UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

BRAIN COMPUTER INTERFACE FOR DETECTING DROWSINESS USING INSTANCE-BASED LEARNING APPROACH

AMIR JALILIFARD

ORIENTADOR: PROF. DR. EDNALDO BRIGANTE PIZZOLATO

São Carlos – SP

October/2016

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

BRAIN COMPUTER INTERFACE FOR DETECTING DROWSINESS USING INSTANCE-BASED LEARNING APPROACH

AMIR JALILIFARD

Qualificação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Computation Methodologies and Techniques
Orientador: Prof. Dr. Ednaldo Brigante Pizzolato

São Carlos – SP
October/2016

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Ciência da Computação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de Dissertação de Mestrado do candidato Amir Jalilifard, realizada em 26/09/2016.

Prof. Dr. Ednaldo Brigante Pizzolatto
(UFSCar)

Prof.Dr. Estevam Rafael Hruschka Jr.
(UFSCar)

****

Prof. Dr. Sérgio Vale Aguiar Campos
(UFPE)

****

Prof. Dr. Zhao Liang
(USP)

Certifico que a sessão de defesa foi realizada com a participação à distância dos membros Prof. Dr. Sérgio Vale Aguiar Campos e Prof. Dr. Zhao Liang. Depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa do aluno Amir Jalilifard.

Prof. Dr. Ednaldo Brigante Pizzolatto
Coordenador da Comissão Examinadora
(UFSCar)

Resumo

Uma interface cérebro-computador (ICC, BCI em inglês), também chamada interface mente-máquina (IMM), e também interface neural direta (IND), interface telepática sintética (ITS) ou interface cérebro-máquina, é um caminho comunicativo direto entre o cérebro e um dispositivo externo. Essa cria a possibilidade de usuário controlar um sistema ou um ambiente sem necessidade de usar os músculos. Além disso, a ICC possibilita a gravar e analisar as atividades neuropsicológicas de um indivíduo. Claramente a aplicação da ICC aumentou significativamente durante a década passada. Entre várias aplicações dele, recentemente, pesquisadores tem se interessado em uso dos sinais EEGs na área de segurança ao volante. A condução sonolenta é uma das maiores causas de acidentes nas rodovias do país. Essa pesquisa tem como objetivo o desenvolvimento de um sistema de detecção de sonolência por meio de uma abordagem eficiente baseada em algoritmo K-nearest neighbors (K-NN). Na primeira fase a distribuição de energia dos sinais EEG foi obtida usando uma transformação de Fourier (STFT) e depois o valor médio de energia em períodos de tempo de 0.5 segundos, desvio padrão e entropia Shanon foram calculados para cada das sub-frequências de EEG. Por fim, 52 características foram extraídas. O algoritmo Random Forest foi aplicado nesses dados afim de os atributos mais informativos entre os demais. Finalmente 11 características foram selecionados foram selecionados para classificar a sonolência e o estado de alerta. O algoritmo KD-tree foi utilizado como algoritmo de busca de vizinhos mais próximo para ter um classificador K-NN mais rápido. Nossos resultados mostram que a sonolência pode ser classificada eficientemente com 91% de precisão usando os métodos e materiais propostos neste trabalho.

Abstract

Brain Computer Interface (BCI) is a way to establish a communication between brain and computers. It allows the users to control a computer system and even an environment without moving a muscle or it allows the computer to record and analyze the user's neuropsychological brain activities. Clearly, the range of BCI applications has increased in the past decade due to the use of modern machine learning and signal processing methods. Among various applications of BCI, lately the use of EEG records for driver safety has been considered by some researchers. Drowsy driving is a major cause of many traffic accidents. The aim of this work is to develop an automatic drowsiness detection system using an efficient k-nearest neighbors (K-NN) algorithm. First, the distribution of power in time-frequency space was obtained using short-time Fourier transform (STFT) and then, the mean value of power during time-segments of 0.5 second was calculated for each EEG sub-band. In addition, standard deviation (SD) and Shanon entropy related to each time-segment were computed from time-domain. Finally, 52 features were extracted. Random forest algorithm was applied over the extracted data, aiming to choose the most informative subset of features. A total of 11 features were selected in order to classify drowsiness and alertness. The Kd-tree algorithm was used as the nearest neighbors search algorithm so as to have a fast classifier. Our experimental results show that drowsiness can be classified efficiently with 91% accuracy using the methods and materials proposed in this paper. We also compared the classification results obtained by K-NN (as an instance-based learning algorithm) with four well-known classifiers including decision tree, support vector machine, logistic regression and naive bayes.

Keywords: Brain Computer Interface, EEG classification, Drowsiness detection, K-Nearest Neighbors, EEG signal processing, Kd-trees, Random forest, Feature selection

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

ANN – *Artificial Neural Network*

BCI – *Brain Computer Interface*

BMI – *Brain Machine Interface*

BTLE – *Bluetooth Low Energy*

CCD – *Charge-Coupled Device*

DFS – *Discrete Fourier Series*

DFT – *Discrete Fourier Transform*

DTFT – *Discrete-Time Fourier Transform*

DVR – *German Road Safety Council*

DWT – *Discrete Wavelet Transform*

ECG – *Electrocardiography*

ECoG – *Electrocorticography*

EEG – *Eletroencefalograma*

EMG – *Electromyogram*

ERP – *Event Related Potential*

EoG – *Electrooculogram*

FFT – *Fast Fourier Transform*

FT – *Fourier Transform*

HF – *High Frequency*

HRV – *Heart Rate Variability*

IR – *infrared*

KNN – *K-Nearest Neighbors*

KSS – *Karolinska Sleepiness Scale*

LDA – *Linear Discriminant Analysis*

LF – *Low Frequency*

LIBLI – *A Library for Large Linear Classification*

MAC – *Medium Access Protocol*

MANET – *Mobile Ad hoc Network*

MEG – *MagnetoEncephaloGraphy*

MI – *Motor Imagery*

MVC – *Motor Vehicle Collision*

NHTSA – *National Highway Traffic Safety Administration*

NREM – *Non-Rapid Eye Movement*

NSF – *National Sleep Foundation*

PERCLOS – *Percentage of Eyelid Closure*

PET – *Positron Emission Tomography*

REM – *Rapid Eye Movement*

SDK – *Software Development Kit*

SDLP – *Standard Level for Normal Drivers*

STFT – *Short-Time Fourier Transform*

SVM – *Support Vector Machine*

SWM – *Steering Wheel Movement*

SWS – *Slow Wave Sleep*

VLP – *Variation of Lane Position*

WKNN – *Weighted K-Nearest Neighbors*

fMRI – *functional Magnetic Resonance Imaging*

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

This chapter presents the general context and motivation of this research, and the goals that the current investigation aims to achieve.

Brain-computer interface (BCI) (or brain-machine interface (BMI) for some researchers) establishes a direct interface between the brain and some external device. The basic idea is to capture the thoughts by inspecting the brain waves. There are basically two means to do so: invasive and non-invasive. Invasive methods rely on objects (like electrodes or chips) or substances (like chemical molecules) introduced into the brain of a subject or animal. On the other hand, non-invasive methods do not require object implantation or drug inoculation into subjects (or animals) brain (or body). Usually, the subject interacts with the computer system through wearable devices. The fundamental concept is that the brain signal bypasses the normal output pathways of the body.

At the beginning, BCI researches focused on fundamental applications such as prompt control. Later on, the researches focused on applications for paralyzed patients or for cases with locked-in syndrome. Nowadays, the area has experienced more alternative applications in healthy human subjects. One could point out the use of modern machine learning and signal processing methods as the reason for the increase of BCI applications in the past decade. Some of the main applications of this area are: communication, prosthetic control, robotics and security.

One of the investigation areas that has always been considered by researchers is safety related subjects where the main focus is on making the daily activities more secure, specially in the environments that tend to involve in a greater than average number of accidents. Since the human errors is one of the major causes of daily accidents, lately BCI researches have considered human-safety as a fertile area to work on, in order to reduce the probability of human-made mistakes that may cause irreparable causalities.

A traffic collision, also known as motor vehicle collision (MVC), traffic accident or car crash occurs when a vehicle collides with another vehicle, pedestrain, animal, road debris or other stationary obstruction that may result in injury, death or property damage. A number of non-human factors contribute to the risk of collision, including vehicle design, speed of operation and road design. However, human-caused factors such as driver skill, impairment due to alcohol or drugs, behavioral causes, the lack of concentration and drowsiness play a significant role in road accidents (SAHAYADHAS; SUNDARAJ; MURUGAPPAN, 2012a).

Human factor in vehicle collision includes all factors related to human ability to control the vehicle, such as visual and authority acuity, psychological behaviors, reaction speed and decision making ability. A report in 1985 relying on the British and American crush data showed that the driver impairment contribute about 93% of vehicle crushes (LUM; REAGAN, 1995). Driver impairment describes the factors that may prevent driving in a normal level of skills. Common impairments include:

1. Alcohol: According to the government of Canada, coroner reports from 2008 reported that almost 40% of fatally injured drivers had consumed some quantity of alcohol.

2. Physical Impairment: Including poor eyesight or other physical impairments that may cause collision.

3. Drug use: Including some prescription drugs, over the counter drugs (notably antihistamines, opioids and muscarinic antagonists) and illegal drugs.

4. Distraction: Any external stimuli that can affect the driver's attention, for example, conversation with passengers, operating the cellphone or distracting sounds are source of distractions that can cause losing the normal driving state.

5. Drowsiness: Falling sleep suddenly as a result of sleep disorders(e.g. Narcolepsy) or sleep deprivation can lead to losing the car control and putting the car passengers life at risk.

6. Combination of factors: Several aforementioned causes can be combined and result in a much worse situation.

The US National Highway Traffic Safety Administration (NHTSA), conservatively estimated that a total of 100,000 vehicle crashes in each year are the direct result of driver drowsiness. These crashes resulted in approximately 1550 deaths, 71000 injuries and 12.5 billion dollar in monetary losses (RAU, 2005). In 2009, the US National Sleep Foundation (NSF) reported that 54% of adult drivers have driven a vehicle while feeling drowsy, and 28% of them actually fell asleep. The German Road Safety Council (DVR), claims that one in four highway traffic fatalities is due to the momentary driver drowsiness (HUSAR, 2012). These statistics suggest that driver's drowsiness is one of the main causes of road accidents.

A driver who falls asleep at the wheel, loses the control of the vehicle, an action which often results in a crash with either another vehicle or stationary objects. In order to prevent these devastating accidents, the state of drowsiness of the driver should be monitored. The following measures have been widely used for monitoring drowsiness:

1. Vehicle-based measures: A number of metrics, including deviations from lane position, movement of the steering wheel, pressure on the acceleration pedal, etc., are constantly monitored and any change that crosses a specified threshold indicates a significantly increased probability that the driver is drowsy (FORSMAN et al., 2013) , (LIU; HOSKING; LENNÉ, 2009).

2. Behavioral measures: The behavior of the driver, including yawning, eye closure, eye blinking, head pose, etc., is monitored through a camera and the driver is alerted if any of these drowsiness symptoms are detected (FAN; YIN; SUN, 2009), (AKIN et al., 2008).

3. Biomeasures: The correlation between biological signals (electrocardiogram (ECG), electromyogram (EMG), electrooculogram (EoG) and electroencephalogram (EEG)) and driver drowsiness has been studied by many researchers (KOKONOZI et al., 2008), (KHUSHABA et al., 2011), (YANG; LIN; BHATTACHARYA, 2010).

Other than these three, researchers have also used subjective measures, where drivers are asked to rate their level of drowsiness either verbally or through a questionnaire. The intensity of drowsiness is determined based on standard rating (KHUSHABA et al., 2011), (TREMAINE et al., 2010).

Vehicles-based measures can function reliably only at particular environments and are too dependent on the geometric characteristics of the road and to a lesser extent on the kinetic characteristics of the vehicle (LEW et al., 2007). Furthermore, several studies showed that they are poor predictors of performance error risk due to drowsiness (SIMONS et al., 2012), (DAS; ZHOU; LEE, 2012) . Moreover, vehicular-based metrics are not specific to drowsiness. As an example, standard deviation of lane position can also be caused by any type of impaired driving, including driving under the influence of alcohol or other drugs (METS et al., 2011a).

The drowsiness detection using behavioral measures have promissing results, although the majority of researches are conducted in a controlled environments like simulations. Unfortunately, the positive detection rate is decreased significantly when the experiment was carried out in a real environment (PHILIP et al., 2005a). Additionally, driver state cannot be correlated to driving performance which is a major shortcoming of these measures.

Regarding the subjective measures, since the level of drowsiness is measured during long periods of time (several minutes), sudden variations cannot be detected using this kind of measures. Another limitation of using subjective ratings is that the self-introspection alerts the driver, thereby reducing their drowsiness level (SAHAYADHAS; SUNDARAJ; MURUGAPPAN, 2012a).

The previously described measures become apparent only after the driver starts to sleep, which is often too late to prevent an accident. Unlike the aformentioned measures, bio-signals start to change in earlier stages of drowsiness. Hence, they are more adequate to detect drowsiness by making it possible to alert a drowsy driver in a timely manner and thereby prevent many road accidents. Various approaches have been selected to detect and prevent the accidents caused by drowsiness or distraction sources (KAPTEIN; THEEUWES; HORST, 1996) (KURT et al., 2009), however emerging electroencephalogram (EEG) opened new research area in terms of analyzing brain activity through different tasks, such as driving. Using EEG, it is possible to investigate almost all the physical and behavioral activities. Accordingly, EEG processing can be utilized to analyze individual's central nervous system activity through a driving task and evaluate the consciousness and attention level in order to predict and prevent the probable risky situations.

Using EEG with several electrodes, has been proven that it is possible to detect and analyze brain activity changes accurately, although the main lack of EEG equipments with many electrodes is their intrusive nature. In order to make EEG equipments more applicable in day to day life experiments, non-intrusive systems were introduced (YU, 2009) (BAEK et al., 2012). The reliability and accuracy of driver drowsiness detection by using bio-signals is high compared to other methods (SAHAYADHAS; SUNDARAJ; MURUGAPPAN, 2012b). However, the intrusive nature of measuring bio-signals remains an issue to be addressed. This is specially challenging because the less electrode, the less information related to brain neural activities is obtained. Consequently, the detection accuracy may decrease significantly.

For preventing car accidents caused by sleepiness, the system should detect drowsiness accuratly and rapidly. Thus, systems by which drowsiness is detected within long periods of time does not seem to be effective in real world use, even though their detection rate is high. An ideal system also is expected to be non-intrusive and uses less electrodes. These factors imply that drowsiness ought to be detected based on few numbers of EEG samples, which means finding the best possible EEG features that carry enough information about neurons activities during somnolence.

For recognizing drowsiness stage, EEG data must be classified. Regarding the fact that microcontroller device would be used in real experiments (because it is not realistic to use a computer in a limited environment like car), preferably the classification algorithm ought to be fast and easily implementable. Unlinke most learning algorithms, case-based, also called instance-based approaches do not construct an abstract hypothesis but instead base classification of test instances on similarity to specific training cases (AHA; KIBLER; ALBERT, 1991). Instance-based approach postpone generalization until a new instance must be classified. Another characteristic of this approach is that rather than estimating the target function for the entire instance space, the target function is estimated locally and differently for each new instance. In instance-based method, the training is typically simple and fast. Also, it has the ability to learn complex target functions and not to lose information. However, the performance is reduced in high dimensional case, due to the curse of dimensionality (AGGARWAL, 2014). While the aformentioned characteristics of instance-based method make it an appropriate candidate to be engaged in drowsiness detection task, at the same time, the decline in performance in high demensional feature spaces makes it questionable whether it is possible to reduce drowsiness EEG features to a number of features so that an instance-based method can perform well.

Several studies have found evidences that neural oscillations and EEG sub-bands' activities have specific patterns during drowsiness and wakefulness (CANTERO et al., 1999), (CANTERO; ATIENZA; SALAS, 2002). Foong et al. reported a higher Alpha and Theta band power, specially in the frontal, central, parietal and occipital areas, during drowsiness (FOONG et al., 2015). Based on these evidences, since the power of EEG sub-bands and other characteristics of EEG activity of drowsiness differ from alertness then, each of these consciousness stages occupies a specific area in feature space. Regarding the fact that k-nearest neighbors, as an instance-based method, approach tends to partition the feature space into sub-spaces and then classifies each point based on its nearest neighbors, therefore, we assume that when the dimensionality isn't high, K-NN along with kd-trees —as an efficient search algorithm for problems with moderate number of dimensions (SKIENA, 1998) —has the ability to classify EEG patterns of drowsiness, rapidly and accurately.

1.1 Summary

In this capter, we introduced the state of problem and justified the general purpose of the current work. The organization of this document is as follows: Section 2, expatiates neural signal processing and EEG classification techniques. Section 3, reviews the related works in BCI, and then those researches in which sleep detection using EEG techniques were considered as the main idea. In section 4, we will talk about the classification algorithms and the way they may classify the EEG dynamics. The proposed method for detecting and classifying low consciousness level is discussed in Section 5. The last section is allocated to describe the experimental results.

Chapter 2

THEORETICAL BASIS

In this chapter, we will go through a number of techniques of EEG signal analysis that helps understanding how to analyse neural time-series data.

2.1 What is Cognitive Electrophysiology?

Cognitive electropsychology is the study of how cognitive functions including memory, perception, language, social cognition, behavioral monitoring and emotions are generated by electrical activity produced through the population of neurons.

Cognitive electropsychology is a field of a vast and fundamentally different scientific area due to its multi-aspect essence. On one end, medical-oriented researches try to analyze cognitive signals, understanding the physiological-related issues. For these scientists, electropsychology is useful because it is much more precise than behavioral measures or self-reports. In these fields of research, understanding neural mechanism is relevant, but ultimately the goal of research is to dissect and understand the cognitive components of physical rather than psychological properties of brain (COHEN, 2014).

On the other end, psychologists are passionate about reading the electropsychological data, being able to find the relation between behavioral activities and brain's neural activities. For these group of researchers, analyzing the data provided by electropsychology can help them to relate the external human behavior with brain neural interactions.

Engineers hold the other end of this conceptual triangle (HASSANIEN; AR, 2015). They are interested in electropsychology in order to developing the Brain-Computer tools and understanding the neural patterns respected to psychological or physical activities. For these scientists, electropsychology is useful, because it provides a precise data of neural population activity and using these precise data, they can link human nerobiological or neuropsychological process to computational models. Here, analyzing the sophisticated data and interpretating neropsychological signals are crucial. At this spectrum, electrical and computer engineers have their own distinct way to analyze and extract the data they need, in order to develop their favorite tools. Electrical engineers mostly tend to use signal processing techniques along with various digital tools to extract the favorite features from signals. Unlike the previous group, computer engineers incline to use basic signal processing techniques and then using machine learning algorithms, with the purpose of relating extracted features from neural interactions to the external psychological or biological phenomenons, and then using the results to develop Brain Computer Interface (BCI) applications.

Every researcher who works in BCI related studies, may find himself somewhere between two extremes(Figure 2.1) and it is useful for him to think where to place himself because it will help knowing how much time should be devoted to reading cognitive psychology papers versus neuroscience papers, and what kind of analysis should be performed on data.

Figure 2.1: Cognitive electrophysiology is a field defined by a spectrum from cognitive to electrophysiology. As soon as a researcher Knows where the working area is on this spectrum, this will guide his experiments, hypotheses, data analysis, target journals and conferences. (COHEN, 2014)

2.2 Diving into EEG

The brain activity provides several types of signals including magnetic, metabolic and electrical. Magnetic fields can be recorded with MagnetoEncephaloGraphy (MEG); brain metabolic activity related to blood flow can be captured with Positron Emission Tomography (PET) or functional Magnetic Resonance Imaging (fMRI); electrical signals can be acquired through Electro-Encephalography (EEG) (MAKEIG et al., 2012). EEG was first recorded by German psychiatrist Hans Berger (BERGER, 1929) through the use of several electrodes attached to the human skull. EEG is regarded to be the fastest and less expensive way to capture the brain signals and the advances in bio-sensor technologies for EEG have opened up brainwave research and application development at an unprecedented level in recent years.

Technically speaking, the brain has several distinct areas and landmarks such as Nasion, Inion and pre-auricular points. Electrodes should be positioned on the target points of interest in order to capture the areas activities. If they are on the scalp, one can get the EEG measures; if they are on the surface of the brain, it is possible to get the ECoG measures (from the cortical surface with higher spatial resolution than EEG and less susceptible to ambient noise). An international system called 10-20 (see 2.2) was created in order to provide a naming and positioning scheme for EEG applications. Originally, the 10-20 system had only 19 electrodes which were later increased to more than 70. This number can be raised up to 210 electrodes.

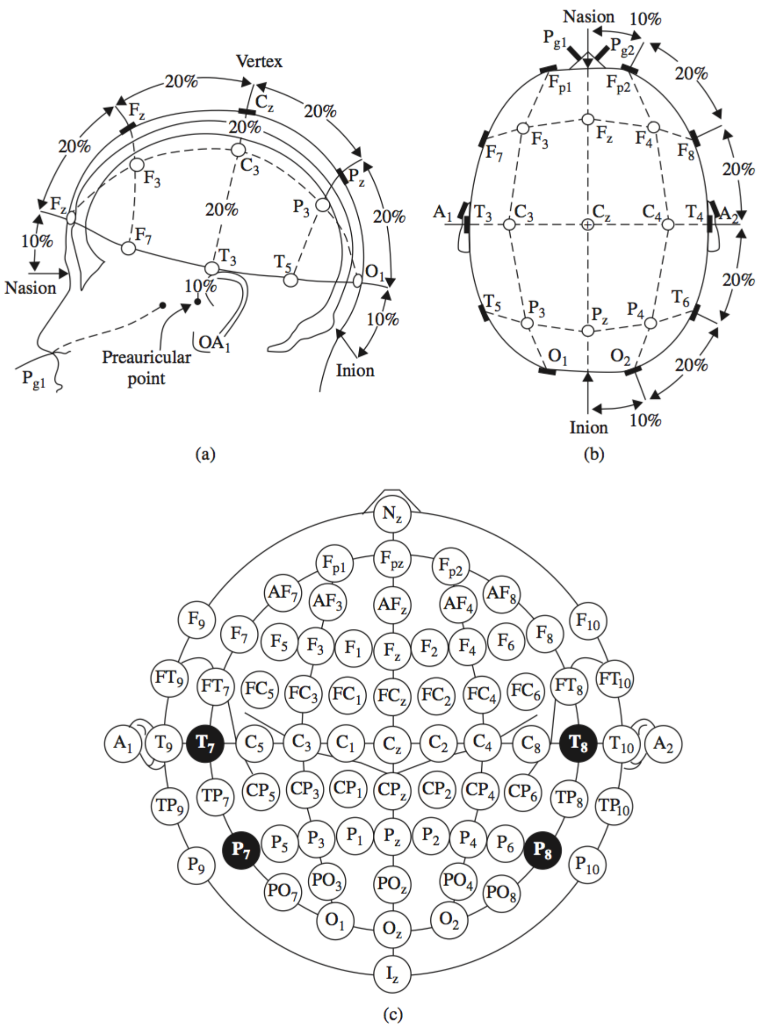

Figure 2.2: A diagrammatic representation of 10-20 electrodes settings for 75 electrodes including the reference electrodes: (a) and (b) represent the tree-dimensional measures, and (c) indicates a two-dimensional view of the electrode setup configuration(COHEN, 2014)

.

