Dissertation

Automatic Annotation of Hyperventilation and Sleep Stages in Electroencephalogram Examination

脳波検査における過呼吸および睡眠段階の自動アノテーション

Graduate School of

Natural Science & Technology

Kanazawa University

Division of Electrical Engineering and Computer Science

Student ID No. 1624042015

Name: Mera Kartika Delimayanti

Chief Advisor: Professor Kenji Satou

Date of Submission: January, 2020

Abstract

The brain is the body's control center, which possesses the ability to regulate thinking, memory, voice, and motion, and also control the function of many organs. The presence of a disease is one of the most complicated disorders in humans, and Electroencephalography (EEG) is a popular method used to make diagnosis in hospitals.

In the diagnosis using EEG, a patient is typically attached more than 20 electrodes to the scalp, then recorded the electrical waves in the brain. During a typical test menu around one hour, a supine patient on the bed with EEG and various equipment receives various instructions and stimulations. EEG automatically records the brain waves, however, such events and specific states of brain wave (e.g. sleep and awake) are annotated manually by technicians.

To reduce the workload of human annotators, we tried two things for automating the annotation. Firstly, we proposed a new approach of the non-contact capturing method of breathing activities using the Kinect depth sensor for automatic annotation of hyperventilation, which is one of the important events in EEG diagnosis. The time-series mean depth value between Kinect and subject’s breast are further processed by feature reduction step, then classified by Support Vector Machine (SVM). This approach achieved 99% accuracy in the classification of three breathing states including hyperventilation.

Secondly, we proposed a method of sleep stage classification. Unlike various existing methods using complicated process of signal filtering, feature extraction, and feature selection, we used high-dimensional features calculated by Fast Fourier Transform (FFT) from single- or multi-channel EEG signals. In the classification of the expanded version of Sleep-EDF dataset with 61 recordings, our method using SVM achieved better or nearly equal performance in comparison with the most recently reported and state of the art method. It means that that our method is useful for the automatic sleep stage annotation in EEG diagnosis.

Keywords: annotation, hyperventilation, sleep stages, classification, fast fourier transform, electroencephalogram

Acknowledgments

Thanks to the God, Allah SWT, Finally, I could finish for writing this thesis. I also would like to express my sincere gratitude to my professor, Professor Kenji Satou, for introducing and assisting me to the world of bioinformatics research. It was only because of his guidance, enthusiasm, and support that I could complete my research.

I am very thankful to all the committee members, Professor Kenji Satou, Professor Takeshi Fukuma, Associate Professor Yoichi Yamada, Associate Professor Hidetaka Nambo, and Associate Professor Makiko Kakikawa, for reading my thesis and giving me valuable remarks and suggestions.

I also would like to express my gratitude to 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) for sponsoring my scholarship. Because of this scholarship, I was able to achieve my dream of studying as a Ph.D. student.

In addition, a thank you to my institution, Politeknik Negeri Jakarta, for giving me permission to leave for pursuing my Doctoral degree.

I would like to thank Kanazawa University for providing me the opportunity to become a Ph.D. student here. My deepest gratitude to all the staff of Kanazawa University who have helped me in my academic life here in Kanazawa.

I would also like to thank my friends in the Bioinformatics Laboratory of Kanazawa University, for all the wonderful memories of working together and all the support. I would like to give my gratitude to my lab mates, Bedy Purnama, Kunti Robiatul Mahmudah and Fatma Indriani, and my senpai, Favorisen Lumbanraja, Nguyen Ngoc Giang, Phan Dau, M. Reza Faisal and Bahriddin Abapihi. Moreover, also thankful for all my friends here in Kanazawa, especially BUDI-LN 2016.

Last but not least, I would like to thank my family, My beloved husband, Artanto Rizky Cahyono, My twin children, Farhan and Farah and my deceased Mom, Dad, Sisters, and Brothers. I am grateful for their love, prayers, understanding, and encouragement.

Thank You!

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background

The brain is the body's control center. It regulates thinking, memory, voice, and motion. It controls the function of many organs. If the brain is healthy, it will function quickly and automatically. If the brain has problems, the effect can be catastrophic. Disease in the brain is one of the most complicated disorders in human. Electroencephalography (EEG) is a popular method to make a diagnosis of brain disease and used in hospitals. An EEG can determine brain activity changes that could be helpful to diagnose brain disorders. An EEG can also be useful for the diagnosis or treatment of sleep disorders, heart stroke, inflammation of the brain or encephalitis, brain damage that may be caused by a variety of causes (encephalopathy), tumor in brain, and brain damage from head injury.

One of the basic physiological needs and an important part of life is an activity called sleeping. One-third of human lifetimes are spent on sleeping. Lack of sleeping may cause health problems, as well as influence mood and cognitive performance[1]. EEG records the activity of the brain over a period of time. The principle of EEG recording was implemented by comparing the electrical voltage or electrical signal from the multiple scalps located in the different areas of the brain. Then, with a neuron that produces a neuronal activity, the potential of extracellular space increases as extracellular currents originates from post-synaptic potentials and inhibitory post-synaptic potentials. The amplitude is higher than usual when more neurons operate actively simultaneously, due to the EEG measurement. The electrodes are usually connected to the skin, so neuronal feedback is very small since the tissues significantly attenuate the signal [2,3]. This occurrence appears in a EEG recording as wavy lines.

EEG waveforms have several kinds of rhythms. These rhythms are remarkably useful for annotation of sleeping score from PSG data. In normal EEG, we differentiate these into five frequency bands. Table 1 shows the frequency and amplitude ranges of the EEG signal[2,4].

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 repeated stages, and the sleep stages are essential sections of activity during sleep. The three main stages of sleep are Awake, Non-REM (NREM) sleep, and Rapid Eye Movement (REM) sleep. During the NREM phase, called dreamless sleep, the breathing is slow, the heart rate is in a natural state, and the blood pressure is average. The NREM sleep will be increasing depth leading to REM sleep. Meanwhile, The REM stage most occurs when dreaming. At the time, the brain briefly controls the arms and legs to prevent the body from playing out these dreams. During the REM sleep, the eyes move quickly in different directions. 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[5].

Usually, technicians annotated the sleep stages detection manually. Hence, it was a time-consuming process. Moreover, it was expensive and dependent on human resources. Because of the time-consuming, expensive, and enormous process, it is not suitable to manually annotate large EEG-data sets for sleep stage screening by the human expert[6]. As a result, it is highly needed to develop automatic sleep stage classification in order to achieve better accuracy.

In addition to the usage of the PSG/EEG system, the patient is also equipped with the surveillance through a camera for monitoring system. EEG records the brain waves and a camera records a video of the patient to catch patients whose condition is deteriorating before their symptoms are obvious. Usually, during a typical test menu around one hour, a supine patient on the bed with EEG and various equipment receives various instructions and stimulations. The instructions are such as to open and close the eyes, breathe deeply and rapidly (hyperventilation) and the stimulations are to look at a flashing light or to hear the loud sound. In fact, such events and specific states of brain wave (e.g. drowsy) are annotated manually by human. To reduce the workload of human annotators, we tried to develop a system for automating the annotation.

The hyperventilation constitutes a condition which starts to breathe very fast. It usually provokes physiological slowing of the brain rhythms and constructs a classic activation procedure of the EEG system. Hyperventilation is an important event in clinical electroencephalography since it sometimes causes epileptic seizures. It is, however, more prevalent and pronounced in patients with epilepsy [7].

During the EEG examination, start time and end time of indicated hyperventilation must be annotated. On the other hand, events about awake and sleep are also important. In addition to recognize patient's awake and sleep, it is also required to recognize stages of sleep (for example, non-REM and REM) since the annotation about these stages could be informative in the diagnosis of various sleep disorders like insomnia, hypersomnia, narcolepsy, and sleep apnea syndrome. Since such events in EEG examination are still annotated manually by experts, it is desired to develop an automatic annotation function in EEG system. To enable the automatic annotation of hyperventilation, we used an RGB-D camera called Kinect. In addition to usual RGB images, it can take infrared (IR) and depth images.

1.2 Objective

In the condition of monitoring patients in the hospital, such events and specific states of brain wave (e.g. drowsy) and sleep stages are annotated manually by the clinician. Moreover, the hyperventilation status of the patient is a critical condition that should be known immediately by the clinician. However, the marking of hyperventilation occurrence and sleep stages condition is commonly annotated manually by the clinician. Since the yielded data from both of the EEG signals and RGB-D camera monitoring for hyperventilation are time-series signals, it will be processed using the machine learning algorithm for classification to get better performance. So, the primary objective of this research is to develop a system for automating the annotation of sleep stages and hyperventilation in Electroencephalogram Examination by implementing machine learning on the classification system.

1.3 Contribution

Machine learning system is one of the most common tools that is usually used in automatic annotation by doing the classification to get the better performance. Many studies of automatic annotation on EEG examination by doing classification system have been explored intensively by researches. This research contributes to the following matters:

1. Propose a novel approach by utilizing the feature extraction algorithm in machine learning method.

Existing research use various classification algorithms to process the time-series signal from the brainwaves and the time-series signal of mean depth value of ROI that yielded from the RGB-D camera. In this research, we found an effective approach to utilize fully thousands of Fast Fourier Transform (FFT) features, or also called high-dimensional FFT features, from the feature extraction algorithm.

