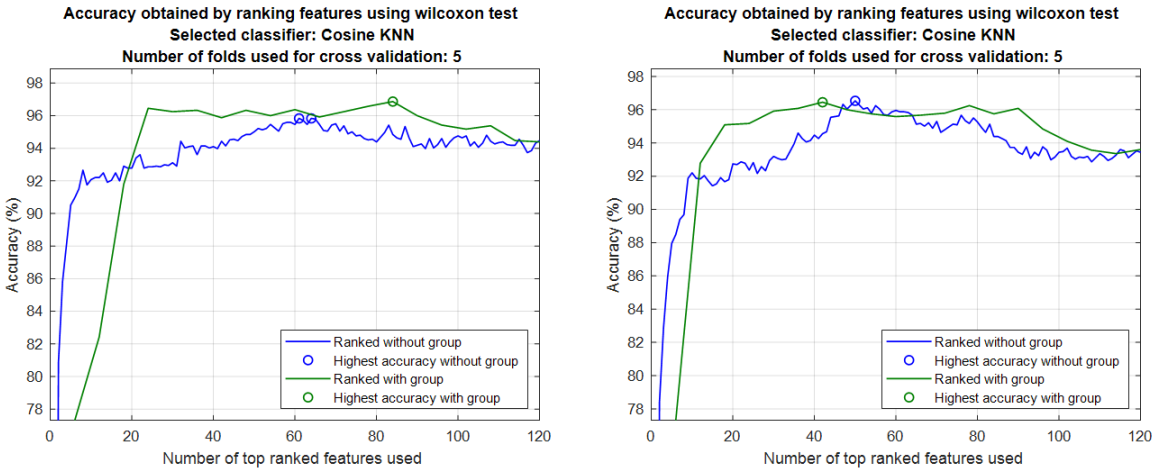


**FIGURE 20.** Accuracy obtained by using GFR and without GFR considering unlabeled sleep stages for (a) F3C3 channel (b) C4P4 channel.

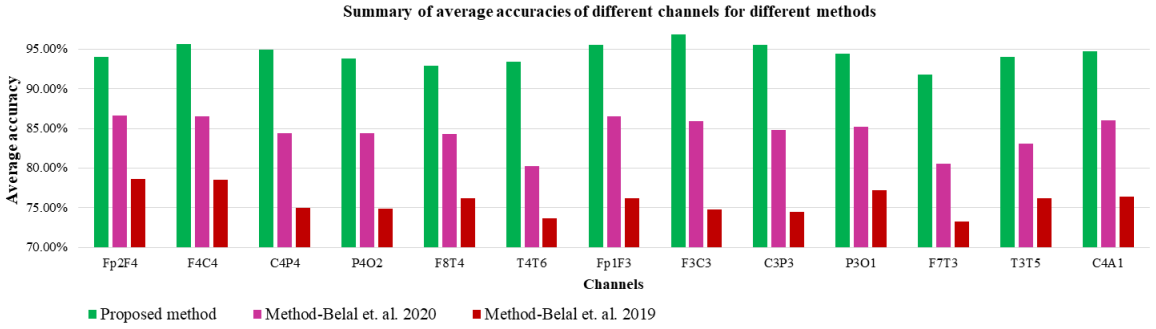


**FIGURE 21.** Overall average of accuracies of different channels using proposed method, method of Heyat et al. [18] and method of Heyat et al. [19].

To check the robustness of proposed GFR technique group wise top ranked features are stacked one after another and checked for accuracy. Accuracy is also checked by stacking top ranked features one after another without grouping. The results are shown in Fig. 20. It is obvious from figure that features ranked with group outperforms features ranked without group. From 24 features to 120 features ranked with group shows higher accuracy than features ranked without group for F3C3. In case of C4P4, features ranked with group reached similar highest accuracy as features ranked without group with less number of features. Throughout the tests green curve is above blue curve which demonstrates overall superiority of GFR. Less correlation exists among intragroup features as they are from different bands. This is the reason for enhanced performance for features ranked with group.

### D. PERFORMANCE COMPARISON

To further examine the performance of our method, the EEG based methods using labeled sleep stage described by Heyat et al. [18] and Heyat et al. [19] are tested using the same data we used. Using similar methods as they used, the results are summarized in Table 14. A comparison of proposed method with their methods using same data is presented in Table 15. Average of accuracies of different channels using proposed method, method of Heyat et al. [18] and method of Heyat et al. [19] using the same data are shown in Fig. 21. It is clear that proposed method outperforms their methods by a good margin. In spite of the dataset being unbalanced, the sensitivity and specificity of our method is much superior to other methods.

