Ali Abdollahi Gharbali

Mestre em Engenharia Electrotécnica e de Computadores

Sleep Stage Classification: A Deep Learning Approach

Dissertação para obtenção do Grau de Doutor em
Engenharia Electrotécnica e de Computadores,
Especialização em Processamento de Sinais

Orientador: Professor Doutor José Manuel Fonseca
Professor Associado com Agregação da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Júri:

Presidente: Doutor Luís Camarinha-Matos

Arguentes: Doutor André Ribeiro Lourenço
Doutor Pedro Almeida

Vogais: Doutor Manuel Martins Barata
Doutor Rui Neves-Silva
Doutor André Damas Mora
Doutor José Manuel Fonseca

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro, 2018

Sleep Stage Classification: A Deep Learning Approach

To my family

Acknowledgments

During the last years, I had the opportunity to embrace some research projects and maintain contact with people that helped me in this long journey. For that, I must thank them all: professors, colleagues, friends, and the most important one family.

I would like to express my deepest gratitude to my advisor, Professor José Manuel Matos Ribeiro da Fonseca for all the patient guidance and assistance. His comments and valuable suggestions made this work possible. Working with him has been a real challenge and definitely made me grow up not only in scientific but also in personal terms. His vision of life, and specifically in science, is absolutely amazing.

I wish to thank the members of the thesis accompanying committee, professor Rui Neves-Silva and professor André Lourenço, for the useful comments and suggestions that were provided during the development of this thesis.

Thank you to the Department of Electrical Engineering and UNINOVA research institute, for making this work possible, providing available facilities and resources.

I also would like to thank my research colleagues at CA3 group of UNINOVA for the support, companionship, and motivation: Professor Rita A. Ribeiro, Professor André Mora, Leonardo Martins. With them it has been like a journey into knowledge and companionship since I started to do research in this centre.

I have to highlight my colleague and friend António Falcão. His motivational support and feedback were not ignorable and effected the progress of this thesis positively. I wish him all the best in the world for his professional and personal life. To him, I would like to specially thank.

To my dearest friends, I have to show my appreciation for the motivation, support, and friendship in finishing this thesis. Friends are the family that we choose…

I would also like to express my most profound gratitude to my wonderful parents, Pooran Valibeygi and Biouk Abdollahi Gharbali. They raised me teaching the meaning of life and the relevance of having a good education. They also motivated me in trying to achieve all the goals that I proposed to, and still do. I could never reach this without them. I really have to thank the patience, companionship, and love that they gave me, without which it would never be possible to finish this work and I feel so blessed that I have both of them. To my sister, Maryam, a very special thank for the continuous motivation and precise word of encouragement or smile in the right moment.

Above all I would like to specially thank my beloved wife Shirin for her love and constant support, for all the late nights and early mornings, and for keeping me sane over the past year. Thank you for being my muse, editor, proof-reader, and beside all of them a great colleague. But most of all, thank you for being my best friend. I owe you a lot.

Abstract

Sleep occupies significant part of human life. The diagnoses of sleep related disorders are of great importance. To record specific physical and electrical activities of the brain and body, a multi-parameter test, called polysomnography (PSG), is normally used. The visual process of sleep stage classification is time consuming, subjective and costly. To improve the accuracy and efficiency of the sleep stage classification, automatic classification algorithms were developed.

In this research work, we focused on pre-processing (filtering boundaries and de-noising algorithms) and classification steps of automatic sleep stage classification. The main motivation for this work was to develop a pre-processing and classification framework to clean the input EEG signal without manipulating the original data thus enhancing the learning stage of deep learning classifiers.

For pre-processing EEG signals, a lossless adaptive artefact removal method was proposed. Rather than other works that used artificial noise, we used real EEG data contaminated with EOG and EMG for evaluating the proposed method. The proposed adaptive algorithm led to a significant enhancement in the overall classification accuracy.

In the classification area, we evaluated the performance of the most common sleep stage classifiers using a comprehensive set of features extracted from PSG signals. Considering the challenges and limitations of conventional methods, we proposed two deep learning-based methods for classification of sleep stages based on Stacked Sparse AutoEncoder (SSAE) and Convolutional Neural Network (CNN). The proposed methods performed more efficiently by eliminating the need for conventional feature selection and feature extraction steps respectively. Moreover, although our systems were trained with lower number of samples compared to the similar studies, they were able to achieve state of art accuracy and higher overall sensitivity.

Keyword: Sleep Stage Classification, Deep Learning, Convolutional Neural Networks, Transfer Learning, Wavelet, Adaptive Filtering, Denoising.

Resumo

O sono ocupa uma parte significativa da vida humana. Por isso, o diagnóstico de distúrbios relacionados com o sono é de grande importância. Para registar as atividades físicas e elétricas específicas do cérebro e do corpo, é normalmente efetuado um teste multiparamétrico durante o sono a que se dá o nome de polissonografia (PSG). No entanto, o processo de classificação visual das diversas fases do sono é demorado, subjetivo e dispendioso. Com o objetivo de melhorar a precisão e eficiência da classificação das fases do sono, diversos algoritmos de classificação automática têm sido desenvolvidos.

Nesta tese o trabalho desenvolvido focou-se principalmente no pré-processamento (faixa de filteragem e algoritmos de cancelamento de ruído) e na classificação automática das fases do sono. A principal motivação para este trabalho foi o desenvolvimento de uma estrutura de pré-processamento e classificação que permita atenuar o ruído do sinal de electroencefalograma (EEG) de entrada sem manipular os dados originais, melhorando assim a possibilidade de aprendizagem dos classificadores de aprendizagem profunda (deep learning).

Para o pré-processamento dos sinais EEG, foi proposto um método de remoção de artefactos adaptativo e sem perdas. Contrariamente a outros trabalhos que usaram ruído artificial para a avaliação dos resultados, foram neste trabalho usados dados reais de EEG contaminados com sinais de electrooculografia (EOG) e eletromiografia (EMG) para a avaliação do método proposto. Como se demonstra ao longo dos ensaios efetuados, o algoritmo adaptativo que aqui se propõe permitiu uma melhoria significativa na precisão geral da classificação.

Na área da classificação, foi avaliado o desempenho dos classificadores de fases do sono mais comuns usando um conjunto abrangente de características extraídas de sinais PSG. Considerando os desafios e limitações dos métodos convencionais, foram propostos dois métodos baseados em aprendizagem profunda para classificar as fases do sono, nomeadamente o Stacked Sparse AutoEncoder (SSAE) e as Convolutional Neural Networks (CNN). Os métodos propostos mostraram-se mais eficientes do que os métodos clássicos, eliminando a necessidade de seleção de características e de extração de características, respetivamente. Além disso, embora os sistemas propostos tenham sido treinados com um menor número de amostras em comparação com os estudos similares, eles foram capazes de atingir uma precisão que se pode considerar ao nível da atingida pelos métodos atualmente mais desenvolvidos, juntando a isso uma maior sensibilidade geral.

Palavras-chave: Classificação das fases do sono, aprendizagem profunda, Convolutional Neural Networks, transferência de aprendizagem, Wavelet, filtragem adaptativa, Denoising.

List of Contents

List of Figures

List of Tables

List of Acronyms

AASM	American Academy of Sleep Medicine
AHC	Ascendant Hierarchical Clustering
ANN	Artificial Neural Network
AR	Autoregressive
BCI	Brain-Computer Interfaces
CAP	Cyclic Alternating Patterns
Chi2	Chi-Square
CMAC	Cerebellar Model Articulation Control
CMIM	Conditional Mutual Information Maximization
CNN	Convolutional Neural Network
CWD	Choi–Williams Distribution
CWT	Continuous Wavelet Transform
DBM	Deep Boltzmann Machine
DBN	Deep Belief Network
DNN	Deep Neural Networks

DS	Deep-Sleep
DSVM	Dendogram based Support Vector Machine
DT	Decision Tree
ECG	Electrocardiogram
EEG	Electroencephalogram
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
EMG	Chin electromyogram
EMG	Leg electromyogram
EOG	Electro-oculogram
ERP	Event Related Potential
FCBF	Fast Correlation Based Filter
FFNN	Feed-Forward Neural Network
FIR	Finite Impulse Response
GMDH	Group-Method of Data Handling
GMM	Gaussian Mixture Model
HHT	Hilbert–Huang Transform
HMM	Hidden Markov Model
HR	Heart Rate
IG	Information Gain
IIR	Infinite Impulse Response
ILSVRC	ImageNet Large-Scale Visual Recognition Challenge
ISD	Itakura Spectral Distance

KMCFW	k-means Clustering Based Feature Weighting
k-NN	*k*-Nearest Neighbourhood
LDA	Linear Discriminant Analysis
LDC	Linear Discriminant Classifier
LMS	Least Mean Squares
LS	Light-Sleep
LSTM	Long Short-Term Memory
LVQ	Learning Vector Quantization
MCDM	Multi-Criteria Decision Making
MEE	Minimum Error Entropy
MLP	Multi-Layer Perceptron
MNN	Mixed Neural Network
mRMR	minimum Redundancy Maximum Relevance
MSE	Mean Square Error
MSE	Multiscale entropy
MT	Movement Time
NLMS	Normalized Least Mean Squares
NN	Neural Net
NREM	Non-Rapid Eye Movement
OAA	One-Against-All
OAO	One-Against-One
PCA	Principal Component Analysis
PLS	Partial Least Squares

PNN	Probabilistic Neural Network
POSTS	Positive Occipital Sharp Transients of Sleep
PSG	Polysomnography
RBF	Radial Basis Function
RBM	Restricted Boltzmann Machine
ReLU	Rectified Linear Units
REM	Rapid Eye Movement
RF	Random Forest
RIP	Respiratory Inductance Plethysmography
RMS	Recursive Least Mean Squares
RNN	Recurrent Neural Network
RRA	Robust Rank Aggregation
S1	Stage 1
S2	Stage 2
S3	Stage 3
S4	Stage 4
SDBN	Sparse Deep Belief Network
SEM	Slow Eye Movement
SGD	Stochastic Gradient Descent
SLFN	Single-hidden Layer Feedforward Neural Network
SNR	Signal to Noise Ratio
SS	Stage Shift
SSAE	Stacked Sparse Autoencoder

SSL	Stacked Sequential Learning
STW	Saw Tooth Waves
SVM	Support Vector Machine
SWS	Slow Wave Stage
WP	Wavelet Packet
ZCR	Zero Crossing Rate

Chapter 1

1 Introduction

1-1 Background Theory

Sleep occupies a very significant part of human life. Therefore, the diagnoses of sleep related disorders are of great importance in both health care and sleep research.

Sleep is a particular condition of the nervous system with noticeable features and brain activity phases. Although most people think sleep is a passive and constant process, as a matter of fact, sleep is an active state. Human bodies move frequently during the night and the human brain is sometimes more active during sleep than in the period of the normal waking state. The human brain travels through several psycho-physiological states during sleep that are quite stable and noticeable. For the sleep study, it is mandatory to record specific physical activities of the brain. For this aim a multiple-parametric test, called polysomnography (PSG), is normally used. Through the American Academy of Sleep Medicine (AAMS) the PSG monitors many body functions such as [1]:

- Electroencephalogram (EEG) derivations;
- Electro-oculogram (EOG) derivations;
- Chin electromyogram (EMG);
- Leg electromyogram (EMG);
- Airflow signals;
- Respiratory effort signals;
- Oxygen saturation;
- Body position;
- Electrocardiogram (ECG);

Sleep study is usually done by analysing macro and microstructures.

1-1-1 Macrostructure

According to the publication "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects", established in 1968 by a committee co-chaired by A. Rechtschaffen and A. Kales, sleep is divided in five stages. Non-Rapid Eye Movement stage (NREM) which includes four stages: S1, S2, S3 and S4. Rapid-Eye Movement stage (REM) and wakefulness that is frequently included in analysis as a sixth stage. NREM itself occupies almost 75% of sleep. In the following, sleep stages will be described by detail.

According to AASM the EEG data should be divided into 30s epochs, with sleep staging done by relating each epoch to the corresponding stage.

Wake:

The waking stage covers from full alertness to the early stage of drowsiness. Since in this stage body prepares to sleep, we can call it a relaxing stage.

Stage 1 of NREM:

Stage 1 is the bridge between wakefulness and sleep (Sleep Onset) and can also be called *light sleep stage*. In this stage, the heart rate starts to slow down and breathing, step by step, becomes more regular. This stage may last for 5 to 10 minutes and the subject can still be easily wake up. Slow eye movement (SEM), which are called sinusoidal eye movements with an initial deflection, regularly lasting more than 500 msec. Low amplitude waves with the frequencies between 4 and 7Hz are noticeable also vertex sharp waves (V waves) are included with a duration less than 0.5 sec.