2. Improve classification performance.

   We found that our approach worked to achieve better result or nearly equal performance of the classification from the breathing activities on hyperventilation and the classification of sleep stages from the EEG signals using the datasets. Moreover, we achieved significant improvement for all cases with the dataset we used. This result has shown that better performance had been achieved with the proposed method in classification to get the automatic annotation in the EEG examination.

1.4 Thesis Organization

This thesis consists of four chapters.

Chapter 1 Introduces the background and the reasons for conducting the research. This chapter also contains objectives and contribution of this research for bioinformatics application.

Chapter 2 explains that the research had been conducted on the case of clustering and classification of breathing activities by depth image for automatic annotation of hyperventilation. The depth image was yielded by calculating the mean depth value of the thorax area. Some steps on machine learning had been selected and executed to get better performance. We will also explain the classification performance evaluation.

Chapter 3 describes the classification of brainwaves for sleep stages automatic annotation by high-dimensional FFT features from EEG Signals. In this chapter, we explain about three main steps in our experiment. The first step is Brainwaves acquisition from the EEG channel by obtaining the time-series signal of EEG signals. The process continues with the preprocessing by doing segmented Brainwaves Epochs. The next step, was continued with feature processing by doing feature extraction using FFT, and it will result in the high dimensional FFT features. Finally, we performed the classification evaluation by doing 10-fold cross-validation and SVM classifier to get classification performance evaluation.

Chapter 4 summarizes the thesis by stating a conclusion of achievements. Suggestions for future work are discussed in this chapter.

Chapter 2 Clustering and Classification of Breathing Activities by Depth Image for Automatic Annotation of Hyperventilation

2.1 Introduction

Breathing is a vital physiological task in living organisms including human, and one of critical indicators of a person's health. There are two methods to monitor breathing rate activities, contact or non-contact method. For the contact method, frequently called invasive method, a sensing device or some parts of this is put on to the subject's organs. For non-contact method, called noninvasive method, there is no direct interaction between the instruments with the subject [8].

Many medical instruments can be categorized as invasive approaches such as Respiratory Inductance Plethysmography (RIP)[9], Thoracic Impedance [10], Impedance Pneumography (IP) [11], Photo plethysmography (PPG) [12], Acoustic Monitoring [13,14], Strain Gauges [15] and Magnetometers [16]. All of the methods are implemented to monitor human breathing activities. Those are state-of-the-art devices especially for breathing activities through direct contact. However, these methods' primary drawback is that they interfere with the natural respiration of the subject.

The microwave-based techniques had been developed for some non-contact respiratory measurements [17,18]. Moreover, the optical-based techniques are refined too includes Structured Light Plethysmography (SLP) [19] and Optoelectronic Plethysmography (OEP) [20,21]. Despite the fact that there is no need to directly contact with the subject while measuring, these instruments tend to have the complicated procedure.

Normal breathing occurs when the balance of breathing in oxygen and breathing out carbon dioxide. The hyperventilation exists if the exhaling more than inhaling; in fact, this causes a rapid reduction in carbon dioxide in the body. A hyperventilation incident is a critical event in clinical electroencephalography because it induces epileptic seizures. The noninvasive method based approach can be implemented to monitor this event by using camera monitoring. In this research, we proposed a method to measure the morphological changes of the subject's chest area in real-time using RGB-D camera namely Kinect V2, which is a commercial depth camera in order to monitor breathing activities. Therefore, we can estimate the activities of the subject based on the monitoring of the subject's breathing without contact directly to the subject.

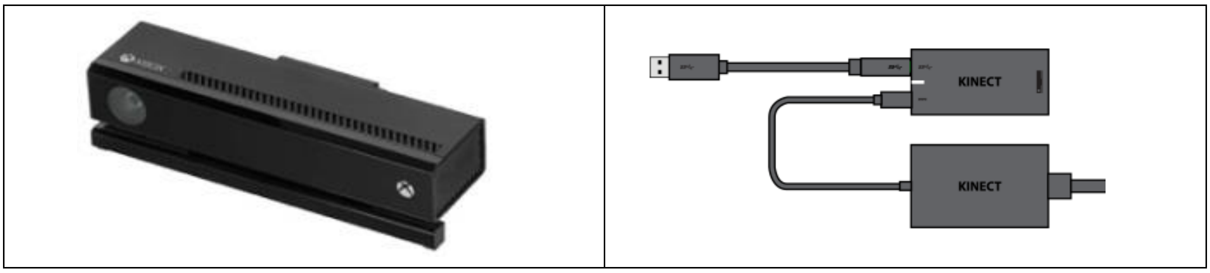

Figure 1. Kinect V2[22]

Microsoft had released two series of Kinect; they were Kinect version 1 (v1) and Kinect version 2 (2) [22]. Kinect v2 applied an active sensor called Time-of-Flight method to measure the distance of a surface by calculating the round-trip time of a pulse of light [23]. In other hand, Kinect v1 does not have this ability to do that. As a result, the depth images resulted from Kinect v2 have better quality compared the other one. Figure 1 shows Kinect v2. The mean depth value from Kinect was reconstructed in time series signal. This signal is used to classify breathing activities by building a machine learning algorithm using SVM. The block diagram of this study is shown in Figure 2. The data input was taken with the RGB-D camera called Kinect Version 2 from Microsoft.

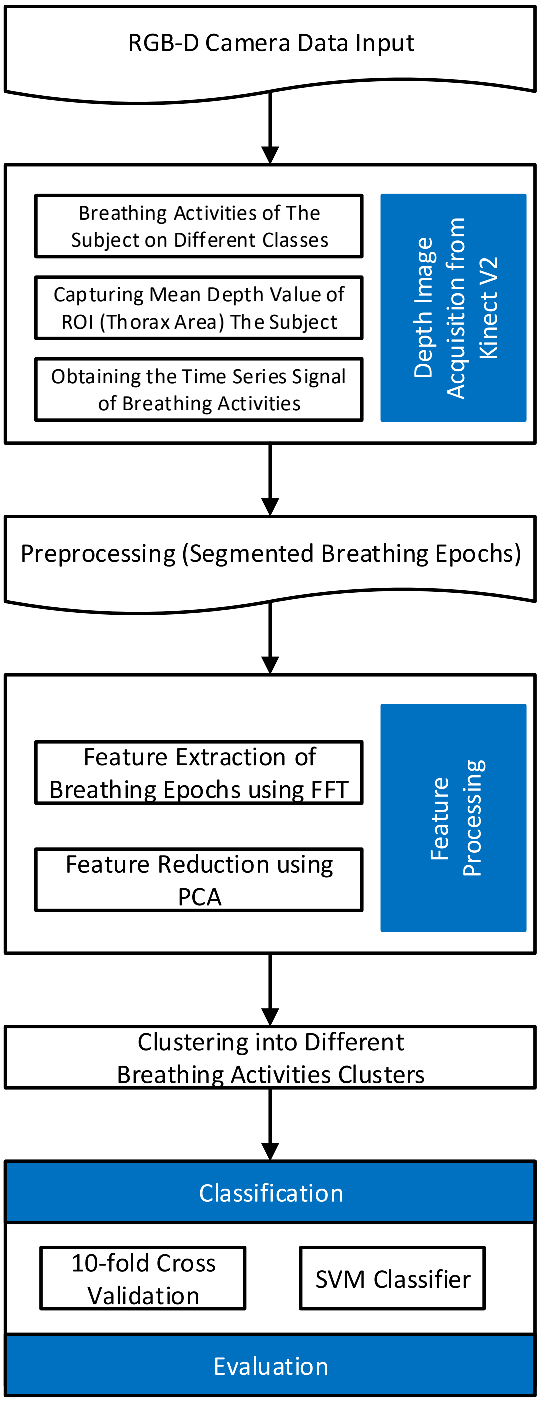

Figure 2. Block diagram of this study

2.2 Materials and Methods

This research had many steps as mentioned previous and conducted on four different subjects of the human. After acquiring the data using Kinect, the RGB-D camera, then the time series waveforms were processed into the features. The feature processing, clustering, classification, and evaluation steps were conducted on R programming.

2.2.1 Acquisition of Depth Image and Capturing Mean Depth Value

In this step, Kinect was utilized to capture the human breathing activities. Kinect can capture depth images at the resolution of 640 x 480 pixels a maximum of 30 fps using IR Receiver. Furthermore, it also can capture color images using an infrared laser emitter combined with a monochrome sensor. The experiment was conducted indoor, and the subjects were asked to sit at a distance of the depth camera (Figure 3). In this research, we recorded four samples three times for different breathing activities in front of MS Kinect. The three different stages of activities can be separated as follows:

- First 60 seconds, deep and fast breathing
- Second 60 seconds, aloud reading the article in the newspaper
- Last 60 seconds, relaxing by listening deep meditation music.

All the depth image data were captured and then continued to calculate the mean depth value from the Kinect to the subject, especially at the Region of Interest (ROI) on the thorax area. The mean depth value was calculated from the perpendicular distance between the subject to the receiver in Kinect. The movement of the breathing activities on the thorax area leads to the changing of mean depth value every time. The calculation can be depicted into time series waveforms or time series signal as pointed out in Figure 4.