For each EEG recording, it is mandatory to choose a reference electrode which should ideally, be affected by global voltage changes in the same manner as all the other electrodes. This reference electrode allows canceling out brain unspecific activity (e.g. slow voltage shifts due to sweating). Usually, a reference can be the ear-lobes, the nose, or the mastoids (i.e. the bone behind the ears). It is also possible to compute the average reference when the system has multi-channels.

2.2.1 Event Related Potential

An Event-Related Potential (ERP) is any stereotyped electrophysiological response to an internal or external stimulus. In simple terms, it is any measured brain response that is the direct result of a thought process or perception. This stimulus can be a specific sensory, cognitive or motor event (LUCK, 2014). The Event Related Potential signals can be negative (e.g. N100, N400) or positive (e.g. P300), and they are named based on the exact time in which they appeared after an internal or external stimuli.

ERPs are measured by EEG signals and it has been found that an event related potential across the parieto-central area of the skull is usually captured around 300ms after the target stimulus, and it is called P300. In other words, P300 is a positive deflection in human event related potential (PICTON, 1992). The P300 potential is characterized by a positive deflection in the EEG amplitude at latency of approximately 300ms after the target stimulus is presented within a random sequence of non-target stimuli. Elicitation time and amplitude are correlated to user's fatigue and to saliency of stimulus (color, contrast, brightness, etc.).

This potential is always presented as long as the user is attending to the process, and its variability among users is relatively low. Also the amplitude of ERP can be different for distinct subjects based on their psychological or physical state. For instance, P300 shows different aplitude for alcoholic and controlled subjects (Figure 2.3) (CHELLA, 2012).

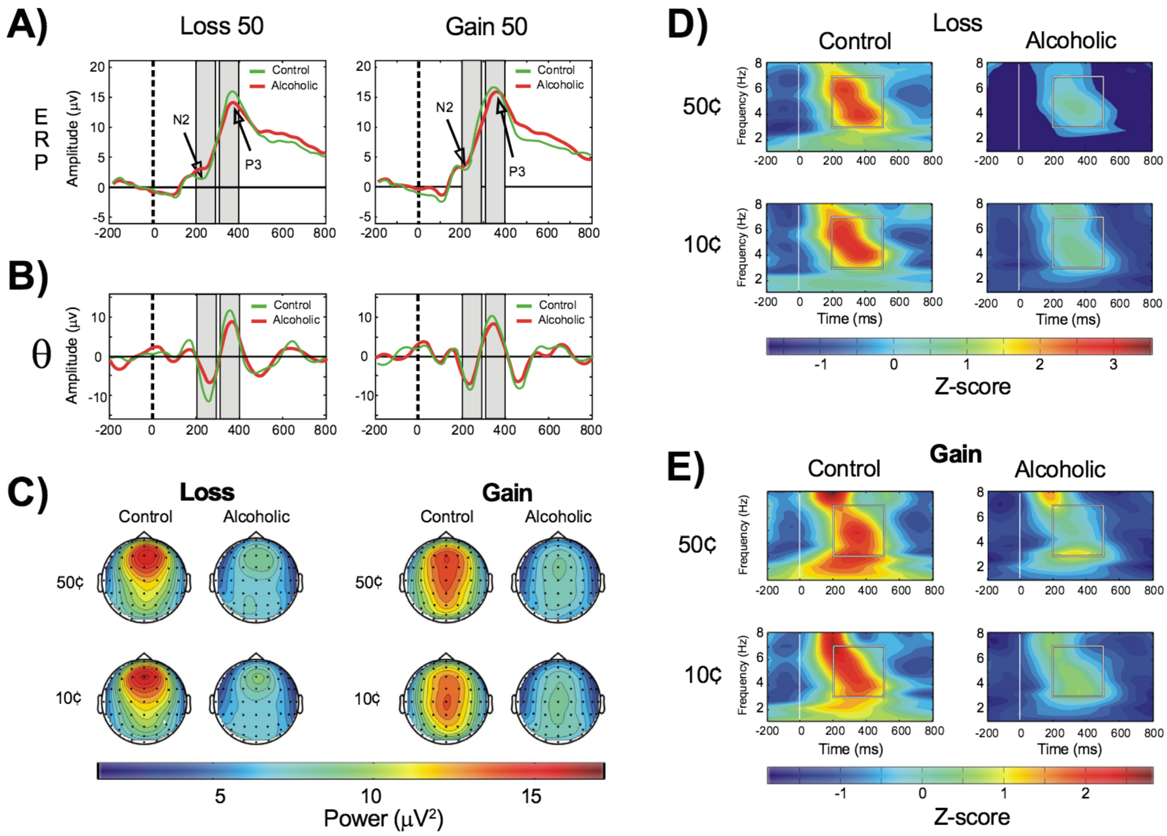

Figure 2.3: A) The grand-averaged ERP waveforms for Loss 50 and Gain 50 conditions; B) Theta bands during Loss 50 at FZ electrode and Gain 50 at CZ electrode are shown. Alcoholics showed decreased amplitude in theta (37 Hz) band more anteriorly (frontal) for the loss condition and more posteriorly (centro-parietal) for the gain condition. C) Topographic maps of theta (3-7 Hz) power in control and alcoholic groups illustrating that the loss condition had an anterior topography and the gain condition had a posterior topography. Decreased theta band power in alcoholics in both conditions is illustrated; D) Time-frequency (TF) plots during the loss condition at the Fz electrode in the alcoholic and control groups; and E) TF plots during the gain condition at the Cz electrode. The square box inside TF plots marks the time-frequency region of interest, namely the time interval of 200-500 ms across the theta frequency range 3-7 Hz for analysis. (CHELLA, 2012)

ERPs are typically quite small (1-30 μv) relative to the background EEG activity. As a result in order to obtain ERP, an averaging filter should operate on EEG signals for several times during different trials, although this can be a disadvantage for ERP due to losing the time-locked activity during the averaging process.

2.2.2 Image Motory

Motor imagery (MI) is a dynamic state during which an action is mentally simulated without any body movement. This technique is a multi-sensorial experience as images can include visual, auditory, tactile and kinesthetic components (GUILLOT; NADROWSKA; COLLET, 2009). Motor imagery has been engaged vastly in BCI especially for sport training or improving the performance of the unhealthy individuals for controlling wheelchair and so on. During the MI process the individual imagines a movement such as moving right hand or left hand (Figure 2.4) in the various directions and the EEG signals of MI are gathered to be applied over BCI. MI is efficient to continuous control commands such as left turn, right turn and stop. Especially for dynamic environments MI-Based BCI has the characteristics of fast feedback and high accuracy (CHOU et al., 2014).

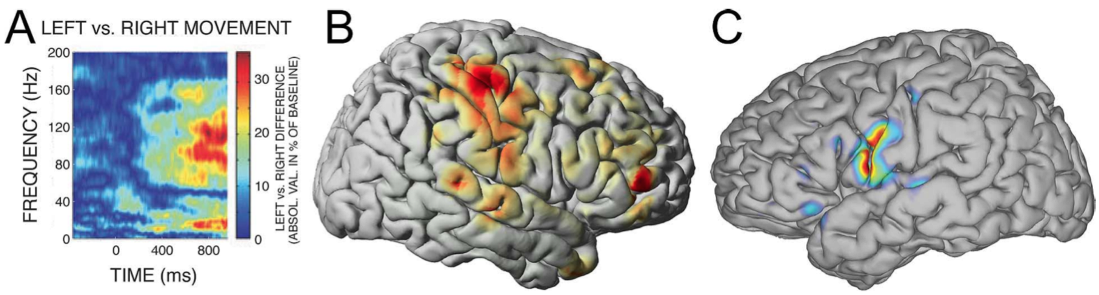

Figure 2.4: Information in EEG about the direction of hand movement and imagined vowels. A) EEG in different frequency bands at one location in the contralateral hand area of motor cortex from one subject differentiates left and right movement directions. B) Color-coded shading of average data from five subjects illustrates the information about hand movement direction provided by EEG recorded over different cortical areas. C) Colored-coded shading of average data from six subjects illustrates the information over production of different vowels provided by EEG recorded over different cortical areas. (LEUTHARDT et al., 2004)

.

2.2.3 EEG Processing Techniques

When EEG raw signal is gathered through various electrodes, the so called data have to be processed. The default EEG data are based on the changes through amplitude of oscillation of each electrode during the time. Using the raw data, it is possible to calculate the voltage using the coefficients, respecting to the device that was used to record EEG signals.

It is worth mentioning that, oscillations are defined by tree information: frequency, power and phase. Frequency, is the speed of oscillation and its unit is Hertz (Hz), which refers to the number of cycles per second in radian-base space. The more cycles a signal can round per second, the faster it is. Moreover, when we are talking about the power of signal, actually, we are talking about the amount of energy which is the squared amplitude of the oscillations. Phase, on the other hand, means, where the signal stands in the current moment along the sin wave (See figure 2.5).

For EEG raw data the time-domain is accessible by default. By utilizing Fourier transform, it is possible to obtain the frequency-domain, which is another powerful mathematical tool for analyzing data.

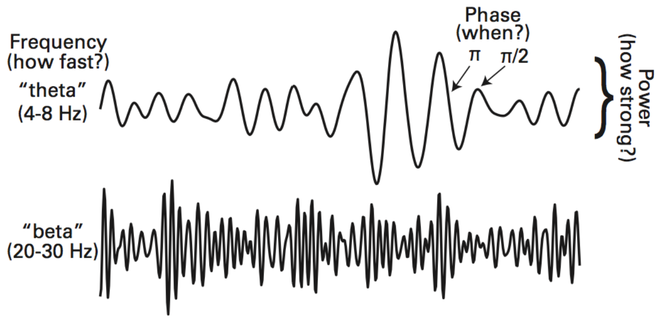

Figure 2.5: The tree dimensions that show frequency, power and phase (COHEN, 2014)

.

2.2.4 Time-Domain

As it was mentioned, the raw EEG data lives in time-domain, where the changes of amplitude during the time is observable. Amplitude is convertible to volt (typically microvolt) which is the measurement of EEG signals(Figure 2.6).

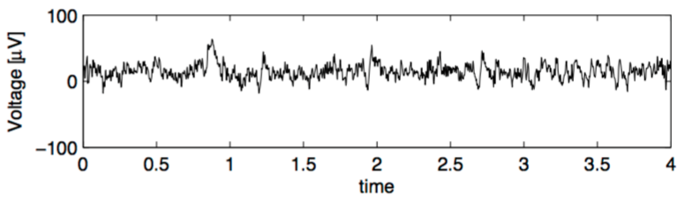

Figure 2.6: Raw EEG data showing oscillation at different speeds and for different lengths of time. Each line corresponds to a specific electrode (SAŁABUN, 2014)

.

There are studies in which the main focus is on time-domain, although the values in this domain can be difficult to interpret due to the following reasons :

1. Time-domain values may differ depending on data processing and analysis decision.
2. Temporal and spatial filters will change the microvolt values.
3. The results are shown in time-domain are too complicated, not beautiful in analysis aspect.
4. When signal is observed in time-domain, values are not periodic. It means it is not possible to sample data, analyze and generalize the results for the future upcoming data.

All the given reasons can lead us to get attracted by something more elegant and easier to understand: the so-called frequency-domain, where the clear results can offer a better environment to analyze the data and understand the mysteries behind EEG signals.

2.2.5 Frequency-Domain and Fourier Transform

Frequency, means how fast a wave round in radian space. When signals are observed in frequency-domain, changes in phase or amplitude are measurable. Frequency-domain provides a better representation of signal. In order to move from time-domain to frequency-domain, a powerful mathematical tool called *Fourier Transform*, is utilized. The goal of Fourier analysis is to express signal in terms of basic sinusoidal components, sine and cosines. By doing so, the hidden signal properties is accessible. This helps have a better understanding about the signal (LYONS, 2010). For converting time-domain to frequency-domain, there are tree major types of Fourier transform as follow :

1. Discrete Fourier Transform (DFT), which maps length-N signals into a set of N discrete frequency components.

2. Discrete Fourier Series (DFS), which maps N periodic sequence into a set of N discrete frequency components.

3. Discrete-Time Fourier Transform (DTFT), which maps infinite sequences into a space of 2 π-periodic function of real-valued argument.

   By sampling EEG data, we can use DFT for moving to frequency-domain. It is possible to use another algorithm to calculate Fourier transform of a finite-length signal called, Fast Fourier Transform (FFT), in order to have an algorithm with better performance. Nevertheless, nowadays computers can use DFT without any problem.

2.2.6 Ways to view Time-Frequency Results

We can imagine time-frequency results as a 3-D cube in which the dimensions are begin time, frequency and space. Then if we want a 2-D view, we can slice this cube from different dimensions as follow:

1. Frequency slice (Figure 2.5A): In this case an specific frequency band is selected and the activities over time (power) is plotted.

2. Time slice (Figure 2.5B): What this slice carry out is to show power as a function of frequency during a period of time.

3. Space slice(Figure 2.5C): Space slice shows data on one time-frequency point over electrodes in a topographical plot.

4. Time-Frequency slice(Figure 2.5D): In time-frequency slice, time is on x-axis and frequency is shown on y-axis. The colors of plot shows various features of signal as though power,phase clustering and so on.

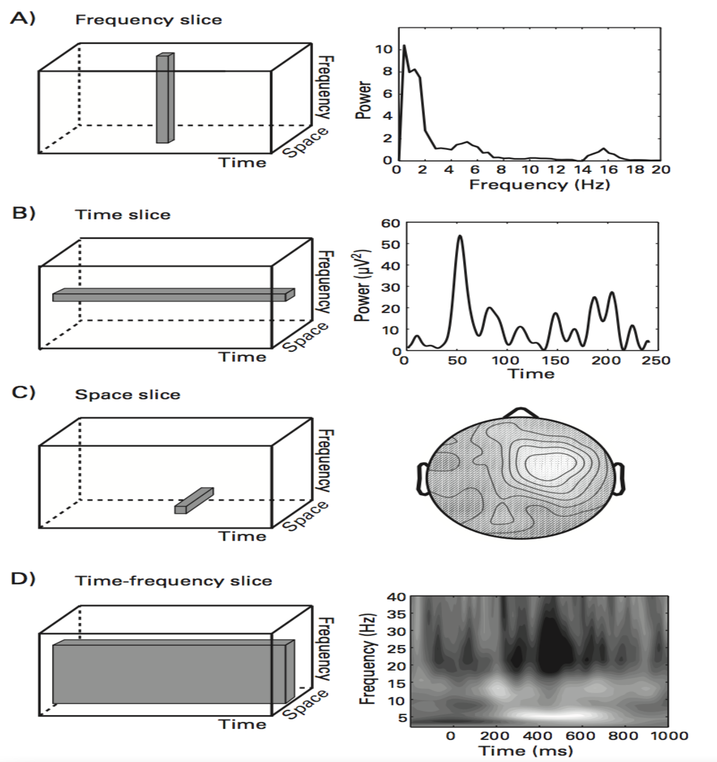

Figure 2.7: The data cube containing information over time, frequency and space. Since it is difficult to view a 3-D cube, it is sliced in four points of view. (COHEN, 2014)

2.2.7 Brain Frequency Bands

EEG data contains rhythmic activity that reflects neural oscillation. Even in raw EEG data, this rhythmic activity is observable. The oscillations can be fast, slow or transient. They can also be changed through the task events. Brain wave frequency bands include delta(2-4 Hz), theta(4-8 Hz), alpha(8-12 Hz), beta(15-30 Hz), lower gamma(30-80 Hz) and upper gamma(80-150 Hz), however, there is no specific boundary defining these bands. As you might see, theta may refer to 3-9 Hz, or 3-7 Hz or even 4-7 Hz. Furthermore, individual differences can impact the aforementioned boundaries. These frequency bands are obtained after changing time-domain to frequency-domain with the purpose of using signal filtering techniques to separate each band. Analyzing the changes in frequency bands, is a way to detect different brain states, like detecting various stages of sleep.

2.3 Sleep EEG

Sleep is the state of natural rest, which is observable in both humans and animals. It is also, an interesting but not perfectly known psychological phenomenon. From physical view, sleep is characterized by reduction in body movement, and decreased reaction for internal or external stimuli. However all these characteristics are same in unconsciousness state, but they are technically different (STERIADE, 1992). Sleep is a stage of unconsciousness by which, a person can be aroused (e.g. sleepwalking and sexual arouse). In this stage the brain is more responsive for the internal stimuli than external one. Therefore, sleep is distinguishable from all the other types of deep unconsciousness like coma in which the individual cannot be aroused.

Sleep is necessary for brain to be healthy. Sleep deprivation leads to have lack of concentration. It also causes developing a major decease in memory-related and calculation tasks. The release of growth hormone in both adults and children takes place during the sleep time. Moreover the body cells show increased production and reduced breakdown during sleep. Therefore, sleep helps maintaining optimal social and emotional functionality in human being, however, when it doesn't takes place in right time and right place, it may cause some problems in form of sleep diseases and unpleasant accidents during day to day life.

2.3.1 Sleep Stages

There are two distinct states that alternate in cycles and reflect differing levels of neuronal activity. Each state is characterized by a different type of EEG activity. Sleep consists of non-rapid eye movement (NREM) and REM sleep (ŠUŠMÁKOVÁ, 2004). NREM is further subdivided into four stages of I (drowsiness), II (light sleep), III (deep sleep), and IV (very deep sleep).

During the night, NREM and REM stages of sleep alternate. Stages I, II, III, and IV are followed by REM sleep. A complete sleep cycle, from the beginning of stage I to the end of REM sleep, usually takes about one and a half hours. However, generally, the ensuing sleep is relatively short and, for most practical purposes, a duration of 10-30 minutes suffices.

Stages I is where drowsiness happens, and it is considered as a transition stage from consciousness to unconsciousness stage when the muscles begin to relax. This stage accounts for 5-10% of sleep time. An individual can be easily awaken in this sleep stage. The slow activity increases as drowsiness get deeper. The findings show that in light drowsiness, the P300 increases in latency and decreases in amplitude (KOSHINO et al., 1993). Also the interhemispheric and intrahemispheric EEG coherence alter (WADA et al., 1996). Figure 3.1 shows the EEG record during drowsiness.

Stage II of sleep occurs throughout the sleep period and represents 40-50% of the total sleep time. During stage II, brain waves slow down with occasional bursts of rapid waves. Eye movement stops during this stage. Slow frequencies ranging from 0.7 to 4 cycles/s are usually predominant; their voltage is high, with a very prominent occipital peak in small children and gradually fall as age increases (ACHERMANN, 2009).

In stage III, delta waves begin to appear. They are interspersed with smaller, faster waves. Sleep spindles are still present at approximately 12-14 cycles/s but gradually disappear as the sleep becomes deeper. In stage IV, delta waves are the primary waves recorded from the brain. Delta or slow wave sleep (SWS) usually is not seen during routine EEG (MARTIN; MARZEC, 2003). Stages III and IV are often distinguished from each other only by the percentage of delta activity.

REM sleep includes 20-25% of the total sleep, follows NREM sleep and occurs 45 times during a normal 8-9 hours sleep period (SIEGEL, 2001). The EEG recorded during REM sleep is similar to one recorded during wakefulness. Evaluation of REM sleep involves a long waiting period since the first phase of REM does not appear before 60-90 minutes after the start of sleep. The EEG in the REM stage shows low voltage activity with a slower rate of alpha.

In general, frequencies from 1 Hz to 16 Hz increase when approaching sleep, while frequencies 17 Hz and higher decrease. This pattern continues after the onset of sleep, except that power in the 8-11 Hz range begins to increase as well. The change in the 8-11 Hz signal is especially noticeable in the occipital region where the signal diminishes and transitions to the frontal lobe. Also in Stage I (that is very important for detecting drowsiness) of sleep the amplitude of raw EEG signal is low and the signal power in higher frequencies attenuate (HAL et al., 2014).

2.4 EEG Equipments and Research Costs

At first glimpse, it may seem that EEG research doesn't cost a lot but this idea comes from not understanding the problem scope that the researcher is going to work with. Buying EEG equipments considering EEG caps in various sizes to be used for different head sizes, electrodes and the other accessories can cost US$ 5000. New amplifiers are expensive too. For university researches buying the license is also mandatory and the license can coast at least US$ 1200 , depending on the EEG cap or headset company. By adding powerful computer, eye tracker and electrode localization equipment, a BCI research can cost over US$ 15000.

The main problem with EEG and BCI research is that, the more the equipments are precise, the more research becomes difficult to be implemented in the real life problems. For example, detecting drowsiness while driving, is hardly possible to be done in a real car using a cap with 128 electrodes, and some computers connected to those electrodes. Another example is using EEG to control a wheelchair, that can be overwhelming due to the use of too many wires, several computers and other equipments.

Sometimes the EEG equipments need to be configured and installed by an expert that adds the configuration and installation cost to the aforesaid equipment expenditure. Furthermore, there is a main question that may come to mind : How can we apply the results of researches using huge laboratory equipments, in the real life? It is clear-cut that, we can't ask the user to wear an elaborate EEG hat in daily life.

The fact that, elaborate EEG equipments are expensive, not easily configurable and not possible to be applied in the daily life, imposes this idea of using some equipments that, despite the less precise data, they are useful enough to be applied in data analysis and development of BCI applications. This is what leads the researchers to use a new generation of EEG device called *single electrode equipment*. Single electrode EEG equipment, comes with lots of defects which may not be acceptable in advanced research fields, but instead, it totally fits the idea of having not very accurate, but, good enough data for our needs. These devices are easily configurable, portable and handy. In the past these devices were not taken seriously, but lately the number of investigations that use single electrode recording as their main EEG data source, has been increased. (KATONA et al., 2014).

2.5 Summary

In this chapter, we took a look at the theorical basis of brain computer interface. Various methods of analyzing EEG signals were reviewed. Afterwards an introduction about sleep EEG and signals and sleep stages was presented and in the end, we had a look on EEG equipment and research cost, and we concluded that it seems necessary to use low-cost single electrode EEG equipments because not only they are cheaper, they are more user-friendly as well . For recognizing EEG signals related to wakefulness and drowsiness, the EEG signals should be classified. To this end, in the next section several well-known classification and feature selection approachs will be described.

Chapter 3

CLASSIFICATION AND FEATURE SELECTION

In this chapter the classification and feature selection algorithms that have been studied in the current research will be explained.

In general, when we talk about classification problems, we tend to go through a decision making process for future coming data in terms of putting them in predefined and distinguishable classes, based on already labeled data called *training* set. Regarding the drowsiness detection problem, some researchers have been interested in using machine learning algorithms for drowsiness detection and several classification methods have been used such as support vector machine (SVM) and artificial neural networks (ANN). This study, besides surveying the ability of instance-based learning method for drowsiness detection, aims at comparing the classification results obtained by other well-known algorithms with the one obtained by instance-based approach. Hence, in this chapter, in addition to instance-based learning method, five common classification algorithms and two feature selection methods are explained.