Stage 2 of NREM:

Stage 2 is still called as light sleep stage but more difficult to be awakened. Blood pressure and body temperature decrease and heart rate slows down in order to prepare body for deep sleep. It is possible to score an epoch stage 2 by observing either or both K-Complex and sleep spindles. These distinguishing phenomena belong to stage 2.

Stage 3 of NREM:

Stage 3 is called deep, delta or slow-wave-stage (SWS). Comparing to stage 1 and 2 it is more difficult to awaken the subject. In deep sleep stages, the body repairs and regrows tissues and relaxes the muscles. Delta brain waves with frequencies of 0.5 to 4Hz along with a small number of spindles (comparing to the stage 2) can be noticed. One epoch can be considered as a stage 3 if more than 20% of it is occupied by SWS.

Stage REM:

This is the final stage of sleep. If a person enters this stage, his/her brain will become more active. It is during this stage of sleep that dreams usually happen. Also, sleep problems such as sleepwalking, bedwetting and sleep talking happen in REM stage. One of the most noticeable and distinguishable changes is rapid eye movement (REM). The similarity between stage REM and stage 1 through the PSG signals can be noticed. Usually, a person can enter stage REM after 90 minutes of sleep with the first period typically lasting for 10 minutes.

NREM and REM sleep occur in irregular cycles, each lasting approximately 90 to 100 minutes, with a total of 4 to 6 cycles. Generally, on a healthy young adult, NREM sleep accounts for 75-90% of total sleep time (3-5% for Stage 1, 50-60% for Stage 2, and 10-20% for Stages 3) while REM sleep accounts for 10-25% of total sleep time [1].

1-1-2 Microstructure

Microstructure of sleep is characterized by EEG patterns of different morphology and short duration. These patterns are called phasic events. Phasic events are brain, muscle, or autonomic related events of a brief and episodic nature occurring in sleep [2], [3]. These events are superimposed on background rhythms and represent sudden changes on the ongoing activity, both in amplitude and morphology. Phasic events appear spontaneously and may be elicited in the same form by different kinds of sensory stimuli, regardless of its modality. Their EEG morphological features depend on intensity and biological meaning of stimulus and level of sleep where the stimulus arrives. Phasic events, common in different parts of sleep, are described below:

Phasic events of the sleep onset:

- Micro-sleep;
- Vertex sharp transients (vertex sharp waves);
- Positive occipital sharp transients of sleep (POSTS);
- Slow-rolling eye movements;
- Blinks and rapid eye movements of sleep onset;
- Theta bursts;

Phasic events of NREM sleep:

- Spindles;
- K-Complexes;
- Delta burst;

Phasic events of REM sleep:

- Rapid eye movements in REM sleep;
- Saw-tooth waves (STWs);
- Twitch;

Phasic events appearing in all sleep stages:

- Arousals;
- Body movement;
- Stage shifts (SS);
- Stage latency;

Sleep Stage	Brain Waves Defining EEG Frequency	Type	Characteristics	First Appearance	Additional Comments	Cognitive Events
Alert wakefulness	Fast, with many waves >13 Hz	Beta 16 – 20 Hz	Low voltage, random pattern, with few rhythmic components			
Relaxed wakefulness	8-13 Hz	Alpha 8 – 12 Hz	Low voltage, rhythmic alpha, with occasional bursts of the alertness pattern	The person is relaxed or drowsy, with eyes closed.		
NREM Stage 1 (S1)	Low amplitude 3-7 Hz	Theta 4 – 8 Hz	Theta waves interspersed with brief periods of alpha waves, Sleep Onset, Vertex sharp waves with duration <0.5 sec distinguishable from the background activity	As soon as alpha waves are < 50% of a 30-second epoch	Reactivity to outside stimuli diminished, while subjectively the sleeper may still feel awake with slow eye movement	Drifting thoughts
NREM Stage 2 (S2)	12-14 lasting >0.5 seconds Isolated slow/high amplitude waves	Theta with sleep *Spindles* and *K-complexes*	At least 1 sleep spindle or K-complex per 30 seconds on a S1 background, Sometimes K-complexes are followed by spindles	When the first sleep spindle or K-complex appears	The most prominent sleep stage, deeper than S1, lighter than S3	
NREM Stage 3 (S3)	< 4 Hz	Delta	High amplitude >75 µv, slow wave activity, Spindles may persist	Occurs within 15-45 minutes after sleep onset	Deepest sleep; duration of SWS depends on age (less in the elderly), Eye movement are not typically seen	Pre-lucid & Non-lucid Dreams 54% of the time
REM sleep	S1 pattern with "saw tooth waves"	Low voltage, random, fast	Conjugate, sharply peaked eyes movement; the autonomic system is activated (e.g. respiratory and cardiac irregularities)	First REM period occurs after ~85 minutes of NREM sleep	A unique state, during which dreams usually occur. The brain is awake; the body is paralyzed (REM-related Atonia), Low EMG, Saw teeth waves	Lucid dreams 60-90% of the time

Table 1. Sleep stages characteristics (AASM V2.1) [1].

Autonomic sleep - related events:

- Modifications of heart rate (HR);
- Blood pressure-related arousals;

Cyclic alternating patterns (CAPs):

- K-Complexes;
- Delta burst;
- Vertex sharp transients (vertex sharp waves);
- Polyphasic burst;
- K-alpha;
- EEG arousals;

1-1-3 Classifiers

Classification is the process of categorizing data into relevant groups. The first step in the classification process is the identification of features or characteristics that will enable the discrimination between the different groups of data. A classification model should be developed in a way that provides a structure for how the classification processes' actions will be realized. Ideally, this model should be chosen to optimize the system performance, although it may need to be revised as the classifier design progresses. A classifier is then implemented and "trained" to recognize the chosen features in the data, or to determine the best input-to-output mapping. Generally, there are two ways to train a classifier: supervised learning and unsupervised learning. A system is called supervised learning if it uses data labelled by the expert to create an optimal response for the system, which is used as feedback to the learning system to increase accuracy. In contrast, unsupervised learning occurs when the system does not use any labelled data to modify its output. Once the system has trained and learned, it is ready to recognize and classify specific inputs. It can be tested and evaluated with such metrics as speed of computation and accuracy of classification.

1-2 Motivation

The function of human body is frequently associated with signals of electrical, chemical, or acoustic origin. Such signals convey information that may not be immediately perceived because it is hidden in the signal's structure. However, signals' complexity is often considerable and therefore, the biomedical signal processing has become a vital tool for extracting clinically significant information hidden in signals. The artefacts such as body movements, sweating and sensor fault can reduce the accuracy in signal processing especially in sleep signal analysis. The conventional solution is to detect the artefacts and denoise the signal by removing corresponding epochs from the sleep signal. However, this way, the EEG signal will be manipulated and may lose important information. One of the motivations of this thesis is to develop and improve noise cancelation method that does not manipulate the signal and protect its originality [4].

Deep learning is an emerging technique that can be applied to a broad field of science in order to improve learning and classification algorithms. Deep learning is rarely used to classify biosignals and still there is a lack of applying this technique to sleep staging problems. It is proven that shallow learning techniques are not adequate tools to discriminate among stages (e.g. still there is accuracy reduction in detecting S1).

In this research work we will focus on pre-processing (filtering boundaries and de-noising algorithms) and classification for automatic sleep stage classification. Since a contaminated EEG signal would negatively affect sleep signal analysis quality, the main motivation for this work is to develop a pre-processing and classification framework to clean the input EEG signal without manipulating the original data and enhance learning stage of classifier through deep learning methods.

Chapter 2

2 Research Question

2-1 Main Research Question

As mentioned above, a new challenge emerges when dealing with "Deep Learning" within sleep stage classification. In addition, several studies reveal that inadequate training of a classifier is very often the cause for misclassification and conflicts that, in some cases, might lead to reduction in the accuracy. Moreover, artefacts always have been an obstacle for signal processing tools in the extraction and understanding of the better features. If we take for instance, the adaptive filtering techniques and develop them to detect and cancel the noise and estimate a clean signal, it is likely that automatic sleep classification could be a trustable tool either for experts to analysis the sleep signal faster and better or for patients themselves to analyse their own sleep quality. If a proper learning method is achieved, the model built by the classifier could be able not only to discriminate better the stages and even the artefacts, but also to detect sleep disorders such as apnea, insomnia, rapid eye movement sleep behaviour disorder, somniphobia, etc. In this way, monitoring human sleep could be done in home environment with lots of benefits such as: 1) each person can evaluate his/her own sleep quality before going to the hospital, 2) a cyber physical system would be organized, 3) huge reduction in the queues of hospitals will be noticeable, 4) sleep monitoring could be done on a much earlier stage due to the easiest and more convenient access to the diagnose tools. Considering these challenges, the main research question for this work is the following:

> How to design a system for automatic sleep stage classification that is capable of denoising EEG signals and has an efficient and fast learning phase?

In addition, five more detailed questions are proposed with the aim of better solidifying the main research question, two of them are related to the artefact cancelation and three related to the hybrid classification structure.

The first question is related to the noise detection:

> How to detect other noises (e.g. sweat and sensor fault) rather than eye blink and body movement to cancel them from EEG signal?

The second question is related to adaptive noise cancelation structures:

What kind of adaptive noise cancelation techniques are more suitable for EEG signals?

The third question is related to the learning phase:

What kind of un-supervised learning method as a first component of the hybrid structure performs better?

The fourth question is related to supervised-classification:

Does Deep Learning method overcome other structures from accuracy point of view?

Finally, considering the feature extraction characteristics of deep learning networks, a question emerges:

Considering conventional feature extraction methods, does deep learning networks extract more discriminative features or not?

2-2 Hypothesis and Approach

The proposed hypothesis to address this question is:

A reliable automatic sleep stage classification with the desired performance can be achieved if;

- Besides the conventional filtering and artifact removal algorithms, a method for canceling the artifacts from EEG signal without manipulating the original data will be designed.
- The hybrid structure of supervised and un-supervised learning is utilized to enhance the accuracy and reduce the learning time.

To give a detailed answer to the previously mentioned research questions, our research methodology is divided into two main sections: 1) Denoising, 2) Classification. In the denoising section, first of all, the structure of the noise will be studied in order to find suitable threshold to detect the noises. Second, different kinds of adaptive filters will be applied to find the best structure.

In the classification section, advanced learning methods such as deep learning will be applied to enhance the performance of sleep stage classification.

2-3 Research Method

This section presents the aimed contribution to answer the proposed main research question. The detailed planed work was developed according to the classical research method as shown in figure 1.

Figure 1. The classical research method. Adapted from the hand-outs of the Scientific Research Methodologies and Technologies course of the PhD program in Electrical and Computer Engineering by Professor Luis Camarinha-Matos.

2-3-1 Aimed Contribution

To give answers to the proposed research questions, the adopted approach will develop a system that will comprise the following main blocks:

i. Noise detection:
   - Study structure of the noises.
   - Building a framework aimed at finding better thresholds to detect the noise.

ii. Denoising:
   - Applying an adaptive filter structure to cancel the noises.
   - Parameter configuration of adaptive filter to find the most efficient denoising.

iii. Classifier:

- Exploring different kind of classifiers in sleep stage classification.
- Applying deep learning methods for sleep stage classification in order to increase the accuracy.

Chapter 3

3 Literature Review

3-1 PSG Channel Selection

In hospital environment doctors use AASM rules for manual sleep scoring. There are a few recommended parameters that should be reported for a PSG study. At least, three EEG channels (frontal, central and occipital derivations) plus two EOG channels (from left and right eyes) and two chin EMG channels are necessary to perform manual sleep scoring. Recommendation for EEG channels by AASM consist of F4-M1, C4-M1 and O2-M1 and in the case which those channels are not available, alternative EEG channels set includes Fz-Cz, Cz-Oz, and C4-M1.

Taking the above paragraph to the consideration, researchers try to emulate the visual sleep scoring process by using a proper subset of PSG recordings in automatic sleep stage classification. The usual subset consists of EEG, submental EMG and EOG. In Table 2, a summary of PSG subsets used in the literature is presented. Papers summarized in this table include studies that classify sleep recordings into 2 stages (REM/Non-REM or Sleep/Wake), 3 stages, 4 stages, 5 stages or 6 stages. Studies that detect patterns such as spindles, k-complex or sleep disorder detection papers are not included in this table.

Table 2. Summary of PSG subsets used in sleep stage classification.