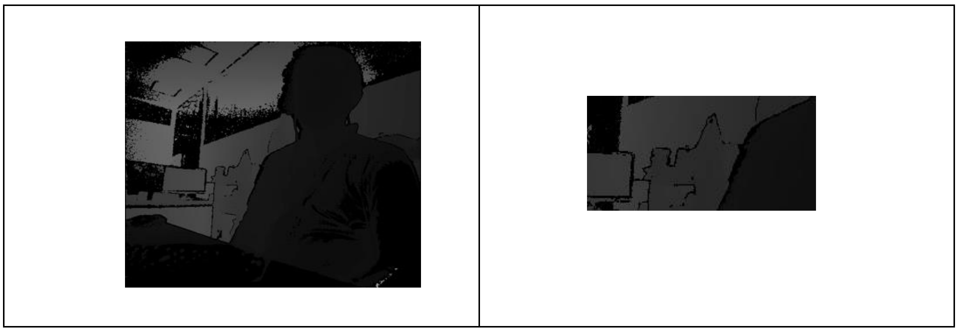

Figure 3. Kinect Depth Image (left) and ROI of Depth Image (right)

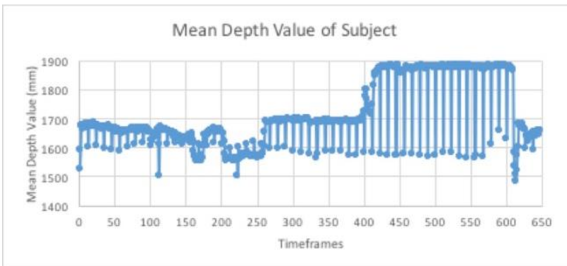

Figure 4. Calculation result of subject 1's mean depth value of ROI

2.2.2 *Feature Extraction and Feature Reduction*

Fast Fourier Transform (FFT) is one of the recognized and useful tools for signal processing in time-series signal. To decompose signals into segmented breathing sequences, we used the equal time intervals called epoch. For calculating, the length of each epoch was set to every 30 timeframes. The epochs were then processed using frequency analysis in which frequency spectra were generated using FFT. The obtained feature extraction using the FFT process was presented in Figure 5. We have carried out FFT to transform the signal from its original time domain to a representation of frequency domain and vice versa [24]. The FFT analysis had been completed for four samples from four subjects, and the process continues to extract all of the principal components (PCs) as the features from the spectra through Principal Component Analysis [25]. PCA is a technique for dimensional reduction that widely used in the time series signal analysis. To measure breathing activities rate and to process feature extraction and feature reduction for each subject, we did separately for each subject in order to validate the clustering and classification algorithm. Figure 6 shows the PCA result from one of the subject.

	comp1	comp2	comp3	comp4	comp5	comp6	comp7	comp8	comp9	comp
e1	1581.665	3.547987	7.110261	13.03341	2.845914	8.988022	11.16255	1.991704	11.35828	8.81
e2	1582.193	3.234049	6.706951	12.68202	2.579994	8.460107	11.29009	1.847521	11.06371	9.25
e3	1588.253	4.444163	6.867366	6.972603	3.699705	7.407252	7.275467	4.213264	7.086674	6.35
e4	1588.121	4.34839	6.765107	7.10337	3.633583	7.282548	7.407025	4.19591	6.963945	6.47
e5	1586.671	3.729344	5.389082	8.226908	4.294416	6.216633	7.915616	5.607716	6.635812	5.76
e6	1586.383	3.720783	5.101115	8.326361	4.5804	6.298001	7.71762	5.71341	6.923347	5.54
e7	1586.215	3.759494	4.951097	8.288269	4.710722	6.460267	7.563084	5.571559	6.92846	5.68
e8	1585.71	4.008624	4.641862	7.923404	4.784439	6.831955	7.601447	5.252572	6.423081	5.87
e9	1585.578	4.102015	4.616052	7.793461	4.709822	6.811942	7.729617	5.34558	6.432263	5.7
e10	1585.798	3.918743	4.572504	7.99422	4.92675	7.013843	7.60829	5.170543	6.2142	5.73
e11	1585.774	3.940987	4.585793	7.981696	4.907524	6.994098	7.593829	5.158776	6.209789	5.76

Figure 5. Feature extraction result by using FFT

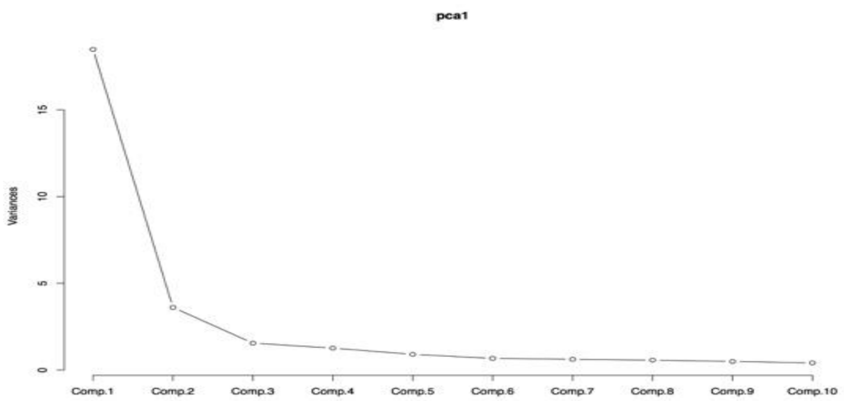

Figure 6. Example of PCA value one of the subject

2.2.3 Clustering The Mean Depth Value of the Subject

After feature reduction, then we apply non-parametric density-based clustering to the features to detect clusters. This step was executed in order to validate the annotation of the class label or labeling events on specific breathing activities for every timeframe [26]. We chose the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for the clustering algorithm. Figure 7, Figure 8, Figure 9, and Figure 10 exhibit the clustering result in two dimensional from the subjects. It can be seen that the clustering process had yielded the separated three of the cluster area, which presented the three different breathing activities.

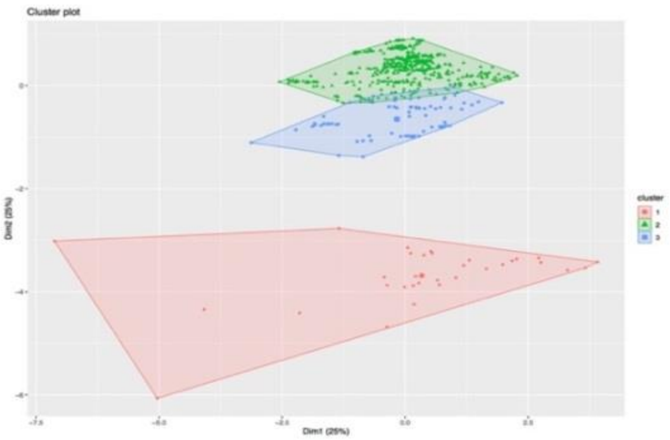

Figure 7. Clustering result of the subject 1

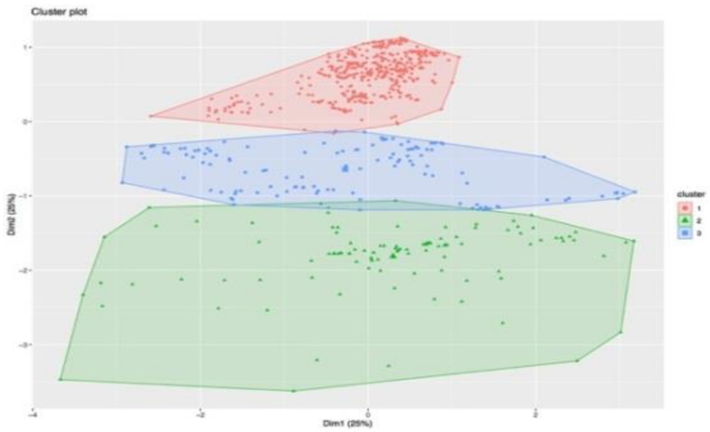

Figure 8. Clustering result of the subject 2

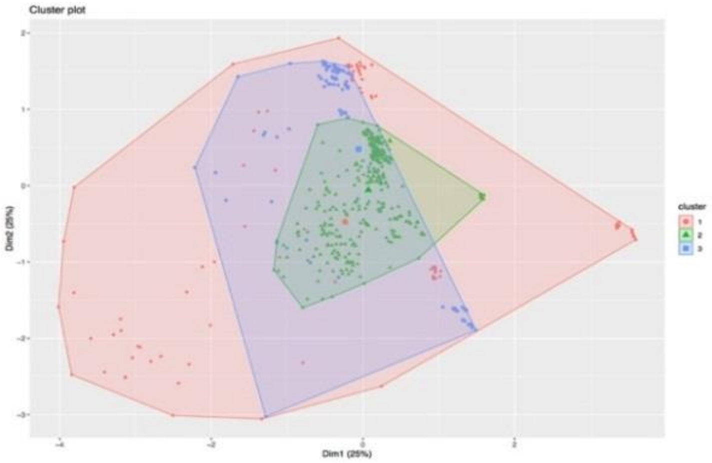

Figure 9. Clustering result of the subject 3

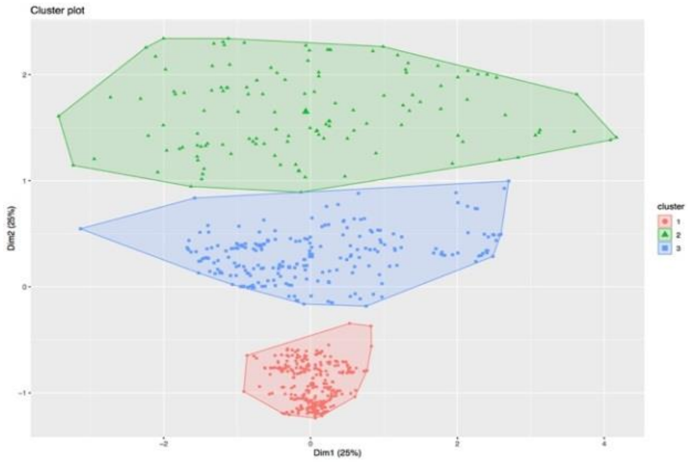

Figure 10. Clustering result of the subject 4

2.2.4 Classification

In addition, the data from the subjects were used to be the dataset. All of the data have been divided into a training set and a test set throughout the experiments. The classification step was completed with the 10-fold cross-validation or jackknife test, each process in this step are repeated 10 times. We trained each fold in order to have a better estimation of the true error rate of each set. The 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. The model is created by using the training data, and the test data is used to evaluate the performance of classification.

