Classification of Brainwaves for Sleep Stages by High-Dimensional FFT Features from EEG Signals

by Mera Delimayanti

Submission date: 14-Jan-2022 09:16PM (UTC+0700)
Submission ID: 1741664317
File name: applsci-10-01797-v2.pdf (1.41M)
Word count: 5658
Character count: 29188

applied sciences

MDPI

Article

Classification of Brainwaves for Sleep Stages by High-Dimensional FFT Features from EEG Signals

Mera Kartika Delimayanti [1,2,*], Bedy Purnama [1,3], Ngoc Giang Nguyen [1], Mohammad Reza Faisal [4], Kunti Robiatul Mahmudah [1], Fatma Indriani [1], Mamoru Kubo [5] and Kenji Satou [5]

1 Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa 9201192, Japan; bedypurnama@telkomuniversity.ac.id (B.P.); giangnn.bkace@gmail.com (N.G.N.); robiatul.mahmudah@gmail.com (K.R.M.); f_indriani@yahoo.com (F.I.)
2 Department of Computer and Informatics Engineering, Politeknik Negeri Jakarta, Depok 16425, Indonesia
3 Telkom of Computing, Telkom University, Bandung 40257, Indonesia
4 Computer Science, Lambung Mangkurat University, Banjarbaru 70714, Indonesia; reza.faisal@gmail.com
5 Institute of Science and Engineering, Kanazawa University, Kanazawa 9201192, Japan; kubom@se.kanazawa-u.ac.jp (M.K.); ken@t.kanazawa-u.ac.jp (K.S.)
* Correspondence: mera.kartika@tik.pnj.ac.id or mera.kartikadelimayanti@gmail.com

Received: 29 November 2019; Accepted: 2 March 2020; Published: 5 March 2020

Abstract: Manual classification of sleep stage is a time-consuming but necessary step in the diagnosis and treatment of sleep disorders, and its automation has been an area of active study. The previous works have shown that low dimensional fast Fourier transform (FFT) features and many machine learning algorithms have been applied. In this paper, we demonstrate utilization of features extracted from EEG signals via FFT to improve the performance of automated sleep stage classification through machine learning methods. Unlike previous works using FFT, we incorporated thousands of FFT features in order to classify the sleep stages into 2–6 classes. Using the expanded version of Sleep-EDF dataset with 61 recordings, our method outperformed other state-of-the art methods. This result indicates that high dimensional FFT features in combination with a simple feature selection is effective for the improvement of automated sleep stage classification.

Keywords: automatic sleep stage classification; electroencephalogram; fast Fourier transform

1. Introduction

Sleep is one of the basic physiological needs, and an important part of life. A typical human spends one-third of his lifetime sleeping. Lack of sleep may cause health issues, influence mood, and interfere with cognitive performance [1,2]. Examination of sleep is usually performed with the aid of polysomnography (PSG). PSG is used to examine multiple parameters that may be useful in the diagnosis of sleep disorders, or may be analyzed in pursuit of a deeper understanding of sleep itself. Hollan, Dement, and Raynal introduced the term Polysomnography in 1974. PSG is performed using an electronic device equipped to monitor multiple physiologic parameters during sleep by recording corresponding electrophysiological signals, for instance: from the brain via electroencephalogram (EEG), from the eyes via electrooculogram (EOG), from the skeletal muscles via electromyogram (EMG), and from the heart via electrocardiogram (ECG) [3]. To collect this data, recording devices are attached to the relevant locations of the body, typically including three EEG electrodes, one EMG electrode and two EOG electrodes. ECG is also a compulsory component of PSG. Additionally, the monitoring of respiratory functions may be desired in the diagnosis of respiratory disorders such as sleep apnea and require the addition of other tools applied in conjunction with the EEG electrodes, most often a pulse oximeter, oral thermometer, nasal cannula, thoracic and abdominal belt, and a throat microphone [4,5].

Figure 1 represents the standard system used for measuring the EEG signal, termed as the 10–20 system, in which the minimum number of electrodes used is 21. This method regulates the physical placement and designations of electrodes on the scalp. The head is divided into proportions from important sites of the skull so that all areas of the brain are adequately covered. The label of 10–20 indicates that the actual distances between neighboring electrodes are either 10% or 20% of the distance from the nasion (front side of the head/anteriorly) to the inion (back side of the head/posteriorly) between the ears and those where electrode points are chosen. Generally, electrodes marked with even numbers are placed on the right side of the head and those marked with odd numbers on the left side. The electrodes are also marked with letters to represent their locations relative to the anatomical divisions of the brain: F (frontal), C (central), T (temporal), P (parietal), O (occipital), and Fp (Frontal pole). A subscript z is used to mark the midline electrodes as zero.

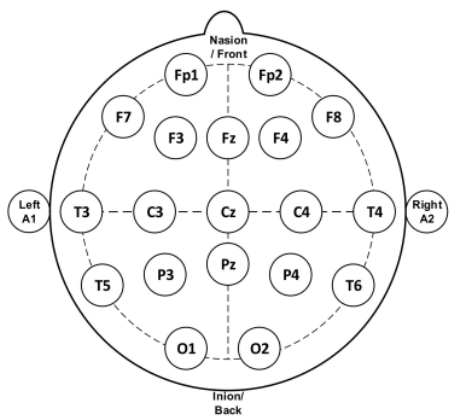

Figure 1. Electrodes placement of electroencephalogram (EEG) measurement [4] (reproduced with permission by Elsevier (License Number 4781771458692)).

The electric signal in the brain is determined by measuring the difference of the electric activity between the two electrodes over a period of time. As it propagates, the signal gradually decays with distance from the source. Finally, the signal has a decreased value since only one of the parallel combinations of electrodes gives precise measurement [4,6].

EEG waveforms have several kinds of rhythms. These rhythms are remarkably useful for classification annotation of sleep score as recorded by PSG. In a normal EEG, we differentiate these rhythms into five frequency bands. Table 1 lists the frequency and amplitude ranges of these bands [4,7]

Table 1. Frequency and amplitude range of the EEG signal.