3.1 Support vector machine (SVM)

Support Vector Machine (SVM) (WESTON; WATKINS, 1998) is a promising method to classify both linear and non-linear datasets. It uses a non-linear mapping in order to transform the data into a higher dimension. The idea is to find the maximum marginal hyperplanes (MMH) or widest street (SOMAN; LOGANATHAN; AJAY, 2009) between separable areas among infinite number of separating lines in order to minimizing the classification error. The area between two separating lines is called plane and the term hyperplane is used to refer to the decision boundary that we are looking for (HAN; PEI; KAMBER, 2011). The separating hyperplane can be written as $W.X + b = 0$ where W is weight vector, X is training tuple and b

is the offset parameter.

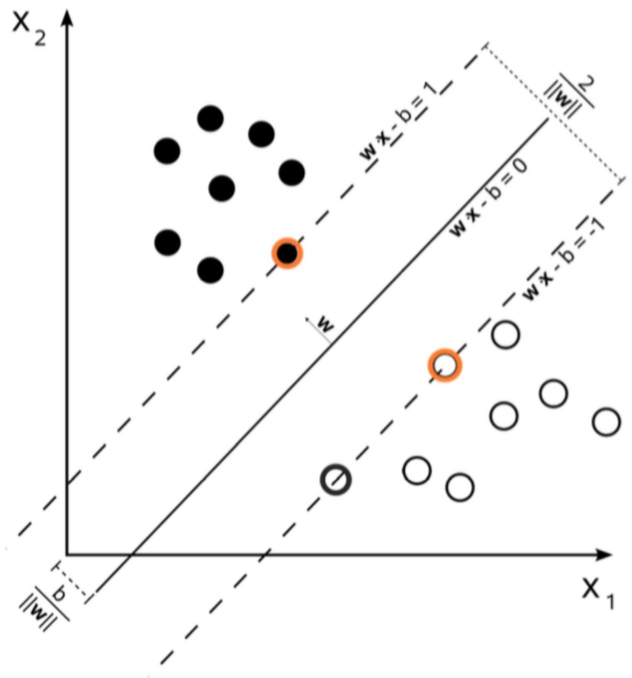

Figure 3.1: SVM finds the widest hyperplane. Support vectors are shown with a red border (XU et al., 2009)

.

Considering b as an additional weight, the points that are above and under the separating plane satisfy

$$y_i(w_0 + w_1x_1 + w_2x_2) \geq 1, \forall i. \tag{3.1}$$

where y_i is +1 if the point is above the hyperplane and -1 when it falls under the hyperplane. Using Lagrangian formulation the MMH is written as a decision boundary.

$$d(X^T) = \sum_{i=1}^{l} y_i \alpha_i X_i X^T + b_0 \tag{3.2}$$

where X_i is support vector, X^T is the test tuple and y_i is the class label of i^{th} support vector.

Given the test tuple X^T, the output tells in which part of the plane the test instance falls. If the result is a black point, then it falls on the top of plane and belongs to the black class and in case of being a white, it falls under hyperplane and belongs to white class. When instances aren't linearly separable, SVM transforms data to a new dimension in which instances related to each class will be separable linearly. Rather than the size of data, the complexity of the learned class is characterized by support vectors. Hence, SVM is less prone to over-fitting problem (HAN; PEI; KAMBER, 2011).

SVM has been used for drowsiness classification in several studies (ABOALAYON; OCBAGABIR; FAEZIPOUR, 2014), (YEO et al., 2009). In spite of the fact that SVM is one of the most powerful classification techniques and it was successfully applied to many real world problems (LECUN et al., 1995), it has some limitations. Perhaps the biggest problem of support vector approach is choosing the best kernel function for the given problem (WANG, 2005). The second drawback of SVM is the speed in both training and test phases (WANG, 2005). Burges et al. conducted a research to improve the accuracy and speed of support vector machine and they achieved a factor of fifty speedup in test phase (BURGES; SCHLKOPF, 1997). However, the training speed is still an unsolved problem (WANG, 2005).

3.2 Naive Bayes

The Naive Bayes (JIAWEI; KAMBER, 2001) is based on the Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Naive Bayes can handle an arbitrary number of independent variables whether continuous or categorical. Given a set of variables, $X = \{x_1, x_2, x_3, ..., x_d\}$, we want to construct the posterior probability for the event C_j among a set of possible outcomes $C_j = \{c_1, c_2, c_3, ..., c_d\}$. In a more familiar way, X is the predictors and C is the set of categorical levels present in the dependent variable. Using Bayes' rule:

$$P(C_j \mid x_1, x_2, x_3, \ldots.., x_d) = P(x_1, x_2, x_3, \ldots.., x_n \mid C_j) \times P(C_j) \tag{3.3}$$

where $P(C_j|x_1, x_2, x_3, ..., x_d)$ is the posterior probability of class membership (i.e., the probability that X belongs to C_j). Since Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent we can decompose the likelihood to a product of terms:

$$P(X \mid C_j) = \prod_{k=1}^{d} P(X_k \mid C_j) \tag{3.4}$$

and rewrite the posterior as:

$$P(C_j \mid X) = P(C_j) \prod_{k=1}^{d} P(X_k \mid C_j) \tag{3.5}$$

Using the above Bayes' rule, we label a new case X with a class level C_j that achieves the highest posterior probability.

Although the assumption that the predictor (independent) variables are independent is not always accurate, it does simplify the classification task dramatically, because it allows the class conditional densities $P(X_k|C_j)$ to be calculated separately for each variable, i.e., it reduces a multidimensional task to a number of one-dimensional ones. As a matter of fact, Naive Bayes reduces a high-dimensional density estimation task to a one-dimensional kernel density estimation. Naive Bayes is simple, easy to implement, robust to noisy data and isn't very sensitive to over-fitting problems due to its probabilistic nature.

3.3 Decision tree

A decision tree (QUINLAN, 1987) is a flowchart like tree structure where each nonleaf node denotes a test on an attribute and each branch represents a result of a test. Each terminal node holds a class label. The topmost node is the root node. Given a tuple, T, for which the class is unknown, the attribute values of tuple are tested against the decision tree. The path is followed from node to leafs where the class prediction for tuple is held. A decision tree works on two different kind of attributes namely numerical and non-numerical. In case of numeric attributes, decision trees can be geometrically interpreted as a collection of hyperplanes, each orthogonal to one of the axes.

Iterative Dichotomiser 3 (ID3) (QUINLAN, 1986) is an algorithm used to generate a decision tree from a dataset. The main choice in ID3 is selecting which attribute to test at each node in the tree. The idea is to select the attribute that is most useful for classifying instances. It defines a statistical property called *Information gain* that measures how well a given attribute separates the training examples according to their target class. In order to define the information gain, it uses a measure of information theory called *Entropy*. ID3 uses this information gain measure to select among the candidate attributes at each step while growing the tree.

In general, for constructing a decision tree no knowledge about domain or no parameter setting is required. Therefore, it is appropriate for exploratory knowledge discovery (HAN; PEI; KAMBER, 2011). Decision tree can handle high dimensional data. Also, since it represents the data in tree-like construction, it is appropriate to be assimilated by humans. In general, not only the learning and classification steps of decision tree induction are simple and fast, also it has good accuracy. Nevertheless, overfitting is a significant practical difficulty for decision tree models.

3.4 Logistic regression

The general form of regression, called generalized linear regression (YAN, 2009), assumes that the data points are coming from a distribution that has a mean that comes from a monotonic nonlinear transformation of a linear function of the predictors. If we can call this transformation g, the equation can be written as:

$$y = g(\alpha + \beta * x) + \varepsilon \tag{3.6}$$

where α is a constant, sometimes also denoted as b_0, β is a vector of the same size as our input variable x and where the error term is ε.

Although the linear regression model is simple and used frequently its not adequate for some purposes. For example, imagine the response variable Y to be a probability that takes on values between 0 and 1. A linear model has no bounds on what values the response variable can take, and hence Y can take on arbitrary large or small values. However, it is desirable to bound the response to values between 0 and 1. Also, for linear regression it is hard to describe a dichotomy outcome with two parallel lines (see Figure 3.2). For this we would need something more powerful than linear regression.

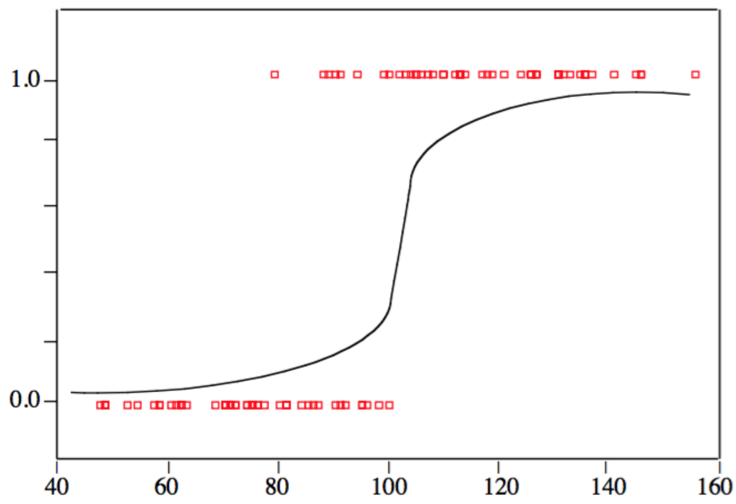

Figure 3.2: Relationship of a dichotomous outcome variable with a continuous predictor (PENG; LEE; INGERSOLL, 2002)

.

Logistic regression (WALKER; DUNCAN, 1967) solves these kinds of problems using logit transformation —natural logarithm of odds ratio —to the dependent variable, where odds are ratios of probability of Y happening. The logistic model can be describe as:

$$logit(Y) = ln(odds) = ln\left(\frac{\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}}{1-(\frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}})}\right) \tag{3.7}$$

where α is the Y intercept and β is the regression coefficient.

Logistic regression is intrinsically simple, it has low variance and so is less prone to over-fitting. It is also fast and reliable when the dimension gets large (TU, 1996). Another advantage is that it is easy to inspect.

3.5 Instance-based learning

Instance-based learning (STANFILL; WALTZ, 1986), (ATKESON, 1989) and (MARON; MOORE, 1993) is an approach including several algorithms that explicitly remembers all the data that they recieve. They usually have no training time and the computation is performed at prediction time. This is what seperate this approach from other conventional machine learning methods where training step occurs between data reception and prediction phases. During the classification process, this group of methods takes a query point and perform a search on the database to locate the similar datapoints. Then, it creates a local model such as local average or local regression so as to predict the output value. There are three main reasons that cause instance-based approach be prefered over other methods (DENG; MOORE, 1995). These reasons are: a) flexible inductive bias; b) learning parameters does not need to be fixed in advance; c) instance-based can cover the global local spectrum.

3.5.1 Flexible inductive bias

When the size of instances is little, an algorithm like K-nearest neighbors yields unreliable prediction, however, as the size of dataset is increased, so the complexity of function that nearest neighbor can approximate. This is a characteristic with which the approximation power is increased locally acording to amount of data, that is not observable, for example, by default in multi-layer neural network.

3.5.2 Learning parameters does not need to be fixed in advance

There are several learning parameters to be configured in instance-based learning. One of the most important parameters is concerned with trading smoothing of noise against fitness suitability. Other parameters include a distance metric (to determine the similarity between existing datapoints in dataset and a query instance), decision of relevant attributes, decision about neighbors range and the importance of datapoints based on their approximity to a query point. Although all of these parameters are configurable, instance-based methods do not need to decide on them in advance. Hence, they can decide to use a desired default parameter for one prediction and another different set of parameters for other one. This can be a very useful feature in autonomous system where it can predict online and at the same time tune the learning parameters for new coming data (MOORE; HILL; JOHNSON, 1992). In contrast, for non-instance-based approaches, setting parameters in order to train is necessary. In case of any need to change one or more parameters, the training phase must be repeated.

3.5.3 Instance-based can cover the global local spectrum

For instance-based methods, it is not necessary to use all the datapoints in feature space to form a prediction (MOORE; HILL; JOHNSON, 1992). This is specially crucial for noisy systems where the underlying function is non-linear but generally smooth. In this case, instance-based methods may use a percentage of nearest datapoints to the query point to form its prediction.

These characteristics make this approach desireable. K-nearest neighbors classification algorithm is a classical example of instance-based method, in which the k nearest neighbors of a classifier are used in order to create a local model for the test instance. This algorithm has been successfully applied to many classification problem (EBRAHIMPOUR; KOUZANI, 2007), (SHI et al., 2011), (DEOLE; LONGADGE, 2014). In the following, this algorithm is described.

3.5.4 K-nearest neighbors (K-NN)

K-nearest neighbors, also known as proximity search, similarity search or closest point search, is the most basic instance-based learning algorithm which can be used for both continuous and discrete values. It assumes that all the instances correspond to a point in the n-dimensional feature space. This algorithm also considers that if two instances are near to each other, then, they belong to a same class. For example, if x is an arbitrary instance defined by feature vector as below:

$$< a_1(x), a_2(x), a_3(x), ..., a_n(x) > \tag{3.8}$$

where $a_r(x)$ indicates the value of rth attribute of x. Therefore the distance is calculated by d(xi, xj) where:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \tag{3.9}$$

Giving labeled training dataset and the above formula, it partitions the feature space to polygonal-like sections where each polygon contains all the nearest neighbors to a specific point in feature space.

This algorithm can be refined by giving a weight coefficient based on the distance of each points to the new instance. Dudani (DUDANI, 1976), first introduced a weighted voting method for KNN, called the distance-weighted k-nearest neighbor rule (WKNN). In WKNN, the closer neighbors are weighted more heavily than the farther ones, using the distance-weighted function. Consequently a neighbor with smaller distance is weighted more heavily than one with greater.

The K-nearest neighbor algorithm is easily adapted to approximating continuous-valued target functions. To accomplish this, there is an approach that considers the mean value of the k nearest training examples rather than calculate their most common value.

In comparison with some classifiers like Support Vector Machine (SVM) and Neural Network (NN), K-nearest neighbors is simple and easy to implement. Therefore this classifier has been considered interesting for some researchers who wanted to classify EEG data. Le. et. al. (LI et al., 2012) used EEG signals to find the study effectiveness in students. They applied Naive Bayes and K-nearest neighbors to the features extracted from EEG data to find out a combination, which may mostly contribute to reflect learner's affect, for example, Attention. They intended to find a relationship between emotional state and characteristic of brain activity. During this investigation they reached the classification rate of 57% using KNN (K=5). They mentioned that attaining to 57% of correctly classified instances was a good result since they could find a correlation between emotional state and brain activity while studying. (LI et al., 2012)

K-NN is a type of instance-based learning algorithm where the function is only locally approximated. As it was explained, it is the simplest classification algorithm when there is little or no knowledge about the distribution of data (DEVROYE, 1981). Unlike ANN and SVM, K-NN has a very fast training, although, its main drawback is running slowly when the size or dimension of dataset is large. This is because all computation is deferred until classification which is referred as laziness problem (BHATIA et al., 2010). This algorithm usually searches under a brute-force approach which compares the query point to each data point in feature space (INDYK; MOTWANI, 1998). Since lots of the real world problems posed with moderate (up to the 10) or large dimensionality (more than 10), it is beyond the capability of such search to be efficient. To tackle the aforementioned problem and having a fast nearest neighbors classifier, several methods such as effective search approaches (PREPARATA; SHAMOS, 1985), (FRIEDMAN; BENTLEY; FINKEL, 1977a) and approximate nearest neighbors (INDYK; MOTWANI, 1998) have been proposed.

3.6 Kd-Tree

The nearest neighbors searching is the following: given a set S of n data points in the feature space, X^D, the task is to process these data points, so that given any query point $q \varepsilon X^D$, the K nearest points $\{s_1, s_2, ..., s_k\} \varepsilon S$ to q be reported as quick as possible. Various studies mentioned that finding nearest neighbors in high-dimensional or large datasets using brute-force search is inefficient (ARYA et al., 1998), (INDYK; MOTWANI, 1998), thus, several methods proposed to alleviate the search problem of K-NN (LIU et al., 2004).

The Kd-tree algorithm in both its original (BURKHARD; KELLER, 1973) and optimized (FRIEDMAN; BENTLEY; FINKEL, 1977b) forms is a data structure that provides an efficient way for examining just the nearest points to any query point rather than searching the whole feature space, and as a result reducing the search time for finding the best match. In fact it is a binary tree that uses a hierarchical data structure and decomposes the feature space into relatively small cells by using a plane through different dimensions. The root node of the tree represents the whole dataset which is available as a global array, and the terminal nodes are small subsets that form a partition of space. The terminal subsets of records are called *buckets*. If the node under investigation is terminal, then all the records in the bucket are examined. Each node in the tree is defined by a plane through one dimension that partitions the point sets into left and right sets. At each node, the partition not only devides the current subfile, but it also defines an upper and lower boundaries on the value of the discriminator key for each record in the two new subsets. Each subset then is split evenly by plane through different dimensions and this procedure is done recursively. Partitioning stops after $\log n$ levels, with each point in its own leaf cell. If the search is done in a one-dimensional space, then a record is defined as a key and the plane (or partitioner) is defined by some values of that key. All the recordes with values less than the plan value is considered as left branch and those one with values larger than plan are known as the right branch. In case of K-dimensional searching, a record is represented by K keys. Any of these keys can play a discriminator for partitioning a subset under a particular node.

The search algorithm is described as a recursive procedure . When kd-trees is used as K-NN search algorithm, it constructs a binary tree structure based on training data and then, navigates down the tree to the region that probably contains nearest neighbors (see Fig. 3.3). Since finding 1 nearest neighbor in a balanced Kd tree with randomly distributed points takes $O(\log n)$ time on average, it is fairly fast, however, its query time grows exponentially when the dimensionality is increased (SPROULL, 1991), yet by choosing the most informative subsets of features and reducing dimensionality of search space and using Kd-tree as search algorithm, the classification time may reduce significantly while the accuracy is preserved.

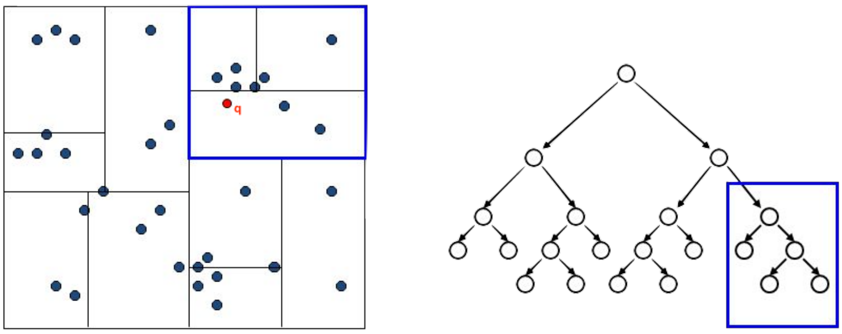

Figure 3.3: K-nearest neighbors search using kd-trees

3.7 Feature selection for demensionality reduction

Dimensionality reduction is a commonly used step in machine learning, especially when dealing with a high dimensional space of features. The original feature space is mapped onto a new, reduced dimensioanllyity space and the examples to be used by machine learning algorithms are represented in that new space. The mapping is usually performed either by selecting a subset of the original features or/and by constructing some new features. In this section the Principal Component Analysis (PCA) and Random Forest algorithms, as two common demensionality reduction techniques, are presented.

3.7.1 Principal component analysis (PCA) for feature selection

PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The question that PCA helps us to answer fundamentally is this: Which of these M parameters explain a signficant amount of variation contained within the data set?

Let X be a zero mean n-dimensional random feature vector and Σ be the covariance matrix of X . Let A be a matrix whose columns are the orthonormal eigenvectors of the matrix Σ:

$$\Sigma = A\Lambda A^T \tag{3.10}$$

where Λ is a diagonal matrix whose diagonal elements are the eigenvalues of $\Sigma, \lambda_1 \geq \lambda_2 \geq \lambda_3 ... \geq \lambda_n$. Let A_q be the first q columns of A and let $V_1, V_2, ..., V_n \varepsilon \mathbb{R}^q$ be the rows of matrix A_q.

Each vector V_i represents the projection of the ith feature (variable) of the vector X to the lower dimensional space, that is, the q elements of V_i correspond to the weights of the ith feature on each axis of the subspace. The key observation is that features that are highly correlated or have high mutual information will have similar absolute value weight vectors V_i. To find the best subset we use the structure of the rows V_i to first find the subsets of features that are highly correlated and follow to choose one feature from each subset. The chosen features represent each group optimally in terms of high spread in the lower dimension, reconstruction and insensitivity to noise.

3.7.2 Random forest algorithm for feature selection

Bagging (BREIMAN, 1996) is a technique that takes high variance (noisy) and low-bias procedures, averages them and reduces the variance of the estimated function. Random forest (BREIMAN, 2001) is a modification of bagging that builds a large collection of random trees and then averages them. The algorithm has two sources of randomness. The first is randomness in data and the second is randomness in split. It provides the class of dependent variable based on many trees. In other words, it is an ensemble learning technique that operates by constructing several trees at training time and outputting the class that is the mean prediction of individual trees and by doing so, it corrects the existing overfitting problem of decision trees and is very effective in eliminating noise in the model input data. Because Random Forest builds many trees using a subset of the available input variables and their values, it inherently contains some underlying decision trees that omit the noise generating variable/feature(s). In the end, when it is time to generate a prediction, a vote among all the underlying trees takes place and the majority prediction value wins.

If there is a training set with size N, the random forest draws a bootstrap sample Z^* of size N from training data, grows a collection of trees T_b to the bootstrapped data, selects m variables randomly and then ranks them. In the end, it outputs an ensemble of B numbers of such trees.

In order to measure the importance of variable X_m to predict class Y, it uses out-of-bag

(OOB) samples to construct a variable importance measure. When the bth tree is grown, the out-of-bag examples are randomly permuted and the data run down the corresponding tree (BREIMAN, 2001). Afterward, the values for the jth variable is permuted and the classification accuracy is saved. Then, the decrease of accuracy as a result of this permuting is averaged for all the trees and is used to evaluate the importance of the variable.

3.8 Summary

In this chapter, five well-known classification algorithms were explained. Additionally, we indicated three facts of instance-based learning that make it to be prefered over other methods. For avoiding the curse of dimensionality, two feature selection algorithms and an optimized nearest neighbors search method were described. In the next section, we will give a general view of the methods that have been suggested and previous researches that have been carried out for drowsiness detection.

Chapter 4

LITERATURE REVIEW

In this chapter the previous works in state of art are discussed in order to acquiring a clear view of general trends toward detecting drowsiness.

There are two types of BCI studies which are worth mentioning: hardware and software. Considering non-invasive methods, from a hardware viewpoint, researchers investigated the replacement of wet electrodes by dry ones. Wet electrodes had some disadvantages like: a) time-consuming to setup; b) long-term monitoring (gel dependent electrode could dry up). The first prototypes of dry electrodes were developed in the late 1960s and early 1970s (LOPEZ; RICHARDSON, 1969) (BERGEY; SQUIRES; SIPPLE, 1971). Recently, the potential of capacitive electrodes and dry sensors made of silicone conductive rubber (OEHLER, 2008) have been demonstrated.