The classification was performed using Support Vector Machine (SVM), a supervised machine learning method with having good accuracy as well as being used for Protein Sequence Classification [27] and this algorithm can be used to solve classification or regression problem. In this part, an overview of the method used in this research was explained.

The SVM is a classifier which separates the data in a different class on a maximal-margin hyperplane. A hyperplane is a line that splits the input variable space. SVM changes the information into a higher dimensional space with the goal that the nonlinear separable problem in the first example space can be changed to a linear separable problem. SVM algorithm implements an implicit mapping Φ of the input data into a high dimensional feature space as a kernel function turning the inner product of $\Phi(x), \Phi(x^i)$ between the images of two data points x, x^i in the feature space. The data points only appear inside dot products with other points and the process took place in the feature space and called The "kernel trick" which introduced by Scholkopf and Smola[28]. More precisely, if a projection $\Phi : X \rightarrow H$ is used, the dot product $\Phi(x), \Phi(x^i)$ can be represented by a kernel function k.

$$k(x,x^i) = [\Phi(x), \Phi(x^i)] \qquad (1)$$

which is computationally simpler than explicitly projecting x and xi into the feature space H[29]. The advantages of SVM implementation are:

1. The efficient classifier in high-dimensional spaces.
   It is especially applicable to text or time-series signal classification problems where the dataset can have a large number of features.
2. The efficient of memory used.
   The current process of appointing new members to a class uses only a subset of the training data. Consequently, only this subset needs to be stored in the memory when making classification decisions.
3. Versatile.
   The process of class separation is often non-linear. Therefore, the ability to implement different kernels allows the flexibility for decision boundaries in order to get the better performance.

The caret package was used to execute this algorithm on R programming. SVM function has a model using linear kernel and non-linear kernel like radial basis function; those were realized to get which one better accuracy in this research. Customization on SVM function by selecting C value (cost) in linear classifier by inputting values in grid search. This step will increase the accuracy result. The results from classifying the breathing activities data captured by MS Kinect v2 for four subjects with 10-fold cross-validation and SVM classifiers are presented in Figure 11 and Figure 12. In this dissertation report, we try to use all the PCs and without using the PCs for feature reduction in order to get the highest accuracy of the classification. The classifier was used with SVM linear and radial basis function with grid or no grid in R programming. Based on the results, SVM radial with grid basis function seemed to be a good choice of classifier among SVM function. Moreover, non-parametric density can be implemented to execute clustering of the breathing activities by using depth image from Kinect v2.

2.2.5 Classification Performance Evaluation

A. Confusion Matrix

The confusion matrix is a table that is used to determine the performance of a classifier[30]. This matrix has four combinations of prediction result as shown in table 2. True positive (TP) and True Negative (TN) occur when the result of the prediction is the same as the outcome of the real observation. False Positive (FP) and False Negative (FN) occur when the result of the prediction is different from the outcome of the real observation.

Table 2. Confusion Matrix

		Predicted Condition	
		Positive	**Negative**
True Condition	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

B. Accuracy

Accuracy is a measurement to calculate the proportion of the number of times the classification predicted the result correctly[30]. The formula to calculate accuracy is shown in below formula.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

C. Sensitivity

Sensitivity is used to measure the proportion of the actual positive result which is classified correctly[30]. The formula to calculate sensitivity is shown in below formula.

$$sensitiviy = \frac{TP}{TP + FN} \tag{2}$$

D. Specificity

Specificity is a used to calculate the classification performance of predicting negative results correctly[30]. The formula to calculate specificity is shown in below formula.

$$specificity = \frac{TN}{TN + FP} \tag{3}$$

This research calculated the classification performance evaluation using the accuracy. From Figure 11 and Figure 12, we had seen that the accuracy got better value when we did not use PCA for feature reduction. The performance reached over 95 % for all subjects using SVM radial with the grid as the classifier. For example, on subject 4, when all components on PCA was used for feature reduction, the accuracy reached 98.80%. Hence, PCA was not used, the accuracy up to 99.5%.

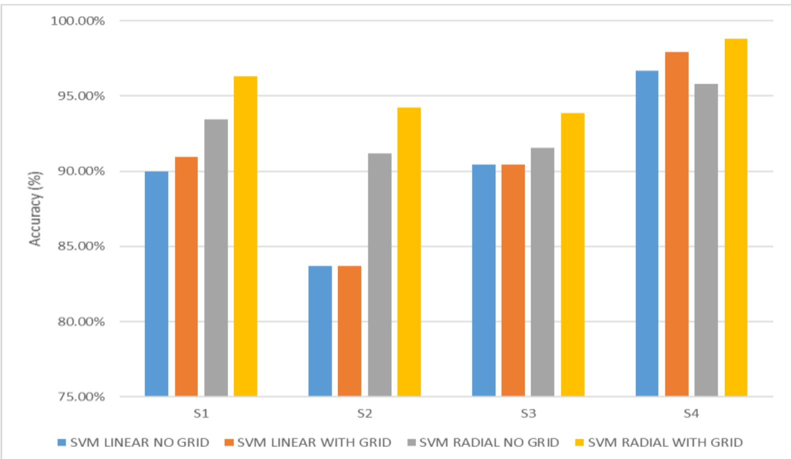

Figure 11. Performance comparison of four subjects using 10-fold cross-validation and SVM classifier with all components of PCA

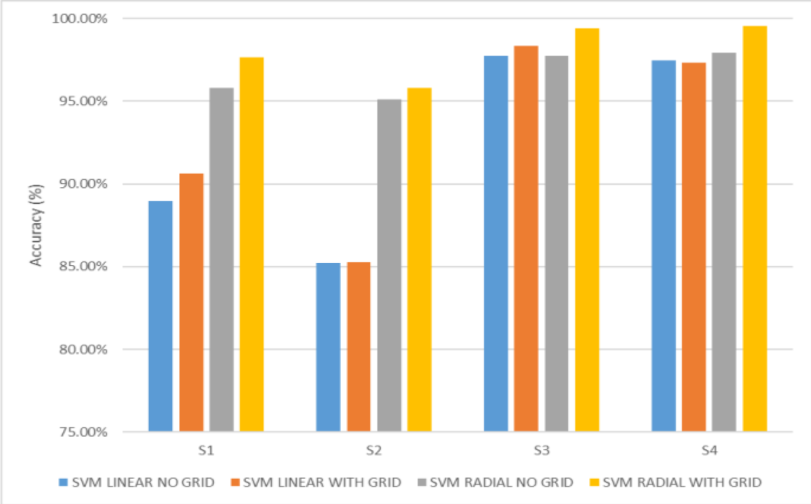

Figure 12. Performance comparison of four subjects using 10-fold cross-validation and SVM classifier without using components from PCA

2.3 Discussion and Conclusion

This paper has presented the method of capturing on breathing activities data from image depth of Kinect v2. This method is the noninvasive mechanism to estimate the activities of the subject from breathing activities monitoring. Those data were used to calculate the mean depth value on thorax area and were displayed on time series signal. FFT had been applied to do the feature extraction from time series into numeric values. PCA is optionally used for feature reduction on this classification, but the result exposed that the highest accuracy was achieved without using PCA components. As a result, we have seen that, feature reduction using PCA is not effective on time series signal in our study. Besides, the process had carried out the clustering using non-parametric density estimation, and the supervised machine learning, classification, the algorithm had been implemented by doing 10-fold cross-validation and using SVM classifier for all four subjects. It has been shown the SVM radial with the grid is the most efficient classifier with the highest accuracy for all the subjects over 99%. The result obtained is promising to predict activities from breathing. However, further work is required, especially for feature selection in order to get better classification results for a larger dataset.

Chapter 3 Classification on Brainwaves for Sleep Stages for Automatic Annotation by High-Dimensional FFT Features from EEG Signals

3.1 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[1,31]. 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)[32]. 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[2,4].

EEG records the electrical activity of the brain over a period of time. The EEG record is derived from multiple scalp electrodes to compare differences in electrical potential in different areas of the brain. The electrical potential in extracellular space increases in conjunction with neuronal activity, likely as a result of the extracellular currents originating from post-synaptic potentials and inhibitory post-synaptic potentials. The amplitude of the tracing that is captured on EEG is greater whenever individual neurons are active co-locally and contemporaneously. Electrodes are usually attached to the scalp and as a result, the contribution of a single neuron to the amplitude of the recorded tracing is quite small as the skull and scalp tissue attenuate the signal significantly[2,3].

Figure 13 represents the standard system used for measuring the EEG signal, termed the 10-20 system, in which the minimum number of electrodes used is 21. This method regulated 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 ears and nose 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), and O (occipital), Fp (Frontal pole). A subscript z is used to mark the midline electrodes as zero.