Bands	Frequencies (Hz)	Amplitude (μV)
Delta	0.5–3.5	20–100
Theta	3.5–7.5	10
Alpha	7.5–12	2–100
Beta	12–30	5–10
Gamma	>31	-

Human sleep consists of cyclic stages, and the sleep stages are essential sections of activity during sleep. Three main stages of the sleep cycle are awake, non-REM (NREM) sleep, and rapid eye movement (REM) sleep. The NREM phase is also called dreamless sleep: breathing is slow and heart rate and blood pressure are normal. NREM sleep eventually deepens and leads to REM sleep. The REM stage occurs most often while dreaming. At the time, the body goes into a temporary paralysis to prevent it from acting out these dreams. However, during REM sleep the eyes move quickly back and forth. The absence of one of these stages or the overabundance of another can lead to the diagnosis of numerous conditions ranging from sleep apnea, hypersomnia, insomnia, or sleep talking [8].

There are two recognized standards for interpreting sleep stages based on sleep recordings: the Rechtschaffen and Kales (R & K) criteria and the American Academy of Sleep Medicine (AASM) criteria. The R & K recommendations classify sleep into seven discrete stages: wake/wakefulness, S1/drowsiness, S2/light sleep, S3/deep sleep, S4/deep or wave sleep, REM and MT/movement time [9]. The AASM criteria are a modified version of the R & K criteria. Some differences between the AASM and R & K criteria are as follows [9,10]:

1. NREM stages in the R & K criteria (S1, S2, S3, and S4) are referred to as stages N1, N2, and N3 in the AASM criteria.
2. In the AASM criteria, deep sleep (N3) is a combination of the S3 and S4 stages of the R & K criteria.
3. Movement time (MT) is eliminated as a sleep stage in the AASM criteria.

The stages of sleep can be thought of as a cyclic alternation of non-rapid eye movement (NREM) and rapid eye movement (REM) stages [11]. It has been recognized that NREM sleep consists of four distinct stages S1, S2, S3, and S4, each with specific characteristics. In S1, the patient is drowsy but still awake. The appearance of sleep spindles, vertex sharp waves, and K complexes mark S2 sleep. Shallow sleep consists of both S1 and S2, while deep sleep consists of S3 and S4 [12].

Conventionally, technicians have interpreted and marked the sleep stages manually. As such, it is a time-intensive process as well as being expensive and dependent on human resources. Because it is time consuming, expensive, and is an enormous process, it is not suitable to hold the large EEG datasets for sleep stages annotation by the human expert [13]. As a result, it has become necessary to develop a sleep stage classification in order to achieve better accuracy.

Previous attempts at automated classification of sleep stages have been based on single-channel as well as multi-channel EEG recordings and various other physiological markers. Ronzhina et al. described a single-channel EEG based scheme utilizing an artificial neural network coupled with power spectrum density analysis of EEG recordings [14]. Zhu et al. analyzed nine features from single-channel EEG recordings and applied an artificial intelligence technique referred to as a support vector machine (SVM) to perform classification [15]. High classification performance has been reported by Huang by applying short-time Fourier transform to a two-channels recording of forehead EEG signals and a relevance vector machine [16]. In addition, Aboalayon et al. have conducted a comprehensive review of automatic sleep stage classification (AASC) systems, which includes a survey of processing techniques including pre-processing, feature extraction, feature selection, dimensionality reduction, and classification. This study evaluated AASC methods against the sleep-EDF database based on single-channel EEG recordings, and is remarkable for having selected 10 s epochs for its analysis. Their model's performance achieved the highest accuracy in comparison to previous results [17]. Braun et al. had applied low dimensional FFT features on the sleep-EDF database with the usage of eight statistical features from the Pz-Oz EEG channel. The classification performance had reached with the accuracy 90.9%, 91.8%, 92.4%, 94.3%, and 97.1% for all 6- to 2-state sleep stages [18].

In this paper, we present a system of sleep stage classification based on EEG signals. Instead of using complicated processes of signal filtering and feature extraction, we utilized high-dimensional features calculated by fast Fourier transform (FFT) from single- or multi-channel EEG signals. FFT is one of the traditional and verified techniques capable of extracting features from EEG signals. If an EEG signal is recorded at a sampling frequency of 100 Hz, the FFT can separate the signal into features in the range of 0–100 Hz. Typically, in previous studies, a small number of FFT features corresponding to the bands shown in Table 1 were extracted and used. However, a sampling window of 30 s at 100 Hz sampling frequency allows for an extraction of at most 3000 features in the range of

0–100 Hz. In this study, we demonstrate that by incorporating high-dimensional FFT features and simple feature selection by random forest algorithm into the analysis, it is possible to outperform state-of-the-art algorithms for the Sleep-EDF database. Our proposed approach consists of three main steps: brainwaves acquisition from EEG channel, feature processing, and finally, the classification evaluation by measuring the accuracy. The flowchart of our approach is shown in Figure 2.

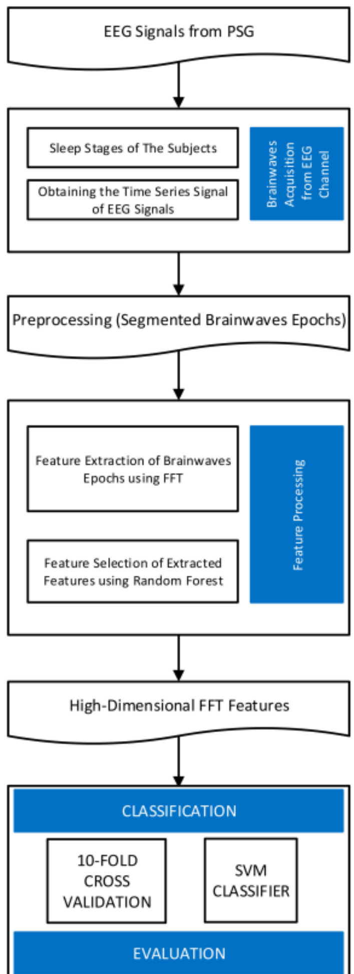

Figure 2. The flowchart of the proposed research.