From the software perspective, the origin of BCI can be traced to work in 1960s by Delgano (DELGADO, 1969) and Fetz (FETZ, 1969). Delgano developed an implantable chip (which he called stimoceiver) that could be used to both stimulate the brain by radio and send electrical signals of brain activity by telemetry, allowing subject to move about freely. In a now-famous demonstration, Delgado used the stimoceiver to stop a charging bull in its track by pressing a remote control button that delivered electrical stimulation to the caudate nucleus in the basal ganglia region of the bull's brain. At around the same time Fetz showed that monkeys can control the activity of single brain cells to control a meter needle and obtain food rewards. Slightly later, Vidal (VIDAL, 1973) explored the use of scalp-recorded brain signals in humans to implement a noninvasive BCI based on visually evoked potentials. In 1976, he provided evidences that brain signals could be used as a communication channel to control a cursor through a two dimensional maze (Vital, 1976). Various works combining hardware and software were conducted between 1980s and the end of the 20th century mainly to confirm that the brain (of humans and animals) could be able to control a prompt on a computer screen or a device.

Lots of studies have been conducted with the purpose of improving life quality for patients with major disabilities, including wheelchair control (TANAKA; MATSUNAGA; WANG, 2005) (PHILIPS et al., 2007a) (VANACKER et al., 2007) (GALÁN et al., 2007) (HEMA et al., 2011) (PHILIPS et al., 2007b) and locked-in syndrome (KAUFMANN; HOLZ; KÜBLER, 2013) (HINTERBERGER et al., 2001) (KUBLER et al., 1998) (BIRBAUMER, 2006) (PFURTSCHELLER; FLOTZINGER; KALCHER, 1993). The potential of BCI caught the military attention and many other applications were investigated under the DARPA sponsorship such as: a) mobile robot controlling; b) telepresence; c) exoskeleton controlling; d) robotic augmentation; and e) silent-talk. Most of them are related to wheelchair control as the purpose is to control robots, drones or exoskeleton. Silent-talk is a bit different as the purpose is human-human communication on the battlefield without the use of vocalized speech. Another important BCI application is cyber security which frequently gets the attention of the banks, e-commerce companies and defense departments.

In the field of drowsiness accidents that can be attributed to driver drowsiness, it may be more devastating than the statistics reveal (COLTEN; ALTEVOGT et al., 2006). Hence, in order to avoid this type of accident, it is necessary to derive effective measures to detect driver drowsiness and alert the driver. Researchers have used various methods to detect drowsiness, but four mostly used methods are as follow:

1. Subjective Measures
2. Vehicle-Based Measures
3. Behavioral Measures
4. Biosignal Measures

The rest of this section will provide a review of the researches have been done through these four measures.

4.1 Subject measures

This measure is found on self estimation of subject (the driver) through verbal expression and lots of scales to rate the expressions to a sleepiness rate. One of the most common drowsiness scales is the *Karolinska Sleepiness Scale (KSS)*, a 9 point scale, based on self-reported subjective statement of subject's level of drowsiness (INGRE et al., 2006).The descriptors are marked from 1 as ”very alert”up to 9 as ”very sleepy”. Its original format only includes scores of 1,3,5,7 and 9, but in the additional version more descriptors were added (See Table 3.1) (REYNER; HORNE, 1998) (OTMANI et al., 2005).

Table 4.1: Karolinska sleepiness scale (KSS).

Rating	Verbal description
1	Extremely alert
2	Very alert
3	Alert
4	Fairly alert
5	Neither alert nor Sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep alert
8	Sleepy, some effort to keep alert
9	Very sleepy, great effort to keep alert, fighting sleep

Hu.et al. (HU; ZHENG, 2009) measured the KSS rating every 5 minutes and used it as a reference of EOG collected signal. Portouli et al. (PORTOULI et al., 2007) evaluated EEG data by conforming the drowsiness through questionnaires and a medical practitioner. Some researches investigated on finding a relationship between KSS which was recorded during driving task and variation of lane position (VLP) and found that these measures weren't in agreement (SOMMER et al., 2009). Ingre et al. (INGRE et al., 2006) realized that there is a relationship between the blinking duration and the explained level of KSS scale. Researchers have determined that major lane departures, high eye blink duration and drowsiness-related physiological signals are prevalent for KSS ratings between 5 and 9 (INGRE et al., 2006).

Since the level of drowsiness is measured approximately every 5 minutes, sudden variations cannot be detected using subjective measures. Another limitation for using subjective ratings is that it is difficult to obtain drowsiness feedback from a driver in a real driving situation. Therefore, although subjective ratings are useful in determining drowsiness in a simulated environment, the remaining measures may be better suited for the detection of drowsiness in a real environment.

4.2 Vehicle-based measures

Another method to measure the drowsiness is based on measuring the signals, which are being sent by various vehicle components. In this case, several sensors are installed on the different vehicle components, then data are analyzed and the level of drowsiness is measured. Liu et al. (LIU; HOSKING; LENNÉ, 2009) published a review on current vehicle-based measures. Also, some researches found that there is a relationship between largely speed variation and the drowsiness (OTMANI et al., 2005). The two most common vehicle-based measures are steering wheel movement (SWM) and standard deviation of lane position (SDLP).

4.2.1 Steering wheel movement (SWM)

In this approach, the steering wheel movement (SWM) is measured using a steering angel sensor. This approach has been used widely so as to measure the drowsiness (OTMANI et al., 2005) (THIFFAULT; BERGERON, 2003) (FAIRCLOUGH; GRAHAM, 1999). By mounting a sensor on steering column the micro corrections are measured. Since the micro corrections while drowsiness are reduced in comparison with a normal driving state (FENG; ZHANG; CHENG, 2009), then significant reduction is a measure to detect the drowsiness of car driver (VURAL, 2009).

Fairclough and Graham found that sleep deprived drivers made fewer steering wheel reversals than normal drivers (FAIRCLOUGH; GRAHAM, 1999). In order to eliminate the effect of lane changes, the researchers considered only small steering wheel movements (between 0.5 and 5), which are needed to adjust the lateral position within the lane (OTMANI et al., 2005). Hence, based on small SWMs, it is possible to determine the drowsiness state of the driver and thus provide an alert if needed. In a simulated environment, light side winds that pushed the car to the right side of the road were added along a curved road in order to create variations in the lateral position and force the drivers to make corrective SWMs (THIFFAULT; BERGERON, 2003).

Car companies, such as Nissan and Renault, have adopted SWMs but it works in very limited situations (VURAL, 2009). This is because they can function reliably only at particular environments and are too dependent on the geometric characteristics of the road. (VURAL, 2009).

4.2.2 Standard Deviation of lane position (SDLP)

Using this approach the deviation of lane position is measured by a camera and if the deviation is more than a standard level for normal drivers (or SDLP level), then a risky situation is predicted. Ingre et al. (INGRE et al., 2006) conducted an experiment using KSS and measured the SDLP. They found out that when the KSS rating is increased, the deviation of lane position is also increased. For instance, KSS ratings of 1, 5, 8, and 9 corresponded to SDLP measurements of 0.19, 0.26, 0.36 and 0.47, respectively. In this experiment, 20 individuals participated and the KSS rate and SDLP measures were recorded. They reported that however there was a relationship between these measures, but for some drivers the SDLP measure didn't raise more than 0.25, even for KSS rating of 9 that can be a weak point of SDLP approach. Another defect of SDLP is being completely dependent o external situations, like road structure, climate, light and resolution. Most of the researches and practical experienced showed that the vehicle-based measures aren't very reliable because they are poor predictors. Moreover a measure like SDLP can be caused by various origins, like impaired driving and not just drowsiness (SIMONS et al., 2012) (DAS; ZHOU; LEE, 2012) (METS et al., 2011b).

4.3 Behavioral measures

During the drowsiness, the individual shows some behaviors with specific characteristics, such as change in blinking duration, frequent yawning (FAN; YIN; SUN, 2009), swinging his/her head, the top down head movements, etc. All these abnormal signs are what behavioral measures focus on, in order to detect the drowsiness(LEW et al., 2007). A vast amount of investigations have been interested in blinking as a behavioral characteristic (DORAZIO et al., 2007) (LIU et al., 2010). PERCLOS which is the percentage of eyelid closure, shows a slow and droop closure rather than blinking, during drowsiness. PERCLOS has been analyzed in many studies (ZHANG; ZHANG, 2010) (DINGES et al., 1998) (ABE et al., 2011) (MCKINLEY et al., 2011) and has been found as an effective measure to predict drowsiness (LEXUS, 2012). Some other studies concentrated on facial characteristics, such as jaw drop along side eye blink detection (LEW et al., 2007) (FAN; YIN; SUN, 2008).

Table 4.2: List of previous works on driver drowsiness detection using behavioral measures.

Ref.	Sensor used	Drowsiness measure	Detection technique	Feature extraction	Classification	Detection rate
(WEIJIE et al., 2012)	CCD micro camera, InfraRed	Pupil	Ada-boost	Red eye effect, Texture detection method	Ratio of eye-height, eye-width	92%
(BERGASA et al., 2006)	Camera, Infra-Red Illuminator	Pupil	PERCLOS, eye closure duration, blink frequency,3 other Two Kalman filters for pupil detection	Modification of the algebraic distance algorithm for conics Approximation	Fuzzy Classifier	100%
(FAN; YIN; SUN, 2009)	CCD camera	Yawning	Ada-boost	Red eye effect, Texture detection method	Ratio of eye-height, eye-width	92%
(LEW et al., 2007)	Digital Video camera	Facial action	Gabor filter	Wavelet Decomposition	SVM	96%
(DORAZIO et al., 2007)	Fire wire camera, webcam	Eye Closure Duration, Freq of eye closure	Ada-boost	Red eye effect, Texture detection method	Ratio of eye-height, eye-width	92%
(FAN; YIN; SUN, 2008)	Camera	Multi Scale dynamic features	Gabor filter	Local Binary Pattern	Ada boost	95%

Table 4.3: List of previous works on driver drowsiness detection using behavioral measures.

Ref.	Sensor used	Drowsiness measure	Detection technique	Feature extraction	Classification	Detection rate
(FLORES; ARMINGOL; ESCALERA, 2010)	IR Camera	Eye State	Gabor filter	Condensation algorithm	SVM	93%
(LIU et al., 2010)	Simple Camera	Eye blink	Cascaded Classifiers Algorithm detects face, Diamond searching algorithm	Duration of eyelid closure, No. of continuous blinks, Frequency of eye blink	Region Mark Algorithm	98%
(ZHANG; ZHANG, 2010)	Camera with IR Illuminator	PERCLOS	Haar Algorithm to detect face	Unscented Kalman filter algorithm	SVM	99%

The major problem with the approaches in which their core idea is capturing and analyzing facial behaviors is lighting problem. It means, if we consider camera as the visioning equipment, most of the cameras have visioning problem at night (BERGASA et al., 2006). In order to overcome this problem, in some investigations, an infrared Light Emitting Diode was utilized (BERGASA et al., 2006). These efforts were somewhat effective to eliminate the night visioning problem, however, using LED cameras, it lacks an important feature which is having a good vision during the day light. This defect is because of the fact that the night cameras don't have a proper vision in the day light (HARTLEY; COMMISSION et al., 2000). In addition, the majority of experiments have been done by mimicking the drowsiness, and not a natural drowsiness behavior. Mostly, image is acquired using simple CCD or web camera during day (WEIJIE et al., 2012) and IR camera during night time (FLORES; ARMINGOL; ESCALERA, 2010) at around 30 fps. After capturing the video, some techniques, including Connected Component Analysis, Cascade of Classifiers or Hough Transform, Gabor Filter and Haar algorithm are applied to detect the face, eye or mouth (ZHANG; ZHANG, 2010) (VURAL et al., 2007) (DORAZIO et al., 2007) (FLORES; ARMINGOL; ESCALERA, 2010). After localizing the favorite region in image, the behavioral features, such as yawning, head angle or PERCLOS are extracted using feature extraction methods like wavelet decomposition and discrete wavelet transform (FAN; YIN; SUN, 2009) (LEW et al., 2007) (DORAZIO et al., 2007) (FLORES; ARMINGOL; ESCALERA, 2010). Afterward the features are classified using classifiers including SVM, fuzzy classifier, neural network and LDA (FAN; YIN; SUN, 2009) (DORAZIO et al., 2007) (FLORES; ARMINGOL; ESCALERA, 2010) (BERGASA et al., 2006). The classes with which the classifiers categorized the features, were either normal, slightly drowsy and high drowsy.

The classification rate during the aforesaid classification techniques were highly dependent on the variety of classes which were used. The determination of drowsiness using PERCLOS and Eye Blink has a success rate of close to 100% (BERGASA et al., 2006) and 98% (LIU et al., 2010), however, it is important to be noted that the high detection rate achieved by bergasa and colleagues (BERGASA et al., 2006) was just obtained when the driver didn't wear the glass. Another important issue that is worth mentioning that in majority of researches which hav been done through behavioral approaches, the experiments have been carried out in simulated environment and testing them in the real environment caused having a significant reduction in detection rate (PHILIP et al., 2005b).

4.4 Biosignal measures

When a driver becomes sleepy, his/her head begins to sway or the head's angel changes, the heart beating rhythm gets slower and the vehicle may wander from the center of lane. These are the signs, behavioral measures lay on, but these approaches lack in detecting the sleepiness soon.

Unlike behavioral signs, biosignal-related signs change in the earlier stages of drowsiness. As a result, biosignal measures are more suitable and safer because they detect the sleepiness soon and thereby prevent road accidents.

The following signals have been engaged in many investigations as Biosignals (SAHAYADHAS; SUNDARAJ; MURUGAPPAN, 2012a): a) Electrocardiogram (ECG); b) Electromyogram (EMG); c) Electroencephalogram(EEG) and d) Electro-oculogram (EoG).

Sleep stages are accompanied by some neural changes in brain activity, eye movement and heart rate. Biomedical signals, such as electrocardiography (ECG), electrooculography (EoG) and Electroencephalography (EEG) have been used to measure these changes. EEG has been used in several researches for detecting drowsiness (CORREA; OROSCO; LACIAR, 2014), (LIN et al., 2012). Some researchers investigated on using EoG to identify drowsiness through eye movement (KHUSHABA et al., 2011) (HU; ZHENG, 2009) (AKIN et al., 2008). The electrical potential between retina and cornea generates an electrical field which reflects the eye movement. This electrical field is named EoG signal. Researchers have investigated the horizontal eye movement by placing an electrode in the outer corner of each eye, and another electrode as a reference electrode on forehead. Using these electrodes, the rapid eye movement (REM) and slow eye movement (SEM) (which occur respectively during drowsiness and when an individual is awake) are detected easily (LAL; CRAIG, 2001).

The heart rate (HR) varies significantly through the different stages of cautiousness level (LIANG et al., 2009) (MIYAJI; KAWANAKA; OGURI, 2009), therefore it has been considered as a proper feature in order to measuring the drowsiness. Heart rate can be determined by ECG signals. Also Heart Rate Variability (HRV) has been utilized in many researches. In HRV, low frequencies (LF) and high frequencies(HF) fall in the range of 0.04–0.15 Hz and 0.14–0.4 Hz, respectively (KHUSHABA et al., 2011) (PATEL et al., 2011). The ratio of LF to HF (LF/HF) decrease as an individual progresses from awake to drowsiness stage (PATEL et al., 2011) (YANG; LIN; BHATTACHARYA, 2010). This change can be used for identifying the transient state, happens between wakefulness and sleep.

The Electroencephalogram (EEG) is a electrophysiological feature which has been used for detecting the drowsiness. As it was already mentioned, EEG signals have various frequency bands. Delta band (0.5-4 HZ) corresponds to sleep activity, the Theta band (4-8 HZ) is related to drowsiness, Alpha band(8-13 HZ) represents relaxation and beta band (13-25 HZ) corresponds to alertness (AKIN et al., 2008) (LIN et al., 2012) (LIU; ZHANG; ZHENG, 2010). The Alpha band increases with meditation and decreases with drowsiness. Additionally, the Theta band increases during somnolence. Consequently, researchers have used the changes that happens in brain's frequency bands for classifying different consciousness stages. Akin et al. observed that the success rate of using a combination of EEG and EMG signals to detect drowsiness is higher than using one signal alone (AKIN et al., 2008) .

EEG signals are often noisy (like Blink, eyelid flutter and lateral eyeball movements artifacts) and this unpleasant noisy characteristic sometimes make the data analysis hard, then noisy data should be eliminated(Figure 4.1).

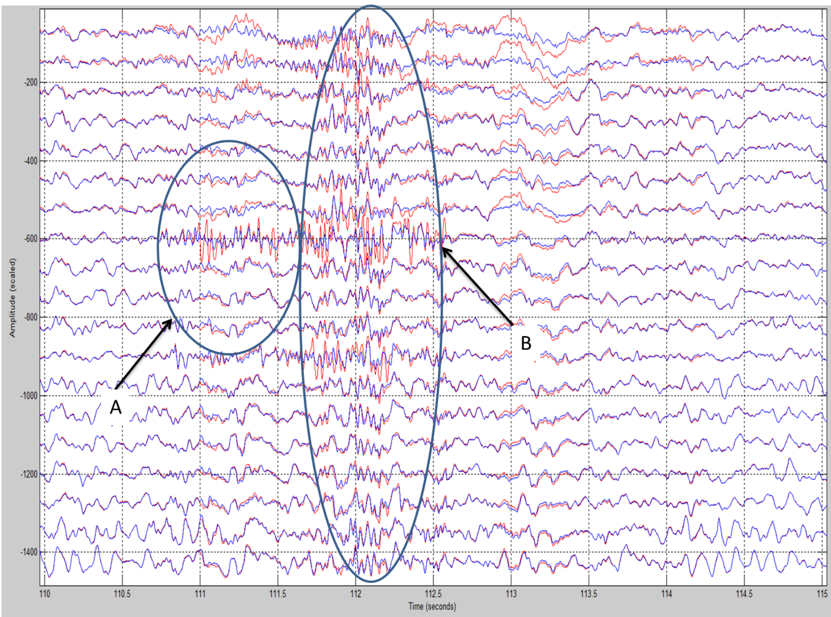

Figure 4.1: EEG signals showed in two points (A and B) before and after noise eliminations. The red signals are noisy and the blue traces indicate the EEG after processing using an artifact removal algorithm (ZEMAN, 2012)

.

Biosignal measures are reliable and accurate, but, they are intrusive and hard to implement in the real environment. This problem can be solved by using single EEG recording which focus on EEG low-cost equipments in order to have a friendly and accurate approach for measuring drowsiness stage.

The advantages and disadvantages of the different type of measures is summarized in Table 4.4.

Table 4.4: Advantages and limitations of various measures.

Measures	Parameters	Advantages	Limitations
Subjective measures	Questionnaire	Subjective	Not possible in real time
Vehicle based measures	Deviation from the lane position Loss of control over the steering wheel movements	Non-intrusive	Unreliable
Behavioral Measures	Yawning, Eye closure, Eye blink, Head pose	Non-intrusive, Ease of use	Lighting, condition, Background
Biosignal measures	Statistical and energy features derived from ECG, EoG and EEG	Reliable, Accurate	Intrusive

Some studies used a combination of biosignals in order to detect sleepiness (see Table4.5). Khushaba et al. developed a fuzzy mutual-information-based wavelet packet transform (FMIWPT) feature-extraction method for classifying the driver drowsiness using 3 EEG channels, EoG signals recorded from left eye and a surface electrode measuring ECG signals. They achieved a classification accuracy of 95%–97% on an average across all subjects (KHUSHABA et al., 2011). Akin et al. (AKIN et al., 2008) used three primary measures to define physiological sleep stages including: (1) The electroencephalogram (EEG) which is popularly known as brain waves; (2) The electromyogram (EMG) which is a record of the electrical activity and emanates from active muscles; (3) The electrooculogram (EoG) which records eye movements. They reached a classification accuracy of 98%–99%. Subasi et al. (SUBASI et al., 2005) used Eog and 5-s long sequences of full-spectrum EEG recordings for classifying alert versus drowsy states in an arbitrary subject. This study used the wavelet-based neural network model trained with the LevenbergMarquardt algorithm to discriminate the alertness level of the subject, and the classification results of this study was 93.3% alert and 96.6% drowsy. Through one of the investigations, concerning drowsiness identification, K-NN has been used alongside a set of psychological measures (3 EEG, 1 ECG and 1 EoG channels) for detecting drowsiness (LIANG et al., 2009) and they achieved a detection rate of 95% correctly classified samples. Kurt et al. (KURT et al., 2009) used 32 EEG, 1 EoG and 1 EMG electrodes for recognizing sleepiness. They used Discrete Wavelet Transform (DWT) for extracting the features from recorded signals. Using ANN they to detect drowsiness by 97% accuracy.

Table 4.5: List of previous works on driver drowsiness detection using biosignal measures.

Ref.	Drowsiness measure	Detection technique	Feature extraction	Classification	Classification accuracy
(KHUSHABA et al., 2011)	EEG, ECG, EoG	Optimal and fuzzy wavelet packet	The Fuzzy MI-based Wavelet-Packet Algorithm	LDA, LIBLINEAR, KNN, SVM	95-97%
(PATEL et al., 2011)	ECG	Band Pass Filter	Fast Fourier Transform (FFT)	Neural Network	90%
(HU; ZHENG, 2009)	EoG, EMG EEG,	Band Pass Filter and Visual Inspection	Discrete Wavelet Transform (DWT)	Artificial Neural Network (ANN) Back Propogation Algorithm (Awake, Drowsy, Sleep)	98-99%
(AKIN et al., 2008)	EEG, EoG	Filter and Visual Inspection	Discrete Wavelet Transform	ANN	97-98%
(LIN et al., 2010)	EEG	Low pass filter 32 Hz	512 point Fast Fourier Transform with 448 point overlap	Mahalanobis distance	88.7%
(LIN et al., 2012)	EEG, EoG, ECG	Independent Component Analysis Decomposition	Fast Fourier Transform	Self-organizing Neural Fuzzy Inference Network	96.7%
	EEG, EoG, EMG	Band Pass Filter	Discrete Wavelet Transform (DWT)	ANN	97%

In general EEG has a high time resolution, is simple in use, cheap and almost does not disturb the subject. EEG can be recorded in any environment without interrupting user's daily tasks. EEG is also a convenient tool for psychophysiological research when the subject has to perform some behavioral tasks or when he/she is out of laboratory (GREEN et al., 1985).

Several studies considered using just EEG records for drowsiness detection. Mervyn et al. used a 32-channel EEG equipment and classified the collected data using support vector machine (SVM) classifier with 99.3% accuracy (YEO et al., 2009). They used the the beta activity in time domain, occipital alpha activity and alpha dropout events as determinitive features for distinguishing between alertness and sleep stage. More recently, some studies (HALICI, 1999) have concentrated on detecting the information on drowsiness available from a full EEG spectrum. Vuckovic et al. (VUCKOVIC et al., 2002) presented a method for classifying alert versus drowsy states from 1-s long sequences of full-spectrum EEG recordings in an arbitrary subject. This method uses time series of interhemispheric and intrahemispheric cross-spectral densities of full-spectrum EEG recordings as the input to an artificial neural network (ANN) with two discrete outputs: drowsy and alert. They achieved a detection accuracy of 94% using 14 electrodes.