In order to demonstrate the ability of the proposed method to perform in the real-time application, its computational complexity is computed. For the purpose of comparing the computational complexity, two other methods are considered. For these methods, the computational complexity is computed using the same data shown in Table 15. The summary of computational complexity comparison is presented in Table 16. Here, preprocessing and feature extraction time/instance (PFET/instance) refers to the time of preprocessing each 60 second segment and extracting the desired features. Testing time/instance (TT/instance) refers to the time to classify these extracted features by the trained classifier/classifiers. Run time/instance (RT/instance) refers to the time from start of preprocessing to final classification of each 60 second segment i.e. sum of PFET/instance and TT/instance. The PFET/instance of the proposed method is

**TABLE 14.** Results using method of Heyat et al. [18] and Heyat et al. [19].

| Accuracies (%) of Each Channel and Each Sleep Stage using method of Belal et al. [18] | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Sleep stage** | **Fp2F4** | **F4C4** | **C4P4** | **P4O2** | **F8T4** | **T4T6** | **Fp1F3** | **F3C3** | **C3P3** | **P3O1** | **F7T3** | **T3T5** | **C4A1** | **Average** |
| **S1** | 78.77 | 79.62 | 73.96 | 72.14 | 76.95 | 68.53 | 73.18 | 71.21 | 72.40 | 72.91 | 66.08 | 67.40 | 68.33 | **72.42±4.23** |
| **S2** | 76.88 | 78.58 | 74.50 | 74.15 | 78.19 | 76.92 | 74.44 | 72.67 | 74.00 | 73.96 | 73.18 | 76.70 | 75.70 | **75.37±1.91** |
| **S3** | 72.33 | 69.48 | 65.61 | 66.05 | 66.14 | 67.79 | 69.78 | 66.25 | 66.00 | 74.13 | 65.80 | 72.33 | 71.93 | **68.74±3.08** |
| **S4** | 79.70 | 81.18 | 82.06 | 80.64 | 81.13 | 75.11 | 81.11 | 83.30 | 80.00 | 82.93 | 82.05 | 85.13 | 83.82 | **81.4±2.45** |
| **REM** | 85.24 | 83.63 | 78.86 | 81.20 | 78.69 | 79.87 | 82.49 | 80.52 | 80.03 | 81.91 | 79.04 | 79.51 | 82.24 | **81.02±2** |
| **Average** | **78.58± 4.68** | **78.5± 5.39** | **75± 6.21** | **74.84± 6.31** | **76.22± 5.84** | **73.64± 5.29** | **76.2± 5.41** | **74.79± 6.99** | **74.49± 5.87** | **77.17± 4.83** | **73.23± 7.38** | **76.21± 6.77** | **76.4± 6.61** | |
| **Overall average accuracy for all channels and all sleep stages** | | | | | | | | | | | | | | **75.79** |
| **Accuracies (%) of Each Channel and Each Sleep Stage using method of Belal et al. [19]** | | | | | | | | | | | | | | |
| **Sleep stage** | **Fp2F4** | **F4C4** | **C4P4** | **P4O2** | **F8T4** | **T4T6** | **Fp1F3** | **F3C3** | **C3P3** | **P3O1** | **F7T3** | **T3T5** | **C4A1** | **Average** |
| **S1** | 76.84 | 81.40 | 74.04 | 72.63 | 80.35 | 69.12 | 75.09 | 78.60 | 76.14 | 77.54 | 75.79 | 74.04 | 76.49 | **76.01±3.23** |
| **S2** | 88.17 | 87.66 | 87.98 | 88.33 | 85.44 | 83.02 | 88.69 | 87.14 | 87.58 | 87.02 | 79.33 | 85.40 | 88.17 | **86.46±2.66** |
| **S3** | 86.44 | 84.46 | 84.46 | 83.19 | 81.64 | 78.95 | 86.86 | 85.17 | 83.05 | 83.90 | 73.73 | 81.21 | 86.86 | **83.07±3.64** |
| **S4** | 90.27 | 88.44 | 89.05 | 89.29 | 88.69 | 85.28 | 92.46 | 90.88 | 90.21 | 89.90 | 87.47 | 88.56 | 90.88 | **89.34±1.79** |
| **REM** | 91.61 | 90.68 | 86.25 | 88.59 | 85.26 | 84.74 | 89.38 | 87.81 | 87.03 | 87.81 | 86.56 | 86.41 | 87.66 | **87.68±2** |
| **Average** | **86.67± 5.84** | **86.53± 3.63** | **84.36± 6.02** | **84.41± 7.02** | **84.28± 3.32** | **80.22± 6.68** | **86.5± 6.69** | **85.92± 4.58** | **84.8± 5.48** | **85.23± 4.81** | **80.58± 6.22** | **83.12± 5.74** | **86.01± 5.53** | |
| **Overall average accuracy for all channels and all sleep stages** | | | | | | | | | | | | | | **84.51** |