2. Materials and Methods

2.1. Experimental Data

Sleep-EDF Dataset

The dataset is open-source, and many previous researchers have utilized this dataset in sleep scoring research [15,17–22]. Among three available versions of the dataset, we used an expanded version containing 61 recordings from 42 Caucasian male and female subjects. The subjects' ages ranged from 18 to 79 years. This dataset was organized into two sub-sets. The first subset with 39 recordings from 20 subjects was EEG data recorded in a study from 1987 to 1991. These subjects were healthy and in ambulatory condition. The second subset with 22 recordings from 22 subjects was EEG data recorded in a study in 1994 and the subjects reported slight difficulty in falling asleep but were otherwise healthy. The EEG data had been collected over 24 h of the daily lives of the subjects. A miniature telemetry system recorded nocturnal EEG data from four subjects in a hospital [19]. The data was collected from just two channels: Fpz-Cz and Pz-Oz, at a sampling frequency of 100 Hz. The previous researchers had established that on single-channel analysis the Pz-Oz channel demonstrated improved performance over the Fpz-Cz channel. Using R & K criteria, EEG recordings of both of the subsets had been annotated by an experienced sleep technician on a 30 s basis. Therefore, the duration of each epoch is established as 30 s and yielded 3000 samples. The epochs had been annotated by sleep technicians as: AWA, REM, S1, S2, S3, S4, "Movement Time" or "Unscored." On the other hand, the annotations using AASM criteria consisted of the designations AWA, REM, N1, N2, N3, and "Unknown sleep stage." The number of samples according to R & K criteria are shown in Table 2. After removing "Movement Time" and "Unscored," total number of the samples is 127,663. The epoch duration was 30 s.

Table 2. The number of samples in Sleep-EDF dataset (R & K criteria).

# of Classes	AWA	REM	S1	S2	S3	S4
6	74,827	11,848	4,848	27,292	5075	3773
5	74,827	11,848	4,848	27,292	8848	
4	74,827	11,848	32,140		8848	
3	74,827	11,848	40,988			
2	74,827	52,836				

As far as features analyzed, we prepared 6000 features extracted from two channels (Pz_Oz and Fpz_Cz). For each channel, 1000 features were extracted at a sampling frequency of 100 Hz and epoch lasted 30 s. The features were used in the experiments separately or in combination.

2.2. Feature Extraction with Fast Fourier Transform (FFT)

The feature represents a differentiating property or an operative component identified in a section of a pattern, and a recognizable measurement. Feature extraction is a critical step in EEG signal processing. Consequently, minimizing the loss of valuable information attached to the signal is one of the goals of feature extraction. Additionally, feature extraction decreases the resources required to describe a vast set of data accurately. When carried out successfully, feature extraction can minimize the cost of information processing, reduce the complexity of data implementation, and mitigate the possible need to compress the information [23].

The extraction of informative statistical features from the EEG signal is necessary to perform sleep stage classification efficiently. In general, the EEG signal is highly complex and non-linear, so it would be better to use a non-linear model [24]. In this study, the fast Fourier transform (FFT) is utilized to extract the features of EEG signal for sleep stage classification. Hence, the values of a given time-series data as a numeric sequence data are converted into a finite set of the frequency domain. Then, to deconstruct signals into segmented EEG signal sequences, we divided them into equal time intervals called epochs. The length of each epoch was set to every 30 s of EEG signal. Accordingly, the epochs were then processed using frequency analysis in which frequency spectra were generated using FFT. We used FFT to convert a signal from its original, time domain signal to a representation in the frequency domain signal [25]. Figure 3 represents in the form of time-domain signal and frequency-domain signal.

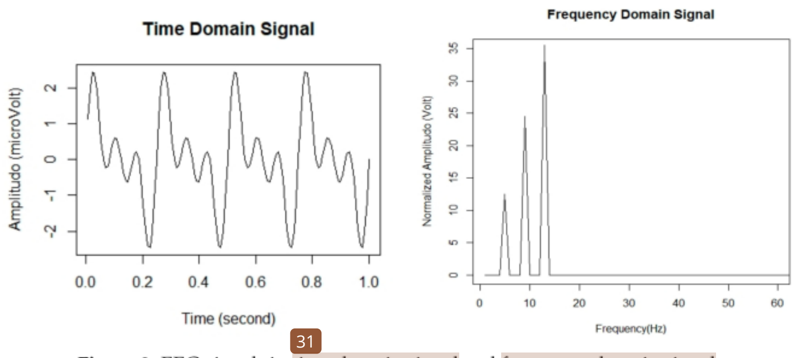

Figure 3. EEG signals in time domain signal and frequency domain signal.

Previous studies have shown that FFT is a promising tool for stationary signal processing, and enjoys a speed advantage over virtually all other available methods in real-time applications, and it is more appropriate for sine waveforms such as in EEG signals. However, the disadvantage is that it does not have excellent spectral estimation and cannot be employed for analysis of short EEG signals [23].

2.3. Feature Selection and Optimization

Feature extraction is an effective way of recognizing and visualizing significant data. This process shortens the time for training and application, as well as reducing demands for data calculation and storage. Some researchers combine several feature extraction techniques in order to achieve better data analysis. Consequently, application of multiple processes may often affect feature redundancy and expansion of feature dimension. Feature selection reduces the dimension of feature space and minimizes the data training and application [26]. In this study, we conducted a simple feature selection based on the importance of each feature evaluated by random forest algorithm. Mean decrease in Gini was calculated by using the randomForest package for R, then all features were sorted in the descending order of this value. For Sleep-EDF dataset, we examined the number of important features to be selected with 50 increments in between 500 and 2500, and the most appropriate feature subset were determined for each number of classes. More details about the feature selection in this study are found in [27] which describes basically the same feature selection method.