The electric signal in the brain is determined by measuring the difference of the electric activity between two electrodes over period of time. As it propagates, the signal gradually decays with distance from the source. Eventually, the signal has decayed to the point where gets precise measurement only from one of the parallel combinations of electrodes[2,3].

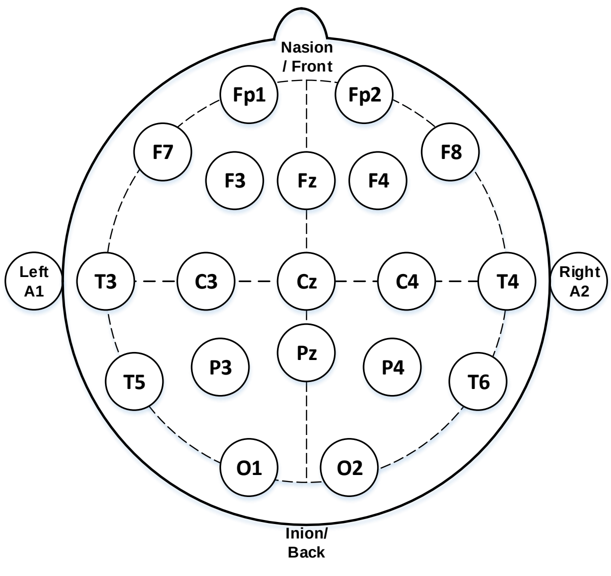

Figure 13. Electrodes placement of EEG measurement [2]

During PSG measurement, EEG electrodes can be affixed either to the scalp into the skin or within the skull (intracranially). Currently there are several methods to affix the scalp electrodes efficiently. The electrodes are typically made from silver chloride or gold, and are cup-shaped, designed to hold conducting paste. For this reason, the electrodes were occasionally glued in place to provide better result during recordings that lasted more than 24 hours[2,3].

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 3 lists the frequency and amplitude ranges of these bands[2,33].

Table 3. 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. The 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 the 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[5].

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[34]. The AASM criteria are a modified version of the R&K criteria. Some differences between the AASM and R&K criteria are as follows[34,35]:

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 [36]. 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[37].

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 of the consuming-time process, expensive, and enormous process, it is not suitable to hold the large EEG datasets for sleep stage annotation by the human expert[6]. 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 stage 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 [38]. 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[39]. 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[40]. Lajnef et al. incorporated a large number of EMG, EOG, and EEG signal features into their analysis using a multi-class support vector machine for computer-assisted sleep scoring[41]. Other recent works have utilized multivariate linear regression[42], linked component analysis[43], sparse Bayesian learning[44], Bayesian machine learning approaches[45,46], and probabilistic common spatial patterns on multichannel EEG [47] to perform feature extraction and classification of EEG signals. In addition, a study of multi-class sleep stage analysis has been performed using the Bayesian neural network classifier model, achieving an accuracy of greater than 88%[48]. 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 second epochs for its analysis. Their model's performance had achieved the highest accuracy in comparison to previous results[49]. 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[50].

In this report, we present a system of sleep stage classification based on EEG signals. Instead of using complicated processes of signal filtering, feature extraction, and feature selection, and we used high-dimensional features calculated by Fast Fourier Transform (FFT) from single- or multi-channel EEG signals. FFT is one of the traditional, 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 3 were extracted and used. However, a sampling window of 30 seconds at 100 Hz sampling frequency allows for an extraction of at most 3,000 features in the range of 0-100 Hz. In this study, we demonstrate that by incorporating high-dimensional FFT features by utilizing thousands of features into the analysis, it is possible to outperform state-of-the-art algorithms for the Sleep-EDF database. In addition, we report the relationship between the number of FFT features incorporated into the analysis and overall performance of our model. We applied this method to the DREAMS database. Our proposed approach consists of the main three steps. That are 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 14.

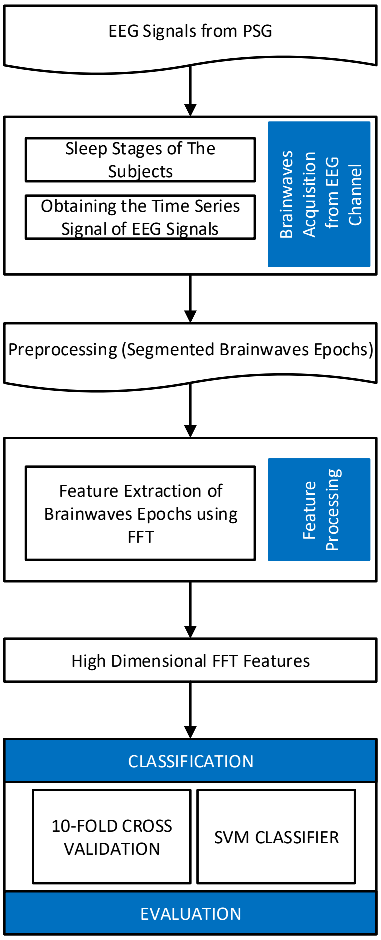

Figure 14. The flowchart of the proposed research

3.2 Materials and Methods

3.2.1 Experimental Data

A. Dreams Dataset

The first dataset used in this study is the DREAMS subject database. This dataset comprises 20 healthy subjects, including 16 females and four males, aged 20 – 65 years. The subjects who participated in this study were not taking any medication at the time. All twenty recordings were conducted in the sleep lab of Andr Vsale hospital (Montigny-le-Tilleul, Belgium), and spanned 7 to 9 hours each. The EEG data was collected using a 32-channel digital polygraph (BrainnetTM System, MEDATEC, Brussels, Belgium). In addition, the DREAMS subject database includes PSG data, comprising three channels of EEG data (C3-A1 or CZ-A1, O1-A1, and FP1-A1), one channel of EMG data, and two channels of EOG data. The data had been recorded in European Data Format (EDF) at a sampling frequency of 200Hz.

These recordings come from people who are free of drugs and participants in other research projects that performed in the sleep laboratory and was chosen with their consistency.

The sleep records were scored using both AASM and R&K annotation criteria by an expert from the sleep laboratory. Notably, since the R&K criteria designate the sleep stages as AWA, REM, S1-S4, 'sleep stage movement', and 'unknown sleep stage', the unknown sleep stages and movement times have been excluded from this analysis. Instead, the analysis has been conducted on the five states corresponding to AWA, N1, N2, N3, and REM of the AASM criteria[51]. The classification of sleep stage according to the R&K criteria is based on epochs of 20 seconds, whereas scoring according to the AASM criteria has been completed (nearly two years later) based on epochs of 30 seconds. A single expert of the sleep laboratory had annotated by visually.

Hassan et al. and Seifpour et al. had conducted the studies using the Dreams dataset with R&K criteria[52-54]. Since we did not use the AASM criteria, we have chosen to quantify annotate the samples using only the R&K criteria (4). The total number of the samples is 20,143 and it is close to the number of samples 20,257 used in the other study. Table 4 presents the characteristics of the subjects in Dreams Database.

Table 4. The characteristic of the subjects on Dreams Dataset

Name	Sampling Frequency	Age	Sex	Recording duration (hh.mm.ss)
Subject1	200 Hz	23	Woman	08:00:40
Subject2	200 Hz	47	Woman	08:12:30
Subject3	200 Hz	24	Woman	08:24:20
Subject4	200 Hz	48	Woman	08:46:30
Subject5	200 Hz	46	Woman	08:51:30
Subject6	200 Hz	65	Woman	08:18:40
Subject7	200 Hz	45	Woman	08:26:00
Subject8	200 Hz	22	Woman	08:05:00
Subject9	200 Hz	21	Woman	09:18:40
Subject10	200 Hz	20	Woman	08:36:20
Subject11	200 Hz	30	Woman	08:24:10
Subject12	200 Hz	54	Woman	08:00:40
Subject13	200 Hz	23	Woman	09:15:40

Subject14	200 Hz	57	Woman	08:22:10
Subject15	200 Hz	20	Woman	07:00:00
Subject16	200 Hz	27	Woman	08:03:50
Subject17	200 Hz	23	Man	08:18:30
Subject18	200 Hz	27	Man	08:30:40
Subject19	200 Hz	27	Man	08:36:50
Subject20	200 Hz	20	Man	09:32:30

Table 5. The number of samples in DREAMS dataset (R&K criteria)

# of classes	AWA	REM	S1	S2	S3	S4
6	3,733	3,034	1,181	8,823	1,423	1,949
5	3,733	3,034	1,181	8,823	3,372	
4	3,733	3,034	10,004		3,372	
3	3,733	3,034	13,376			
2	3,733	16,410				

We analyzed 42,000 features extracted from seven channels available in the dataset (CZ2_A1, CZ_A1, FP1_A2, FP2_A1, NAF2P_A1, O1_A2, and O2_A1). For each channel, 6,000 features were extracted at a sampling frequency of 200 Hz over a 30 second epoch. These features were examined in combination or individually by channel.

B. Sleep-EDF Dataset

The second dataset is open-source, and many previous researchers have utilized this dataset in sleep scoring research[39,49,50,53,55-57]. 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 21 to 101 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 1989. These subjects were healthy and in ambulatory condition. The second subset with 22 recordings from 22 subjects was EEG data recorded in a study from 1994 and the subjects reported feeling slight difficulty in falling asleep but were otherwise healthy. The EEG data had been collected over 24 hours of the daily lives of the subjects. A miniature telemetry system recorded nocturnal EEG data from the four subjects in a hospital [53]. 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 have been annotated by an experienced sleep technician in 30 seconds basis. Therefore, the duration of each epoch is established as 30 seconds and yielded 3,000 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, and N3 and 'Unknown sleep stage'. The number of samples according to R&K criteria are shown in table 6. After removing 'Movement Time' and 'Unscored', total number of the samples is 127,663. The epoch duration was 30 seconds.