Even though acquiring data using EEG equipments with large number of channels or a combination of biomedical signals, leads having a high detection accuracy, but it lacks the ease of use in real life experiments and involves complex mathematical computations (CORREA; OROSCO; LACIAR, 2014). This problem has been addressed in some studies (see Table 4.6 and 4.7). In a work by Ko et al. a single channel wireless EEG device was used for real-time fatigue level detection (KO et al., 2015). Some other investigations also used wireless and Bluetooth technologies (KOBAYASHI, 2012) (KLINGEBERG; SCHILLING, 2012). It is expected that the use of these sensors for estimating the drowsiness becomes a field of interest in the future researches. (LEE; CHUNG, 2012) (SLOTEN et al., 2008).

Table 4.6: List of previous works on driver drowsiness detection using machine learning approach and just EEG signal.

Ref.	Features	Number of electrodes	Classification method	Accuracy
(YEO et al., 2009)	beta activity, occipital alpha activity and alpha dropout events	32 channels	SVM	99%
(VUCKOVIC et al., 2002)	cross-spectral densities	14 channels	ANN	94%
(CORREA; OROSCO; LACIAR, 2014)	Central Frequency (CF), Peak frequency (PF), Ratio H/L, (RH/L), First and Third Quartile Frequency (Q1F and Q3F), Spectral Standard Deviation (SSD)	1 channel	ANN	83%
(ABOALAYON; OCBAGABIR; FAEZIPOUR, 2014)	Entropy, total energy	1 channel	SVM	92.5%
(LIU; ZHANG; ZHENG, 2010)	approximate entropy, Kolmogorov complexity	32 channels	Hidden Markov Model	84%

Table 4.7: List of previous works on driver drowsiness detection using machine learning approach and just EEG signal.

Ref.	Features	Number of electrodes	Classification method	Accuracy
(KIYMIK; AKIN; SUBASI, 2004)	Power Spectrum Density	32 channels	ANN	95%
(SUBASI et al., 2005)	Ratio of absolute mean values of adjacent sub-bands, avarage power of wavelet coefficients and Standard Deviation	1 channels	ANN	93%

As it is shown in Table 4.6 several studies considered using just EEG signals for drowsiness detection. Correa et al. (CORREA; OROSCO; LACIAR, 2014) developed an automatic method to detect drowsiness in single EEG records using time, spectral and wavelet analysis. They selected 19, 13 and 7 parameters to feed an artificial neural network (ANN) and their method achieved 87.4% and 83.6% classification rates for alertness and drowsiness, respectively. In this study they used Central Frequency (CF), Peak frequency (PF), Ratio H/L, (RH/L), First and Third Quartile Frequency (Q1F and Q3F), Spectral Standard Deviation (SSD), Interquartile Range (IR), Maximum Frequency (MaxF), Asymmetry Coefficient, (AC) and Kurtosis Coefficient (KC) as features. Aboalayon et al. (ABOALAYON; OCBAGABIR; FAEZIPOUR, 2014) used EEG sub-bands' features to fed a SVM and achieved 92.5% classification accuracy. Kiymik et al. (KIYMIK; AKIN; SUBASI, 2004) used an equipment with 32 channels for obtaining EEG data. They used power spectrum density as EEG feature and using ANN their method achieved a classification rate of 95%. Subasi et al. (SUBASI et al., 2005) proposed an ANN-based method for classifying drowsiness using ratio of absolute mean values of adjacent, avarage power of wavelet coefficients and Standard Deviation, and their method managed to classify drowsiness by 93% accuracy. In this study the power of signal and entropy in a 60 seconds period were used as EEG features.

4.5 Summary

In this chapter the variety of drowsiness measurements were reviewed. It was mentioned that since measures like behavioral and vehicle-based measures are not always correlated with drowsiness stage, and considering the fact that biosignals are changed in the first moments of drowsiness (or even before the subject become fully drowsy), biosignal measures can be considered as better measures for detecting sleepiness. Several studies which utilized a combination of biosignals were reviewed. It was then told that because EEG has a high time resolution, is simple in use, cheap and almost does not disturb the subject, it is a proper biosignal for measuring conciousness stages. Researches in which just EEG signals were used for drowsiness recognition were reviewed. Finally, it was emphasized that for using automatic drowsiness systems in daily experiments, it is important to use more user-friendly solutions such as EEG equipments with only one electrode. Two studies which have utilized single channel EEG recording were mentioned as well. In the next chapter, the whole research process is explained. In the next chapter, we will talk about our research goals and then we will describe the research process step by step.

Chapter 5

RESEARCH PROCESS

In the previous chapters, the research motivation, theoretical foundations and related researches were explained in detail. In this chapter, firstly the questions this research intends to answer will be enumerated. Afterward the methodology and materials will be explained.

In the current chapter, firstly the research goals will be mentioned one by one. Later, the materials and methods including the data aquisiotion, signal processing and feature extraction, some of our interesting observations about the extracted EEG features related to alertness and drowsiness, and feature selection will be explained. We will also describe why we chose some specific features like mean energy, Standard Deviation and entropy. Finally, we will show an ordered list of the selected features —based on their importance —which carry enough information and allows us for detecting drowsiness with high classification rate.

5.1 Goals

This research aims to improve the previous works regarding the following particular aspects:

1. Investigating the use of single-channel EEG for detecting sleepiness using short EEG epochs.

2. Investigating the use of a simple classification algorithm instead of the statistical approaches (HAL et al., 2014) or sophisticated learning algorithms.

3. Introducing the mean energy of EEG signal as an adequate feature for drowsiness detection.

4. Comparing the classification rate of the well-known classification algorithms such as SVM, logistic regression, decision tree and K-NN in order to find the best classifier using the data obtained in this study.

The specific questions we would like to answer by carrying out this investigation are:

1. Can K-nearest neighbors classify the brain frequencies using single channel EEG recording, as good as when it was used with 5 channels (3 EEG, 1 ECG and 1 EoG channels) (KHUSHABA et al., 2011)?

2. Considering the disadvantages of K-NN, can optimized search and dimension reduction techniques result in a significant improvement in classification time while preserving the accuracy of correctly detected drowsiness instances?

3. How short EEG epochs can be so as to detect drowsiness accurately (in order to use the least possible EEG data)?

From the accuracy and simplicity perspectives, the answers of first question can give us a clue that, instead of using an elaborate classification algorithm like SVM, for drowsiness detection (YEO et al., 2009), we can use an algorithm that not only is fast and simple, but also it may perform better than most of the elaborate algorithms —in this specific research field and with this dataset characteristics. This idea comes from the fact that there are several investigations in which it was shown that in the competition among sophisticated algorithms(such as Neural Network and Support Vector Machine) and simple ones (like Naive Bayes and K-Nearest Neighbors), the clear winner aren't necessarily the more elaborate ones (ASHARI; PARYUDI; TJOA, 2013) (COLAS; BRAZDIL, 2006).

Using dimension reduction techniques alongside the other nearest search algorithms may affect the classification accuracy. Considering that when the size of dataset is large, K-NN is a slow classifier and the fact that using short epochs leads having a big dataset, answering the second question may give a new prespective for the future investigations by finding that whether using K-NN alongside Kd-trees —with a dataset which its dimension has been reduced —has the ability of detecting drowsiness accurately or not —when epochs are very short.

Since detecting drowsiness as soon as possible to avoid car accident is critical, just having a higher classification rate doesn't guarantee having an efficient approach. The system must also detect the drowsiness within a shorter period of time. Consequently, using long time segments (for example, epochs of 60 seconds) cannot be very useful. Using short epochs will result in less calculation. Consequently, we may have a faster drowsiness detection system.

5.2 Methodology & materials

In general, the following tasks have been done during the current study:

(A) Consultancy to neurologists in order to acquiring knowledge about the sleep EEG recording prerequisites aiming to know the best possible approaches for obtaining sleep stages data.

(B) Studying the neurological manuals related to sleep stages such as Rechtschaffen and Kales scores; (RECHTSCHAFFEN; KALES, 1968), hypnograms data, etc.

(C) Programming the system modules such as EDF and hypnogram file reader, random EEG epoch selector from the labeled (awake or drowsiness stages) EEG data, sleep stage extractor, implementing the short-time Fourier transform and Morelet wavelet in order to get time-frequency domain data and implementing feature extraction module.

(D) Final EEG data processing and feature extraction using the obtained EEG data;

(E) Using waikato environment for knowledge analysis(WEKA) and Matlab for classifying the extracted features.

5.2.1 Data acquisition

The Sleep-EDF [Expanded] Database from PhysioNet was used in this study (KEMP; ZWINDERMAN; TUK HAC KAMPHUISEN, 2000). The data sampled at 100 Hz using Fpz-Cz and Pz-Oz channels. The recorded data comes with a hypnogram file that contains sleep patterns corresponding to each subject, scored according to Rechtschaffen & Kales. These patterns consist of sleep stages W, 1, 2, 3, 4, R, M and 'not scored' which are assigned binaries 0, 1, 2, 3, 4, 5, 6 and 9 respectively. Each scored instance in hypnogram file is associated with an EEG epoch with 30 seconds duration. The recordings have been obtained from Caucasian males and females (21-35 years old). The EEG data of 11 subjects recorded from Fpz-Cz (Fpz as a single recording electrode and Cz as reference electrode) channel, were used in order to be processed.

5.2.2 Signal processing and feature extraction

All the signal processing, feature extraction and classification tasks were done using developed scripts in our laboratory, running under Matlab 8.3 (The MathWorks, Inc.) and Weka

3.8, in a computer with 8GB ram, 2.4 GHz Intel Core i5 CPU and Macintosh operating system. Figure.5.1 presents the block diagram of our method.

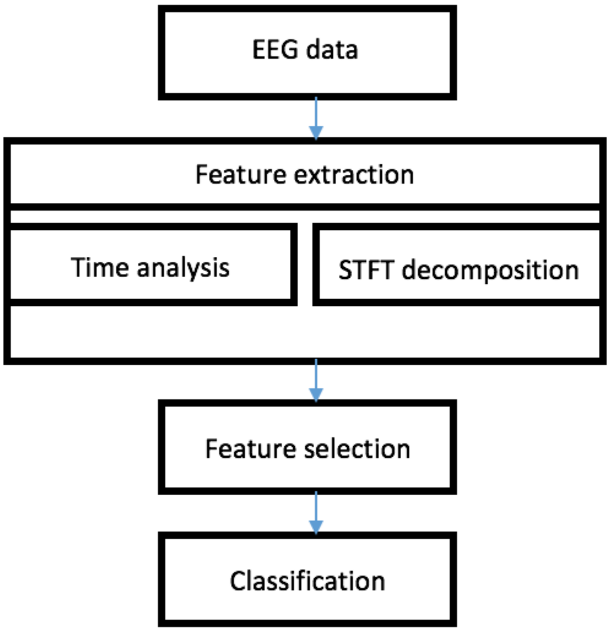

Figure 5.1: The block diagram of the our method

First, 326 sleep epochs, each with 30 seconds duration, labeled with either 'W' (wakefulness stage) or '1' (drowsiness stage), were chosen from database. Then, the EEG data corresponding to each epoch was extracted from EDF files.

In order to obtain the EEG dynamics related to different consciousness stages, EEG subbands consisting of Delta (0.5–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Beta (12–30 Hz) and Gamma (30–50 Hz) were decomposed using short-time Fourier transform. Using STFT, short pieces of signal with window function are extracted and then its frequency representation is computed:

$$F_x(t,f;h) = \int_{-\infty}^{+\infty} x(u)h^*(u-t)e^{-i2\pi uf}du \tag{5.1}$$

where *h(t)* is the sliding analysis window of *STFT*. If the window has a finite energy then, it can be represented as:

$$x(t) = \frac{1}{E_h}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F_x(u,f;h)h(t-u)e^{i2\pi tf}dudf \tag{5.2}$$

where

$$E_h = \int_{-\infty}^{+\infty} |h(t)|^2 dt \tag{5.3}$$

In order to find how EEG data related to drowsiness differ from those ones of wakefulness we calculated the mean energy, maximum energy, minimum energy, Standard Deviation, non-linear energy operator and Entropy. It was found out that for all the randomly selected subjects, the mean energy of majority of frequencies related to Delta, Theta, Beta and Gamma sub-bands of wakefulness is higher than drowsiness (see Figure 5.2 , Figure 5.3, Figure 5.4 and Figure 5.5). Moreover, a higher value of Entropy and Standard Deviation were observed for drowsiness and wakefulness stages respectively (see Figure 5.6 and Figure 5.7).

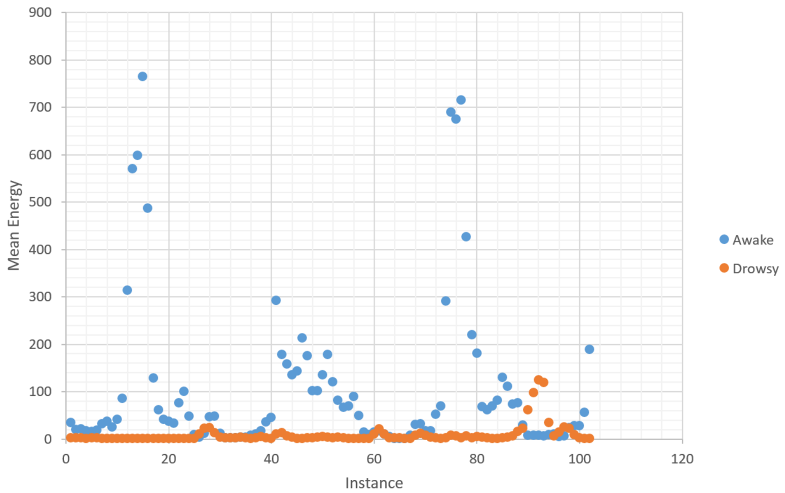

Figure 5.2: The mean energy of EEG sub-bands related to drowsiness and wakefulness of 11 subjects. A higher mean energy of Frequency=1 Hz is observable.

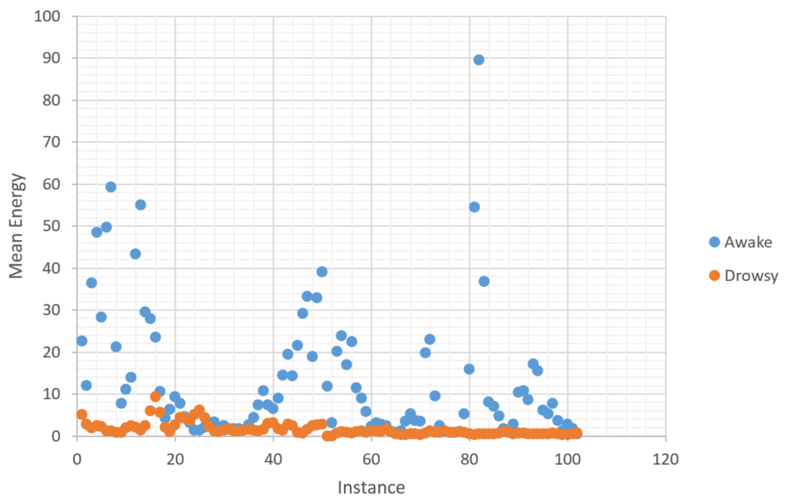

Figure 5.3: The mean energy of EEG sub-bands related to drowsiness and wakefulness of 11 subjects. A higher mean energy of Frequency=6 Hz is observable.

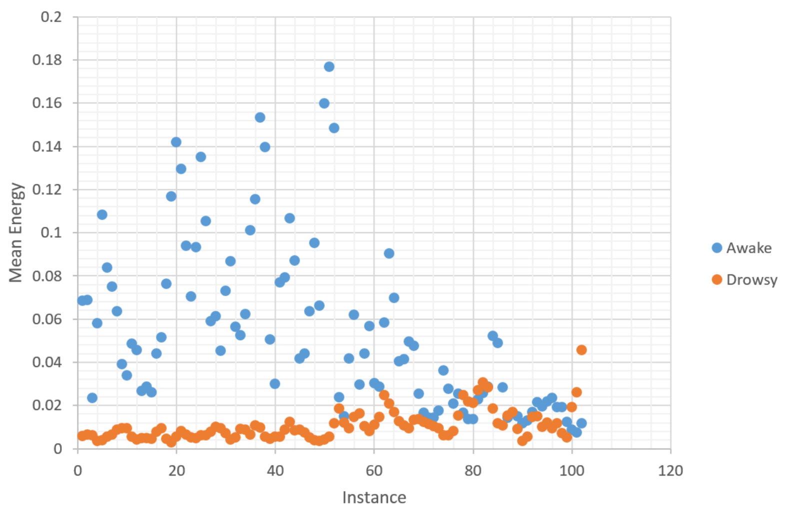

Figure 5.4: The mean energy of EEG sub-bands related to drowsiness and wakefulness of 11 subjects. A higher mean energy of Frequency=49 Hz is observable.

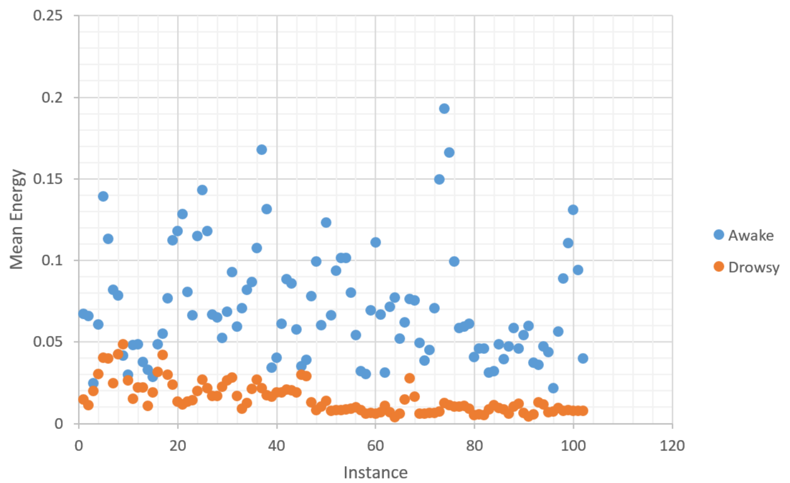

Figure 5.5: The mean energy of EEG sub-bands related to drowsiness and wakefulness of 11 subjects. A higher mean energy of Frequency=50 Hz is observable.

Figures 5.2, 5.3, 5.4 and 5.5 show the mean energy of 102 randomly selected instances from 11 subjects —related to frequencies of F=1 Hz, F=6 Hz, F=49 Hz and F=50 Hz —where the horizontal axis represents the frequency instance and the vertical axis shows the mean energy related to each instance. As it is seen, for the majority of instances a higher mean energy is observable. This implies that even though there are few instances of wakefulness with the same enegy level of drowsiness, the mean energy may be used as an appropriate feature for drowsiness and alertness distiction. In order to obtain even a better accuracy, more appropriate features are supposed to be considered.

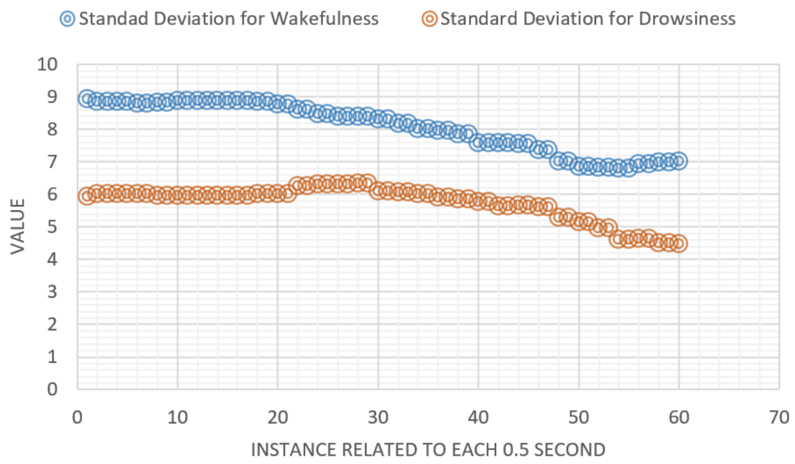

Figure 5.6: The values of Standard Deviation for drowsiness and wakefulness stages.

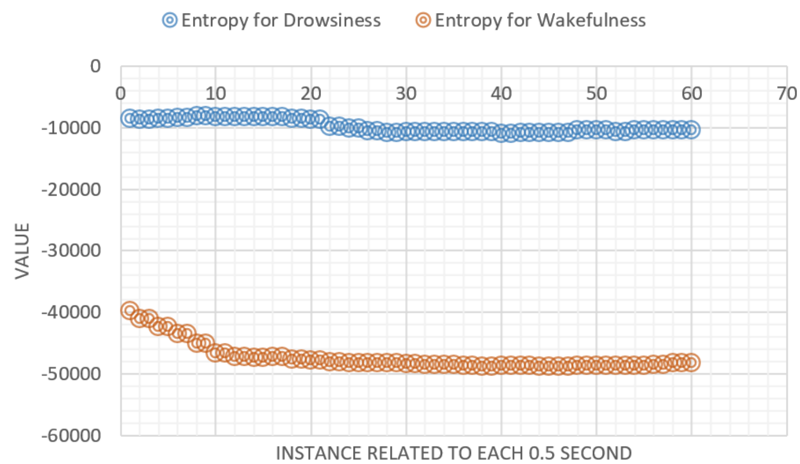

Figure 5.7: The values of Entropy for drowsiness and wakefulness stages.

Figures 5.6 and 5.7 show the values of Standard Deviation and Entropy for wakefulness and drowsiness stages —where the horizontal axis represents the selected instances and the vertical axis shows the value of entropy and Standard Deviation. It is observable that for the randomly selected instances, Standard Deviation and entropy of alertness is higher than drowsiness. Although for the selected instances and for the majority of existing instances in the dataset the Standard Deviation and entropy values have seperable margins, there are some instances with overlapping values for both entropy and Standard Deviation. This means that Standard Deviation and entropy can be considered as adequate features for drowsiness detection although, just utilizing these features do not guarantee a high classification accuracy. Then, using these features alongside mean energy of frequency sub-bands may yield a high classification result.

By that means, each input time-series related to the drowsiness and wakefulness epochs was divided into a sequence of discrete segments of 0.5 second and then, the mean energy was calculated for each frequency band during each of the time segments.

In time-domain, Shanon entropy and Standard Deviation (SD) related to each time-segment were calculated. Entropy is a non-linear measure quantifying the complexity in a time series and in brain computer interface (BCI) systems it can be used to evaluate the level of chaos (PHUNG et al., 2014). If X is a set of finite discrete variables $X = \{x_1, x_2, ..., x_n\}$ and $x_i \varepsilon R^d$, the Shanon entropy, $H(X)$, can explicitly be written as:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i) \tag{5.4}$$

where $P(x_i)$ is the probability that X is in the state x_i.

Standard deviation is a measure that quantifies the amount of variation of a set of data values and is expressed as:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \tag{5.5}$$

where μ is the mean value of population in a set of data.

Finally, 52 features —including 50 features as the mean power of frequency bands between 0.5 and 50 Hz, standard deviation and entropy —containing 17437 instances, obtained from 11 subjects and labeled with either 'Awake' or 'Drowsy', were collected as the final set of data.