2.4. Classification Evaluation

The classification step was completed with the 5- or 10-fold cross validation. This means for each process, this step is repeated 5 or 10 times per sample. The 5- or 10-fold cross-validation comes from the cross validation technique to evaluate prediction performance from classification model. This technique splits the dataset into training and test data. We trained each fold in order to have a better estimation of the true error rate of each set. The model is created by using the training data, and the test data is used for evaluating the performance of prediction.

Among the various classification algorithms, we adopted the multiclass support vector machine (SVM) algorithm, a supervised machine learning method, implemented in the kernlab package for R. The SVM classifier is a popular algorithm widely applied to various problems in machine learning. SVM constructs the maximum margin around the separating hyperplane between the classes. In this study, we utilized a Gaussian or Radial Basis Function (RBF) Kernel. One of the advantages of the SVM method is that this method is effective when the number of features is greater than the number of samples. In addition, the model is sufficient as a classification model of the EEG signal.

3. Experimental Results

3.1. Classification of Sleep-EDF Dataset

In this experiment, 2–6 classes defined by R & K criteria were selected as class labels. Features from two channels were analyzed separately or in combination. From one channel, 3000 features (i.e., 100 Hz × 30 s) were extracted. For performance evaluation, five-fold cross-validation was conducted. The results obtained using our method are compared with the results acquired using other state-of-the-art methods in. In this Table 3, it can be seen that our method with 6000 features from two channels (Pz_Oz and Fpz_Cz) outperformed all other methods in the classification of 6 to 2 classes.

Table 3. Performance comparison on Sleep-EDF dataset (R & K criteria, 2–6 classes).

Method	Length of Epoch (s)	# of Epochs	# of Classes	Accuracy (%)
Nakamura et al. 2017	30	126,699	6	86.60
			5	88.60
			4	91.00
			3	94.50
			2	97.40
Yildirim et al. 2019 [22]	30	127,512	6	89.43
			5	90.48
			4	92.24
			3	94.23
			2	97.85
Our method (Pz_Oz)	30	127,663	6	90.17
			5	91.42
			4	92.24
			3	94.36
			2	97.79
Our method (Fpz_Cz)	30	127,663	6	89.70
			5	88.57
			4	90.02
			3	92.69
			2	97.13
Our method (Pz_Oz and Fpz_Cz)	30	127,663	6	90.77
			5	91.73
			4	92.82
			3	94.41
			2	97.88

3.2. Classification of Sleep-EDF Dataset Expanded (197 Recordings)

The results of applying our method against the latest, extended version of the Sleep-EDF database, show that in contrast to the first version of the database which consisted of 61 recordings (version 1), the latest version consists of 197 recordings (version 2, released in 2018). It has been studied in many recent papers (e.g., [27,28], however, because of its large size, it is rarely studied as a whole (many papers which classified it are using only a small subset of it). Therefore, it is hard to compare the performance on it under the same or similar conditions. Instead of comparing performances, in this subsection we mainly analyzed the relationship between classification performance and the balance of classes.

Table 4 contains the result of the classification experiment using our method. "SC" and "ST" found in the recording ID. The prefixes "SC" and "ST" stand for "Sleep Cassette" and "Sleep Telemetry", respectively. In this experiment, ten-fold cross validation was conducted for each recording. The average, highest, and lowest accuracies were 87.84%, 96.54%, and 37.03%, respectively. Since the accuracies are greatly affected by the degree of sample distribution among the classes in each recording, a large discrepancy exists between the highest and lowest accuracies. For example, in the recording SC4201, which achieved the highest accuracy, the AWA class occupies ~73% of the recording. In contrast, the lowest accuracy was achieved by ST7151 with a more even distribution between the classes (AWA:REM:S1:S2:S3:S4 = 104:143:78:304:142:126). This is even more clearly demonstrated in Figure 4, where we show the relationship between accuracy and degree of class imbalance (represented in this experiment by the standard deviation of class sizes in a recording). There appears to be an almost linear relationship (correlation coefficient was 0.9857).

Table 4. Performance of classification for each recording in Sleep-EDF database (version 2).

ID	Accuracy (%)	ID	Accuracy (%)	ID	Accuracy (%)	ID	Accuracy (%)
SC4001	94.17	SC4252	92.77	SC4522	94.24	SC4812	91.26
SC4002	92.49	SC4261	89.25	SC4531	89.22	SC4821	92.67
SC4011	94.27	SC4262	92.43	SC4532	92.68	SC4822	89.17
SC4012	93.43	SC4271	90.37	SC4541	93.68	ST7011	75.56
SC4021	94.11	SC4272	91.59	SC4542	90.28	ST7012	81.39
SC4022	92.86	SC4281	91.27	SC4551	90.97	ST7021	83.62
SC4031	95.88	SC4282	91.42	SC4552	94.29	ST7022	79.29
SC4032	94.44	SC4291	91.54	SC4561	84.57	ST7041	58.73
SC4041	90.51	SC4292	91.67	SC4562	90.25	ST7042	65.79
SC4042	91.58	SC4301	91.48	SC4571	89.66	ST7051	43.61
SC4051	95.29	SC4302	92.94	SC4572	92.47	ST7052	84.77
SC4052	92.51	SC4311	91.99	SC4581	89.89	ST7061	80.09
SC4061	95.33	SC4312	89.40	SC4582	88.17	ST7062	85.53
SC4062	94.27	SC4321	88.92	SC4591	91.41	ST7071	79.06
SC4071	93.71	SC4322	92.26	SC4592	85.54	ST7072	81.05
SC4072	93.70	SC4331	91.51	SC4601	92.68	ST7081	83.10
SC4081	92.56	SC4332	94.24	SC4602	86.79	ST7082	81.77
SC4082	90.92	SC4341	89.21	SC4611	87.08	ST7091	75.45
SC4091	91.59	SC4342	96.42	SC4612	93.10	ST7092	77.99
SC4092	90.59	SC4351	94.34	SC4621	84.74	ST7101	80.34
SC4101	93.38	SC4352	90.59	SC4622	91.20	ST7102	75.67
SC4102	94.84	SC4362	92.08	SC4631	91.06	ST7111	82.40
SC4111	92.12	SC4371	91.13	SC4632	93.04	ST7112	83.43
SC4112	95.32	SC4372	86.79	SC4641	94.62	ST7121	79.76
SC4121	92.54	SC4381	93.39	SC4642	92.58	ST7122	82.65
SC4122	91.56	SC4382	93.23	SC4651	89.64	ST7131	85.80
SC4131	92.97	SC4401	91.94	SC4652	85.62	ST7132	76.47
SC4141	95.12	SC4402	93.40	SC4661	85.77	ST7141	75.50