Table 6. 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	5,075	3,773
5	74,827	11,848	4,848	27,292	8,848	
4	74,827	11,848	32,140		8,848	
3	74,827	11,848	40,988			
2	74,827	52,836				

As far as features analyzed, we prepared 6,000 features extracted from two channels (Pz_Oz and Fpz_Cz). For each channel, 1,000 features were extracted at a sampling frequency of 100 Hz and epoch lasted 30 seconds. The features were used in the experiments separately or in combination.

3.2.2 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[58].

The extraction of remarkable 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[59]. 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 seconds 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 and vice versa[24]. The FFT method employs mathematical techniques for EEG data analysis. FFT is applied n each epoch and Figure 15 represents in the form of time-domain signal and frequency-domain signal.

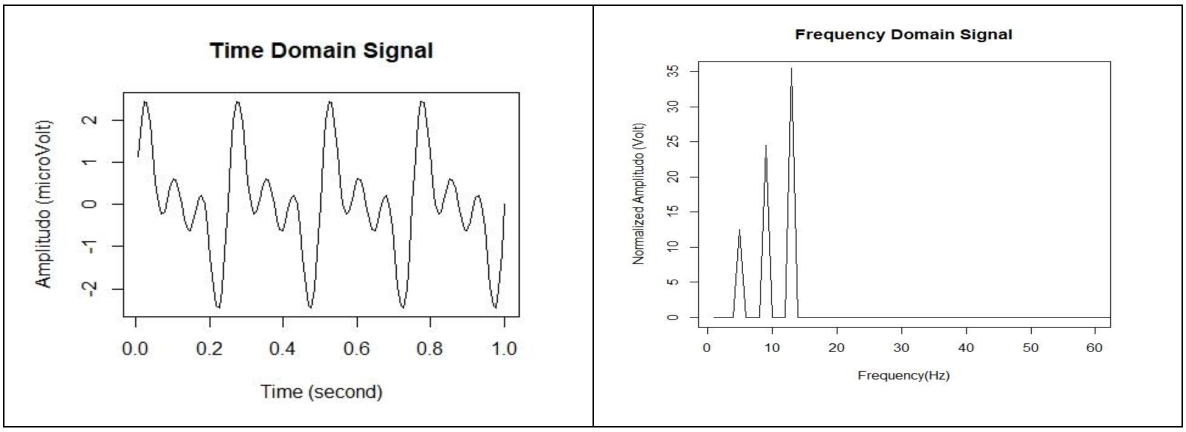

Figure 15. EEG signals in time domain signal and frequency domain signal

Previous studies have shown that FFT is a promising tool for stationary signal processing over virtually all other available methods in real-time applications, and it is more appropriate for sine waveforms such as in EEG signals. enjoys a speed advantage. However, the disadvantage is that it does not have excellent spectral estimation and cannot be employed for analysis of short EEG signals[58].

3.2.3 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[60]. In this study, feature selection was simply conducted by selecting EEG channel(s).

3.2.4 Classification Evaluation

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

From among 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.

3.3 Experimental Results

3.3.1 Classification of DREAMS dataset

First, we applied our method of classification to the DREAMS dataset. Seven EEG channels described in this dataset were first analyzed separately (i.e. the number for feature extracted from one channel was 200 Hz × 30 seconds = 6,000), then analyzed in combination (42,000 features in total). The 6 classes defined by R&K criteria were used. The accuracies calculated through 10-fold cross-validation are shown in table 7. In addition, due to the large amount of data in our study (a maximum of 42,000 features from 20,143 samples), we conducted cross-validation subject by subject, and averaged 20 accuracies. However, we were able to demonstrate that combination of features from seven channels improved the performance ~5 %.

Table 7. Performance comparison on DREAMS dataset (R&K criteria, 6 classes)

Method	Length of epoch (sec)	# of epochs	Accuracy (%)
Hassan et al.2016 [52]	30	20,257	68.74

Hassan et al. 2017 [53]	30	20,257	76.39
Seifpour et al. 2018 [54]	30	20,257	**83.40**
Our method (CZ2_A1)	30	20,143	69.97
Our method (CZ_A1)	30	20,143	70.22
Our method (FP1_A2)	30	20,143	70.88
Our method (FP2_A1)	30	20,143	70.14
Our method (NAF2P_A1)	30	20,143	70.01
Our method (O1_A2)	30	20,143	69.72
Our method (O2_A1)	30	20,143	70.02
Our method (all seven channels)	30	20,143	75.18

3.3.2 Classification of Sleep-EDF dataset

Secondly, the Sleep-EDF dataset was classified. 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, 3,000 features (i.e. 100 Hz × 30 seconds) were extracted. For performance evaluation, 5-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 table 8. In this table, it can be seen that our method with 6,000 features from two channels (Pz_Oz and Fpz_Cz) slightly outperformed Yildirim's method in the classification of 6 and 5 classes and achieved nearly equal performance in 4, 3, and 2 classes.

Table 8. Performance comparison on Sleep-EDF dataset (R&K criteria, 2-6 classes)

Method	Length of epoch (sec)	# of epochs	# of classes	Accuracy (%)
Nakamura et al. 2017 [56]	30	126,699	6	86.60
			5	88.60
			4	91.00
			3	94.50
			2	97.40
Yildirim et al. 2019[57]	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	88.56
			5	89.93
			4	91.04
			3	93.31
			2	97.59

Our method (Fpz_Cz)	30	127,663	6	88.38
			5	86.21
			4	87.85
			3	91.03
			2	96.73
Our method (Pz_Oz and Fpz_Cz)	30	127,663	6	**89.76**
			5	**91.02**
			4	92.17
			3	94.06
			2	97.82

To clarify the relationship between the number of included FFT features and performance, we conducted similar performance evaluations after removing features at fixed intervals. The results are shown in Table 9 and Figure 16, and it was revealed that more than one thousand features can contribute to improve the performance of classification. It also suggests the possibility of further improvement of performance by incorporating more features.

Table 9. Effect of reduced number of features on Sleep-EDF dataset (R&K criteria, 2-6 classes)

# of classes	6,000 features	3,000 features	2,000 features	1,000 features	500 features	250 features
6	89.76	89.76	89.60	89.39	88.87	88.15
5	91.02	91.02	90.87	90.68	90.22	89.53
4	92.17	92.17	91.98	91.81	91.32	90.54
3	94.06	94.06	94.00	93.87	93.61	93.19
2	97.82	97.82	97.80	97.76	97.64	97.48

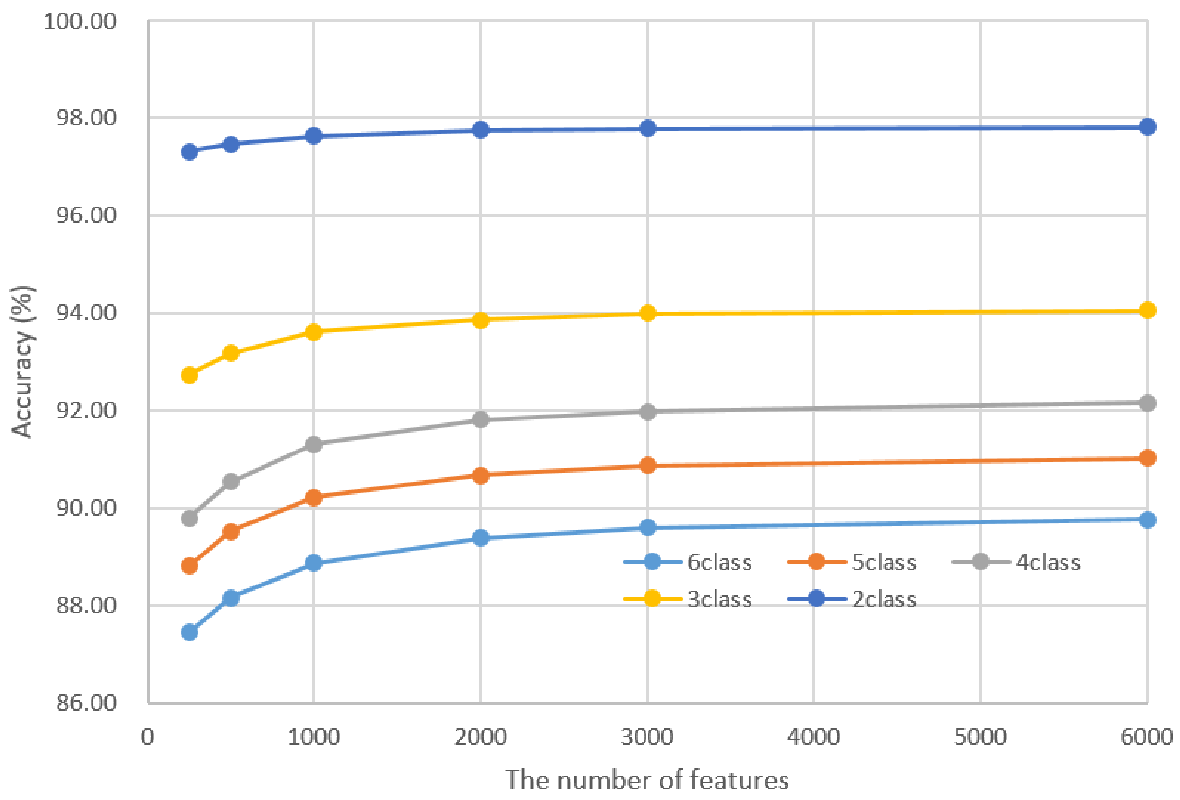

Figure 16. Effect of reduced number of features on Sleep-EDF dataset (R&K criteria, 2-6 classes)

3.3.3 Classification of Sleep-EDF dataset expanded (197 recordings)

Finally, we show the results of applying our method against the latest, extended version of the Sleep-EDF database. 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). Table 10 contains the result of the classification experiment using our method. "SC" and "ST" found in the recording ID prefix stand for "Sleep Cassette" and "Sleep Telemetry", respectively. In this experiment, 10-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 17, 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 appear to be an almost linear relationship (correlation coefficient was 0.9857).