Subset Type	Signal	Channels	References
Single Channel	EEG	C3-A2	[5]–[22]
		C4-A1	[23][16][20][24]
		C3-A1	[25]
		Fpz-Cz/Pz-Oz	[11], [13], [18], [26]–[43]
		F3-A2	[20]
		F4-A1	[20]
		O1-A2	[20]
		Cz-Pz	[44]
		A1-A2	[45]
		Cz-A1	[46]
	EOG	Left EOG	[47]
		E2-E1	[48]
	ECG		[49]–[51]
Multi-Channel	EEG, EOG, and EMG	EEG (C3-A2), Left and Right EOG, and chin EMG	[52]
		Six EEG (F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, O2-A1), Left and Right EOG, and chin EMG	[53], [54]
		Four EEG channels (C3-A2, P3-A2, C4-A1, and P4-A1), one horizontal EOG and one chin EMG	[55]

	EEG and EOG	EEG (C3 and Cz), Left and Right EOG	[56]
		Six EEG channels (F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, O2-A1) and two EOG channels (Left and Right)	[57][58]
		EEG (Pz-Oz) and Horizontal EOG	[59]
		Two EEG (Fz and Oz) and two EOG (Left and Right) Channels	[60]
		EEG (C4-M1), EOG	[61]
	Heart Rate, Breathing Rate and Movement Information	Heart Rate, Breathing Rate and Movement Information	[62]
	EEG, ECG and Respiration Features	EEG (C1-A2), ECG and Respiration Features	[63]
	ECG and Respiration	ECG and respiratory inductance plethysmography (RIP)	[64]
	EEG and EMG	EEG (C4-M1) and chin EMG	[65]
		EEG (C3-A2) and chin EMG	[66] [67]
	ECG, Respiratory and actigraphy and signals	ECG, Respiratory and actigraphy and signals	[68]
	EEG	Fp1-C3, Fp2-C4, Fp1-T3 and Fp2-T4	[69]
		Pz, Cz, Pz, T3, T4	[70]
		Six EEG channels (Fp1–M2, C3–M2, O1–M2, Fp2–M1, C4–M1, and O2–M1)	[71]
		Fpz-Cz and Pz-Oz	[72], [73]
		C3-A2 and C4-A1	[74] [75]
		Six EEG Channels (F3-A2, C3-A2, O1-A2, F4-A1, C4-A1 and O2-A1)	[76]
		C4-A1, O2-A1 and C3-O1	[77]
	EOG	Left and Right	[78][79]

	EEG, EOG, EMG and ECG	Six EEG channels (Fp1–M2, C3–M2, O1–M2, Fp2–M1, C4–M1, and O2–M1), two EOG channels (Left and Right), one chin EMG channel and ECG.	[71][80]

3-2 Pre-Processing

3-2-1 Introduction

The presence of artefacts might lead to the misapprehension, low-accuracy and distorted quantitative results. Therefore, a pre-processing step is necessary to cancel artefacts and remove cropped epochs to magnify informative components of raw EEG, EOG and EMG signals prior to any further analysis [28], [81].

Considering the fact that nowadays portable devices for patient monitoring and automatic sleep stage classification could be a helpful assistance for experts on the analysis of sleep signals, the main motivation for the current work is the lack of a systematic method for automatic artefact detection and cancellation which leads to an improvement in the automatic stage classification accuracy compared to the original acquired data.

In this section of the thesis two of main issues will be discussed: the filtering criteria for bio-signals (i.e. EEG, EOG and EMG) and the denoising of EEG signals. The artefacts diffused on EEG channel may mask cerebral activity and simulate sleep phasic events such as sharp vertex waves, K-complexes and spindles. Additionally, some other noises and artefacts are introduced by the measurement devices as well. The power line interference (the 50/60 Hz components) or movements of electrodes can impose variation in the baseline characteristics of PSG signals. Also, some artefacts are created from non-cerebral sources such as eye/muscle movement and cardiac sources. In addition, considering the relatively high amplitude level of ECG signals, the QRS complex regularly interfere in the EEG signals causing spiky patterns [82]–[84].

3-2-2 Filtering the data

Filtering is an important step in the pre-processing of any sleep signal analysis algorithm. It means that, if the signal is not filtered properly, then the artefacts with low frequency and high frequency range (i.e. as it is proved that the main energy of the EEG is between 0.3-35 Hz according to AASM) will affect the classification results.

The AASM manual provides some recommendations which are summarized in Table 3.

Table 3. Filtering rules considering AASM manual [1].

Filter Settings	Low Frequency	High Frequency
EEG	0.3 Hz	35 Hz
EOG	0.3 Hz	35 Hz
EMG	10 Hz	100 Hz
ECG	0.3 Hz	70 Hz

These rules are usually considered in papers. Here the big challenge is to follow the AASM rules or not. There are some papers which authors created their own rules to pre-process the data in order to increase accuracy. However, a comprehensive study to compare different filtering rules is still missing. Table 4 shows a survey of the recent papers from the filtering-band point of view.

Table 4. Different band-pass filtering.

Authors	EEG	EOG	EMG
Maja Čić [69]	0.3 – 70 Hz	-	-
Sheng-Fu Liang [17]	0.1 – 75 Hz	0.1 – 75 Hz	10 – 75 Hz
Mohamed ElMessidi [47]	-	0.5 – 30 Hz	-
Han G. Jo [21]	0.5 – 100 Hz	0.5 – 100 Hz	0.5 – 100 Hz
Sirvan Khalighi [85]	0.3 – 35 Hz	0.3 – 35 Hz	10 – 70 Hz
Jun Shi [75]	0.3 – 64 Hz	-	-
Sheng-Fu Liang [79]	0.5 – 30 Hz	0.5 – 30 Hz	5 – 100 Hz
Jing Zhou [86]	0.1 – 40 Hz	-	-
Tarek Lajnef [56]	0.2 – 40 Hz	-	-
Salih Gunes [24]	0.1 – 60 Hz	-	-
Baha Şen [6]	0.1 – 60 Hz	-	-
Seung-Hyeon Oh [87]	0.5 – 50 Hz	-	-
Florian Chapotot [65]	0.3 – 35 Hz	0.3 – 35 Hz	0.3 – 35 Hz
Jose L. R. Sotelo [72]	0.5 – 100 Hz	-	-
Teresa Sousa [57]	0.5 – 45 Hz	0.5 – 45 Hz	-
Pedro Pinero [88]	0.5 – 30 Hz	0.5 – 30 Hz	10 – 100 Hz
Seral Ozsen [89]	0.5 – 35 Hz	0.5 – 35 Hz	10 – 70 Hz
Jinwoo Kim [5]	0.3 – 50 Hz	-	-
Mustafa Radha [20]	0.6 – 27 Hz	-	-
Junming Zhang [90]	0.5 – 32 Hz	0.5 – 32 Hz	5 – 32 Hz

3-2-3 Data Denoising

PSG recordings include EEG, EMG, EOG, ECG, snoring and other physiological signals to detect body movements. EEG is widely adopted for the automatic detection of sleep stages and neuronal activity evaluation during sleep. However, EEG is usually contaminated with several artefacts such as power line noise, EMG, EOG and electrode movements. Removal or attenuation of noise and unwanted signals is a prerequisite for most of the EEG signal processing applications. The presence of artefacts makes the EEG analysis difficult, since it may introduce spikes that can be confused with the original EEG trend, decreasing the reliability of the subsequent processing stages.

Some of these artefacts are easily removed by a finite impulse response (FIR) or infinite impulse response (IIR) filter if their power spectrum doesn't have overlap with the EEG power spectrum. However, EMG and EOG, which have significant power spectrum overlap with EEG are not easy to remove requiring careful consideration [91].

A commonly used method for avoiding artefacts is the rejection of the contaminated segments of the recorded EEG [92]–[94]. This method, although simple, results in a huge data loss and was an early technique of management artefacts. However, nowadays, with the emerging of recent signal processing techniques, the preference is for artefact cancelling or correcting techniques, instead of rejecting the data epoch [95]. Consequently, denoising the contaminated EEG segments would not only preserve the amount of data, but would also contribute to the increase of the accuracy in the automatic sleep stage classification [96]. Following this idea, an extensive number of studies have tried to extract the clean EEG out of the contaminated recording in different research areas, but still no optimal method is agreed upon [95], [97].

EEG is mainly intended for recording cerebral activity, yet other extra electrical activities are also recorded. These extra activities are usually considered harmful artefacts that can be either physiological like EOG, EMG and ECG or extra physiological like power line interference. EOG measures are captured mainly by frontal electrodes, but they are strong enough to also affect other electrodes. About the EMG, the degree and type of contamination depends on the contracted muscle, the recording purpose and the environment [91]. However, EMG and EOG, which have significant power spectrum overlap with EEG, are not easy to remove requiring careful consideration [91]. The state of the art in EEG denoising is quite broad. The most classic methods used in the last years are regression (especially for ocular interferences), blind source separation or component base techniques [98] and Wiener and Bayes filtering methods [97]. Also adaptive filtering [97], wavelet denoising [99] and empirical mode decomposition (EMD) [100] are among the most widely used denoising techniques. A great range of studies pay particular attention to the improvement of the existing methods or using more objective performance criteria [101].

3-2-3-1 Adaptive Filtering for Denoising

An adaptive filter is a linear system that iteratively models its transfer function according to the relationship between the input and the output signals. The parameters are adjusted according to an optimization method in adaptive algorithm [102]. The filter weights can be adapted based on the feedback from the output of the filter, but it needs a reference input to compare the desired output with the observed output.

The output of an adaptive filtering system is given by:

$$y(n) = \mathbf{w}(n)^H .u(n) \tag{1}$$

where $\mathbf{w}$ is the weight vector. The Wiener-Hopf solution gives the optimum weight vector for (1). However, it introduces a high order of computational complexity. A simple recursive solution to the classical wiener filtering problem is the gradient-based algorithms such as steepest descent and Least-Mean-Squares (LMS) optimization techniques. In the steepest descent optimization method, the weight vector is made to evolve in the direction of the negative gradient

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{2}\left[-\nabla\left(E\left\{\varepsilon^2(n)\right\}\right)\right] \tag{2}$$

where $\varepsilon(n)$ is the error signal given by:

$$\varepsilon(n) = d(n) - \mathbf{w}^H .\mathbf{u}(n) \tag{3}$$

and $d(n)$ is the desired signal (given as clean signal). The gradient vector is computed as:

$$-\nabla\left(E\left\{\varepsilon^2(n)\right\}\right) = 2\mathbf{R}.\mathbf{w}(n) - 2\mathbf{r} \tag{4}$$

where $\mathbf{R} = E\left[\mathbf{u}(n).\mathbf{u}^H(n)\right]$ is the autocorrelation matrix of the $\mathbf{u}(n)$ and $\mathbf{r} = E\left[d^*(n).\mathbf{u}^H(n)\right]$ is the cross-correlation vector between $\mathbf{u}(n)$ and $d(n)$. The main disadvantage of this method is the computational complexity involved in the real-time manipulation of the matrices $\mathbf{R}$ and $\mathbf{r}$. The LMS algorithm is basically a simplification of the steepest descent method where instantaneous values of $\mathbf{R}$ and $\mathbf{r}$ are used, and the weight vector is updated as:

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu.\mathbf{u}(n).\varepsilon^{*}(n) \tag{5}$$

The block diagram of an adaptive filter is depicted in figure 2.

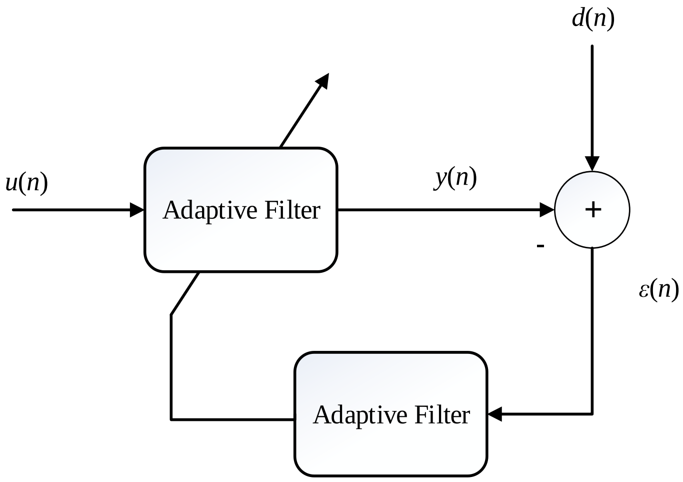

Figure 2. Structure of Adaptive filter configuration

The main drawback of LMS is being too sensitive to the scaling of its input as well as choosing μ directly effects the convergence rate and stability of the filter [102], [103]. The Normalized-Least-Mean-Square (NLMS) filter is a variant of LMS filter which handles this problem by normalizing with squared-norm of the input $u(n)$. Therefore, a time variant step-size $\mu(i)$ is described as follows:

$$\mu(i)=\frac{\mu}{c+\|u(i)\|^{2}} \tag{6}$$

where c is a small constant applied to avoid the probability of getting $\mu(i)$ instable in the case of having value of $\|u(i)\|^{2}$ near to zero.