**TABLE 15.** Result comparison of proposed method with other methods using same data for classification with labeled sleep stage.

| **Author** | **Year** | **Feature extraction method** | **Classifier used** | **Channel** | **Sleep Stage** | **Best Accuracy obtained (%)** | **Sensitivity at best Accuracy (%)** | **Specificity at best Accuracy (%)** | **Major fidelities** | **Major limitations** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Proposed Method** | **Present** | **New GFR method** | **Cubic SVM** | **F3C3** | **REM** | **98.39** | **95.04** | **99.29** | **Better accuracy** | **Higher computation time** |
| Md Belal Bin Heyat et al. [18] | August 12, 2019 | Power spectral density | DT | Fp2F4 | REM | 85.24 | 63.21 | 91.13 | Less computation time | Inferior accuracy |
| Md Belal Bin Heyat et al. [19] | October 2020 | Power spectral density | Voting with 10 classifiers | Fp1F3 | S4 | 92.46 | 72.02 | 97.71 | Less computation time | Inferior accuracy |

found higher due to the use of more features which require more computational time and the use of IDWT-RBL signals. The TT/instance is longer in the method reported in [19] as it requires voting after prediction of 10 different classifiers. The TT/instance of the method reported in [18] is bit shorter than that is found by the proposed method. They used DT classifier which uses a simple tree structured algorithm to classify. The use of Cubic SVM classifier in the proposed method is one of the reasons behind the greater computational complexity. However, the relatively longer computational complexity is still within the acceptable limit considering significantly better classification performance with respect to other existing methods.

The proposed method has potential in detection of bruxism by using single EEG signal with high accuracy. It may be used in self-applicable, wearable, single channel EEG based bruxism diagnostic system.

### *E. APPLICATION OF DEEP LEARNING TO DETECT BRUXISM*

Applications of deep learning methods are also introduced in this study to detect bruxism. Deep learning models with three convolution-2D layers and four convolution-2D layers are used. They are termed as first and second model respectively for ease of description. Their schematics are shown in Fig. 22. The input consists of 60 second segments of EEG of all the

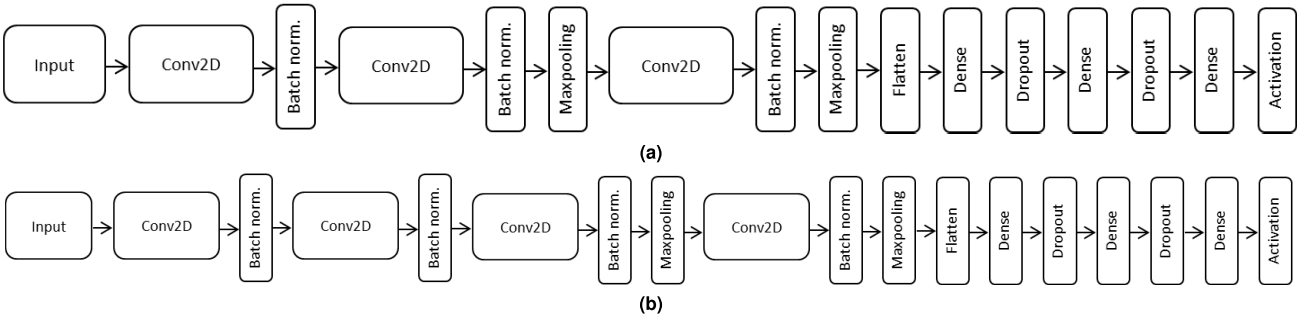


**FIGURE 22.** Schematic diagrams of 2D-CNN used. (a) 2D-CNN using 3 convolution-2D layers. (b) 2D-CNN using 4 convolution-2D layers.

**TABLE 16.** Computational complexity comparison of proposed method with other methods using same data for classification with labeled sleep stage.

| Author | Preprocessing and feature extraction time/instance (in seconds) | Testing time/instance (in seconds) | Run time/instance (in seconds) |
|---|---|---|---|
| **Proposed method** | **0.566070** | **$1.6246\times10^{-5}$** | **0.566080** |
| Md Belal Bin Heyat et al. [18] | 0.096759 | $1.1429\times10^{-5}$ | 0.096771 |
| Md Belal Bin Heyat et al. [19] | 0.131280 | $1.4286\times10^{-2}$ | 0.145560 |

channels. In conv-2D layers valid padding is used with stride 1. In the dense layers, 16 and 8 dense layers are used with 0.1 dropout rate. In final dense layer sigmoid activation is used. In the first model 32, 16 and 8 convolution filters are used and in the second model 64, 32, 16 and 8 convolution filters are used. 80% data are used for training and remaining 20% data are used for validation. Sleep stage independent classification is performed. Best model in 100 epoch is chosen for final evaluation. The results are summarized in Table 17. The accuracies are 90.93% and 90.72% for first model and second model respectively. The accuracies are similar in both cases. Specificity values are very good. Sensitivity values decreased from 76.67% to 67.50% from first to second model. Low sensitivity scores of these models resulted in increased Type I error. This is not at all desired for disease classification. However, further study can be done using various deep learning techniques to increase sensitivity as well as accuracy.