Table 4. *Cont.*

ID	Accuracy (%)	ID	Accuracy (%)	ID	Accuracy (%)	ID	Accuracy (%)
SC4142	95.31	SC4411	92.92	SC4662	88.53	ST7142	72.74
SC4151	92.97	SC4412	89.87	SC4671	91.07	ST7151	37.03
SC4152	93.25	SC4421	95.23	SC4672	93.35	ST7152	79.94
SC4161	90.29	SC4422	92.07	SC4701	88.10	ST7161	48.93
SC4162	91.04	SC4431	91.50	SC4702	92.39	ST7162	74.06
SC4171	92.20	SC4432	92.08	SC4711	88.52	ST7171	79.37
SC4172	86.52	SC4441	88.84	SC4712	93.23	ST7172	77.20
SC4181	92.34	SC4442	90.77	SC4721	83.95	ST7181	82.92
SC4182	91.00	SC4451	91.24	SC4722	87.05	ST7182	54.30
SC4191	90.25	SC4452	90.94	SC4731	88.42	ST7191	47.16
SC4192	91.21	SC4461	94.06	SC4732	87.96	ST7192	87.10
SC4201	96.54	SC4462	93.78	SC4741	92.12	ST7201	66.16
SC4202	95.04	SC4471	90.90	SC4742	90.71	ST7202	69.62
SC4211	92.40	SC4472	85.81	SC4751	94.07	ST7211	79.40
SC4212	93.67	SC4481	90.11	SC4752	87.98	ST7212	77.28
SC4221	88.17	SC4482	93.31	SC4761	92.82	ST7221	82.95
SC4222	89.88	SC4491	93.88	SC4762	89.00	ST7222	82.78
SC4231	93.08	SC4492	92.37	SC4771	90.33	ST7241	65.62
SC4232	86.76	SC4501	91.16	SC4772	90.19	ST7242	62.94
SC4241	92.30	SC4502	93.77	SC4801	91.42		
SC4242	94.92	SC4511	90.53	SC4802	91.43		
SC4251	96.34	SC4512	92.85	SC4811	91.79		

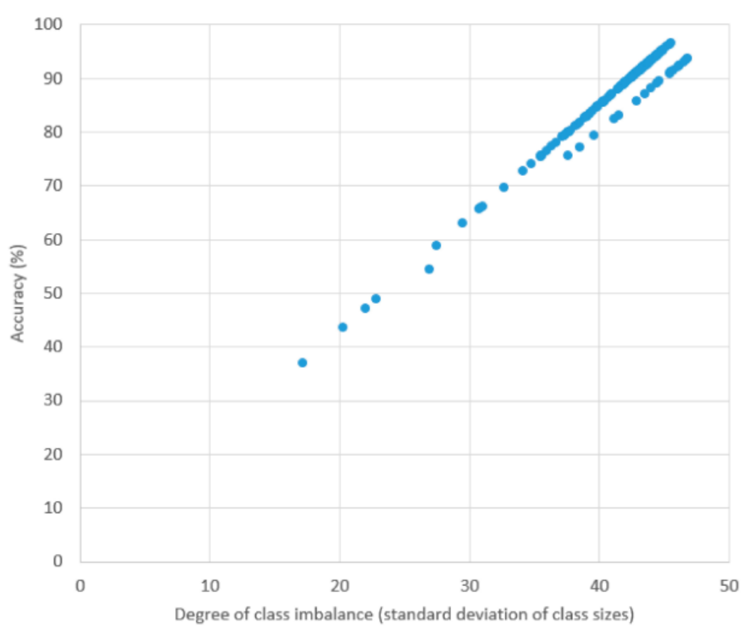

Figure 4. Plot of accuracy and standard deviation of class sizes in each recording.

4. Discussions and Conclusions

In order to improve the performance of sleep stage classification, previous work has mainly focused on the following points:

- More effective methods of feature extraction from the original EEG signal (e.g., wavelet transform)
- Application of filters (e.g., band-pass filter) and noise reduction algorithms
- Identification of better classifier algorithms (e.g., random forest, adaptive boosting, and convolutional neural network)
- Improvement of class imbalance by under- and/or over-sampling (e.g., SMOTE)

In contrast, we have demonstrated in this paper that fully utilizing thousands of FFT features extracted from single- and multi-channel EEG signals in combination with simple feature selection is an effective means of improving the performance of automated sleep stage classification. In our experiment on 6- to 2-class classification against the Sleep-EDF dataset, our method outperformed other recent and advanced methods.

Additionally, we demonstrated the result of application of our method to the classification of the recording included in the latest version of Sleep-EDF database. We clearly showed that accuracy in classifying a recording is highly influenced by the degree of class imbalance.

The differences between the amounts of the majority class and the minority class in the datasets leads to an imbalanced dataset. In other words, balanced class distributions are essential in supervised learning as standard classification. One of the methods for solving this problem is by doing oversampling and it aims to achieve a balanced class distribution by creating an artificial data. SMOTE (synthetic minority over-sampling technique) is an over-sampling method that is typically used to balance an imbalanced data as a part of machine learning. New instances are created as minority class instances from minority class neighbors that performed like the original instances of the minority class [29]. Related to sleep stage classification, our experimental results suggested that by combining our method with under- and/or over-sampling methods like SMOTE, we may achieve better classification performance of the recordings in the latest Sleep-EDF database.