GuruvaReddy et al. [104] applied LMS and NLMS adaptive filters to remove the power line and baseline noise from EEG. Although the reason of using adaptive filter instead of notch filter to remove 50 Hz noise is not described, they put some results to prove that adaptive filters can remove above mentioned noises. But still a comparison between the two kind of adaptive filters performance with the signal to noise ratio (SNR) is missing.

Senthil Kumar et al. [105] removed EOG artefacts from EEG by adaptive filtering techniques through wavelet transform. LMS filter was chosen and the database was provided from Physionet. The results illustrated that artefacts are removed from the EEG signal successfully, but for the validation part of the paper, just a suppression ratio was introduced which seems that is not significant.

Ahirwal et al. [106] exploited adaptive filtering techniques to remove noise for EEG signal. The proposed method tested with LMS and recursive LMS (RLS) algorithm over electroencephalography/event related potentials (EEG/ERP) noise removal to build up relation between input SNR and output fidelity parameters. Their proposed method got a better SNR and mean square error (MSE) with RLS.

Darroudi et al. instead of MSE which is a conventional criterion [106]–[108] selected error entropy in order to eliminate noise from EEG signals. In this work it is claimed that by deploying MSE, only second-order moment of the error distribution is optimized, which is not adequate for the noisy EEG signal. But minimizing error entropy, all moments of the error distribution are minimized; hence, using the Minimum Error Entropy (MEE) algorithm instead of MSE-based adaptive algorithms will improve the performance of noise elimination. They reached higher SNR with the proposed method for different input SNRs (5, 2, -2 and -5).

3-2-3-2 Wavelet Denoising

In the context of sleep stage classification, Estrada et al. [109] proposed a denoising method based on applying a predefined threshold to the wavelet coefficients of noisy EEG with the objective of finding the best threshold defining rule and value. The noisy signal is constructed by adding white noise ($\mathrm{E}[x]=0, \sigma=20$) and 50 Hz power line sinusoidal noise. The coefficients that contribute to the noise components are zeroed out using a threshold discrimination filter. Preliminary results showed that the combination of a soft thresholding rule applied to the detailed wavelet coefficients with the Universal threshold value produced better performance measures such as a smaller MSE and a larger SNR. Similarly, improved results were obtained for S1, S2, S3, stage 4 (S4) and REM stage EEG signals using this combination.

Noviyanto et al. [50] evaluated the contribution of temporal pattern in sleep stages classification results based on the fact that sleep stages are time series data. Ten datasets of single lead ECG signal from healthy subjects have been collected and fifteen features extracted from raw ECG signal to describe the sleep stages. The smoothing process was applied to the datasets using stationary wavelet transform with "*coiflet*" mother wavelet, four level decomposition, and universal hard thresholding method. An example of denoising process is shown in figure 3.

Still effect of wavelet denoising in sleep classification accuracy is missing which is one of objectives of this thesis.

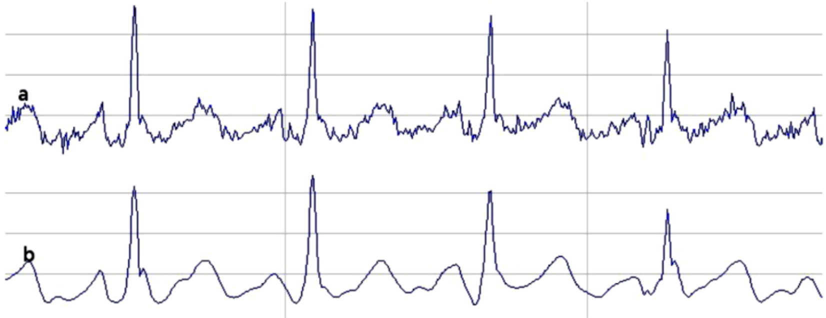

Figure 3. Denoising using wavelet; a) original signal, b) signal after denoising.

3-3 Classification

This section is aimed to give a comprehensive overview about conventional classifiers and present the most recent works in a table to provide a summary and discuss about deep learning techniques which are applied to several classification issues. In case of sleep stage classification, there are few studies applying deep learning techniques.

3-3-1 Conventional Classifiers

Several papers provide evidence for the high performance of Support Vector Machine (SVM) specifically for high dimensional classification problems [56], [110]–[113]. In principle, SVMs are designed for binary classification problems (discrimination between two classes). However, as in many classification tasks, automatic sleep scoring requires discrimination between multiple classes (Awake, S1, S2, S3 and REM). Hence, for getting benefit from the assumed advantages of SVM classification, a multi class SVM framework needs to be implemented. Two of the most widely used approaches for multi-class SVM classification are the One-Against-All (OAA) and the One-Against-One (OAO) approaches. The OAA framework consists of using a binary SVM to distinguish each class from all other classes and the decisions obtained by applying a winner-takes-all strategy. By contrast, in the OAO multi-SVM approach a dedicated classifier is trained for each of all possible pairs of classes. In other words, for a total of N classes, it needs to train $N(N-1)/2$ classifiers and the decision is then obtained by performing a majority vote (max-wins voting).

Benabdeslem et al. [114] proposed a new approach called dendogram based support vector machine (DSVM). Although DSVMs are not as well-known and established as OAO and OAA multi-class SVM methods, decision-tree-based multi-SVM classification has been explored in the machine learning and computer science literature [114]–[117].

Lajnef et al. [56] proposed a sleep staging framework based on a DSVM classification. Its performance was evaluated using polysomnographic data from 15 subjects' EEG, EOG and EMG recordings. They mentioned that the DSVM method is a decision-tree classification framework where each binary classification node is fulfilled by a binary SVM. First of all, a procedure applied consists in defining the tree with its binary branching, i.e. the structure of the dendrogram. This was done by computing the mean values of the features for each class. So, they applied ascendant (agglomerative) hierarchical clustering (AHC). Hierarchical clustering is a cluster analysis method that seeks to build a hierarchy of clusters. Strategies for hierarchical clustering divide into two basic paradigms: agglomerative (bottom-up) and divisive (top-down). Agglomerative strategies start at the bottom and at each level recursively merge a selected pair of clusters into a single cluster. This produces a grouping at the next higher level with one less cluster. The pair chosen for merging consists of the two groups with the smallest intergroup dissimilarity. Divisive methods start at the top and at each level recursively split one of the existing clusters at that level into two new clusters. The split is chosen to produce two new groups with the largest inter-group dissimilarity. With both paradigms, there are $N - 1$ levels in the hierarchy. In the ascendant approach used here, observations that stem from each final class are sequentially merged as one move up the hierarchy [118], [119]. As well as achieving high sensitivity, specificity and accuracy, DSVM outperforms the two standard multi-class procedures (OAA-SVM and LDA). They suggested that the generation of an optimized dendrogram structure using ascending hierarchical clustering is an important source of performance enhancement. It reduces both the number of classifiers and the execution time. In particular, compared to the one-against-one SVM approach that would require 10 SVMs to run a 5-class classification, the dendrogram approach boils down the classification to four binary SVMs. Another advantage of this method is that there is no need for a classification decision stage; any given sample is assigned at the end of the tree to one class.

To better understand multi-channel SVM, all its structures are illustrated in figures 4, 5 and 6.

Pascualvaca et al. [120] used two techniques to improve the accuracy of sleep stage classifiers based on SVMs from EEG, EOG and EMG signals. Moreover, three different SVM multi-classifiers have been tested to evaluate and compare their performance. Just three feature groups are extracted from EEG on their work. For EMG, average signal power and for EOG, saccades density are extracted. Although three different SVM classifiers are analysed (as it can be seen in figures 4 to 6), one-vs-all overcame the others, but still lack of using deep learning techniques and not enough features are noticeable.

As mentioned in [120] Herrera et al. [19] introduced Stacked Sequential Learning (SSL), a meta-learning method in which the base classifier is augmented by making it aware of the labels of nearby patterns, was successfully used for improving the accuracy of the classifier. Nevertheless, in this study only 4 sleep stages were detected (Awake, REM, S1 and S2 considered as Light-Sleep (LS), and S3 and S4 considered as Deep-Sleep (DS)) and EEG as a single source of information was analysed.

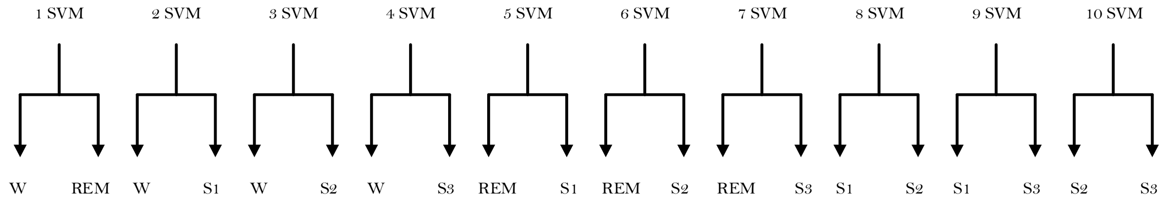

Figure 4. One-Against-One structure.

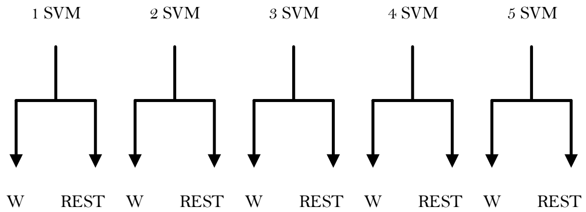

Figure 5. One – Against – All structure.

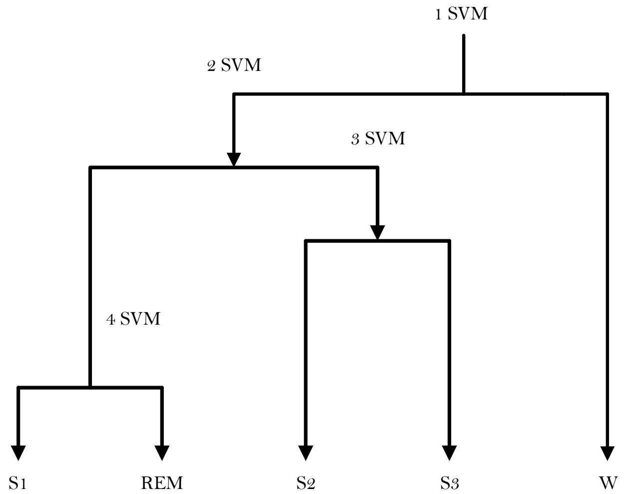

Figure 6. Dendogram based SVM structure.

Wang et al. [45] studied the performance of automated sleep–wake detection algorithm (Z-ALG) used in the Zmachine® (a portable, single-channel, EEG acquisition and analysis system) against laboratory PSG using a consensus of expert visual scorers. Instead of using the channels recommended in AASM, they used A1-A2, EEG channel. Especially this channel, selected because it is located completely outside of the hairline, allows an easy patient self-application and removal of the sensors. Sensors A_1 and A_2 are also located away from the face making it comfortable and unobtrusive to wear during sleep, eliminating the possibility of leaving noticeable marks on the face and forehead. However, the Z-ALG was only able to detect sleep and wake states. Almost one year later they developed another algorithm named Z-PLUS which was able to detect Light Sleep, Deep Sleep, and REM, but still doesn't satisfy the demand of experts to evaluate the sleep data and a comparison between the specified EEG channel with others which are recommended by AASM (e.g. C3-A2) is missing.

Banaee et al. [121] made a review of the latest methods and algorithms used to analyse data from wearable sensors for physiological monitoring of vital signs in healthcare services. In their framework, common data mining tasks have been applied such as anomaly detection, prediction and decision making when considering continuous time series measurements.