5.2.3 Feature selection and classification

Random forest and PCA algorithms were applied over dataset so as to calculate the feature importance. Afterward, the dataset containing selected optimal features (see Table 5.1) was delivered to the classification algorithm. The K-nearest neighbors with kd-trees search algorithm was used for classification task. For measuring the overfitting rate of each algorithm that has been used in this research, we first used the same exisiting data in training set for classification task and then, the data were separated into test set and training set. The train and test sets were randomly chosen from 11 subjects. To do so, the data of 8 subjects (12669 instances) were used for training and data of 3 subjects (4768 instances) were chosen as test sets. Classification accuracy was calculated for K values, $1 \leq K \leq \sqrt{n}$, where n is the number of instances in the dataset (DUDA; HART; STORK, 2001). We also chose the K parameter to be odd numbers in order to avoiding ties (HASSANAT et al., 2014).

Table 5.1: The ordered list of 11 most important features. Each feature is related to time segments of 0.5 second

Feature	Meaning
Entropy	The entropy of signal in time-domain
Standard Deviation	The standard deviation of signal in time-domain
F52	The mean energy for frequency=52 Hz
F49	The mean energy for frequency=49 Hz
F1	The mean energy for frequency=1 Hz
F6	The mean energy for frequency=6 Hz
F5	The mean energy for frequency=5 Hz
F32	The mean energy for frequency=32 Hz
F17	The mean energy for frequency=17 Hz
F9	The mean energy for frequency=9 Hz
F28	The mean energy for frequency=28 Hz

Table 5.1 shows the ordered list of selected features by Random Forest. As it is seen, Entropy is considered as the most informative feature. Standard deviation has been ranked as the second best feature for destinguishing between drowsiness and alertness. It is observable that two frequency bands related to Gamma (F=52 Hz and F=49 Hz), one frequency of Delta (F=1 Hz) and two frequencies of Theta sub-bands (F=5 Hz and F=6 Hz) carry more adequate information to recognize drowsiness. Beta band with three frequencies (F=17 Hz, F=32 Hz and F=28 Hz) and Alpha band with one frequency (F=9 Hz) are the next best features for identifying drowsiness.

5.3 Summary

In the next chapter, the experimental results are explained in details. We will show the classification performance of selected features. Also, it will be shown that how using feature selection and an optimized search algorithm like Kd-trees improves the classification time. Afterwards the details of drowsiness and alertness classification results will be shown and then, they will be compared with the similar studies. Finally, we compare the results of four well-known classification algorithms —including SVM, decision tree, logistic regression and Naive Bayes —with those ones obtained by K-NN as an instance-based approach.

Chapter 6

EXPERIMENTAL RESULTS AND DISCUSSION

In this chapter, the results obtained through this study are discussed compared with the results of the similar works.

We carried out five experiments including demensionality reduction and feature selection, investigating the classification accuracy and classification time, comparing the results with other similar researches —which used single EEG channel —from literature, comparing the classification performance of other four classification algorithms with K-NN, and finally measuring the classification rate across all subjects. In the next sections, each experiment will be described.

6.1 Experiment 1 - Measuring the performance of selected features for drowsiness detection

In this section, we will first show how classification rate is change for each feature dimension. Afterwards, we explain how the selected signal features including Standard Deviation, mean energy of frequency sub-bands and entropy affect the detection accuracy.

6.1.1 Classification accuracy for distinct feature dimension

In order to choose the most informative features, we tested the classification rate for the ranked features for $1 \leq K \leq 127$. The best results were obtained for $K = 3$, $K = 5$ and $K = 7$.

Figure 6.1 presents the classification accuracy for a variety of feature dimensions, when K ranges between 3 and 7. As it is observable, the first 11 features selected by random forest (see Table 5.1) provide the greatest amount of information.

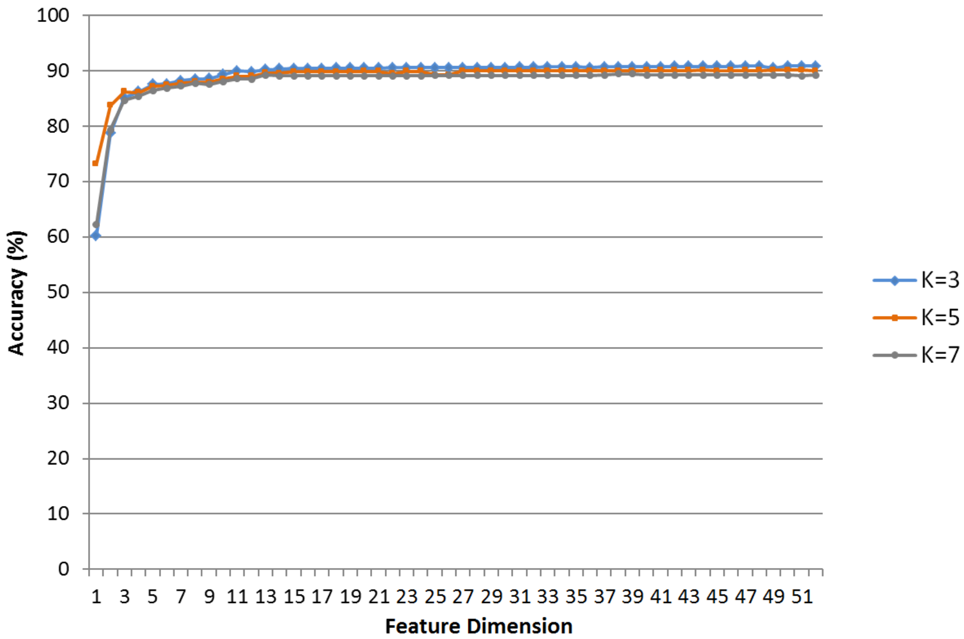

Figure 6.1: The classification performance of feature dimension for correctly classified instances.

Table 6.1: The classification accuracy of drowsiness based on the number of features engaged in classification task.

Feature dimension	Accuracy for K=3	Accuracy for K=5	Accuracy for K=7
1	60.24%	73.2%	62.24%
2	78.75%	83.72%	79.36%
3	85.14%	86.27%	84.69%
4	86.27%	86.08%	85.46%
5	87.57%	87.22%	86.46%
6	87.64	87.44	86.89%
7	88.26%	87.8%	87.25%
8	88.55%	88.01%	88.79%
9	88.6%	87.92%	87.58%
10	89.34%	88.48%	88.07%
11	90%	89.03%	88.57%
12	89.81%	89.036%	88.56%
13	90.22%	89.61%	89.3%
14	90.38%	89.62 %	89.33%
15	90.44%	89.88%	89.12%
52	90.98%	90.01%	89.29%

As it is shown in Table 6.1 and Figure 6.1, when the features were engaged one after another in classification task, the increase in drowsiness identification accuracy is significant for feature dimension between 1 and 11. Although, for feature dimensions more than 11, the increase in classification accuracy is negligible. Consequently, we chose 11 features to be engaged in classification task.

6.1.2 The effectiveness of mean energy, Standard Deviation and entropy

It is worth to mention that using just mean energy values of EEG sub-bands (the first 9 features in Table.5.1), a classification rate of 77% was obtained. Adding Standard Deviation increased the accuracy with 11% (accuracy = 88%). When Entropy was added to the previous features (Standard Deviation and mean energy) the detection accuracy increased by 3% (see Figure) and it reached the classification accuracy of 91%.

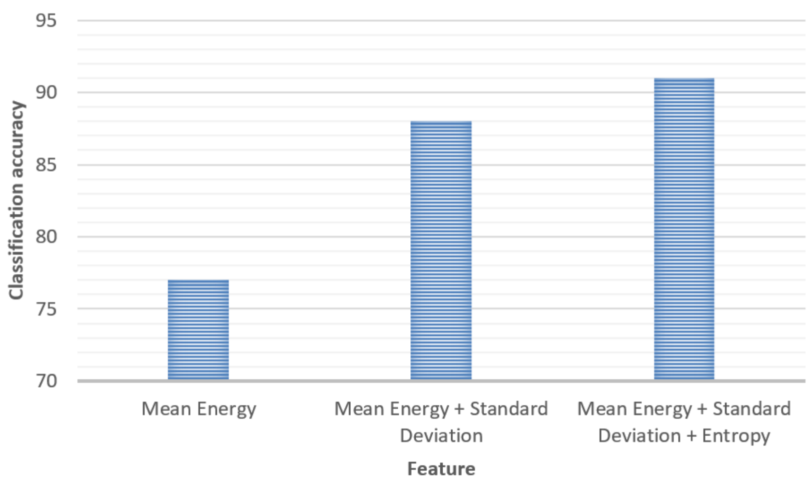

Figure 6.2: The classification accuracy of mean energy, mean energy and Standard Deviation, and mean energy, Standard Deviation and Entropy together.

In this experiment, the classification accuracy of all the features was measured for K=1, K=3 and K=5 (as they yeild the best classification accuracies). It was found out that the first 11 features carry out the most informative data for drowsiness detection. In order to find the effectiveness of each feature —including the mean energy, Standard Deviation and entropy —, we first used features related to mean energy. It was found out that using just the values of mean energy, we can classify drowsiness and alertness by 77% accuracy. In the second step, we added Standard Deviation and as a result the detection rate increased 11%. Finally, by adding entropy as final feature, we reached a classification rate of 91%.

6.2 Experiment 2 - Classification results

In this experiment, the classification performance is measured using the proposed approach and then, it will be compared with the situation when no optimized search and demesion reduction technique is applied. Then, we will take a look at the classification performance after using PCA and Random Forest. Afterwards, the classification results —using the proposed method —for both drowsiness and alertness will be shown. Finally, the detection performance by collecting the classification vote during time segments of 5 seconds will be discussed.

6.2.1 Classification performance with and without using demensionality reduction and optimized nearest neighbors search

In order to find the effectiveness of the techniques which were used for dimensionality reduction and optimizing nearest neighbors search, first both datasets with 11 and 52 features were classified using K-NN with linear search (see Table 6.2). In order to show that there is a slight difference in classification accuracy when more than 11 features are used, we also added the results of dataset with 13 features. Afterwards the classification task was repeated using kd-trees search.

Table 6.2: The classification accuracy and time for different values of feature dimensions, using linear search. Letter *d* represents the feature dimension

K	Correctly classified instances (%)			Classification time(sec)		
	d=52	d=11	d=13	d=52	d=11	d=13
k=3	91%	90%	90.08%	67.94	26.73	26.81
k=5	90.01%	89.03%	89.37%	73.80	27.33	27.37
k=7	89.29%	88.578%	88.62%	77.13	34.17	34.41

Table 6.3: The classification accuracy and time for different values of feature dimensions, using kd-trees search algorithm. Letter *d* represents the feature dimension

K	Correctly classified instances(%)			Classification time(sec)		
	d=52	d=11	d=13	d=52	d=11	d=13
k=3	91%	90%	90.08%	15.90	2.09	2.1
k=5	90.016%	89.03%	89.37%	17.18	2.17	2.24
k=7	89.29%	88.57%	88.62%	17.67	2.29	2.31

6.2.2 Classification performance using the dataset provided by Random Forest and PCA

As it is shown in Table 6.3, using 11 features provided by random forest improves the classification time significantly, without having a tangible decrease in classification accuracy. Since PCA chooses the best values related to attributes (14153 values were chosen out of 17437 instances) and reduces the dataset size, it helped for classification time reduction although, the classification accuracy decreased -6% (classification accuracy of 86%), which can be considered significant in road security domain (see Table 6.4).

Table 6.4: The accuracy and classification time after using Random Forest and PCA

Feature selection algorithm	Accuracy after applying	Classification time (sec)
Random forest	91%	2.1
PCA	86%	1.9

As it is seen in Table 6.4, using PCA the classification time was decreased slightly by -0.3 second, however the classification accuracy is reduced as well. Thereby, we chose the features selected by Random Forest without decreasing the number of instances.

6.2.3 Alertness and drowsiness classification

The final classification results for both drowsiness and wakefulness are shown in details in Table.6.5. The overall accuracy, sensitivity and precision of each sleep stage along with the confusion matrix of the epochs correctly and incorrectly classified can be calculated through the following equations:

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{6.1}$$

$$Precision = \frac{t_p}{t_p + f_p} \tag{6.2}$$

$$Sensitivity = \frac{t_p}{t_p + f_n} \tag{6.3}$$

Where t_p, t_n, f_p, f_n represent true positive, true negative, false positive and false negative, respectively. Sensitivity, Specificity, and Accuracy are the terms which are most commonly associated with a Binary classification test and they statistically measure the performance of the test. In a binary classification, we divide a given data set into two categories on the basis of whether they have common properties or not by identifying their significance and in a binary classification test, as the name itself conveys, we deal with two datasets. Of these two categories, in general, Sensitivity indicates, how well the test predicts one category and Specificity measures how well the test predicts the other category. Whereas Accuracy is expected to measure how well the test predicts both categories.

Table 6.5: Classification accuracy, precision and sensitivity for both drowsiness and wakefulness stages

K	Accuracy(%)		Precision		Sensitivity	
	Drowsy	Awake	Drowsy	Awake	Drowsy	Awake
k=3	91%	89%	0.894	0.906	0.909	0.891
k=5	90%	88%	0.884	0.9	0.901	0.88
k=7	90%	87%	0.876	0.896	0.901	0.886

As it is shown in Table 6.5, the best drowsiness and wakefulness accuracy was obtained when $K = 3$ with 91% and 89% correctly classified instances related to drowsiness and wakefulness, respectively. For $K = 5$ and $K = 7$ the accuracy of correctly classified drowsiness instances is equal, however the classification rate of instances related to wakefulness is decreased by -1% when $K = 7$.

6.2.4 Drowsiness detection using the majority vote

We also collected the vote of 10 classifications results during each 5 seconds and then used the majority of votes for final decision (see Figure 6.3).

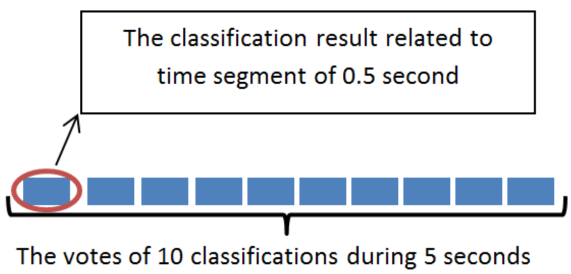

Figure 6.3: The votes of 10 classifications during 5 seconds are gathered and the majority vote is considered as the conciousness state during each specific time period.

As it is seen in Figure 6.3, each blue piece represents the classification vote related to time segment of 0.5 second. These votes are collected and the majority vote during each 5 seconds (of 10 votes) is considered as the corresponding conciousness state. Using this approach we had an average detection accuracy of 92.68% over all instances in dataset.

6.2.5 Effectiveness of proposed approach by means of accuracy and classification time

In order to find how effective the feature selection and Kd-tree search algorithm were in classification time reduction, we compared the classification time of proposed approach (using Kd-trees and optimal features) with the non-optimal search and non-optimal features. Figure 6.4 shows the classification time using linear and Kd-tree searches for the dataset before and after dimension reduction.

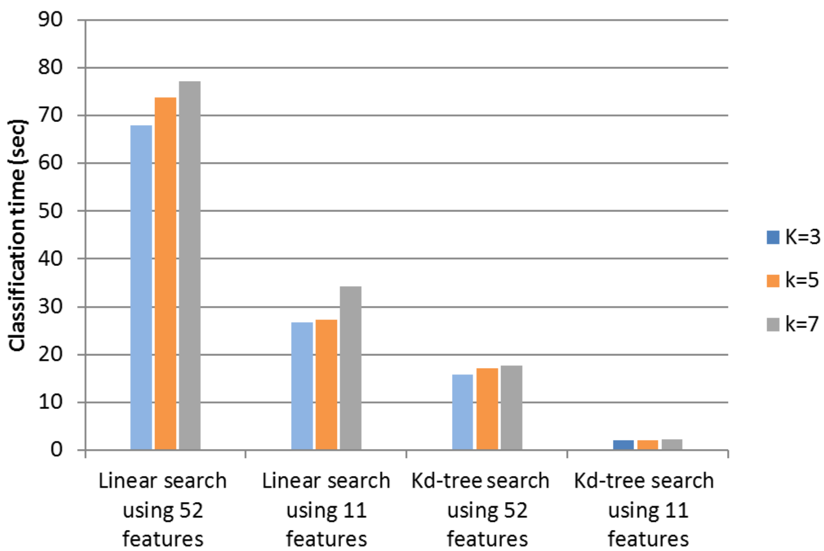

Figure 6.4: K-NN classification time using linear and kd-trees search

Decreasing dimensionality by choosing the most informative features and using an appropriate search algorithm, improves the classification time significantly while it preserves the classification accuracy.

In this section, the classification accuracy and classification time of datasets before and after using dimensionality reduction and optimized search was discussed. It was found out that by reducing dimension based on the selected features and using Kd-tree as an efficient search algorrithm, the classification time is reduced significantly while preserving the classification rate. By measuring the Random Forest and PCA classification performance, since a significant decrease in classification rate was observed, we decided to choose the selected features by Random Forest. We also used time segments of 5 seconds in order to improve classification accuracy by collecting the votes of 10 classification result and choosing the vote of majority decisions during 5 seconds —as a result of 10 classifications —as the conciousness state during an specific period of time.

6.3 Experiment 3 - Comparing results with others from the literature

Although our approach succeeded to detect drowsiness with a high accuracy, it is important to draw an analogy between the effectiveness of our approach with other studies. In order to know how effective our approach is, we compared our results with those ones obtained in other works where machine learning methods were engaged in sleepiness detection by using single EEG recording. An efficiency comparison of the current approach versus the preliminary proposals is shown in Table 6.6.

Table 6.6: An efficiency comparison of the current approach and the preliminary proposals that used single EEG electrode

Study	Classifier	Accuracy (%)	Epoch	EEG samples
Correa et al. (CORREA; OROSCO; LACIAR, 2014)	ANN	83.6%	5 sec	1375
Aboalayon et al. (ABOALAYON; OCBAGABIR; FAEZIPOUR, 2014)	SVM	92.5%	60 sec	6000
Subasi et al. (SUBASI et al., 2005)	ANN	93%	5 sec	5000
The current study	K-NN	91%	0.5 sec	50
The current study	K-NN	92.68%	5 sec	500

In comparison with the study carried out by Correa et al. that proposed using artificial neural network with 19, 13 and 7 features and achieved 83.6% and 87.4% of drowsiness and alertness correct detection, respectively (CORREA; OROSCO; LACIAR, 2014), the current approach achieves a better classification accuracy with 91% and 89% correctly classified instances for drowsiness and alertness, respectively. Nevertheless, each instance in their dataset is related to the time segments of 5 seconds (sampling rate= 275 Hz), which is longer than the time duration used in this work (0.5 second). The SVM classifier proposed by Aboalayon et al. (ABOALAYON; OCBAGABIR; FAEZIPOUR, 2014) —which uses time segments of 60 seconds with sampling rate of 100 Hz —has +1.5% better classification rate than the proposed approach in this paper, however, a simpler classifier and shorter epochs were used in the current study (see Table.6.6). Additionally, by considering the classification results of our approach using majority votes in time periods of 5 seconds, our approach achieves an even better classification rate by 92.68% in comparison with that study. Finally, in comparison with the results obtained by Khushaba et al. (KHUSHABA et al., 2011) —that uses three EEG, one ECG and one EoG channels and K-NN classifier —our mathod has -4% less accuracy, although we use just one electrode for acquiring EEG data. The method proposed by Subasi et al. (SUBASI et al., 2005) —which uses ANN as classifier —with 93% has a slightly better classification rate than our approach, although they have used 5000 EEG samples (time segments of 5 seconds with sampling rate of 1000 Hz) for identifying drowsiness.

In this section, a comparison of our results with those ones obtained in similar studies was presented. Because the proposed approach calculates the EEG features based on the information obtained from short time-segments (0.5 second with sampling rate of 100 Hz) then, less information is needed to decide whether the subject is drowsy or not. As a result, the proposed method offers a high classification accuracy using the data obtained from reasonably short EEG epochs, so that drowsiness can be detected early enough in order to avoid car accidents.

6.4 Experiment 4 - Comparing the classification results of K-NN with four other classification algorithms

In order to support the effectiveness of the current method and comparing the classification accuracy, we repeated the classification task using four well-known algorithms: a) SVM b) decision tree; c) logistic regression and d) Naive Bayes (see Table 6.7 and Figure 6.5). For all the classification algorithms, the dataset was devided into train set and test set. The train set contains the data of 8 subjects with 12669 and the test set contains the data of 3 subjects with 12669 instances. In the this section the experiment of each classifier and results will be explained. In order to measuring overfitting rate, the classification was also done using train set. In the next sub-sections, the results which have been shown in Table 6.7 and Figure 6.5 will be explained.

Table 6.7: A comparison of accuracy, overfitting rate and classification time of five well-known classification algorithms

Classification algorithm	Accuracy	Overfitting	Classification time (sec)
SVM	86%	1.1%	1.6
Decision tree	88%	8%	1.6
Logistic regression	80%	0.25%	1.9
K-NN	91%	7.4%	2.1
Naive Bayes	67%	2.4%	0.8

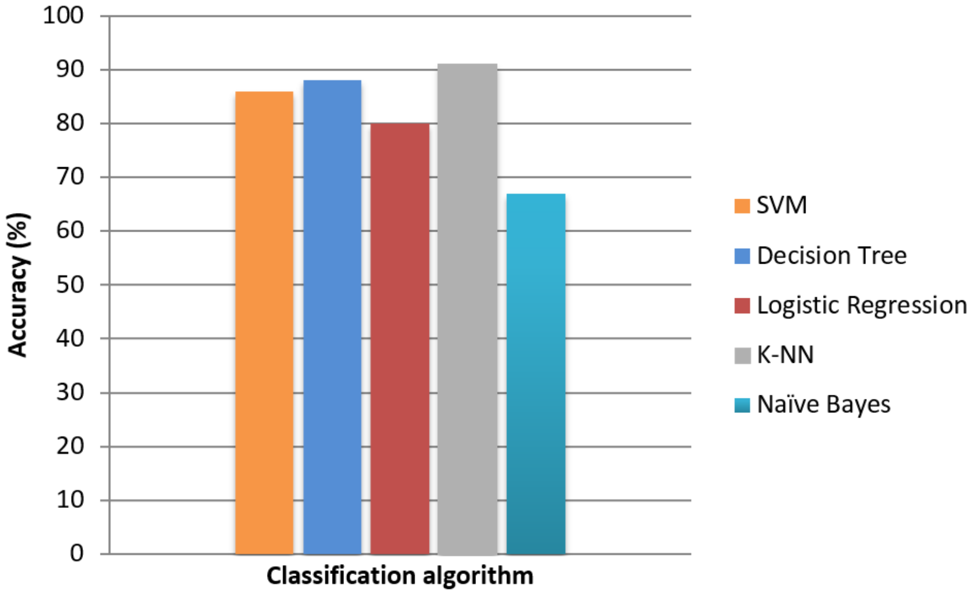

Figure 6.5: An accuracy comparison between five different classifiers using the data obtained through the proposed approach.

6.4.1 SVM

SVM with the following parameters was used to be trained by train set and to classify the test set: a) Pearson VII function-based universal kernel (as it yielded the best result); b) the cost parameter $C : C\varepsilon\{10^{-5}, 10^{-4}, ..., 10^{2}\}$, c) random seed equals to 1; and d) tolerance parameter=0.001. SVM obtains the classification rate of 86% , however, it is worth mentioning that it took nearly 8 hours for training however, after training, SVM could classify the test set in 1.6 seconds. This algorithm had an overfitting rate of 1.1%.