One of the disadvantages in our method is the intensive computational requirements in memory and processor. However, it also means that if the available resource of computing is rich, its performance can be further improved. In addition to the analysis of relationship between the number of features and performance, we need to conduct future work on the effectiveness of our method in other datasets.

Author Contributions: Conceptualization, M.K.D. and K.S.; formal analysis, B.P., K.R.M., and F.I.; funding acquisition, K.S.; investigation, M.K.D.; methodology, K.S.; software, N.G.N., M.R.F., and M.K.; supervision, K.S.; writing—original draft, M.K.D.; writing—review and editing, B.P., K.R.M., F.I., and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by JSPS KAKENHI Grant Number JP18K11525 and Kanazawa University CHOZEN project.

Acknowledgments: The first and second authors would like to gratefully acknowledge the BUDI-LN scholarship from Indonesia Endowment Fund for Education (LPDP), Ministry of Education and Culture Republic of Indonesia (KEMENDIKBUD), and Ministry of Research and Technology of Republic Indonesia (KEMENRISTEK). In this research, the super-computing resource was provided by Human Genome Center, the Institute of Medical Science, the University of Tokyo. Additional computation time was provided by the super computer system in Research Organization of Information and Systems (ROIS), National Institute of Genetics (NIG).

Conflicts of Interest: The authors declare no conflict of interest

References

1. Touchette, É.; Petit, D.; Seguin, J.R.; Boivin, M.; Tremblay, R.E.; Montplaisir, J.Y. Associations Between Sleep Duration Patterns and Behavioral/Cognitive Functioning at School Entry. *Sleep* **2007**, *30*, 1213–1219. [CrossRef] [PubMed]
2. Walker, M.P.; Stickgold, R. Sleep, Memory, and Plasticity. *Annu. Rev. Psychol.* **2006**, *57*, 139–166. [CrossRef] [PubMed]

3. Álvarez-Estévez, D.; Moret-Bonillo, V. Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 54–63. [CrossRef] [PubMed]
4. Keenan, S.A. *Handbook of Clinical Neurophysiology, An Overview of Polysomnography*; Elsevier B.V.: Amsterdam, The Netherlands, 2005; Volume 6, Chapter 3, p. 18.
5. Billiard, M.; Bae, C.; Avidan, A. *Sleep Medicine*; Smith, H.R., Comella, C.L., Hogl, B.L., Eds.; Cambridge University Press: Cambridge, UK, 2012.
6. Kaniusas, E. *Biomedical Signals and Sensors I*; Springer: Berlin, Germany; pp. 1–26. [CrossRef]
7. Aboalayon, K.A.I.; Faezipour, M. Multi-Class SVM Based on Sleep Stage Identification Using EEG Signal. Proceedings of the 2014 IEEE Healthcare Innovation Conference (HIC), Piscataway, NJ, USA, 8–10 October 2014; pp. 181–184.
8. Thorpy, M.J. *The International Classification Of Sleep Disorders: Diagnostic and Coding Manual. Rev.Ed*; One Westbrook Corporate Center, Suite 920: Westchester, IL, USA, 2001.
9. Rechtschaffen, A.; Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*; National Government Publication: Los Angeles, CA, USA, 1968.
10. Iber, C.; Medicine, A.A.o.S. *The AASM Manual for The Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*; American Academy of Sleep Medicine: Darien, IL, USA, 2007.
11. Hobson, J.A. Sleep is of The Brain, by The Brain and For The Brain. *Nature* **2005**, *437*, 1254–1256. [CrossRef] [PubMed]
12. Marshall, L.; Helgadottir, H.; Molle, M.; Born, J. Boosting Slow Oscillations During Sleep Potentiates Memory. *Nature* **2006**, *444*, 610–613. [CrossRef] [PubMed]
13. Norman, R.G.; Pal, I.; Stewart, C.; Walsleben, J.A.; Rapoport, D.M. Interobserver Agreement Among Sleep Scorers From Different Centers in a Large Dataset. *Sleep* **2000**, *23*, 901–908. [CrossRef] [PubMed]
14. Ronzhina, M.; Janousek, O.; Kolarova, J.; Novakova, M.; Honzik, P.; Provaznik, I. Sleep Scoring Using Artificial Neural Networks. *Sleep Med. Rev.* **2012**, *16*, 251–263. [CrossRef] [PubMed]
15. Zhu, G.; Li, Y.; Wen, P.P. Analysis and Classification of Sleep Stages Based on Difference Visibility Graphs From A Single-Channel EEG Signal. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1813–1821. [CrossRef] [PubMed]
16. Huang, C.-S.; Lin, C.-L.; Ko, L.-W.; Liu, S.-Y.; Su, T.-P.; Lin, C.-T. Knowledge-based Identification of Sleep Stages Based on Two Forehead Electroencephalogram Channels. *Front. Neurosci.* **2014**, *8*, 263. [CrossRef] [PubMed]
17. Aboalayon, K.A.I.; Faezipour, M.; Almuhammadi, W.S.; Moslehpour, S. Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation. *Entropy* **2016**, *18*, 272. [CrossRef]
18. Braun, E.T.; Silvera, T.L.T.D.; Kozakevicius, A.D.J.; Rodrigues, C.R.; Giovani, B. Sleep Stages Classification Using Spectral Based Statistical Moments as Features. *Revista de Informática Teórica e Aplicada* **2018**, *25*. [CrossRef]
19. Hassan, A.R.; Subasi, A. A Decision Support System for Automated Identification of Sleep Stages from Single-Channel EEG Signals. *Knowl.-Based Syst.* **2017**, *128*, 115–124. [CrossRef]
20. Liang, S.-F.; Kuo, C.-E.; Hu, Y.-H.; Pan, Y.-H.; Wang, Y.-H. Automatic Stage Scoring of Single-Channel Sleep EEG by Using Multiscale Entropy and Autoregressive Models. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 1649–1657. [CrossRef]
21. Nakamura, T.; Adjei, T.; Alqurashi, Y.; Looney, D.; Morrell, M.J.; Mandic, D.P. Complexity science for sleep stage classification from EEG. In Proceedings of the International Joint Conference on Neural Networks, Anchorage, AK, USA, 14–19 May 2017.
22. Yildirim, O.; Baloglu, U.B.; Acharya, U.R. A Deep Learning Model for Automated Sleep Stages Classification Using PSG Signals. *Int. J. Environ. Res. Public Health* **2019**, *16*, 599. [CrossRef] [PubMed]
23. Al-Fahoum, A.S.; Al-Fraihat, A.A. Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neurosci.* **2014**, *2014*, 730218. [CrossRef] [PubMed]
24. Freeman, W.J.; Skarda, C.A. Spatial EEG patterns, non-linear dynamics and perception: the neo-Sherringtonian view. *Brain Res.* **1985**, *357*, 147–175. [CrossRef]
25. Nussbaumer, H.J. *Fast Fourier Transform and Convolution Algorithms*; Springer: Berlin/Heidelberg, Germany, 1981. [CrossRef]
26. Wen, T.; Zhang, Z. Effective and Extensible Feature Extraction Method Using Genetic Algorithm-based Frequency-Domain Feature Search for Epileptic EEG Multiclassification. *Medicine (Baltimore)* **2017**, *96*, e6879. [CrossRef] [PubMed]