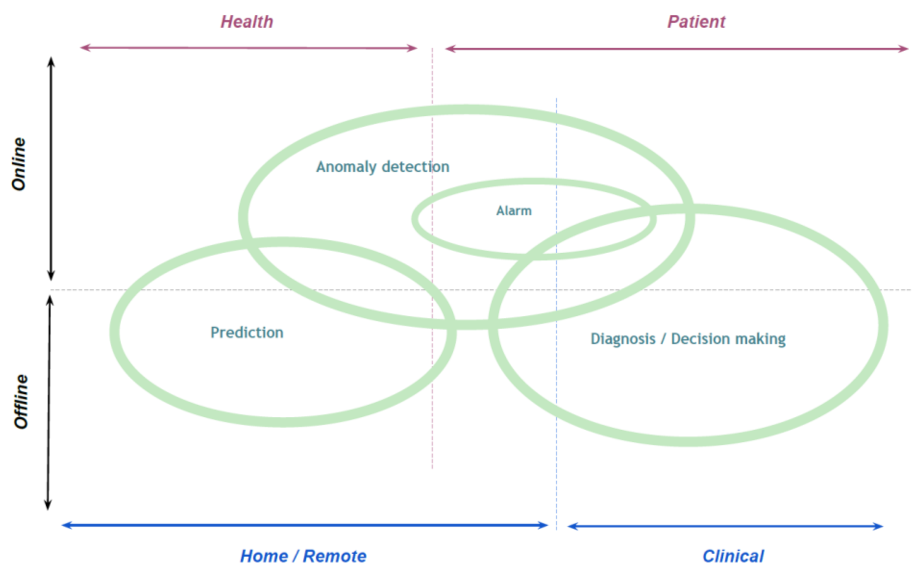

Figure 7. Illustration of each task in relation to the three dimensions (adapted from [121]).

Figure 7 provides an illustration of each task in relation to the three dimensions. The first dimension is dealing with the setting considering the environment. Most monitoring applications which consider home settings or remote monitoring deal predominantly with prediction and anomaly detection whereas the applications in clinical settings are typically focused on diagnosis [121], [122]. This can be easily explained by considering the growing desire to have a more preventative approach by using wearable sensors and increase the sense of security (alarm) in home environment. Likewise, in clinical settings for diagnosis and assist in decision making much more information is available.

The second dimension considers the type of subject for data mining. For patients with medical records, both diagnosis and possibility to have alarms are important whereas, for healthy subjects ensuring the maintenance of good health, prediction and anomaly detection are essential.

The third dimension describes the method of processing the data. All three tasks (anomaly detection, prediction and decision making) have been addressed both in online and offline scenarios. Alarm related tasks are usually used in the context of online continuous monitoring. In their work, usability, efficiency and challenges of each technique in the medical domain are indicated. Banaee et al. [121] outline the most common algorithms used with wearable sensor data.

SVM techniques are often proposed for anomaly detection and decision-making tasks in healthcare services. However, SVM is not an appropriate method to integrate domain knowledge to use metadata or symbolic knowledge seamlessly with the measurements from the sensors. Moreover, like other classifiers, SVM cannot be applied to find the unexpected information from unlabelled data.

Since the progress of learning in Neural Network (NN) would be complex, the method is regularly used for decision-making in clinical conditions with large and complicated data sets. But same as SVM, NN is not able to handle domain knowledge to improve the results. Furthermore, as the modelling process in NN is a black box progress, NN method needs to justify for each input data. So, NN is not counted as a portable technique to easily apply for diverse data sets.

Decision tree (DT) methods are restricted to the space of the constructed features as the inputs of the model. So, finding hidden information out of constricted features would not be recognizable. Furthermore, since the number of features can impact on the efficiency of the method, DT models are not usually applied to big and complex physiological data. Nonetheless, DT is one of the preferable classifiers considering the fact that it is simple and easy to implement and understand.

Even though Gaussian Mixture Model (GMM) is able to detect unseen information in physiological data, it has been rarely used for prediction tasks. Since the computation time of constructing the models is high, applying GMM in real-time is usually not affordable. Moreover, the initialization step in GMM does not necessarily improve the modelling (as GMM approached assume data as a combination of Gaussian distinction) due to the character of health parameters in real world situations.

The Hidden Markov Model (HMM) is used for anomaly detection rather than any other tasks. Moreover, this method has not been applied to multi health parameters analysis and big data sets. However, based on the abilities of HMM to model the unexpected behaviours of the data, it is applicable to use the modified versions of HMM in more problems in healthcare domains.

Through the mentioned review paper [121], the techniques which are most applied to ECG analysis as data mining tools are SVM, NN and DT, respectively and for HR are SVM and statistical methods, respectively. Also, they provide an outline of the three most commonly used data mining tasks in relation to the vital signs that can be measured by wearable sensors. ECG provides mostly the rich data, which is predominantly used for all tasks in comparison to the other types of sensors. Next, with a huge difference, HR is used to analyse the subject.

ElMessidi et al. [47] introduce a statistical based solution which explores a simple rule based method to accurately predict the occurrence of S3. This automatic method detects S3 epochs using a single-channel EOG based on a simple rule-based algorithm with adaptive thresholds. They evaluated their method through 9 healthy subjects and the results are compared to the clinical visual scoring. The agreement of the detection method for the validation data was 90.0%, the sensitivity was 90.5%, the specificity was 89.9% and the kappa value was 0.74. The results look promising but only S3 and non-S3 epochs are detected.

Fraiwan et al. [25] used single channel EEG to classify sleep stages. Three time–frequency techniques were deployed for the analysis of the EEG signal: Choi–Williams distribution (CWD), continuous wavelet transform (CWT), and Hilbert–Huang transform (HHT). The classification of the extracted features was done using Random Forest (RF) classifier. Among all time–frequency techniques, they reached the best performance by using RF classifier by choosing CWT and RF classifier from Weka classification software package.

Breiman [123], which consists of many individual classification trees (in their work they chose 10 trees), where each tree is a classifier by itself that is given a certain weight for its classification output (in Weka all trees are given the same weight). The classification outputs from all trees are used to determine the overall classification output which is done by choosing the mode (the output with most votes) of all tree's classification output (figure 8). Although the authors claimed that RF outperforms the other classifiers, this comparison is missing.

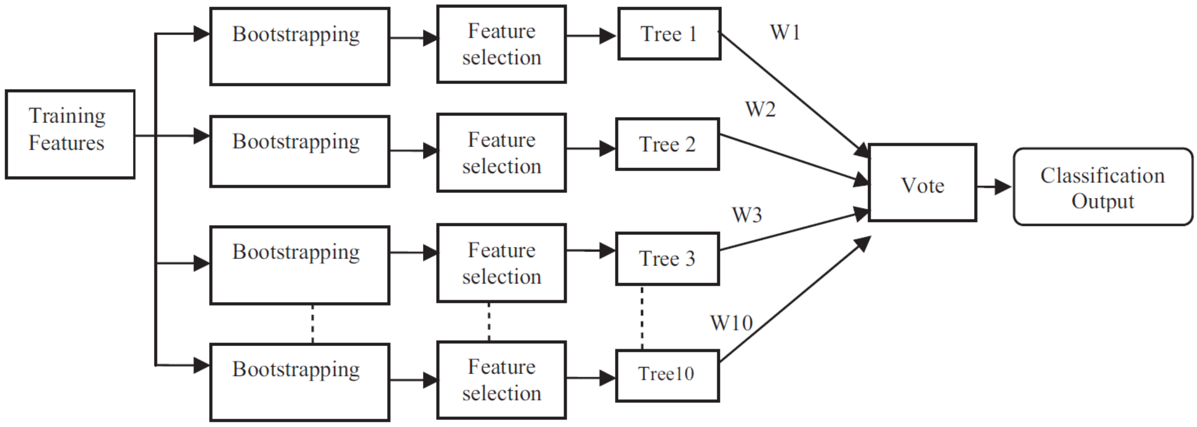

Figure 8. Flowchart of the RF classifier showing the different steps of the classification procedure.

As mentioned before, one of the future works proposed in this thesis is the reduction of the necessary channel numbers to extract the features and classify the sleep stages, which will result in the reduction of the computational complexity and hardware requirement.

Moreover, looking from another point of view and considering home-based settings, data acquisition instruments are also important in order to make a comfortable environment for the subject to sleep and achieve user friendly devices (which anyone can use without the support of an expert).

Considering this aspect, Fonseca et al. [64] presented a method to identify overnight sleep stages using cardiorespiratory features extracted from ECG and RIP signals. After extracting feature and post-processing them, they used a multi-class Bayesian linear discriminant with time-varying prior probabilities [124], to classify the stages. By choosing this specific classifier considering the non-linear structure of ECG, multi-class Bayesian linear discriminant couldn't be a proper classifier in this case. It can be described simply as following: non-linear classifiers will have a higher risk of overfitting, since they have more dimensions of freedom. They can suffer from learning just good set of data points, rather than coming up with a good generalization. On the other hand, a linear classifier has less freedom to fit, and in the case of non-linearly separable data, will fail to find a good decision function and suffer from high error rates.

Liang et al. [79] designed their own eye mask to record EOG (right and left) using conventional electrodes which can be easily worn by a patient at home. The idea was to integrate a sleep eye mask with EOG electrodes. Conventional EOG electrodes are made of adhesive pads or conductive paste and are neither reusable nor easy to use. Therefore, they replaced conventional electrodes with a material that can be reused and easily combined with the cloth eye mask. Multiscale entropy (MSE) and autoregressive (AR) coefficients of the theta band signals are extracted from the EOG. Linear discriminant analysis (LDA) were adopted to classify the five sleep stages based on the extracted MSE values and the AR coefficients. After classifying the sleep stage by LDA [125], a smoothing technique [126] was used to improve the classification accuracy. LDA finds a hyperplane that best separates two or more classes of objects or events by adjusting the linear weighting of features. They performed two experiments. The first experiment was designed to compare the eye mask system with manual scoring by experts during overnight sleep. These EOG signals included several sleep cycles of five sleep stages (Wake/S1/S2/S3/REM).

The second experiment was designed to evaluate, online, the overall agreement during nap sleep. They woke the subjects up when they first reached an S3, or after a 30-min naptime. The recordings contained only three sleep stages (Wake/S1/S2). In the overnight experiment, the overall agreement between the computer scoring and the manual scoring was 84.33%. In the nap experiment, the overall agreement was 83.08%.

Partial least squares regression (PLS) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of minimum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables into a new space [127]. PLS is an extended class of methods for modelling relations between sets of observed variables by means of latent variables. By encoding the class membership in a suitable indicator matrix, PLS can also be applied to classification problems [39]. Kayikcioglu et al. compared PLS performance by means of accuracy and time with *k*-nearest neighbourhood (*k*-NN), linear discriminant classifier (LDC) and Bayes. Figure 9 shows the flowchart of the proposed classification of single channel EEG sleep method that includes three parts: (1) pre-processing; (2) feature extraction; and (3) classification. They proved that for two-class staging (Sleep and Wake) PLS overcome other algorithms and is more suitable for real-time application. It is necessary to consider two points: first the performance of LDC is almost similar to PLS and secondly (to the best of our knowledge) still there is no comparison between PLS and other classifiers for full stage classification (W, S1, S2, S3 and REM).

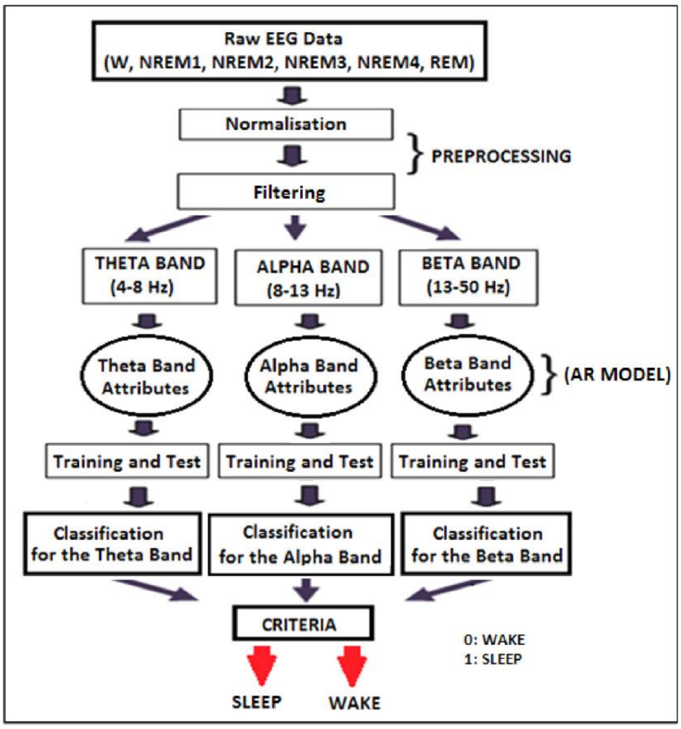

Figure 9. Flowchart of the proposed single channel EEG classification (adapted from [39]).

Huang et al. [128] introduced a novel method for training to overcome the typical slowness of the artificial neural network (ANN). It is clear that the learning of feedforward neural networks is in general far slower than required and it has been a major bottleneck in their application during the past decades. Two key reasons behind may be: (1) the slow gradient-based learning algorithms that are extensively used to train neural networks, and (2) all the parameters of the networks are tuned iteratively by using such learning algorithms. Unlike these conventional implementations, Huang et al. proposes a new learning algorithm called extreme learning machine (ELM) for single-hidden layer feedforward neural networks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs.