Table 10.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
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		

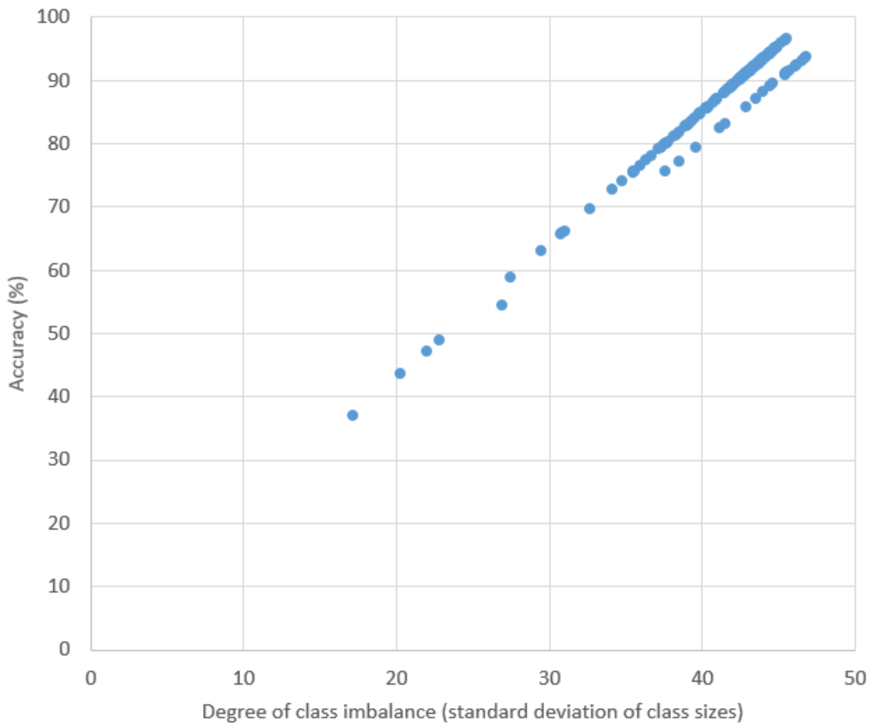

Figure 17. Plot of accuracy and standard deviation of class sizes in each recording

3.4 Discussions and Conclusion

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)
- Removing useless or harmful features by feature selection (e.g. selection by feature importance)

In contrast, we have demonstrated in this paper that fully utilizing thousands of FFT features extracted from single- and multi-channel EEG signals is an effective means of improving the performance of automated sleep stage classification. In our experiment 6- and

5-class classification against the Sleep-EDF dataset, our method outperformed that of Yildirim et al. 2019[57], which is the most recent and advanced methods that has demonstrated the highest-level performance till date.

As our method relies on high-dimensional FFT features, the effect of reducing the number of FFT features was also evaluated. Throughout the experiments, we demonstrated that most high-dimensional FFT features could contribute to the improvement of classification performance. By combining these high-dimensional FFT features with other features studied in previous works, further improvement of performance may be possible.

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. It suggests 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. Selective inclusion and exclusion of the thousands of features that we have collected without decreasing classification performance is our next challenge.

Chapter 4 Summary and Future Work

4.1 Summary

We developed a simple and powerful approach for automatic annotation of hyperventilation and sleep stages in Electroencephalogram examination. There are important keys in our research:

1. We already developed the two methods for automatic annotation of hyperventilation in the case on the classification of breathing activities by depth image from Kinect.
2. Those data were used to calculate the mean depth value on thorax area and were displayed on time series signal and the FFT had been applied to do the feature extraction from time series into numeric values.
3. It has been shown the SVM radial with the grid is the most efficient classifier with the highest accuracy for all the subjects over 99%. So, this obtained result is promising to predict activities from breathing.
4. We have already generated the automatic annotation of brainwaves for sleep stages classification from the EEG signals datasets.
5. We have demonstrated in this paper that fully utilizing thousands of FFT features extracted from single- and multi-channel EEG signals is an effective means of improving the performance of automated sleep stage classification into 2-6 classes.
6. Using the expanded version of Sleep-EDF dataset with 61 recordings, our method achieved better or nearly equal performance in comparison with the most recently reported and state of the art method.
7. Both of the two methods that have been developed have similarities in the types of data carried out in the experiments. Those data are in time-series data that will be executed in the next process using machine learning algorithms.

4.2 Future Work

To improve the performance of hyperventilation and 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)
- Removing useless or harmful features by feature selection (e.g. selection by feature importance)

Bibliography

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, doi:10.1093/sleep/30.9.1213.
2. Keenan, S.A. *Review of Sleep Medicine*; Barkoukis, T.J., Avidan, A.Y., Eds.; BUTTERWORTH HEINEMANN ELSEVIER: 2011; 10.1016/B978-0-7506-7563-5.00038-9pp. 143-167.
3. Kaniusas, E. *Biomedical Signals and Sensors I*; Springer: 2012; 10.1007/978-3-642-24843-6pp. 1-26.
4. Billiard, M.; Bae, C.; Avidan, A. *Sleep Medicine*; Smith, H.R., Comella, C.L., Hogl, B., Eds.; Cambridge: 2012; Vol. 66, pp. 1-46.
5. Thorpy, M.J. *The International Classification Of Sleep Disorders: Diagnostic and Coding Manual. Rev.Ed*; One Westbrook Corporate Center, Suite 920, Westchester, IL 60154-5767, U.S.A: 2001.
6. 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, doi:10.1093/sleep/23.7.1e.
7. Guaranha, M.S.B.; Garzon, E.; Buchpiguel, C.A.; Tazima, S.; Yacubian, E.M.T.; Sakamoto, A.e.C. Hyperventilation Revisited: Physiological Effects and Efficacy on Focal Seizure Activation in the Era of Video-EEG Monitoring. *Epilepsia* **2005**, *46*, 69-75.
8. AL-Khalidi, F.Q.; Saatchi, R.; Burke, D.; Elphick, H.; Tan, S. Respiration rate monitoring methods: a review. *Pediatr Pulmonol* **2011**, *46*, 523-529, doi:10.1002/ppul.21416.
9. Retory, Y.; Niedzialkowski, P.; Picciotto, C.d.; Bonay, M.; Petitjean, M. New Respiratory Inductive Plethysmography (RIP) Method for Evaluating Ventilatory Adaptation during Mild Physical Activities. *PLoS One* **2016**, *11*, e0151983, doi:10.1371/journal.pone.0151983.
10. Houtveen, J.H.; Groo, P.F.C.; Geus, E.J.C.d. Validation of the thoracic impedance derived respiratory signal using multilevel analysis. *International Journal of Psychophysiology* **2006**, *59*, 97-106, doi:10.1016/j.ijpsycho.2005.02.003.
11. Seppa, V.-P.; Viik, J.; Hyttinen, J. Assessment of pulmonary flow using impedance pneumography. *IEEE Trans Biomed Eng* **2010**, *57*, 2277-2285, doi:10.1109/TBME.2010.2051668.
12. Moody, G.B.; Mark, R.G.; Zoccola, A.; Mantero, S. Derivation of Respiratory Signals from Multilead ECGs. *Computers in Cardiology* **1985**, *12*.