27. Huang, W.; Guo, B.; Shen, Y.; Tang, X.; Zhang, T.; Li, D.; Jiang, Z. Sleep staging algorithm based on multichannel data adding and multifeature screening. *Comput Methods Programs Biomed.* **2019**, *187*, 105253. [CrossRef] [PubMed]
28. Timplalexis, C.; Diamantaras, K.; Chouvarda, I. Classification of Sleep Stages for Healthy Subjects and Patients with Minor Sleep Disorders. In Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, 28–30 October 2019; pp. 344–351.
29. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intel. Res.* **2002**, *16*, 321–357. [CrossRef]

Classification of Brainwaves for Sleep Stages by High-Dimensional FFT Features from EEG Signals

ORIGINALITY REPORT

19%	7%	15%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ahnaf Rashik Hassan, Mohammed Imamul Hassan Bhuiyan. "Dual tree complex wavelet transform for sleep state identification from single channel electroencephalogram", 2015 IEEE International Conference on Telecommunications and Photonics (ICTP), 2015 Publication	1%
2	Le, Tu Kien T., Osamu Hirose, Lan Anh T. Nguyen, Thammakorn Saethang, Vu Anh Tran, Xuan Tho Dang, Duc Luu Ngo, Mamoru Kubo, Yoichi Yamada, and Kenji Satou. "Inference of domain-domain interactions by matrix factorisation and domain-level features", International Journal of Functional Informatics and Personalised Medicine, 2014. Publication	1%
3	Submitted to Unviersidad de Granada Student Paper	1%
4	ebin.pub Internet Source	

		1%
5	www.coursehero.com Internet Source	1%
6	Sagar Santaji, Snehal Santaji, Veena Desai. "Automatic sleep stage classification with reduced epoch of EEG", Evolutionary Intelligence, 2021 Publication	1%
7	seer.ufrgs.br Internet Source	1%
8	Divya Acharya, Medha Tiwari, Bullepallii Bhulakshmi Devi, Sadaf Shaikh, Kalaivaani Natarajan, Anu Maria Babu, Arpit Bhardwaj. "Chapter 29 Feature Extraction for Classification Methods of EEG Signals", Springer Science and Business Media LLC, 2021 Publication	1%
9	"Image Analysis and Recognition", Springer Science and Business Media LLC, 2018 Publication	1%
10	WWW.MDPI.COM Internet Source	1%
11	web.archive.org Internet Source	<1%

12	Tingxi Wen, Zhongnan Zhang. "Effective and extensible feature extraction method using genetic algorithm-based frequency-domain feature search for epileptic EEG multiclassification", Medicine, 2017 Publication	<1%
13	Huang, Chih-Sheng, Chun-Ling Lin, Li-Wei Ko, Shen-Yi Liu, Tung-Ping Su, and Chin-Teng Lin. "Knowledge-based identification of sleep stages based on two forehead electroencephalogram channels", Frontiers in Neuroscience, 2014. Publication	<1%
14	Peyman Ghasemzadeh, Hashem Kalbkhani, Shadi Sartipi, Mahrokh G. Shayesteh. "Classification of sleep stages based on LSTAR model", Applied Soft Computing, 2019 Publication	<1%
15	export.arxiv.org Internet Source	<1%
16	Submitted to University of Leeds Student Paper	<1%
17	Ye Yuan, Kebin Jia, Fenglong Ma, Guangxu Xun, Yaqing Wang, Lu Su, Aidong Zhang. "A hybrid self-attention deep learning framework for multivariate sleep stage classification", BMC Bioinformatics, 2019 Publication	<1%

18	giapjournals.com Internet Source	<1 %
19	Huaming Shen, Feng Ran, Meihua Xu, Allon Guez, Ang Li, Aiying Guo. "An Automatic Sleep Stage Classification Algorithm Using Improved Model Based Essence Features", Sensors, 2020 Publication	<1 %
20	M. Murugappan, S. Murugappan. "Human emotion recognition through short time Electroencephalogram (EEG) signals using Fast Fourier Transform (FFT)", 2013 IEEE 9th International Colloquium on Signal Processing and its Applications, 2013 Publication	<1 %
21	Weiyi Yang, Yujuan Si, Gong Zhang, Di Wang, Meiqi Sun, Wei Fan, Xin Liu, Liangliang Li. "A novel method for automated congestive heart failure and coronary artery disease recognition using THC-Net", Information Sciences, 2021 Publication	<1 %
22	Submitted to European University of Lefke Student Paper	<1 %
23	Wolfe, Jeremy, Kluender, Keith, Levi, Dennis. "Sensation and Perception", Sensation and Perception, 2021	<1 %