ELM was originally inspired by biological learning and proposed to overcome the challenging issues faced by Back Propagation learning algorithms. It also aims to provide a biologically inspired simple and efficient unified learning framework filling the gap between artificial learning methods and biological learning mechanisms.

It was demonstrated that ELM has several interesting and significant features different from traditional gradient-based learning algorithms for feedforward neural networks such as:

i. The learning speed of ELM is extremely fast.

ii. In most cases the proposed ELM has better generalization performance than gradient-based learning algorithms such as backpropagation.

iii. ELM tends to reach solutions straightforward overcoming the traditional gradient-based learning algorithms issues such as like local minima, improper learning rate and overfitting.

iv. The analysis of the SLFNs performance with different activation functions is not done and the authors mention that as a future work.

Several years later Huang [129] extended ELM to SLFNs with radial basis function (RBF) kernels which randomly generate the centres and impact widths of RBF kernels and simply analyse the output weights instead of iteratively tuned.

Chen et al. [60] compared SVM with two kinds of ELM: basic ELM with sigmoid activation function and ELM with RBF kernels. The ELM achieved the best classification results with 20 hidden neurons and there was one node in the output layer whose target value was defined as: 1 – drowsiness, 0 – alertness.

It is almost impossible to discuss about classifiers without mentioning the ensemble learning technique [130]. This technique is derived from the principles of regular machine learning topic and uses multiple learning algorithms to obtain better results than conventional learning algorithms. Ensemble learning is a technique for combining a number of base learners in an attempt to produce better results. Basic machine learning classifiers such as bayes, adaboost, DTs, etc. are called base learners or weak classifiers in ensemble methods. Avci et al. [131] provided an efficient and robust method classifying the minute based occurrence of sleep apnea with respiration signals. They implemented three different ensemble learning classifiers such as: adaboost (as an ensemble learning method), RF and Random Subspace. However, the best result was obtained analysing nasal based respiratory signal by using the RF method. In this case, accuracy, F-measure and kappa obtained 98.68%, 0.981 and 0.972, respectively.

Gunes et al. [24] introduced a novel pre-processing method for features called k-means clustering based on feature weighting (KMCFW) and combined it with *k*-NN and C4.5 DT classifiers to discriminate six stages from sleep EEG including Wake, S1, S2, S3, REM and non-sleep (Movement time). Finally, the weighted sleep features have been automatically classified into six sleep stages using *k*-NN and C4.5 DT classifiers. In the classification of sleep stages, the *k* values of 10, 20, 30, 40, 50, and 60 in *k*-NN classifier have been used and compared with each other. In the experimental results, while sleep features have been classified with 55.88% success rate using *k*-NN classifier (*k* value of 40), the weighted sleep features with KMCFW has been recognized with 82.15% success rate *k*-NN classifier (*k* value of 40). They claimed that their system could be used as an online system automatic scoring of sleep stages, which that is arguable from at least two points of view:

i. Classifier:

   Taking into account that *k*-NNs are *k* values based, there are other classifiers which can easily overcome it like RF, DSVM, etc.

ii. Bioelectrical signal:

   Nowadays, bioelectrical signals are preferable which are easily applicable[1] and comfortable from the subject point of view[2].

Generally, neural networks can be divided into two main classes. One class contains feedforward neural networks (FFNNs), and the other contains recurrent neural networks (RNNs). The essential difference between FFNNs and RNNs is the presence of a feedback mechanism among the neurons in the latter. FFNN is a network without any feedback connections among its neurons, while an RNN has at least one feedback connection. Since RNNs allow feedback connections in neurons, the network topology can be very general: any neuron can be connected to any other, even to itself. Allowing the presence of feedback connections among neurons has an advantage: it leads naturally to an analysis of the network as a dynamic system, in which the state of the network, at one moment in time, depends on its state at a previous moment in time [132].

Examples of feedforward networks include the multi-layer perceptron (MLP) [133], the learning vector quantization (LVQ) network [134], the cerebellar model articulation control (CMAC) network [135] and the

[1] Considering that currently people prefer to analyse their sleep signals at home rather than waiting in huge list for hospitals, sleep scoring and analysis devices should be easily used by ordinary patients.
[2] Reducing the number of sensor connected to the patient will allow him/her to sleep better.

group-method of data handling (GMOH) network [136]. Feedforward networks can most naturally perform static mappings between an input space and an output space: the output at a given instant is a function only of the input at that instant and FFNs are extensively used in pattern recognition.

Examples of recurrent networks include the Hopfield network [137], the Elman network [138] and the Jordan network [139]. Recurrent networks have a dynamic memory: their outputs at a given instant reflect the current input as well as previous inputs and outputs. The comparison between feedforward and feedback-ward neural networks always has been an attractive task for researchers in various applications such as time series forecasting, equalizers, speech denoising, etc. [140]–[142]. Therefore, it can be proven that RNNs outperforms the FFNNs in the case of training and classification accuracy.

Hsu et al. [37] used energy features extracted from characteristic waves of EEG signals. The classification performance of a FFNN and a probabilistic neural network (PNN) were compared to proposed classifier. They utilized Elman network as a recurrent neural classifier for categorizing human sleep stages. For classification problems, RNNs make efficient use of temporal information in the input sequence. The classification rate of the recurrent neural classifier (with 6 hidden neurons) was found to be better (87.2%) than those of the two neural classifiers (81.1% for FFNN with 6 hidden neurons and 81.8% for PNN).

Şen et al. [6] did a comparative study on the classification of sleep stages. They compared RF, FFNN, DT, SVM and RBF neural networks from accuracy, time consumption and feature selection points of view. They have provided 6 experiments and in each one a specific feature selection algorithm was used to feed features to the classifiers as follows:

Exp. 1: Fisher score feature selection algorithm,

Exp. 2: mRMR feature selection algorithm,

Exp. 3: *t*-test feature selection algorithm,

Exp. 4: ReliefF feature selection algorithm,

Exp. 5: Fast Correlation Based Filter (FCBF) feature selection algorithm,

Exp. 6: A Hybrid Approach.

Figure 10 to 15 display the classification accuracy rates and the computation times for experiments 1 to 6.

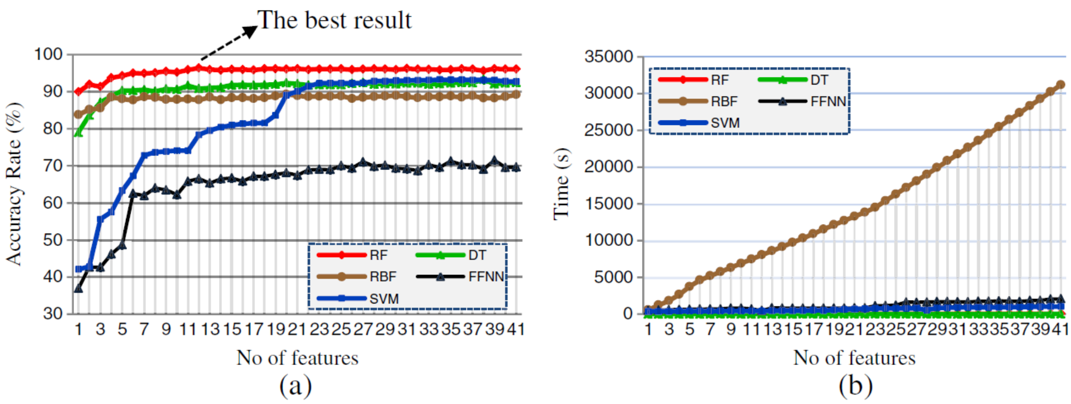

Figure 10. Results for experiment 1.

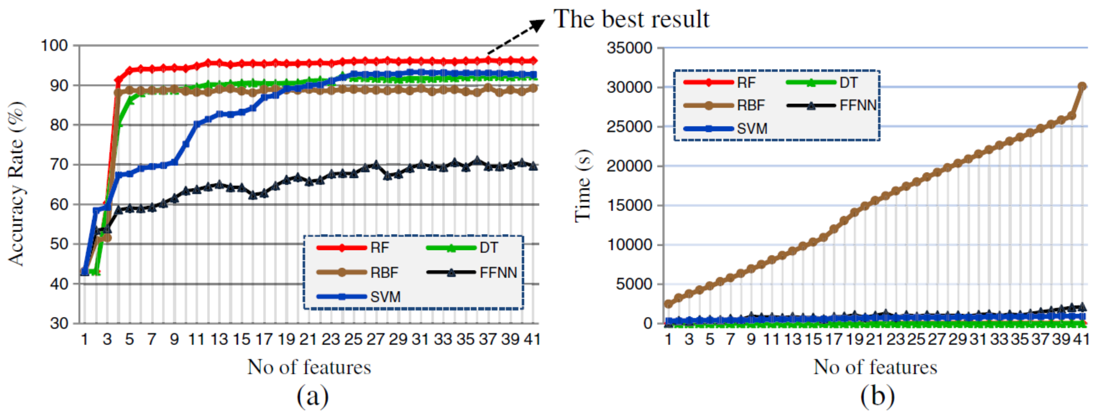

Figure 11. Results for experiment 2.

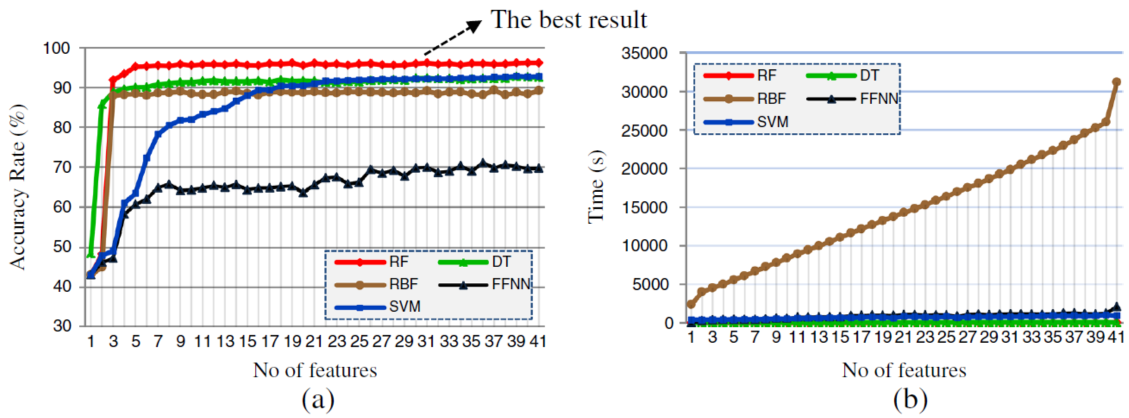

Figure 12. Results for experiment 3.

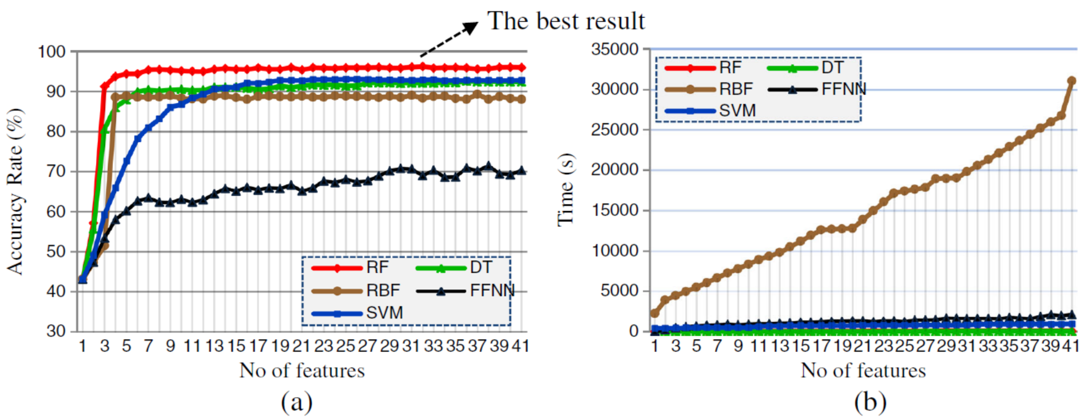

Figure 13. Results for experiment 4.