**TABLE 17.** Results using deep learning models.

| | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| First model | 90.93 | 76.67 | 95.62 |
| Second model | 90.72 | 67.50 | 98.36 |

### F. APPLICATION OF PROPOSED METHOD ON OTHER DATA

To the best of our knowledge, the used dataset is the only publicly available dataset that contains whole night sleep EEG data of bruxism patients. No other such dataset of bruxism patients is found online. Therefore, to demonstrate the versatility of the proposed GFR method, it is applied to detect sleep disordered breathing (SDB) from normal subjects.

**TABLE 18.** Summary of data for SDB detection with unlabeled sleep stage.

| Number of available segments | Segments from patients | Segments from normal subjects | Duration of available segments (Seconds) | Duration of available segments (Hours) |
|---|---|---|---|---|
| **3351** | **979** | **2372** | **201060** | **55.85** |

**TABLE 19.** Result using proposed method to detect SDB.

| Best Accuracy obtained (%) | Sensitivity at best Accuracy (%) | Specificity at best Accuracy (%) |
|---|---|---|
| 99.14 | 98.45 | 99.43 |

Bruxism and SDB have a firm association [45]. Bruxism and SDB occur concurrently in around 20% to 40% of circumstances. Both the diseases have some common signs and symptoms, such as, reduced quality of sleep, mouth breathing, nighttime snoring, a fondness for a supine position during sleep, headaches at morning etc. Rhythmic masticatory muscle activity (RMMA) occurs during bruxism events. RMMA precedes an apnea or hypopnea event in 55% of cases. In 25% of bruxism patients, an apnea or hypopnea event precedes RMMA. As bruxism and SDB have some similarity and both are sleep disorders, SDB is chosen for demonstration of proposed method on other data.

The data are collected from CAP sleep database [25]. Three SDB and five normal subjects are considered. The summary of the data is presented in Table 18. F4C4 channel is chosen and features are ranked using the proposed GFR algorithm and shown in Fig. 23. Performance using top ranked features is tested. Cubic SVM classifier is used with 5 fold cross validation scheme. Accuracies with top ranked features are demonstrated in Fig. 24. With top 10 feature groups the maximum accuracy of 99.14% is obtained. Table 19 summarizes the results. Sensitivity of 98.45% and specificity of 99.43% are obtained at maximum accuracy. The performance metrics attained to detect SDB clearly demonstrate the effectiveness of proposed GFR technique.

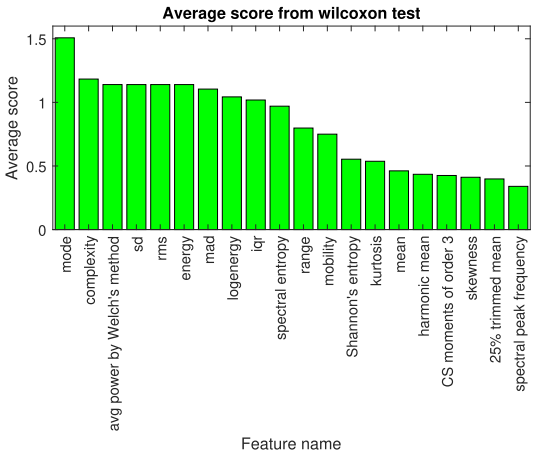


**FIGURE 23.** Average score of each group of features used for classification of SDB with unlabeled sleep stage.

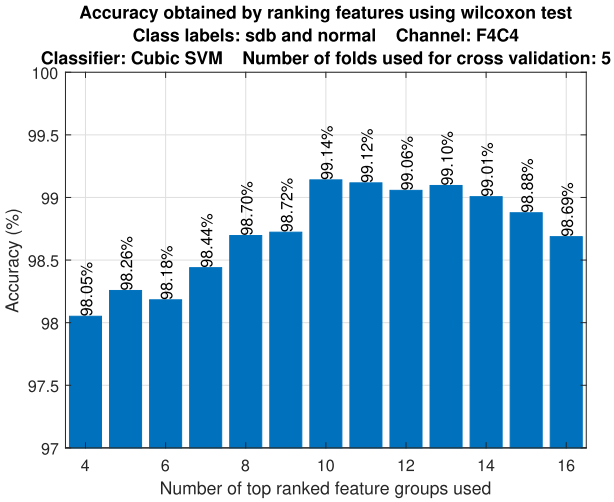


**FIGURE 24.** Average accuracies of F4C4 channel for SDB detection with unlabeled sleep stage using top ranked feature groups.