Publication

24	Ahnaf Rashik Hassan, Mohammed Imamul Hassan Bhuiyan. "Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting", Computer Methods and Programs in Biomedicine, 2017 Publication	<1 %
25	Chen, Shih-Jui, Chia-Ju Peng, Yi-Chun Chen, Yean-Ren Hwang, Ying-Sian Lai, Shou-Zen Fan, and Kuo-Kuang Jen. "Comparison of FFT and marginal spectra of EEG using empirical mode decomposition to monitor anesthesia", Computer Methods and Programs in Biomedicine, 2016. Publication	<1 %
26	Submitted to Queensland University of Technology Student Paper	<1 %
27	Gustavo E. A. P. A. Batista. "A study of the behavior of several methods for balancing machine learning training data", ACM SIGKDD Explorations Newsletter, 6/1/2004 Publication	<1 %
28	Peyman Ghasemzadeh, Hashem Kalbkhani, Mahrokh G. Shayesteh. "Sleep stages classification from EEG signal based on	<1 %

Stockwell transform", IET Signal Processing, 2019
Publication

29	Submitted to University of Technology, Sydney Student Paper	<1 %
30	Yang Li, Momoka Kasuya, Kazuo Sakiyama. "Comprehensive Evaluation on an ID-Based Side-Channel Authentication with FPGA-Based AES", Applied Sciences, 2018 Publication	<1 %
31	A.H. Muqaibel, A. Safaai-Jazi. "A new formulation for characterization of materials based on measured insertion transfer function", IEEE Transactions on Microwave Theory and Techniques, 2003 Publication	<1 %
32	Georgios Douzas, Fernando Bacao, Felix Last. "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE", Information Sciences, 2018 Publication	<1 %
33	Hojat Ghimatgar, Kamran Kazemi, Mohammad Sadegh Helfroush, Ardalan Aarabi. "An automatic single-channel EEG-based sleep stage scoring method based on	<1 %

	hidden Markov Model", Journal of Neuroscience Methods, 2019 Publication	
34	IFMBE Proceedings, 2014. Publication	<1 %
35	Submitted to University of Nottingham Student Paper	<1 %
36	Amjed S. Al-Fahoum, Ausilah A. Al-Fraihat. "Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains", ISRN Neuroscience, 2014 Publication	<1 %
37	journals.plos.org Internet Source	<1 %
38	www.tandfonline.com Internet Source	<1 %
39	Lecture Notes in Computer Science, 2015. Publication	<1 %
40	Luigi Fiorillo, Alessandro Puiatti, Michela Papandrea, Pietro-Luca Ratti et al. "Automated sleep scoring: A review of the latest approaches", Sleep Medicine Reviews, 2019 Publication	<1 %

41 S. Rajalakshmi, R. Venkatesan. "Chapter 3 Exploring Cepstral Coefficient Based Sleep Stage Scoring Method for Single-Channel EEG Signal Using Machine Learning Technique", Springer Science and Business Media LLC, 2018
Publication

<1 %

42 Kenji Sugioka, Bruce Bowerman. "Combinatorial Contact Cues Specify Cell Division Orientation by Directing Cortical Myosin Flows", Developmental Cell, 2018
Publication

<1 %

43 Mahdi Rahbar Alam, Reza Sameni. "Automatic Wake-Sleep Stages Classification using Electroencephalogram Instantaneous Frequency and Envelope Tracking", Cold Spring Harbor Laboratory, 2020
Publication

<1 %

44 Pejman Memar, Farhad Faradji. "A Novel Multi-Class EEG-Based Sleep Stage Classification System", IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018
Publication

<1 %

45 Peker, Musa. "An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms", Neurocomputing, 2016.

<1 %

Publication

46	iosrjournals.org Internet Source	<1 %
47	"Abstracts", Journal of Sleep Research, 08/31/2010 Publication	<1 %
48	Hogeon Seo, Seunghyeok Back, Seongju Lee, Deokhwan Park, Tae Kim, Kyoobin Lee. "Intra- and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG", Biomedical Signal Processing and Control, 2020 Publication	<1 %
49	Mikito Ogino, Yasue Mitsukura. "Portable Drowsiness Detection through Use of a Prefrontal Single-Channel Electroencephalogram", Sensors, 2018 Publication	<1 %
50	Monika Prucnal, Adam G. Polak. "Effect of Feature Extraction on Automatic Sleep Stage Classification by Artificial Neural Network", Metrology and Measurement Systems, 2017 Publication	<1 %
51	Oliver Faust, Hajar Razaghi, Ragab Barika, Edward J Ciaccio, U Rajendra Acharya. "A review of automated sleep stage scoring	<1 %

based on physiological signals for the new millennia", Computer Methods and Programs in Biomedicine, 2019
Publication

52	Rajeev Sharma, Ram Bilas Pachori, Abhay Upadhyay. "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals", Neural Computing and Applications, 2017 Publication	<1%
53	Yong Jae Lee, Young Shin Chung, Jung-Yun Lee, Eun Ji Nam, Sang Wun Kim, Sunghoon Kim, Young Tae Kim. "Impact of increased utilization of neoadjuvant chemotherapy on survival in patients with advanced ovarian cancer: experience from a comprehensive cancer center", Journal of Gynecologic Oncology, 2018 Publication	<1%
54	elearning.medistra.ac.id Internet Source	<1%
55	waset.org Internet Source	<1%
56	www.ncbi.nlm.nih.gov Internet Source	<1%
57	Ali Abdollahi Gharbali, Shirin Najdi, José Manuel Fonseca. "Investigating the	<1%

contribution of distance-based features to automatic sleep stage classification", Computers in Biology and Medicine, 2018
Publication

58	Pan Tian, Jie Hu, Jin Qi, Xian Ye, Datian Che, Ying Ding, Yinghong Peng. "A hierarchical classification method for automatic sleep scoring using multiscale entropy features and proportion information of sleep architecture", Biocybernetics and Biomedical Engineering, 2017 Publication	<1 %
59	Emina Alickovic, Abdulhamit Subasi. "Ensemble SVM Method for Automatic Sleep Stage Classification", IEEE Transactions on Instrumentation and Measurement, 2018 Publication	<1 %
60	repo.ur.krakow.pl Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		