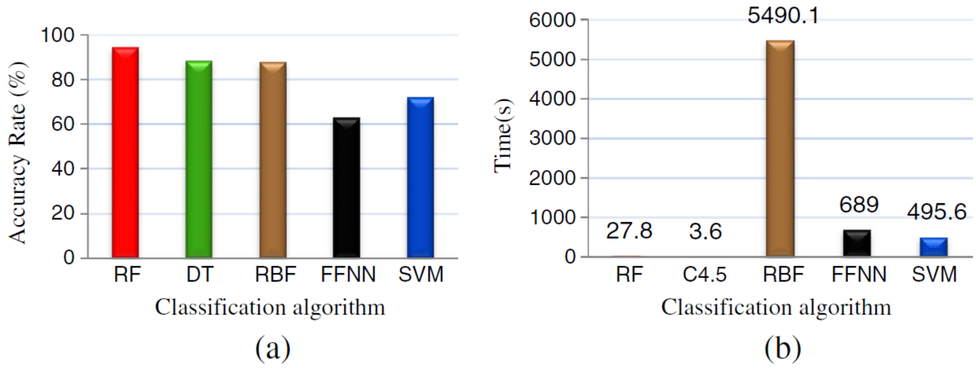

Figure 14. Results for experiment 5.

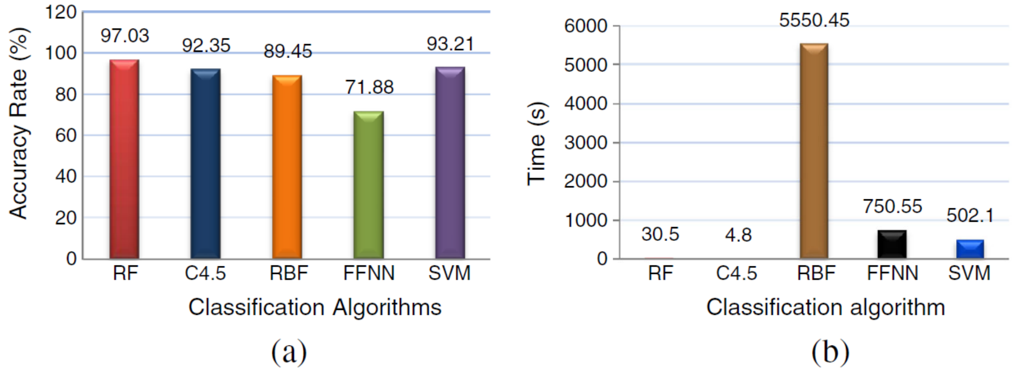

Figure 15. Results for experiment 6.

According to the results, the RF algorithm was the best algorithm in terms of success, considering that RF, SVM, DT and RBF algorithms were found to be effective in different feature clusters. Similar results were obtained in different experiments computational time. The DT algorithm was found to be the one with the lowest computation time followed by RF, SVM, FFNN and RBF algorithms respectively.

In table 5, a number of recent works about sleep stage classification are listed considering the input channels, classifiers, accuracy and capability of discriminating sleep stages.

Table 5. Recent works for automatic sleep stage scoring.

Authors	Detected Sleep Stages	Classifiers	Classification Accuracy	Input Channels
Šušmáková et al. [80]	W, S1, S2, S3, REM	Discriminant Analysis	77%	EEG, EMG, EOG and ECG signals
Chapotot et al. [65]	W, S1, S2, S3, REM, Movement Time (MT)	FFNN (with Back Propagation Algorithm)	W: 34%, S1: 43%, S2: 51%, S3: 82%, REM: 82%, MT: 13%	EEG and EMG signals
Zoubek et al. [55]	W, S1, S2, S3, REM	FFNN (with Back Propagation Algorithm)	71% (EEG only), 80% (EEG, EOG and EMG)	EEG, EMG and EOG signals
Tagluk et al. [9]	REM, S1 (Drowsy), S2 (LS), S3 and S4 (DS), W, Sleep spindles, REM	FFNN (with Back propagation Algorithm)	74.7%	EEG, EMG, REOG and LEOG signals
Sinha [143]	W, Sleep spindles, REM	FFNN (with Back Propagation Algorithm), Combined with Content rules	95.35%	EEG signal
Fraiwan et al. [144]	W, N-REM1, N-REM2, N-REM3, N-REM 4, REM	LDA	84%	EEG Signal
Subasi et al. [8]	Alert, drowsy, sleep	FFNN (with Levenberg–Marquardt Algorithm)	94,03%	EEG signal
Ebrahimi et al. [145]	W, S1, S2, S3, REM	FFNN (with Back Propagation Algorithm)	93%	EEG signal

Doroshenkov et al. [73]	W, S1, S2, S3, S4, REM	HMM	92%	EEG signal
Gunes et al. [24]	W, N-REM1, N-REM2, N-REM3, N-REM 4, REM, Non-Sleep	DT (C4.5)	92.40%	EEG signal
Ozsen [89]	W, N-REM1, N-REM2, N-REM3, REM	ANN	90.93%	EEG, EMG and EOG signals
Hsu et al. [37]	W, N-REM 1, N-REM 2, S3, REM	Elman Recurrent Neural Classifier	87.2%	EEG signal
Baha Şen et al. [146]	W, N-REM 1, N-REM 2, N-REM 3, N-REM 4, REM	RF	97.03%	EEG signal
Baha Şen et al. [146]	W, S1, S2, S3, REM	RF	98.02%	EEG signal
Jinwoo Kim et al. [5]	W, S1, S2, S3, REM	SVM	86%,	EEG signal
Kurihara et al. [147]	W, S1, S2, S3, S4, REM	Rule Based	51.6%, 56.2%, and 77.5% (six, five, and three stages)	Heartbeat and EMG signals
Agarwal et al. [148]	W, S1, S2, S3, S4, REM	Clustering	76.8%	EEG, EMG and EOG signals
Vivaldi et al. [149]	REM, S1, S2, S3	Clustering	Visualized	EEG signal
Hese et al. [150]	REM, S1, S2, S3, Spindles	*K*-means Clustering	Visualized	EEG signal
Pacheco et al. [151]	W, S1, S2, S3, S4, REM	MLP + Rules	93.6%	EEG, EOG and EMG signals

Schwaibold et al. [152]	W, S1, S2, S3, S4, REM	MLP + Neuro Fuzzy System	Obstructive Sleep Apnea (OSA) subjects) 70.7% - (Healthy subjects) 79.8%	EEG, EOG and EMG signals
Virkkala et al. [78]	W, S1, S2, S3, REM	DT	72.5%	EOG L-M1, EOG R-M1 and the calculated EOG L–R signals
Khalighi et al. [53]	W, S1, S2, S3, S4, REM	Importance Weighted Kernel Logistic Regression (IWKLR)	(Sleep/Wake) 73.75% - (All Stages) 96.55%	EEG, EOG and EMG signals
Gudmundsson et al. [153]	W, (Light Sleep) LS, S3, REM	SVM, k-NN	81% - (K=30) 80%	EEG signal
McGrogan et al. [154]	W, S1, S2, S3, S4, REM, MT	ANN	(consensus scores) 72.2% - (single expert Scores) 63.3%	EEG signal
Alvarez-Estevez et al. [155]	W, S1, S2, S3, REM	Fuzzy Logic (111 Fuzzy Rules)	90%	EEG, EOG and EMG signals
Flexer et al. [156]	W, S1, S2, S3, S4, REM	HMM	(S2) 14% - (Wake) 86%	EEG and EMG signals
Fraiwan et al. [25]	W, S1, S2, S3, REM	RF	83%	EEG signal
Liang et al. [126]	W, S1, S2, S3, REM	DT	86.68%	EEG, EOG and EMG signals
Krakovska et al. [71]	W, S1, S2, S3, S4, REM	Rules	81%	EEG, EOG, EMG and ECG signals vs EEG alone
Koley et al. [23]	W, S1, S2, S3, S4, REM	SVM	96.4%	EEG signal
Hanaoka et al. [157]	W, S1, S2, S3, S4, REM, MT	DT	80%	EEG, EOG and EMG signals

Tagluk et al. [9]	S1, S2, S3, S4, REM	ANN	74.7%	EEG, EMG, REOG and LEOG signals
Wang et al. [158]	W, LS, DS, REM	Conditional Probability of the Knowledge Base	W (88.6%) REM (68.1%) LS (82.4%) DS (96.3%)	EEG, EOG and EMG signals
Park et al. [159]	W, S1, S2, S3, S4, REM	Case Based Reasoning and Hybrid Rule	87.5 %	EEG, EOG and EMG signals
Anderer et al. [160]	W, S1, S2, S3, S4, REM	Rule-Based and LDA	80%	EEG, EOG and EMG signals
Gabran et al. [161]	S1, S2, S3, S4, REM	ANN	85%	EEG signal

3-3-2 Deep Learning Approach

Unlike some of the machine learning areas such as natural language processing and object classification, the potential of deep learning techniques is not fully explored in automatic sleep stage classification. This fact is also noticeable when it comes to the feature transformation for sleep scoring. To the best of our knowledge, there are few researches works in this area and in the following we will try to review them.

Since 2006, deep structured learning, or more commonly called deep learning, has emerged as a new area of machine learning research [162]. During the past several years, the techniques developed from deep learning research have already been impacting a wide range of signal and information processing work within the traditional and the new, widened scopes including key aspects of machine learning and artificial intelligence.

By the commonly adopted machine learning tradition naturally deep learning techniques classify into deep discriminative/supervised models (e.g., deep neural networks (DNNs), RNNs, convolutional neural networks (CNNs), etc.) and generative/unsupervised models (e.g., restricted Boltzmann machine (RBMs), deep belief networks (DBNs), deep Boltzmann machines (DBMs), regularized autoencoders, etc.). The third category belongs to the class of hybrid deep network structures, which refers to the deep architecture that either comprises or makes use of both generative and discriminative model components. This two-way classification scheme, however, misses a key insight gained in deep learning research about how generative or unsupervised-learning models can greatly improve the training of DNNs and other deep discriminative or supervised-learning models via better regularization or optimization. deep learning [163]–[165] has been proposed for unsupervised feature learning and has been applied to many domains, such as biomedical signals [90], [166]–[169]. One of the major advantages of deep learning compared to traditional approaches is that they can work directly on raw data and do not require any tuning or hand-crafted features. Instead, they can learn their own feature representations.

Längkvist et al. [169] explored the feasibility of applying a DBN to sleep data. In general, DBN tends to learn distributed, non-sparse representations [170]. However, sparse representation models resemble biological visual system characteristics, and they are able to learn more complex features than simple oriented bars. Lee et al. [165] considering this part of the biological visual system characteristics, presented a sparse variant of the deep belief network (SDBN) model. Deep learning have been successfully applied to detect anomalies related to epilepsy in EEG recordings [167] and to classify sleep stages from EEG as well as recordings of eye movements and skeletal muscle activity [169]. Zhang et al. [90] presented an automatic sleep stage method which includes a SDBN for extracting feature from EEG, EOG and EMG and a combination of multiple classifiers. Moreover, a voting principle based on classification entropy was proposed to enhance the classification performance further by harnessing the complementary information provided by the individual classifier. The flowchart of the proposed technique is shown in figure 16.

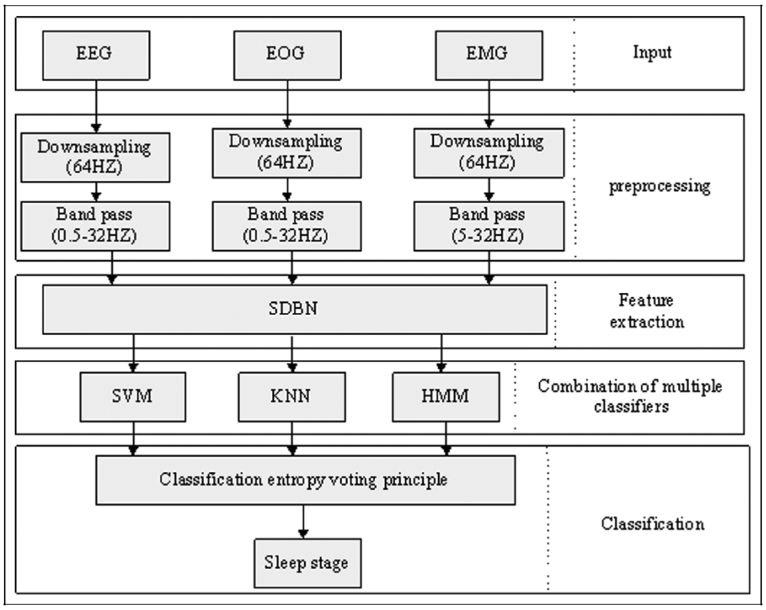

Figure 16. Flow-chart of multi-parameter sleep staging method [90].