## V. LIMITATIONS OF THE STUDY

There are only 6 subjects, two subjects belong to bruxism class and four subjects belong to healthy class. All EEG segments collected from a particular subject are labeled as the same class, i.e. the class of that particular subject. All samples of a subject possess same class. In this case, subject independent classification problem would be formulated as to classify a subject as one of the two classes, bruxism or healthy class. One can take decision on a subject based on the identified classes of all segments of that subject (majority voting could be a simple solution). In this case, to get a better classification performance, there must be a large number of subjects in the training phase. Since the dataset contains only 2 bruxism subjects, subject independent evaluation under leave-one-subject-out cross validation may not be a suitable measurement.

In some medical applications, not necessarily all samples of a particular subject belong to the same class. In that case, leave-one-subject-out cross validation will provide samples of one particular subject for testing purpose, where the problem would be formulated as to classify a sample as one of the two classes. This type of classification is not applicable in our case as a particular subject of the dataset contains only one class.

The cross-database evaluation is very crucial for any clinical application. Unfortunately, we failed to get any publicly available database which contains sleep EEG data of bruxism and normal subjects. Hence, we could not evaluate the performance of the proposed method on other datasets.

There are 2 patients' data and 16 normal subjects' data in the CAP Sleep Database in PhysioNet database. To reduce the class imbalance, we considered 2 patients' data and 4 normal subjects' data in our study. In spite of this, the data used is not balanced. This is a limitation of the dataset. This same data that we used is tested on different methods reported in literatures and our method performed well in terms of accuracy, sensitivity and specificity compared to other methods. However, the method should be tested on more balanced data for real world application.

The run time/instance of the proposed method is higher due to using more features, the use of IDWT-RBL signals and the use of Cubic SVM classifier. The relatively longer computational complexity is within the acceptable limit considering its better classification performance than other existing methods.

The database has only small number of patient data i.e. only two patients' data are available. If sufficient numbers of data are available, the method could be generalized for more standard statistical assessment. Therefore further evaluation is necessary.

## VI. CONCLUSION

The proposed GFR technique and IDWT-RBL signal based feature extraction approach have shown an overall satisfactory performance. One of the major advantages of the method is that optimal feature subset is obtained using group wise ranking. This subset of features resulted in improved accuracy with lower feature dimension. In this study, both unlabeled and labeled sleep stages are considered. The classification performance is deteriorated while removing the features from the non-traditional high frequency bands. It indicates that these high frequency bands contain useful information along with customary bands. Moreover, use of band-limited signals confirms frequency resolution of the system. Unlike other study, all available data segments are considered. Hence, it can be considered as bias free collection of sample data. Comparison of proposed method with other recent methods is done. The proposed method shows classification accuracy of 97.83% and 98.39% in 5-fold cross validations with unlabeled and labeled sleep stage respectively. This automated approach can help medical practitioners in the related field as a convenient tool for aiding proper diagnosis of the disease. Only EEG data is used for the experiment as our aim is to use only one EEG channel to make the system simple. There might be contribution of other types

of data such as EMG, ECG etc. on the results as heart rate and muscle movement are involved in the symptoms of the disease. This sort of multimodal data analysis could lead to other algorithm and should be tested on larger dataset for practical application. However, the proposed method is tested on small EEG database and therefore further evaluation is necessary for more standard statistical assessment.

## ACKNOWLEDGMENT

The authors express gratitude to Information and Communication Technology Division, Government of the People's Republic of Bangladesh for granting fellowship to Ainul Anam Shahjamal Khan.

## REFERENCES

[1] M. Maeda, T. Yamaguchi, S. Mikami, W. Yachida, T. Saito, T. Sakuma, H. Nakamura, M. Saito, M. Mizuno, K. Yamada, and K. Satoh, "Validity of single-channel masseteric electromyography by using an ultraminiature wearable electromyographic device for diagnosis of sleep bruxism," *J. Prosthodontic Res.*, vol. 64, no. 1, pp. 90–97, Jan. 2020.

[2] B. K. Fehlberg, M. B. D. A. Barros, and M. G. Lima, "Health behaviors and multimorbidity associated with bruxism: Population-based study," *Oral Diseases*, vol. 29, no. 1, pp. 245–253, Jan. 2021.

[3] T. Bornhardt and V. Iturriaga, "Sleep bruxism: An integrated clinical view," *Sleep Med. Clinics*, vol. 16, no. 2, pp. 373–380, 2021.

[4] J.-Y. Song, "Implant complications in bruxism patients," *J. Korean Assoc. Oral Maxillofacial Surgeons*, vol. 47, no. 2, pp. 149–150, Apr. 2021.