6.4.2 Decision tree

Decision tree (C4.5 as an extension of Quinlan's earlier ID3 algorithm) was engaged in classification task being set to the following parameters: a) prinnuing set to be true; b) number of folds equals to 3 and c) confidence factor equals to 0.25. Decision tree managed to classify the test set with 88% correctly classified instances. It also had the highest overfitting rate with +8% of difference in detection accuracy when the trainset was engaged in classification process. Moreover, it takes 14 seconds for classifying data using decision tree.

6.4.3 Logistic regression

Logistic regression with the following parameters was used to classify test set after being trained by train set: a) huristicStop equals to 50; b) max boosting iterations set to be 500; and c) weight trim beta equals to 0. Logistic regression achieved a classification rate of 80%. Notstandingwith the lower classification rate of logistic regression in comparison with SVM and decision tree, it had the least overfitting problem with only +0.25% better classification rate when the same training data was used for classification. The classification time of logistic regression was 1.9 seconds.

6.4.4 Naive Baye

Naive Bayes was train with the data of 8 subjects (the rain set) and then it was used to classify the data existing in test set. This algorithm achieved a classification accuracy of 67% which is the worst classification accuracy between the other methods. This might be caused by the fact that Naive Bayes assumes that attributes are independent given a specific class, while in the current problem EEG features vary dependently. Naive bayes reached a classification time of 0.8 second which is the best time between other algorithms studied in this research. It also had 2.4% overfitting rate.

In this section, the classification performance of SVM, decision tree, logistic regression and Naive Bayes was investigated. Based on the aforementioned results, we can see that K-NN —with classification rate of 91% and classification time of 2.1 seconds —with 7.4% has the second high overfitting rate. Based on the data illustrated in Table 6.7, it is seen that K-NN has the worst classification time between the other algorithms, although the time difference is little. It is also revealed that even though the training time of SVM is too long, it is fairly fast in classifying the test set. In the next section, the data related to each subject will be classified independently so as to examine the possibility of developing user-adapted drowsy detection system in the future works.

6.5 Experiment 5 - Classifying data of each subjects independently

Even though the proposed generalized pattern can help predicting the future coming data related to drowsiness for other subjects, adapting system for specific subjects may yeild even better results that may be used in the future works. In this section classification results related to each subject are discussed.

In order to find the detection performance using the proposed approach, we tested the classification results over all subjects. The detection accuracy, classification time and number of instances related to each subject, and classification accuracy for subject-dependent and subject-independent models with and without using the proposed majority vote method, have been illustrated in Table 6.8 and Figure 6.6.

Table 6.8: The subject-dependent classification accuracy for 11 subjects for K-NN where K=3

Subject	Accuracy	Classification time (sec)	Number of instances
1	98.7%	0.3	1857
2	98.14%	0.2	1617
3	98.96%	0.4	2514
4	96.47%	0.2	1732
5	99.09%	0.3	1555
6	94.37%	0.2	1370
7	95.1%	0.2	1431
8	99.3%	0.2	1332
9	98.11%	0.3	1376
10	99.36%	0.2	1254
11	97.2%	0.3	1399
Average:	**97.7%**		

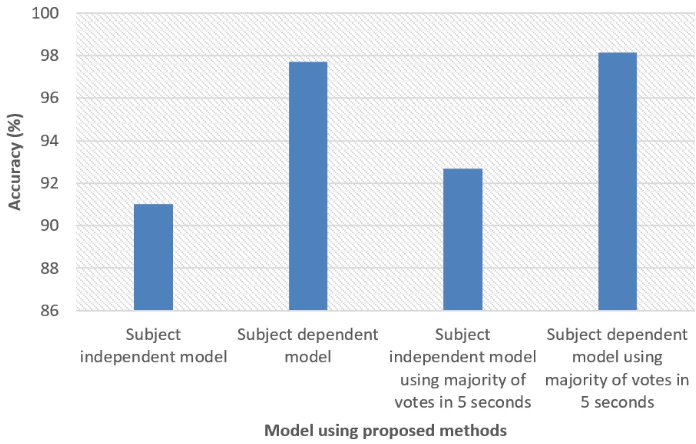

Figure 6.6: An accuracy comparison between subject-dependent and subject-independent models using and without using majority votes

Table 6.8, illustrates the classification accuracy, classification time and the number of instances which exist in dataset of each subjects. As it is observable, in comparison with subject-independent results which have been explained in the previous sections, the subject-dependent classification accuracy for both single and average (97.7%) accuracy over all the 11 subjects is higher. Furthermore, when the majority votes were used during time segments of 5 seconds, a classification accuracy of 98.2% was achieved in average over all subjects. This suggests that if we customize the system and train the classification algorithm with data obtained from a specific subject, we may have an even higher detection accuracy and consequently, an even more accurate and faster automatic drowsiness detection system.

In this section, we compared the results obtained in our study with those ones acquired through the similar studies which used single-EEG recording beside machine learning approach for identifying drowsiness. Using a simple instance-based classification algorithm, features that are easy to be calculated and much less EEG samples, our approach achieves an acceptable sleepiness detection accuarcy in comparison with similar studies.

6.6 Summary

In this chapter, the experimental results of this study were explained in details. Firstly, we measured the effectuality of each features. It was perceived that the first 11 features selected ranked by Random Forest carry enough information for detecting sleepiness with 91% accuracy. After selecting the most informative features, we compared the accyracy and classification time with and without using dimencionality reduction and optimized search algorithm. It was revealed that using the proposed method, the classification time of K-NN is decreased significantly while preserving the accuracy. We also proposed getting the votes of classification results during a time period of 5 seconds and using the majority of votes for final decision about the subject's consiousness stage. Additionally, the classification performance of five well-known (but less used in drowsiness detection using EEG) algorithms compared. Finally, the subject-dependent classification accuracy was computed over 11 subjects. The result show that with average accuracy of 97.7%, we had an even better result when system is adapted with each individual. In the next chapter a conclusion of the achieved results is presented.

Chapter 7

CONCLUSION

In this chapter a resumed review of activities carried out during this study, the conclusion we came to through this work, papers published based on this work and the statement of future works are presented.

Using a fast, accurate and simple classification algorithm is crucial in drowsiness detection systems. We assumed that since some of the EEG characteristics related to drowsiness and alertness share distinct areas in feature space, consequently, they can be classified by K-NN (as an instance-based method) with a high accuracy. More specifically, we chose instance-based approach for its advantage over other methods in terms of its flexibility of inductive bias, easiness of the choice of learning parameters and its simlicity. Since using brute-force searching is inefficient, Kd-tree was used as nearest neighbors search aiming to reduce the search time. As it was mentioned, the query time of Kd-tree grows exponentially when the dimensionality is increased. Thereby, we decided to reduce the dimensionality in order to use the most informative set of features aiming at preventing search inefficiency issues.

As it was explained, when EEG equipments with large number of electrodes were used in previous researches the discrimination rate was very high, although these kinds of EEG caps are not usable in day to day life because they have three main defects. The first shorcoming of these equipments is that they are not affordable. The second drawback concerns the configuration complexity. Being intrusive is the third shortcoming which prevents the ease of use in daily experimets. For avoiding these defects, the data of a single EEG electrode was gathered from a publicly available EEG dataset. Then, using the simplest source of EEG data (just one electrode) while classifying data by a simple classification algorithm, was the most challenging task of this work. More specifically, since we insisted on both simplicity of use and implementation, it was important that our approach has also an acceptable performance in comparison with preliminary proposed methods in state of art.

in order to find which signal features are more adequate for sleepiness detection, we calculated the mean energy of Delta, Theta, Alpha, Beta and Gamma sub-bands from frequency-time domain. In the next step, we also calculated the values of Standard Deviation and Entropy from time-domain. A higher mean energy of Delta, Theta and Gamma bands related to wakefulness was observed in comparison with drowsiness sub-bands. This results are consistent with those features that Random Forest selected as the most informative features. Moreover, a higher value of Entropy and Standard Deviation were observed for drowsiness and wakefulness stage respectively. Thus, the mean energy, Standard Deviation and entropy were used as a deteminitive features for drowsiness detection. We observed that by using just sub-bands' mean energy, it was possible to classify drowsiness by 77%. Also adding entropy and Standard Deviation increased the detection accuracy by 14%. Eventually, a number of 326 sleep EEG epochs obtained from 11 subjects were processed and finally 11 features and 17437 instances were extracted. Each instance is related to a time-segment of 0.5 second. Because the time segments are short, the decision making process can be done within a shorter period of time and with less calculation. The features are also easy to calculate and can be obtained in real time. Random forest and kd-trees were used to decrease dimensionality and to increase the search speed, respectively. Drowsiness and alertness were detected efficiently, with 91% and 89% accuracies, respectively. Our findings show that K-NN along with an adequate search algorithm —considering the size and dimension of dataset —like kd-tress, can be used as an accurate and fast classifier for automatic drowsiness detection.

Since this research is about road safety, we took some aditional steps aiming to achieve even better results. First, we proposed collecting 10 classification results during time periods of 5 seconds and using the vote of majority classification results and we reached a better result with 92.68% accuracy. Moreover, the classification task —using the proposed approach which suggest using specific features alongside an instance-based algorithm with optimized search method —for each subject. The classification rate for each subject and the average rate over all subjects (97.7%) shows that if the system is costumized for each user, we may have the most accurate results while using data from single EEG electrode. By applying the majority votes during time segmants of 5 seconds, our approach yeild an even better result by 98.2% accuracy in average over all subjects.

Finally, a Matlab library was developed with the purpose of helping the other researchers being able to extrcat and analyse data from Sleep-EDF [Expanded] database (or any other recorded sleep database based on Rechtschaffen & Kales scoring), readily. This framework containes stationary wavelet denoising classes for the artifactual recorded data and two different time-frequency decomposers including Morlet wavelet and short-time Fourier transform.

7.1 Contributions

The major comtributions of this study are the followings:

1. Significantly decreasing the required EEG epoch for deciding about the sleep stage. As it was described, using EEG epoch of 0.5 (in other words, 50 EEG samples considering that the sampling rate is 100 Hz), leads having less calculation and detecting drowsiness faster;

2. Introducing the mean energy as an adequate EEG characteristic for drowsiness detection;

3. Classifying the drowsiness EEG data recorded through a single-channel EEG, with a high accuracy by using K-NN as a simple classification algorithm. Using simple classification algorithm is crucial since it is easily implementable in simple devices like microcontrollers;

4. Comparing the classification rate of five well-known algorithms including K-NN, SVM, decision tree, logistic regression and Naive bayes;

5. Developing an open-source library for extracting the EDF and hypnogram data, denoising the recorded EEG signals —if it is required, spectral analysis and frequency decomposition using Morlet Wavelet and short-time Fourier transform and feature extraction using random forest algorithm. There are many aspects in sleep research that require standardization, however, the toolbox does provide a good starting point for, at least, using the signals from a popular database in a uniform manner. By making the code open source, and available on GitHub, it is hoped that other researchers can contribute to it by adding more functionality to the toolbox;

7.2 Publication

The results of this study have been published in 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'16) (Qualis = B1).

7.3 Future works

The future development of this research will be focused on using a more accurate spectral decomposer —for example, Wavelet transform (GIURGIUTIU; YU, 2003) —and finding better features, in order to reach an even better classification rate. Also, in the future works, adaptable drowsy detection systems may be developed. Since drowsiness detection for single subject yields more accurate results, a system which is trained by data of a specific user may guarantee the driver safety through a more accurate detection. Lately single-EEG Bluetooth headsets like Neurosky (FIOLET, 2011) have been used in some BCI researches. Unfortunately these headsets are very noisy and are sensitive to artifacts caused by muscular movement. As a result, if these headsets are used for data acquisition, an extra step ough to be added to the procedure for removing artifacts. This additional phase may impact the whole drowsiness detection time negatively. Thus, it seems necessary to apply the current method over data obtained using this type of EEG equipment through an online experiment in order to evaluate its performance and to tackle the possible obstacles. Finally, we intend to extend the functionality of the developed Matlab toolbox by adding a variety of Wavelet decomposers and also, improving the performance issues.

REFERENCES

ABE, T.; NONOMURA, T.; KOMADA, Y.; ASAOKA, S.; SASAI, T.; UENO, A.; INOUE, Y. Detecting deteriorated vigilance using percentage of eyelid closure time during behavioral maintenance of wakefulness tests. *International Journal of Psychophysiology*, Elsevier, v. 82, n. 3, p. 269–274, 2011.

ABOALAYON, K. A.; OCBAGABIR, H. T.; FAEZIPOUR, M. Efficient sleep stage classification based on eeg signals. In: IEEE. *Systems, Applications and Technology Conference (LISAT), 2014 IEEE Long Island*. [S.l.], 2014. p. 1–6.

ACHERMANN, P. Eeg analysis applied to sleep. *Epileptologie*, Citeseer, v. 26, p. 28–33, 2009.

AGGARWAL, C. C. *Instance-Based Learning: A Survey.* 2014.

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. *Machine learning*, Springer, v. 6, n. 1, p. 37–66, 1991.

AKIN, M.; KURT, M. B.; SEZGIN, N.; BAYRAM, M. Estimating vigilance level by using eeg and emg signals. *Neural Computing and Applications*, Springer, v. 17, n. 3, p. 227–236, 2008.

ARYA, S.; MOUNT, D. M.; NETANYAHU, N. S.; SILVERMAN, R.; WU, A. Y. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM*, ACM, New York, NY, USA, v. 45, n. 6, p. 891–923, nov. 1998. ISSN 0004-5411. Disponível em: ⟨http://doi.acm.org/10.1145/293347.293348⟩.

ASHARI, A.; PARYUDI, I.; TJOA, A. M. Performance comparison between naive bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. *International Journal of Advanced Computer Science and Applications*, Citeseer, v. 4, n. 11, 2013.

ATKESON, C. G. Using local models to control movement. In: *NIPS*. [S.l.: s.n.], 1989. p. 316–323.

BAEK, H. J.; CHUNG, G. S.; KIM, K. K.; PARK, K. S. A smart health monitoring chair for nonintrusive measurement of biological signals. *IEEE transactions on Information Technology in Biomedicine*, IEEE, v. 16, n. 1, p. 150–158, 2012.

BERGASA, L. M.; NUEVO, J.; SOTELO, M. A.; BAREA, R.; LOPEZ, M. E. Real-time system for monitoring driver vigilance. *Intelligent Transportation Systems, IEEE Transactions on*, IEEE, v. 7, n. 1, p. 63–77, 2006.

BERGER, H. ber das elektrenkephalogramm des menschen - archive fr psychiatre und nervenkrankheiten. 1929.

BERGEY, G. E.; SQUIRES, R. D.; SIPPLE, W. C. Electrocardiogram recording with pasteless electrodes. *Biomedical Engineering, IEEE Transactions on*, IEEE, n. 3, p. 206–211, 1971.

BHATIA, N. et al. Survey of nearest neighbor techniques. *arXiv preprint arXiv:1007.0085*, 2010.

BIRBAUMER, N. Breaking the silence: brain–computer interfaces (bci) for communication and motor control. *Psychophysiology*, Wiley Online Library, v. 43, n. 6, p. 517–532, 2006.

BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996.

BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.

BURGES, C. J.; SCHLKOPF, B. Improving the accuracy and speed of support vector machines. In: *Advances in Neural Information Processing Systems 9*. [S.l.]: MIT Press, 1997. p. 375–381.

BURKHARD, W. A.; KELLER, R. M. Some approaches to best-match file searching. *Communications of the ACM*, ACM, v. 16, n. 4, p. 230–236, 1973.

CANTERO, J. L.; ATIENZA, M.; SALAS, R. M. Human alpha oscillations in wakefulness, drowsiness period, and rem sleep: different electroencephalographic phenomena within the alpha band. *Neurophysiologie Clinique/Clinical Neurophysiology*, Elsevier, v. 32, n. 1, p. 54–71, 2002.

CANTERO, J. L.; ATIENZA, M.; SALAS, R. M.; GÓMEZ, C. M. Brain spatial microstates of human spontaneous alpha activity in relaxed wakefulness, drowsiness period, and rem sleep. *Brain Topography*, Springer, v. 11, n. 4, p. 257–263, 1999.

CHELLA, K. Event-related oscillations in alcoholism research: A review. *Journal of Addiction Research & Therapy*, 2012.

CHOU, H.-C.; PRATAKSITA, N.; LIN, Y.-T.; KUO, C.-H. P300 and motor imagery based brain-computer interface for controlling wheelchairs. *Journal of Medical Devices*, American Society of Mechanical Engineers, v. 8, n. 3, p. 030906, 2014.

COHEN, M. X. *Analyzing neural time series data: theory and practice*. [S.l.]: MIT Press, 2014.

COLAS, F.; BRAZDIL, P. *Comparison of SVM and some older classification algorithms in text classification tasks*. [S.l.]: In Artificial Intelligence in Theory and Practice, 2006. 169–178 p.

COLTEN, H. R.; ALTEVOGT, B. M. et al. Committee on sleep medicine and research. National Academies Press (US), 2006.

CORREA, A. G.; OROSCO, L.; LACIAR, E. Automatic detection of drowsiness in eeg records based on multimodal analysis. *Medical engineering & physics*, Elsevier, v. 36, n. 2, p. 244–249, 2014.

DAS, D.; ZHOU, S.; LEE, J. D. Differentiating alcohol-induced driving behavior using steering wheel signals. *Intelligent Transportation Systems, IEEE Transactions on*, IEEE, v. 13, n. 3, p. 1355–1368, 2012.

DELGADO, J. M. R. *Delgado Physical Control of the Mind, Toward a Psychocivilized Society Harper and Row*. [S.l.]: New York, 1969.

DENG, K.; MOORE, A. W. Multiresolution instance-based learning. In: *IJCAI*. [S.l.: s.n.], 1995. v. 95, p. 1233–1239.

DEOLE, P. A.; LONGADGE, R. Content based image retrieval using color feature extraction with knn classification. *IJCSMC*, v. 3, n. 5, p. 1274–80, 2014.

DEVROYE, L. On the equality of Cover and Hart in nearest neighbor discrimination. *IEEE Trans. Pattern Anal. Mach. lntell*, v. 3, p. 75–78, 1981.

DINGES, D. F. . M.; MAISLIN, M. M. .; POWELL, G. .; W., J. *Final Report: Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and as the Basis for Alertness Management; Report for NHTSA: Washington, DC, USA*. [S.l.: s.n.], 1998.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. [S.l.]: Wiley, 2001.

DUDANI, S. A. The distance-weighted k-nearest-neighbor rule. *Systems, Man and Cybernetics, IEEE Transactions on*, IEEE, n. 4, p. 325–327, 1976.

DORAZIO, T.; LEO, M.; GUARAGNELLA, C.; DISTANTE, A. A visual approach for driver inattention detection. *Pattern Recognition*, Elsevier, v. 40, n. 8, p. 2341–2355, 2007.

EBRAHIMPOUR, H.; KOUZANI, A. Face recognition using bagging knn. In: *International Conference on Signal Processing and Communication Systems (ICSPCS2007) Australia, Gold Coast*. [S.l.: s.n.], 2007. p. 17–19.

FAIRCLOUGH, S. H.; GRAHAM, R. Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Sage Publications, v. 41, n. 1, p. 118–128, 1999.

FAN, X.; YIN, B.; SUN, Y. Dynamic features based driver fatigue detection. In: *Rough Sets and Knowledge Technology*. [S.l.]: Springer, 2008. p. 684–691.

FAN, X.; YIN, B.-c.; SUN, Y.-f. Yawning detection based on gabor wavelets and lda. *Journal of Beijing university of technology*, v. 35, n. 3, p. 409–413, 2009.

FENG, R.; ZHANG, G.; CHENG, B. An on-board system for detecting driver drowsiness based on multi-sensor data fusion using dempster-shafer theory. In: IEEE. *Networking, Sensing and Control, 2009. ICNSC'09. International Conference on*. [S.l.], 2009. p. 897–902.

FETZ, E. E. Operant conditioning of cortical unit activity. *Science*, American Association for the Advancement of Science, v. 163, n. 3870, p. 955–958, 1969.

FIOLET, E. Neurosky mindwave brings brain-computer interface to education. *)ˆ(Eds.):Book NeuroSky MindWave Brings Brain-Computer Interface to Education(2011, edn.)*, 2011.

FLORES, M. J.; ARMINGOL, J. M.; ESCALERA, A. de la. Driver drowsiness warning system using visual information for both diurnal and nocturnal illumination conditions. *EURASIP journal on advances in signal processing*, Hindawi Publishing Corp., v. 2010, p. 3, 2010.

FOONG, R.; ANG, K. K.; QUEK, C.; GUAN, C.; WAI, A. A. P. An analysis on driver drowsiness based on reaction time and eeg band power. In: IEEE. *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. [S.l.], 2015. p. 7982–7985.

FORSMAN, P. M.; VILA, B. J.; SHORT, R. A.; MOTT, C. G.; DONGEN, H. P. V. Efficient driver drowsiness detection at moderate levels of drowsiness. *Accident Analysis & Prevention*, Elsevier, v. 50, p. 341–350, 2013.

FRIEDMAN, J. H.; BENTLEY, J. L.; FINKEL, R. A. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw.*, ACM, New York, NY, USA, v. 3, n. 3, p. 209–226, set. 1977. ISSN 0098-3500. Disponível em: ⟨http://doi.acm.org/10.1145/355744.355745⟩.

FRIEDMAN, J. H.; BENTLEY, J. L.; FINKEL, R. A. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, ACM, v. 3, n. 3, p. 209–226, 1977.

GALÁN, F.; NUTTIN, M.; LEW, E.; FERREZ, P. W.; VANACKER, G.; PHILIPS, J.; BRUSSEL, H. V.; MILLÁN, J. d. R. An asynchronous and non-invasive brain-actuated wheelchair. In: *Proceedings of the 13th International Symposium on Robotics Research*. [S.l.: s.n.], 2007.

GIURGIUTIU, V.; YU, L. Comparison of short-time fourier transform and wavelet transform of transient and tone burst wave propagation signals for structural health monitoring. In: *présenté à 4 th International Workshop on Structural Health Monitoring, Stanford, CA*. [S.l.: s.n.], 2003.

GREEN, R. M.; MESSICK, W. J.; RICOTTA, J. J.; CHARLTON, M. H.; SATRAN, R.; MCBRIDE, M. M.; DEWEESE, J. A. Benefits, shortcomings, and costs of eeg monitoring. *Annals of surgery*, Lippincott, Williams, and Wilkins, v. 201, n. 6, p. 785, 1985.

GUILLOT, A.; NADROWSKA, E.; COLLET, C. Using motor imagery to learn tactical movements in basketball. *Journal of Sport Behavior*, Journal of Sport Behavior, v. 32, n. 2, p. 189, 2009.

HAL, B. V.; RHODES, S.; DUNNE, B.; BOSSEMEYER, R. Low-cost eeg-based sleep detection. In: IEEE. *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. [S.l.], 2014. p. 4571–4574.

HALICI, U. *Lecture notes on neural networks, the dynamic of brain multidisipliner*. [S.l.]: Diyarbakir, 1999.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.