13. Corbishley, P.; Rodríguez-Villegas, E. Breathing Detection: Towards a Miniaturized, Wearable, Battery-Operated Monitoring System. *IEEE Trans Biomed Eng* **2008**, *55*, 196-204, doi:10.1109/TBME.2007.910679.
14. Harper, V.P.; Pasterkamp, H.; Kiyokawa, H.; Wodicka, G.R. Modeling and measurement of flow effects on tracheal sounds. *IEEE Trans Biomed Eng* **2003**, *50*, 1-10, doi:10.1109/TBME.2002.807327.
15. GROOTE, A.D.; VERBANDT, Y.; PAIVA, M.; MATHYS, P. Measurement of thoracoabdominal asynchrony: importance of sensor sensitivity to cross section deformations. *Journal of Applied Physiology* **2000**, *88*, 1295-1302.
16. LEVINE, S.; SILAGE, D.; HENSON, D.; WANG, J.; KRIEG, J.; LAMANCA, J.; LEVY, S. Use of a Triaxial Magnetometer for Respiratory Measurements. *Journal of Applied Physiology* **1991**, *70*, 2311-2321.
17. Singh, A.; Lubecke, V.; BoricLubecke, O. Pulse Pressure Monitoring Through Non-Contact Cardiac Motion Detection Using 2.45 GHz Microwave Doppler Radar. In Proceedings of 33rd Annual International Conference of the IEEE EMBS, Boston, Massachusetts, USA; pp. 4336-4339.
18. Dei, D.; Grazzini, G.; Luzi, G.; Pieraccini, M.; Atzeni, C.; Boncinelli, S.; Camiciottoli, G.; Castellani, W.; Marsili, M.; Dico, J.L. Non-Contact Detection of Breathing Using a Microwave Sensor. *Sensors (Basel)* **2009**, *9*, 2574-2585, doi:10.3390/s90402574.
19. AOKI, H.; KOSHIJI, K.; NAKAMURA, H.; TAKEMURA, Y.; NAKAJIMA, M. Study on Respiration Monitoring Method Using Near-infrared Multiple Slit-lights Projection. In Proceedings of INTERNATIONAL SYMPOSIUM, Micro-nanomechatronics and human science, Nagoya.
20. ALIVERTI, A.; DELLACÀ, R.; PELOS, P.; CHIUMELLO, D.; PEDOTTI, A.; GATTINONI, L. Optoelectronic Plethysmography in Intensive Care Patients. *American Journal of Respiratory and Critical Care Medicine* **2000**, *161*, 1546-1552.
21. CALA, S.J.; KENYON, C.M.; FERRIGNO, G.; CARNEVALI, P.; ALIVERTI, A.; PEDOTTI, A.; MACKLEM, P.T.; ROCHESTER, D.F. Chest wall and lung volume estimation by optical reflectance motion analysis. *Journal of Applied Physiology* **1996**, *81*, 2680-2689.
22. Microsoft. Kinect for Windows V2. Availabe online: https://support.xbox.com/ja-JP/xbox-on-windows/accessories/kinect-for-windows-v2-setup (accessed on November, 10).
23. Kolb, A.; Barth, E.; Koch, R.; Larsen, R. Time-of-Flight Sensors in Computer Graphics. In Proceedings of EUROGRAPHICS 2009, Munich, Germany.
24. Nussbaumer, H.J. *Fast Fourier Transform and Convolution Algorithms*; 1981; 10.1007/978-3-662-00551-4.
25. Jolliffe, I.T. *Principal Component Analysis, Second Edition*; Springer: New York, 2002; pp. 518.

26. Azzalini, A.; Torelli, N. Clustering via nonparametric density estimation. *Statistics and Computing* **2007**, *17*, 71-80, doi:10.1007/s11222-006-9010-y.
27. Faisal, M.R.; Abapihi, B.; Nguyen, N.G.; Purnama, B.; Delimayanti, M.K.; Phan, D.; Lumbanraja, F.R.; Kubo, M.; Satou, K. Improving Protein Sequence Classification Performance Using Adjacent and Overlapped Segments on Existing Protein Descriptors. *Journal of Biomedical Science and Engineering* **2018**, *11*, 126-143, doi:10.4236/jbise.2018.116012.
28. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; The MIT Press: Cambridge, Massachusetts, 2002.
29. Karatzoglou, A.; Meyer, D.; Hornik, K. Support Vector Machines in R. *Journal of Statistical Software* **2006**, *15*, 1-28.
30. APIs, T. Classification: True vs. False and Positive vs. Negative. Availabe online: https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative (accessed on September 11, 2019).
31. Walker, M.P.; Stickgold, R. Sleep, Memory, and Plasticity. *Annu Rev Psychol* **2006**, *57*, 139-166, doi:10.1146/annurev.psych.56.091103.070307.
32. Álvarez-Estévez, D.; Moret-Bonillo, V. Identification of Electroencephalographic Arousals in Multichannel Sleep Recordings. *IEEE Trans Biomed Eng* **2011**, *58*, 54-63, doi:10.1109/TBME.2010.2075930.
33. Aboalayon, K.A.I.; Faezipour, M. Multi-Class SVM Based on Sleep Stage Identification Using EEG Signal. In Proceedings of 2014 IEEE Healthcare Innovation Conference (HIC); pp. 181-184.
34. Rechtschaffen, A.; Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*; Bethesda, Md. : U.S. National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968: 1968.
35. 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* **2007**.
36. Hobson, J.A. Sleep is of The Brain, by The Brain and For The Brain. *Nature* **2005**, *437*, 1254-1256, doi:10.1038/nature04283.
37. Marshall, L.; Helgadottir, H.; Molle, M.; Born, J. Boosting Slow Oscillations During Sleep Potentiates Memory. *Nature* **2006**, *444*, 610-613, doi:10.1038/nature05278.
38. 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, doi:10.1016/j.smrv.2011.06.003.

39. 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, doi:10.1109/JBHI.2014.2303991.
40. 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, doi:10.3389/fnins.2014.00263.
41. Lajnef, T.; Chaibi, S.; Ruby, P.; Aguera, P.-E.; Eichenlaub, J.-B.; Samet, M.; Kachouri, A.; Jerbi, K. Learning Machines and Sleeping Brains: Automatic Sleep Stage Classification Using Decision-Tree Multi-Class Support Vector Machines. *J Neurosci Methods* **2015**, *250*, 94-105, doi:10.1016/j.jneumeth.2015.01.022.
42. Wang, H.; Zhang, Y.; Waytowich, N.R.; Krusienski, D.J.; Jin, J.; Wang, X.; Cichocki, A. Discriminative Feature Extraction via Multivariate Linear Regression for SSVEP-Based BCI. *IEEE Trans Neural Syst Rehabil Eng* **2016**, *24*, 532-541, doi:10.1109/TNSRE.2016.2519350.
43. Zhou, G.; Zhao, Q.; Zhang, Y.; Adali, T.; Xie, S.; Cichocki, A. Linked Component Analysis From Matrices to High-Order Tensors: Applications to Biomedical Data. *Proceedings of the Ieee* **2016**, *104*, 310-331, doi:10.1109/Jproc.2015.2474704.
44. Zhang, Y.; Wang, Y.; Jin, J.; Wang, X. Sparse Bayesian Learning for Obtaining Sparsity of EEG Frequency Bands Based Feature Vectors in Motor Imagery Classification. *Int J Neural Syst* **2017**, *27*, 1650032, doi:10.1142/S0129065716500325.
45. Wu, W.; Nagarajan, S.; Chen, Z. Bayesian Machine Learning: EEG/MEG Signal Processing Measurements. *Ieee Signal Processing Magazine* **2016**, *33*, 172-172, doi:10.1109/Msp.2016.2585746.
46. Qian, D.; Wang, B.; Qing, X.; Zhang, T.; Zhang, Y.; Wang, X.; Nakamura, M. Drowsiness Detection by Bayesian-Copula Discriminant Classifier Based on EEG Signals During Daytime Short Nap. *IEEE Trans Biomed Eng* **2017**, *64*, 743-754, doi:10.1109/TBME.2016.2574812.
47. Wu, W.; Chen, Z.; Gao, X.; Li, Y.; Brown, E.N.; Gao, S. Probabilistic Common Spatial Patterns for Multichannel EEG Analysis. *IEEE Trans Pattern Anal Mach Intell* **2015**, *37*, 639-653, doi:10.1109/TPAMI.2014.2330598.
48. Procházka, A.; Kuchyňka, J.; Vyšata, O.; Cejnar, P.; Vališ, M.; Mařík, V. Multi-Class Sleep Stage Analysis and Adaptive Pattern Recognition. *Applied Sciences* **2018**, *8*, doi:10.3390/app8050697.
49. 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*, doi:10.3390/e18090272.
50. Braun, E.T.; Silvera, T.L.T.D.; Kozakevicius, A.D.J.; Rodrigues, C.R.; Giovani Baratto. Sleep Stages Classification Using Spectral Based Statistical Moments as Features. *Revista de Informática Teórica e Aplicada* **2018**, *25*, doi:10.22456/2175-2745.74030.

51. Devuyst, S. The Dreams Subjects Database. Availabe online: https://web.archive.org/web/20180122055737/http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSubjects/ (accessed on 2019-06-01).
52. Hassan, A.R.; Bhuiyan, M.I.H. A Decision Support System for Automatic Sleep Staging from EEG Signals Using Tunable Q-factor Wavelet Transform and Spectral Features. *J Neurosci Methods* **2016**, *271*, 107-118, doi:10.1016/j.jneumeth.2016.07.012.
53. Hassan, A.R.; Subasi, A. A Decision Support System for Automated Identification of Sleep Stages from Single-Channel EEG Signals. *Knowledge-Based Systems* **2017**, *128*, 115-124, doi:10.1016/j.knosys.2017.05.005.
54. Seifpour, S.; Niknazar, H.; Mikaeili, M.; Nasrabadi, A.M. A New Automatic Sleep Staging System based on Statistical Behavior of Local Extrema Using Single Channel EEG Signal. *Expert Systems with Applications* **2018**, *104*, 277-293, doi:10.1016/j.eswa.2018.03.020.
55. 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 Transactions on Instrumentation and Measurement* **2012**, *61*, 1649-1657, doi:10.1109/Tim.2012.2187242.
56. 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.
57. 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*, doi:10.3390/ijerph16040599.
58. 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, doi:10.1155/2014/730218.
59. Freeman, W.J.; Skarda, C.A. Spatial EEG patterns, non-linear dynamics and perception: the neo-Sherringtonian view. *Brain Res* **1985**, *357*, 147-175, doi:10.1016/0165-0173(85)90022-0.
60. 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, doi:10.1097/MD.0000000000006879.