[5] D. J. Burke, A. Seitz, O. Aladesuru, M. S. Robbins, and J. H. Chang, "Bruxism in acute neurologic illness," *Current Pain Headache Rep.*, vol. 25, no. 6, pp. 1–8, Jun. 2021.

[6] M. Carrillo-Diaz, A. R. Ortega-Martínez, M. Romero-Maroto, and M. J. González-Olmo, "Lockdown impact on lifestyle and its association with oral parafunctional habits and bruxism in a Spanish adolescent population," *Int. J. Paediatric Dentistry*, vol. 32, no. 2, pp. 185–193, Mar. 2022.

[7] C. Pinzan-Vercelino, K. Freitas, V. Girão, D. da Silva, R. Peloso, and A. Pinzan, "Does the use of face masks during the COVID-19 pandemic impact on oral hygiene habits, oral conditions, reasons to seek dental care and esthetic concerns?" *J. Clin. Experim. Dentistry*, pp. e369–e375, 2021.

[8] L. Pigozzi, D. Rehm, S. Fagondes, E. Pellizzer, and M. Grossi, "Current methods of bruxism diagnosis: A short communication," *Int. J. Prosthodontics*, vol. 32, no. 3, pp. 263–264, May 2019.

[9] J. Razjouyan, H. Lee, S. Parthasarathy, J. Mohler, A. Sharafkhaneh, and B. Najafi, "Improving sleep quality assessment using wearable sensors by including information from postural/sleep position changes and body acceleration: A comparison of chest-worn sensors, wrist actigraphy, and polysomnography," *J. Clin. Sleep Med.*, vol. 13, no. 11, pp. 1301–1310, Nov. 2017.

[10] C. D. Toedebusch, J. S. McLeland, C. M. Schaibley, I. R. Banks, J. Boyd, J. C. Morris, D. M. Holtzman, and B. P. Lucey, "Multi-modal home sleep monitoring in older adults," *J. Visualized Experiments*, vol. 2019, no. 143, Jan. 2019, Art. no. e58823, doi: 10.3791/58823-v.

[11] M. Hirshkowitz, "Polysomnography challenges," *Sleep Med. Clinics*, vol. 11, no. 4, pp. 403–411, Dec. 2016.

[12] T. Castroflorio, L. Mesin, G. M. Tartaglia, C. Sforza, and D. Farina, "Use of electromyographic and electrocardiographic signals to detect sleep bruxism episodes in a natural environment," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 6, pp. 994–1001, Nov. 2013.

[13] J. Martinot, N. Le-Dong, V. Cuthbert, S. Denison, D. Gozal, and J. Pepin, "Mandibular movement monitoring with artificial intelligence analysis for the diagnosis of sleep bruxism," *Sleep*, vol. 43, no. 1, pp. A301–A302, May 2020.

[14] T. Yamaguchi, S. Mikami, M. Saito, K. Okada, and A. Gotouda, "A newly developed ultraminiature wearable electromyogram system useful for analyses of masseteric activity during the whole day," *J. Prosthodontic Res.*, vol. 62, no. 1, pp. 110–115, Jan. 2018.

[15] J. Gao, L. Liu, P. Gao, Y. Zheng, W. Hou, and J. Wang, "Intelligent occlusion stabilization splint with stress-sensor system for bruxism diagnosis and treatment," *Sensors*, vol. 20, no. 1, p. 89, Dec. 2019, doi: 10.3390/s20010089.

[16] T. Miettinen, K. Myllymaa, S. Westeren-Punnonen, J. Ahlberg, T. Hukkanen, J. Töyräs, R. Lappalainen, E. Mervaala, K. Sipilä, and S. Myllymaa, "Success rate and technical quality of home polysomnography with self-applicable electrode set in subjects with possible sleep bruxism," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1124–1132, Jul. 2018.

[17] D. Lai, M. B. B. Heyat, F. I. Khan, and Y. Zhang, "Prognosis of sleep bruxism using power spectral density approach applied on EEG signal of both EMG1-EMG2 and ECG1-ECG2 channels," *IEEE Access*, vol. 7, pp. 82553–82562, 2019.

[18] M. B. B. Heyat, D. Lai, F. I. Khan, and Y. Zhang, "Sleep bruxism detection using decision tree method by the combination of C4-P4 and C4-A1 channels of scalp EEG," *IEEE Access*, vol. 7, pp. 102542–102553, 2019.

[19] M. B. Bin Heyat, F. Akhtar, A. Khan, A. Noor, B. Benjdira, Y. Qamar, S. J. Abbas, and D. Lai, "A novel hybrid machine learning classification for the detection of bruxism patients using physiological signals," *Appl. Sci.*, vol. 10, no. 21, p. 7410, Oct. 2020, doi: 10.3390/app10217410.