HARTLEY, L.; COMMISSION, N. R. T. et al. *Review of fatigue detection and prediction technologies*. [S.l.]: National Road Transport Commission Melbourne, 2000.

HASSANAT, A. B.; ABBADI, M. A.; ALTARAWNEH, G. A.; ALHASANAT, A. A. Solving the problem of the k parameter in the knn classifier using an ensemble learning approach. *arXiv preprint arXiv:1409.0919*, 2014.

HASSANIEN, A. E.; AR, A. A. T. A. A. *Brain-Computer Interfaces*. [S.l.]: Springer, 2015.

HEMA, C. R.; PAULRAJ, M.; YAACOB, S.; ADOM, A. H.; NAGARAJAN, R. Asynchronous brain machine interface-based control of a wheelchair. In: *Software Tools and Algorithms for Biological Systems*. [S.l.]: Springer, 2011. p. 565–572.

HINTERBERGER, T.; KAISER, J.; KÜBLER, A.; NEUMANN, N.; BIRBAUMER, N. The thought translation device and its applications to the completely paralyzed. *Diebner, Druckrey, and Weibel, editors, Sciences of the Interfaces. Genista-Verlag, Tubingen*, p. 232–240, 2001.

HU, S.; ZHENG, G. Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Systems with Applications*, Elsevier, v. 36, n. 4, p. 7651–7658, 2009.

HUSAR, P. *Eyetracker Warns against Momentary Driver Drowsiness*. Available online: http://www.fraunhofer.de/en/press/research-news/2010/10/eye-tracker-driver-drowsiness.html (accessed on 27, 2012. Disponível em: ⟨http://www.fraunhofer.de/en/press/research-news/2010/10/eye-tracker-driver-drowsiness.html⟩.

INDYK, P.; MOTWANI, R. Approximate nearest neighbors: Towards removing the curse of dimensionality. In: *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*. New York, NY, USA: ACM, 1998. (STOC '98), p. 604–613. ISBN 0-89791-962-9. Disponível em: ⟨http://doi.acm.org/10.1145/276698.276876⟩.

INGRE, M.; ÅKERSTEDT, T.; PETERS, B.; ANUND, A.; KECKLUND, G.; PICKLES, A. Subjective sleepiness and accident risk avoiding the ecological fallacy. *Journal of sleep research*, Wiley Online Library, v. 15, n. 2, p. 142–148, 2006.

JIAWEI, H.; KAMBER, M. *Data Mining Concepts and Techniques. Morgon Kanfmann Pub lishers*. [S.l.]: Inc, 2001.

KAPTEIN, N.; THEEUWES, J.; HORST, R. V. D. Driving simulator validity: Some considerations. *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, n. 1550, p. 30–36, 1996.

KATONA, J.; FARKAS, I.; UJBANYI, T.; DUKAN, P.; KOVARI, A. Evaluation of the neurosky mindflex eeg headset brain waves data. In: IEEE. *Applied Machine Intelligence and Informatics (SAMI), 2014 IEEE 12th International Symposium on*. [S.l.], 2014. p. 91–94.

KAUFMANN, T.; HOLZ, E. M.; KÜBLER, A. Comparison of tactile, auditory, and visual modality for brain-computer interface use: a case study with a patient in the locked-in state. *Front Neurosci*, v. 7, n. 129, p. 1–12, 2013.

KEMP, B.; ZWINDERMAN, A.; TUK HAC KAMPHUISEN, J. O. B. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE-BME*, v. 47, n. 9, p. 1185–1194, 2000.

KHUSHABA, R. N.; KODAGODA, S.; LAL, S.; DISSANAYAKE, G. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *Biomedical Engineering, IEEE Transactions on*, IEEE, v. 58, n. 1, p. 121–131, 2011.

KIYMIK, M. K.; AKIN, M.; SUBASI, A. Automatic recognition of alertness level by using wavelet transform and artificial neural network. *Journal of neuroscience methods*, Elsevier, v. 139, n. 2, p. 231–240, 2004.

KLINGEBERG, T.; SCHILLING, M. Mobile wearable device for long term monitoring of vital signs. *Computer methods and programs in biomedicine*, Elsevier, v. 106, n. 2, p. 89–96, 2012.

KO, L.-W.; LAI, W.-K.; LIANG, W.-G.; CHUANG, C.-H.; LU, S.-W.; LU, Y.-C.; HSIUNG, T.-Y.; WU, H.-H.; LIN, C.-T. Single channel wireless eeg device for real-time fatigue level detection. In: IEEE. *Neural Networks (IJCNN), 2015 International Joint Conference on*. [S.l.], 2015. p. 1–5.

KOBAYASHI, H. Emg/ecg acquisition system with online adjustable parameters using zigbee wireless technology. *IEEJ Transactions on Electronics, Information and Systems*, v. 132, p. 632–639, 2012.

KOKONOZI, A.; MICHAIL, E.; CHOUVARDA, I.; MAGLAVERAS, N. A study of heart rate and brain system complexity and their interaction in sleep-deprived subjects. In: IEEE. *Computers in Cardiology, 2008*. [S.l.], 2008. p. 969–971.

KOSHINO, Y.; NISHIO, M.; MURATA, T.; OMORI, M.; MURATA, I.; SAKAMOTO, M.; ISAKI, K. The influence of light drowsiness on the latency and amplitude of p300. *Clinical EEG and Neuroscience*, SAGE Publications, v. 24, n. 3, p. 110–113, 1993.

KUBLER, A.; KOTCHOUBEY, B.; GHANAYIM, N.; HINTERBERGER, T.; PERELMOUTER, J.; SCHAUER, M.; FRITSCH, C.; BIRBAUMER, N. A thought translation device for brain computer communication. *Studia Psychologica*, SLOVAK ACADEMIC PRESS, v. 40, p. 17–32, 1998.

KURT, M. B.; SEZGIN, N.; AKIN, M.; KIRBAS, G.; BAYRAM, M. The ann-based computing of drowsy level. *Expert Systems with Applications*, Elsevier, v. 36, n. 2, p. 2534–2542, 2009.

LAL, S. K.; CRAIG, A. A critical review of the psychophysiology of driver fatigue. *Biological psychology*, Elsevier, v. 55, n. 3, p. 173–194, 2001.

LECUN, Y.; JACKEL, L.; BOTTOU, L.; BRUNOT, A.; CORTES, C.; DENKER, J.; DRUCKER, H.; GUYON, I.; MULLER, U.; SACKINGER, E. et al. Comparison of learning algorithms for handwritten digit recognition. In: *International conference on artificial neural networks*. [S.l.: s.n.], 1995. v. 60, p. 53–60.

LEE, B.-G.; CHUNG, W.-Y. Multi-classifier for highly reliable driver drowsiness detection in android platform. *Biomedical Engineering: Applications, Basis and Communications*, World Scientific, v. 24, n. 02, p. 147–154, 2012.

LEUTHARDT, E. C.; SCHALK, G.; WOLPAW, J. R.; OJEMANN, J. G.; MORAN, D. W. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng*, p. 63–71, 2004.

LEW, M. . S.; HUANG, N. .; BAKKER, T. .; VURAL, E. .; CETIN, E. .; ERCIL, M. .; LITTLEWORT, A. .; BARTLETT, G. .; MOVELLAN, M. .; J. Drowsy driver detection through facial movement analysis. In: *Human-Computer Interaction; Springer: Berlin, Germany*. [S.l.: s.n.], 2007. v. 4796, p. 618.

LEXUS, L. X. D. M. S. *Available online: http://www.lexus.eu/range/ls/ key-features/safety/safety-driver-monitoring-system.aspx (accessed on 21 November 2012)*. [S.l.: s.n.], 2012.

LI, X.; ZHAO, Q.; LIU, L.; PENG, H.; QI, Y.; MAO, C.; FANG, Z.; LIU, Q.; HU, B. Improve affective learning with eeg approach. *Computing and Informatics*, v. 29, n. 4, p. 557–570, 2012.

LIANG, W. C.; YUAN, J.; SUN, D. C.; LIN, M. H. Changes in physiological parameters induced by indoor simulated driving: Effect of lower body exercise at mid-term break. *Sensors*, Molecular Diversity Preservation International, v. 9, n. 9, p. 6913–6933, 2009.

LIN, C.-T.; CHANG, C.-J.; LIN, B.-S.; HUNG, S.-H.; CHAO, C.-F.; WANG, I.-J. A real-time wireless brain–computer interface system for drowsiness detection. *Biomedical Circuits and Systems, IEEE Transactions on*, IEEE, v. 4, n. 4, p. 214–222, 2010.

LIN, F.-C.; KO, L.-W.; CHUANG, C.-H.; SU, T.-P.; LIN, C.-T. Generalized eeg-based drowsiness prediction system by using a self-organizing neural fuzzy system. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, IEEE, v. 59, n. 9, p. 2044–2055, 2012.

LIU, C. C.; HOSKING, S. G.; LENNÉ, M. G. Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of safety research*, Elsevier, v. 40, n. 4, p. 239–245, 2009.

LIU, D.; SUN, P.; XIAO, Y.; YIN, Y. Drowsiness detection based on eyelid movement. In: IEEE. *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*. [S.l.], 2010. v. 2, p. 49–52.

LIU, J.; ZHANG, C.; ZHENG, C. Eeg-based estimation of mental fatigue by using kpca–hmm and complexity parameters. *Biomedical Signal Processing and Control*, Elsevier, v. 5, n. 2, p. 124–130, 2010.

LIU, T.; MOORE, A. W.; YANG, K.; GRAY, A. G. An investigation of practical approximate nearest neighbor algorithms. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2004. p. 825–832.

LOPEZ, A.; RICHARDSON, P. C. Capacitive electrocardiographic and bioelectric electrodes. *Biomedical Engineering, IEEE Transactions on*, IEEE, n. 1, p. 99–99, 1969.

LUCK, S. J. *An introduction to the event-related potential technique*. [S.l.]: MIT press, 2014.

LUM, H.; REAGAN, J. A. Interactive highway safety design model: accident predictive module. *Public Roads*, v. 58, n. 3, 1995.

LYONS, R. G. *Understanding digital signal processing*. [S.l.]: Pearson Education, 2010.

MAKEIG, S.; KOTHE, C.; MULLEN, T.; BIGDELY-SHAMLO, N.; ZHANG, Z.; KREUTZ-DELGADO, K. Evolving signal processing for brain–computer interfaces. *Proceedings of the IEEE*, IEEE, v. 100, n. Special Centennial Issue, p. 1567–1584, 2012.

MARON, O.; MOORE, A. W. Hoeffding races: Accelerating model selection search for classification and function approximation. *Robotics Institute*, p. 263, 1993.

MARTIN, C.; MARZEC, M. Sleep scoring: 35 years of standardized sleep staging. *RT Magazine.(On-Line)*, 2003.

MCKINLEY, R. A.; MCINTIRE, L. K.; SCHMIDT, R.; REPPERGER, D. W.; CALDWELL, J. A. Evaluation of eye metrics as a detector of fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Sage Publications, v. 53, n. 4, p. 403–414, 2011.

METS, M. A.; KUIPERS, E.; DOMIS, L. M. S.; LEENDERS, M.; OLIVIER, B.; VERSTER, J. C. Effects of alcohol on highway driving in the stisim driving simulator. *Human Psychopharmacology: Clinical and Experimental*, Wiley Online Library, v. 26, n. 6, p. 434–439, 2011.

METS, M. A.; KUIPERS, E.; DOMIS, L. M. S.; LEENDERS, M.; OLIVIER, B.; VERSTER, J. C. Effects of alcohol on highway driving in the stisim driving simulator. *Human Psychopharmacology: Clinical and Experimental*, Wiley Online Library, v. 26, n. 6, p. 434–439, 2011.

MIYAJI, M.; KAWANAKA, H.; OGURI, K. Driver's cognitive distraction detection using physiological features by the adaboost. In: IEEE. *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*. [S.l.], 2009. p. 1–6.

MOORE, A. W.; HILL, D. J.; JOHNSON, M. P. An empirical investigation of brute force to choose features, smoothers and function approximators. In: CITESEER. *Computational learning theory and natural learning systems*. [S.l.], 1992.

OEHLER, M. A multichannel portable ecg system with capacitive sensors, physiological measure. v. 29, n. 7, 2008.

OTMANI, S.; PEBAYLE, T.; ROGE, J.; MUZET, A. Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers. *Physiology & behavior*, Elsevier, v. 84, n. 5, p. 715–724, 2005.

PATEL, M.; LAL, S.; KAVANAGH, D.; ROSSITER, P. Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert systems with Applications*, Elsevier, v. 38, n. 6, p. 7235–7242, 2011.

PENG, C.-Y. J.; LEE, K. L.; INGERSOLL, G. M. An introduction to logistic regression analysis and reporting. *The journal of educational research*, Taylor & Francis, v. 96, n. 1, p. 3–14, 2002.

PFURTSCHELLER, G.; FLOTZINGER, D.; KALCHER, J. Brain-computer interfacea new communication device for handicapped persons. *Journal of Microcomputer Applications*, Elsevier, v. 16, n. 3, p. 293–299, 1993.

PHILIP, P.; SAGASPE, P.; MOORE, N.; TAILLARD, J.; CHARLES, A.; GUILLEMINAULT, C.; BIOULAC, B. Fatigue, sleep restriction and driving performance. *Accident Analysis & Prevention*, Elsevier, v. 37, n. 3, p. 473–478, 2005.

PHILIP, P.; SAGASPE, P.; MOORE, N.; TAILLARD, J.; CHARLES, A.; GUILLEMINAULT, C.; BIOULAC, B. Fatigue, sleep restriction and driving performance. *Accident Analysis & Prevention*, Elsevier, v. 37, n. 3, p. 473–478, 2005.

PHILIPS, J.; MILLÁN, J. d. R.; VANACKER, G.; LEW, E.; GALÁN, F.; FERREZ, P. W.; BRUSSEL, H. V.; NUTTIN, M. Adaptive shared control of a brain-actuated simulated wheelchair. In: IEEE. *Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on*. [S.l.], 2007. p. 408–414.

PHILIPS, J.; MILLÁN, J. d. R.; VANACKER, G.; LEW, E.; GALÁN, F.; FERREZ, P. W.; BRUSSEL, H. V.; NUTTIN, M. Adaptive shared control of a brain-actuated simulated wheelchair. In: IEEE. *Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on*. [S.l.], 2007. p. 408–414.

PHUNG, D.; TRAN, D.; MA, W.; NGUYEN, P.; PHAM, T. Using shannon entropy as eeg signal feature for fast person identification. In: *ESANN*. [S.l.: s.n.], 2014.

PICTON, T. W. The p300 wave of the human event-related potential. *Journal of clinical neurophysiology*, LWW, v. 9, n. 4, p. 456–479, 1992.

PORTOULI, E.; BEKIARIS, E.; PAPAKOSTOPOULOS, V.; MAGLAVERAS, N. On-road experiment for collecting driving behavioural data of sleepy drivers. *Somnologie-Schlafforschung und Schlafmedizin*, Springer, v. 11, n. 4, p. 259–267, 2007.

PREPARATA, F. P.; SHAMOS, M. *Computational Geometry*. [S.l.]: Springer-Verlag, 1985.

QUINLAN, J. R. Induction of decision trees. *Machine learning*, Springer, v. 1, n. 1, p. 81–106, 1986.

QUINLAN, J. R. Simplifying decision trees. *International journal of man-machine studies*, Elsevier, v. 27, n. 3, p. 221–234, 1987.

RAU, P. S. Drowsy driver detection and warning system for commercial vehicle drivers: field operational test design, data analyses, and progress. *National Highway Traffic Safety Administration*, p. 05–0192, 2005.

RECHTSCHAFFEN, A.; KALES, A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. US Government Printing Office, US Public Health Service, 1968.

REYNER, L.; HORNE, J. Falling asleep whilst driving: are drivers aware of prior sleepiness? *International journal of legal medicine*, Springer, v. 111, n. 3, p. 120–123, 1998.

SAHAYADHAS, A.; SUNDARAJ, K.; MURUGAPPAN, M. Detecting driver drowsiness based on sensors: a review. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 12, n. 12, p. 16937–16953, 2012.

SAHAYADHAS, A.; SUNDARAJ, K.; MURUGAPPAN, M. Detecting driver drowsiness based on sensors: a review. *Sensors*, Multidisciplinary Digital Publishing Institute, v. 12, n. 12, p. 16937–16953, 2012.

SAŁABUN, W. Processing and spectral analysis of the raw eeg signal from the mindwave. *Przeglad Elektrotechniczny*, v. 90, n. 2, p. 169–174, 2014.

SHI, K.; LI, L.; LIU, H.; HE, J.; ZHANG, N.; SONG, W. An improved knn text classification algorithm based on density. In: IEEE. *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*. [S.l.], 2011. p. 113–117.

SIEGEL, J. M. The rem sleep-memory consolidation hypothesis. *Science*, American Association for the Advancement of Science, v. 294, n. 5544, p. 1058–1063, 2001.

SIMONS, R.; MARTENS, M.; RAMAEKERS, J.; KRUL, A.; KLÖPPING-KETELAARS, I.; SKOPP, G. Effects of dexamphetamine with and without alcohol on simulated driving. *Psychopharmacology*, Springer, v. 222, n. 3, p. 391–399, 2012.

SKIENA, S. S. *The algorithm design manual: Text*. [S.l.]: Springer Science & Business Media, 1998. v. 1.

SLOTEN, J. . V.; NYSSEN, P. .; HAUEISEN, M. .; MIZUNO, J. .; OKUMURA, A. .; MATSUMURA, H. .; M. Development of neckband mounted active bio-electrodes for non-restraint lead method of ecg r wave. In: *Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering, Antwerp, Belgium, 2327*. [S.l.: s.n.], 2008. p. 13941397.

SOMAN, K.; LOGANATHAN, R.; AJAY, V. *Machine learning with SVM and other kernel methods*. [S.l.]: PHI Learning Pvt. Ltd., 2009.

SOMMER, D.; GOLZ, M.; TRUTSCHEL, U.; EDWARDS, D. Biosignal based discrimination between slight and strong driver hypovigilance by support-vector machines. In: *Agents and artificial intelligence*. [S.l.]: Springer, 2009. p. 177–187.

SPROULL, R. F. Refinements to nearest-neighbor searching ink-dimensional trees. *Algorithmica*, Springer, v. 6, n. 1-6, p. 579–589, 1991.

STANFILL, C.; WALTZ, D. Toward memory-based reasoning. *Communications of the ACM*, ACM, v. 29, n. 12, p. 1213–1228, 1986.

STERIADE, M. Basic mechanisms of sleep generation. *Neurology*, v. 42, n. 7 Suppl 6, p. 9–17, 1992.

SUBASI, A.; KIYMIK, M. K.; AKIN, M.; EROGUL, O. Automatic recognition of vigilance state by using a wavelet-based artificial neural network. *Neural Computing & Applications*, Springer, v. 14, n. 1, p. 45–55, 2005.

ŠUŠMÁKOVÁ, K. Human sleep and sleep eeg. *Measurement Science Review*, Citeseer, v. 4, n. 2, p. 59–74, 2004.

TANAKA, K.; MATSUNAGA, K.; WANG, H. O. Electroencephalogram-based control of an electric wheelchair. *IEEE transactions on robotics*, IEEE, v. 21, n. 4, p. 762–766, 2005.

THIFFAULT, P.; BERGERON, J. Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, Elsevier, v. 35, n. 3, p. 381–391, 2003.

TREMAINE, R.; DORRIAN, J.; LACK, L.; LOVATO, N.; FERGUSON, S.; ZHOU, X.; ROACH, G. The relationship between subjective and objective sleepiness and performance during a simulated night-shift with a nap countermeasure. *Applied ergonomics*, Elsevier, v. 42, n. 1, p. 52–61, 2010.

TU, J. V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of clinical epidemiology*, Elsevier, v. 49, n. 11, p. 1225–1231, 1996.

VANACKER, G.; MILLÁN, J. del R.; LEW, E.; FERREZ, P. W.; MOLES, F. G.; PHILIPS, J.; BRUSSEL, H. V.; NUTTIN, M. Context-based filtering for assisted brain-actuated wheelchair driving. *Computational intelligence and neuroscience*, Hindawi Publishing Corp., v. 2007, p. 3–3, 2007.

VIDAL, J.-J. Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 2, n. 1, p. 157–180, 1973.

VUCKOVIC, A.; RADIVOJEVIC, V.; CHEN, A. C.; POPOVIC, D. Automatic recognition of alertness and drowsiness from eeg by an artificial neural network. *Medical Engineering & Physics*, Elsevier, v. 24, n. 5, p. 349–360, 2002.

VURAL, E. *Video based detection of driver fatigue*. Tese (Doutorado) — Sabanci University, 2009.

VURAL, E.; CETIN, M.; ERCIL, A.; LITTLEWORT, G.; BARTLETT, M.; MOVELLAN, J. Drowsy driver detection through facial movement analysis. In: *Human–computer interaction*. [S.l.]: Springer, 2007. p. 6–18.

WADA, Y.; NANBU, Y.; KOSHINO, Y.; SHIMADA, Y.; HASHIMOTO, T. Inter-and intrahemispheric eeg coherence during light drowsiness. *Clinical EEG and Neuroscience*, SAGE Publications, v. 27, n. 2, p. 84–88, 1996.

WALKER, S. H.; DUNCAN, D. B. Estimation of the probability of an event as a function of several independent variables. *Biometrika*, Biometrika Trust, v. 54, n. 1-2, p. 167–179, 1967.

WANG, J. *Encyclopedia of data warehousing and mining*. [S.l.]: IGI Global, 2005.

WEIJIE, S.; HAIXIN, S.; CHENG, E.; QINGKUN, Z.; LI, Q.; WEIJUN, S. Effective driver fatigue monitoring through pupil detection and yawing analysis in low light level environments. *International Journal of Digital Content Technology & its Applications*, v. 6, n. 17, 2012.

WESTON, J.; WATKINS, C. *Multi-class support vector machines*. [S.l.], 1998.

XU, Q.; ZHOU, H.; WANG, Y.; HUANG, J. Fuzzy support vector machine for classification of eeg signals using wavelet-based features. *Medical engineering & physics*, Elsevier, v. 31, n. 7, p. 858–865, 2009.

YAN, X. *Linear regression analysis: theory and computing*. [S.l.]: World Scientific, 2009.

YANG, G.; LIN, Y.; BHATTACHARYA, P. A driver fatigue recognition model based on information fusion and dynamic bayesian network. *Information Sciences*, Elsevier, v. 180, n. 10, p. 1942–1954, 2010.

YEO, M. V. M.; LI, X.; SHEN, K.; WILDER-SMITH, E. P. V. Can svm be used for automatic eeg detection of drowsiness during car driving. *J. Safety Science*, v. 47, n. 1, p. 115–124, 2009.

YU, X. Real-time nonintrusive detection of driver drowsiness. University of Minnesota Center for Transportation Studies, 2009.

ZEMAN, P. New technique to remove tonic muscle artifact (emg) from eeg. *Applied Brain and Vision Science*, 2012.

ZHANG, Z.; ZHANG, J. A new real-time eye tracking based on nonlinear unscented kalman filter for monitoring driver fatigue. *Journal of Control Theory and Applications*, Springer, v. 8, n. 2, p. 181–188, 2010.