[20] Z. Jia, Y. Lin, J. Wang, X. Ning, Y. He, R. Zhou, Y. Zhou, and L. H. Lehman, "Multi-view spatial–temporal graph convolutional networks with domain generalization for sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1977–1986, 2021.

[21] B. D. Yetton, E. A. McDevitt, N. Cellini, C. Shelton, and S. C. Mednick, "Quantifying sleep architecture dynamics and individual differences using big data and Bayesian networks," *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0194604.

[22] F. Weber and Y. Dan, "Circuit-based interrogation of sleep control," *Nature*, vol. 538, no. 7623, pp. 51–59, Oct. 2016.

[23] Z. Jia, J. Ji, X. Zhou, and Y. Zhou, "Hybrid spiking neural network for sleep electroencephalogram signals," *Sci. China Inf. Sci.*, vol. 65, no. 4, Apr. 2022, Art. no. 140403.

[24] Z. Jia, X. Cai, and Z. Jiao, "Multi-modal physiological signals based squeeze-and-excitation network with domain adversarial learning for sleep staging," *IEEE Sensors J.*, vol. 22, no. 4, pp. 3464–3471, Feb. 2022.

[25] *The CAP Sleep Database*. Accessed: Jan. 11, 2020. [Online]. Available: https://physionet.org/physiobank/database/capslpdb/

[26] M. M. Rahman, M. A. Chowdhury, and S. A. Fattah, "An efficient scheme for mental task classification utilizing reflection coefficients obtained from autocorrelation function of EEG signal," *Brain Informat.*, vol. 5, no. 1, pp. 1–12, Mar. 2018, doi: 10.1007/s40708-017-0073-7.

[27] M. M. Rahman and S. A. Fattah, "An efficient feature extraction scheme for classification of mental tasks based on inter-channel correlation in wavelet domain utilizing EEG signal," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102033, doi: 10.1016/j.bspc.2020.102033.

[28] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. New York, NY, USA: Academic, 2005.

[29] G. M. Macaluso, P. Guerra, G. Di Giovanni, M. Boselli, L. Parrino, and M. G. Terzano, "Sleep bruxism is a disorder related to periodic arousals during sleep," *J. Dental Res.*, vol. 77, no. 4, pp. 565–573, Apr. 1998.

[30] L. Parrino, R. Ferri, O. Bruni, and M. G. Terzano, "Cyclic alternating pattern (CAP): The marker of sleep instability," *Sleep Med. Rev.*, vol. 16, no. 1, pp. 27–45, Feb. 2012.

[31] M. C. Carra, P. H. Rompré, T. Kato, L. Parrino, M. G. Terzano, G. J. Lavigne, and G. M. Macaluso, "Sleep bruxism and sleep arousal: An experimental challenge to assess the role of cyclic alternating pattern," *J. Oral Rehabil.*, vol. 38, no. 9, pp. 635–642, Sep. 2011.

[32] S. Yazdani, S. Fallet, and J.-M. Vesin, "A novel short-term event extraction algorithm for biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 4, pp. 754–762, Apr. 2018.

[33] Karadeniz, Ondze, Besset, and Billiard, "EEG arousals and awakenings in relation with periodic leg movements during sleep," *J. Sleep Res.*, vol. 9, no. 3, pp. 273–277, Sep. 2000.

[34] G. J. Lavigne, P. H. Rompré, F. Guitard, B. J. Sessle, T. Kato, and J. Y. Montplaisir, "Lower number of K-complexes and K-alphas in sleep bruxism: A controlled quantitative study," *Clin. Neurophysiol.*, vol. 113, no. 5, pp. 686–693, May 2002.

[35] M. R. Azevedo, R. Sena, A. M. D. Freitas, A. N. Silva, E. A. L. Júnior, and A. B. Soares, "Neuro-behavioral pattern of sleep bruxism in wakefulness," *Res. Biomed. Eng.*, vol. 34, no. 1, pp. 1–8, Feb. 2018, doi: 10.1590/2446-4740.06617.

[36] G. R. Bayar, R. Tutuncu, and C. Acikel, "Psychopathological profile of patients with different forms of bruxism," *Clin. Oral Investigations*, vol. 16, no. 1, pp. 305–311, Feb. 2012.
[37] J. Ahlberg, F. Lobbezoo, K. Ahlberg, D. Manfredini, C. Hublin, J. Sinisalo, M. Kononen, and A. Savolainen, "Self-reported bruxism mirrors anxiety and stress in adults," *Medicina Oral Patología Oral Y Cirugia Bucal*, vol. 18, no. 1, pp. e7–e11, 2013.
[38] R. Majid Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," *IEEE Access*, vol. 5, pp. 14797–14806, 2017.
[39] Y. Cheng, W. Jia, R. Chi, and A. Li, "A clustering analysis method with high reliability based on Wilcoxon-Mann-Whitney testing," *IEEE Access*, vol. 9, pp. 19776–19787, 2021.
[40] P. Sprent and N. C. Smeeton, *Applied Non Parametric Statistical Methods*. Boca Raton, FL, USA: CRC Press, 2007.
[41] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*. Hoboken, NJ, USA: Wiley, 2013.
[42] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*. Cham, Switzerland: Springer, 2006.
[43] S. Yılmaz, "To see bruxism: A functional MRI study," *Dentomaxillofacial Radiol.*, vol. 44, no. 7, Sep. 2015, Art. no. 20150019, doi: 10.1259/dmfr.20150019.
[44] T. Wieczorek, M. Wieckiewicz, J. Smardz, A. Wojakowska, M. Michalek-Zrabkowska, G. Mazur, and H. Martynowicz, "Sleep structure in sleep bruxism: A polysomnographic study including bruxism activity phenotypes across sleep stages," *J. Sleep Res.*, vol. 29, no. 6, pp. 1–19, Dec. 2020, doi: 10.1111/jsr.13028.
[45] A. González González, J. Montero, and C. Gómez Polo, "Sleep apnea–hypopnea syndrome and sleep bruxism: A systematic review," *J. Clin. Med.*, vol. 12, no. 3, p. 910, Jan. 2023, doi: 10.3390/jcm12030910.

**AINUL ANAM SHAHJAMAL KHAN** (Graduate Student Member, IEEE) received the B.Sc.Engg. and M.Sc.Engg. degrees from the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2001 and 2008, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology (CUET). He is an Associate Professor with the Department of Electrical and Electronic Engineering, CUET. His research interests include biomedical engineering, signal processing, and machine learning.

**SHAIKH ANOWARUL FATTAH** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from Bangladesh University of Engineering and Technology (BUET), Bangladesh, and the Ph.D. degree in electronics and communication engineering (ECE) from Concordia University, Canada. He was a visiting postdoctoral position with Princeton University, Princeton, NJ, USA. He is currently a Professor with the Department of Electrical and Electronic Engineering, BUET. He has published more than 215 international journal papers/conference papers with some best paper awards and delivered more than 80 keynote/invited talks in many countries. His research interests include signal processing, machine learning, and biomedical engineering. He is a Fellow of the Institution of Engineers, Bangladesh (IEB). He is a member of the IEEE PES LRP, the IEEE Public Visibility Committee, and the IEEE Smart Village Education Committee. He received several awards, namely the Concordia University's Distinguished Doctoral Dissertation Prize in ENS, in 2009, the 2007 URSI Canadian Young Scientist Award, the Dr. Rashid Gold Medal (in M.Sc.), the BAS-TWAS Young Scientists Prize, in 2014, the 2016 IEEE MGA Achievement Award, the 2017 IEEE R10 HTA Outstanding Volunteer Award, and the 2018 IEEE R10 Outstanding Volunteer Award. He served as the Chair/the Founding Chair for different IEEE Society Chapters in Bangladesh, such as IEEE SPS, EMBS, RAS, and SSIT. He was the IEEE Bangladesh Section Chair, from 2015 to 2016. He is the Chair of the IEEE PES-HAC and the IEEE SSIT Chapters Committee. He served on IEEE HAC and various committees for IEEE R10, IEEE EAB, and IEEE SIGHT. He served key positions in many international conferences, such as the General Chair for the IEEE R10-HTC2017 and the TPC Chair for the IEEE TENSYMP2020. He is the Editorial Board Member of IEEE Access, *IEEE Potentials*, *BioMed Research International*, and *Journal of Electrical Engineering* (IEB), and the Editor-in-Chief of IEEE PES Enews.

**MUHAMMAD QUAMRUZZAMAN** received the B.Sc.Engg., M.Sc.Engg., and Ph.D. degrees from the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 1996, 2002, and 2013, respectively. He has been a Faculty Member with the Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology (CUET), since 1997. Currently, he is a Professor with the Department of Electrical and Electronic Engineering, CUET. His research interests include biomedical instrumentation, power electronics, static power converters, and power system protection.

**MOHAMMAD SAQUIB** (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Bangladesh University of Engineering and Technology, Bangladesh, in 1991, and the M.S. and Ph.D. degrees in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1995 and 1998, respectively. He was a member of the Technical Staff with Massachusetts Institute of Technology Lincoln Laboratory, and an Assistant Professor with Louisiana State University, Baton Rouge, LA, USA. He is currently a Professor with the Electrical Engineering Department, The University of Texas at Dallas, Richardson, TX, USA. His current research interests include the various aspects of wireless data transmission, radio resource management, and signal processing techniques for low-cost radar applications.