Each data channel was divided into segments of 30s with zero overlap for feature extraction and imported to the SDBN which contains two major points: DBN and sparse coding. The features extracted from SDBN were classified into W, S1, S2, S3 and REM. The classification performance of the selected subjects using the combination of classifiers was 91.31%, with accuracy of 83.98% for the SVM, 82.05% for the *k*-NN and 85.55% for the HMM. Considering the total accuracy, which is higher than each classifier by itself, this technique can be considered successful. However, from another point of view, computational complexity is increased which is not an ignorable objection specially on real-time sleep classification tasks.

In [171], the main idea is to use hybrid deep learning models to increase the performance of sleep stage classification. DBNs are applied on 28 hand-crafted feature set for unsupervised generation of higher-level features. For classification, another deep structure, namely, Long Short-Term Memory (LSTM) is used. In this work, sleep stage classification is regarded as a time series and sequence classification problem. Therefore, the ability of LSTM models for recognizing the patterns from a sequence of events is mentioned as the reason for using this classifier.

The proposed algorithm is tested on two sleep recording datasets and the features are extracted from EEG, EOG and EMG. The performance of the proposed algorithm (DBN+LSTM) is compared to three other sleep stage classification algorithms, namely DBN only, LSTM only and DBN with HMM. Simulation results show that two hybrid methods (DBN+LSTM and DBN+HMM) have significantly better performance than single DBN and single LSTM, while DBN+LSTM performs better than DBN+HMM for both datasets. It has been concluded that LSTM boosted the performance of DBN much better than HMM.

Dong et al. in [172], proposed a practical approach for mitigating the limitations of single-channel automatic sleep stage classification using mixed neural network (MNN). MNN is a deep learning-based feature transformation and classification technique and is composed of an RNN, a LSTM and a Softmax regression. The input to this system is a feature vector with time-frequency domain, statistical and time domain features. Considering temporal dependency of sleep stages to each other, in addition to the features of the current epoch, the features from previous EEG epochs are also fed to the system. In this paper, several alternative electrode placements are explored and finally a convenient single forehead EEG channel together with an EOG channel configuration is proposed for the low-cost, at-home sleep monitoring applications.

Tsinalis et al. [173] presented a time-frequency analysis based feature extraction to feed the stacked sparse autoencoders (SSAE) for classifying the sleep stages. SSAEs are a specific type of neural network model. The key difference between stacked autoencoders and standard neural networks is layer-wise pre-training using unlabelled data (i.e., without class labels) before fine-tuning the network as a whole. Autoencoders were trained using iterative optimization with backpropagation. They got overall accuracy of 78% with S3 being the most correctly classified sleep stage with around 90% correctness. Stages S2, R and W follow, with around 80% of the epochs correctly classified for each stage. The most misclassified stage was S1 with only 60% correctness. Most misclassifications occurred between the pairs S1-W and S1-R (about 15 and 13% respectively), followed by pairs S1–S2 and S2–S3 (about 8%), and S2-R and R-W (about 4%). The remaining pairs had either misclassification rates smaller than 4% (S2-W and S3-W) or almost no misclassifications at all (S1–S3 and S3-R). Although they reached a considerable value of accuracy and F1-score their method is not the most accurate and efficient one because:

i. No feature selection methods were used;
ii. Using a single channel is essential for future monitoring devices but it seems that they did not apply any pre-processing technique to prepare data in order to reach higher accuracy;
iii. Considering the numbers of open-access sleep databases, they could evaluate their method by applying it to other databases as well.

In [174], an end-to-end deep learning method was proposed to perform temporal sleep stage classification using multivariate sleep signals, i.e. EEG, EOG, and EMG. This algorithm basically had three key steps. First, linear spatial filtering was applied to the input signals with the aim of enhancing the information contained in the data. Second, feature extraction architecture, where temporal convolution, rectified linear operator, and max pooling were applied to signals serially. The outputs of the second step were concatenated to form the feature space. Finally, the features were fed to nonlinear Softmax classifier. The results of this work were compared to other methods in the following works:

i. Tsinalis et al. [175],
ii. Lajnef et al. [176],
iii. Supratak et al. [177].

Comparison of this method with three state of the art methods yielded comparable performance as well as low run time and computational cost.

With rapid development of wearable devices and increased accessibility of EEG signal, single channel deep learning models are getting more attention. For instance, Supratak et. al. [177] proposed DeepSleepNet. This model is based on raw sleep EEG signal and contains two different CNNs to extract time-invariant features and one bidirectional LSTM for sequence residual learning [178]. The performance evaluation with two different datasets demonstrated that this method was able to learn features from raw EEG signal and classify them efficiently compared to the methods that use hand-crafted features.

In another work, Vilamala et. Al. [179] developed the idea of single channel system combined with transfer learning. They tested the hypothesis that sleep stage classification can be supported by transfer learning on the data obtained from Physionet Sleep-EDF database [180]. Spectral images were created from each window of EEG signal using a method called multitaper spectral estimation [181]. These images were fed to a pre-trained CNN, VGGNet [182], to be classified into one of five sleep stages. Empirical evaluation results showed that this method reaches a performance comparable to the state of the art methods.

Recently, Ren et al. [183] applied CNN to the feature learning of EEG data and valuated it on the datasets from previous brain-computer interfaces (BCI) competitions. Compared with other state-of-the-art feature extraction methods, the learned features using CNN shown better performance.

3-4 Brief Summary

When relating all the above different existing solutions in the literature with the proposed research work, it is noticeable that deeper research work is required in sleep stage classification to apply these methods as a reliable tool in clinical environments. Despite several works have already proposed methods, further research is however mandatory in particular regarding denoising the EEG signal without manipulation and improving the classifier performance. Some of the challenges are related to the identification of the noise structure to efficiently detect and cancel it from original data. The next challenge is enhancing the generalization ability of the classifier to unseen data by improving the learning phase. The potential of applying advanced learning techniques such as deep learning is not explored adequately in sleep stage classification.

Chapter 4

4 Research Methods

In this section the methods and databases used to achieve the goals of this thesis will be described. Also, a brief description of the methods applied for data pre-processing/denoising, feature extraction/selection and classification procedures can be found.

4-1 Data

Details of all the databases utilized in this thesis will be discussed in this section. For the investigation of sleep stage classification based on PSG signals and evaluation of the developed algorithm, annotated data is needed. Therefore, two different open-access databases annotated under the AASM rules were obtained from online sources.

The following databases were used:

i. *Sleep-EDF Database [Expanded]* [180];

- The collection of data in this database comes from two studies. PSG recordings of the first study are named SC files (SC = Sleep Cassette). PSG recordings of the second study are named ST files (ST = Sleep Telemetry). SC files were not used, since EMG data for first study was a zero-amplitude or no-data recording. Therefore, only ST files were chosen, which are a collection of 22 PSG signals recorded in the hospital during two nights for about 9 h in 1994. EOG, EMG and EEG signals were sampled at 100 Hz, and the event marker at 1 Hz. Except for slight difficulty in falling asleep, subjects were healthy without any sleep related medication. The data were segmented into 30s epochs and all epochs were scored according to R&K guidelines for human sleep staging. These recordings include EEG (Fpz-Cz and Pz-Oz), EOG (horizontal), submental chin EMG, together with the corresponding hypnograms.

ii. *ISRUC_Sleep Database* [85];

- This dataset includes data from healthy subjects as well as subjects with sleep disorders and subjects under the effect of sleep medication. PSG recording was performed using a bio-signal acquisition equipment namely, SomnoStar Pro sleep system, in the sleep medicine centre of the hospital of Coimbra University (CHUC) between 2009 and 2013. Each PSG signal was recorded over a whole night of sleep (approximately eight hours) according to the recommendations of AASM. The sampling frequency was 200 Hz for all EEG, EOG, chin EMG and ECG signals. After data segmentation into 30 second epochs, two different experts performed manual sleep scoring using AASM. To improve the quality of the recordings, a pre-processing technique was applied by the providers of the database:

i. A notch filter was applied to eliminate the 50 Hz electrical noise from EEG, EOG, chin EMG and ECG.

ii. EEG and EOG recordings were filtered using a bandpass Butterworth filter with a lower cut-off frequency of 0.3 Hz and higher cut-off frequency of 35 Hz.

iii. EMG channels were filtered using a bandpass Butterworth filter with a lower cut-off frequency of 10 Hz and higher cut-off frequency of 70 Hz.

4-2 Pre-Processing

Because no online database without major or minor issues exits, the following sub-sections will explain the methods applied for preparing the content of the databases. These preparation procedures effect the performance of classification dramatically.

4-2-1 Database Pruning and Synchronization

In this sub-section, an approach to tackle the initial problems of the databases will be discussed which include cropped epochs identification and pruning. The main causes for the appearance of those kinds of epochs are sensor disconnection, body movement or recording device issues. After analysing two databases two kinds of data corruption was observed:

- In the beginning or at the end of recordings;
- In the middle of the recordings;

In both above mentioned cases data loss occurs which could dramatically affect the number of specific epochs for each stage. Considering the fact that the number of S1 and REM stages in both of the databases is low, it could affect negatively the classification accuracy. After removing the corrupted epochs in one or all the recorded signals (EEG, EOG, ECG and EMG), corresponding epochs are removed from the hypnogram in order to synchronize the signals.

4-2-2 Filtering and Windowing

In order to guarantee the reliability of biomedical signal analysis, artefact free data is necessary. In this thesis, for reducing the artefact, the cropped epochs were automatically detected and eliminated. Then EEG and EOG data were band-pass filtered between 0.3 Hz and 35 Hz and EMG was filtered between 10 Hz to 100 Hz according to the AASM manual for the sleep scoring [1]. Then, the EEG, EOG and EMG signals were windowed in 30s intervals and grouped stage wise by considering the corresponding hypnogram. Next, in contrast with conventional approaches in the literature, which imports all the existing epochs to the classifier, we used a quantity of epochs selected out of each subject. In this method, selected epochs from each subject should have two characteristics. First, the number of epochs must be the same for all the subjects. Second, the number of epochs for each stage should be dependent on the number of occurrences of that stage for each subject. The proposed method is suitable for large databases with lower standard deviation in the quantity of each class helping on the computational complexity reduction of the classifier training stage. Considering that sleep databases have S1 deficiency, the training dataset will become highly unbalanced using the conventional method. Therefore, in this thesis proposed method was used to make balanced and fair training datasets from the number of stages point of view.

4-3 Proposed Adaptive Filtering Technique

Rather than other works that used artificial noise, in [184] real EEG data contaminated with EOG and EMG are used for evaluating the proposed artefact removal algorithm's efficiency via classification accuracy. The artefact detection is performed by thresholding the EEG-EOG and EEG-EMG cross correlation coefficients. Then, the segments considered contaminated are denoised by NLMS adaptive filtering technique. Using a single EEG channel, four sleep stages consisting of Awake, Stage1 + REM, Stage 2 and S3 are classified. A wavelet packet (WP) based feature set together with ANN was deployed for sleep stage classification purposes.

Still there are some open challenges such as:

- Using ECG channel for denoising the EEG,
- Extract or simulate the real artefact from EOG, EMG or ECG and add to clean EEG (proved by expert) to better evaluate the denoising performance.

4-3-1 Artefact Detection and Removal

It is usually assumed that the measured EEG is a linear combination of cerebral activity with one or more kind of artefacts. Therefore, in this work, for detecting the EOG and EMG contamination, the filtered EEG, EOG and EMG recordings are divided into 1000-sample segments and then the cross correlation of each EEG segment is calculated with the corresponding EOG and EMG segment. If the absolute value of the EEG-EOG cross correlation coefficients or EEG-EMG cross correlation coefficients is more than threshold 1 or threshold 2 respectively, the corresponding segment will be fed to an artefact removal block which is based on NLMS adaptive filtering. Adaptive filtering [185] has been extensively used in EEG artefact removal algorithms. It uses a recorded reference of the artefact (in our case horizontal EOG and submental chin EMG) to adjust a vector of weights that models the contamination according to an optimization algorithm.

If the thresholding conditions for cross correlation coefficients are not satisfied, the relevant EEG segment will be copied to the output without any change.

4-4 Feature Extraction

In this section, feature extraction from 30-second epochs of PSG data will be explained in detail. The feature set consists of 80 features extracted from EEG, EOG, and EMG signals. We tried to use the most common features to explore the information contained in these signals [6], [23], [186]–[188]. These features can be mainly categorized into temporal, time-frequency domain, entropy-based and non-linear features. In order to extract time-frequency domain features, a WP tree with seven levels of decomposition was utilized to extract the EEG rhythms. For more details about EEG rhythms and features refer to [41], [189]. As a result, each epoch's feature vector contains 41 EEG, 12 EOG, 13 EMG, 6 ECG, 4 EEG-EMG, and 4 EEG-EOG features. Table 6 summarizes the conventional features and table 7 summarizes the distance-based features used in this thesis along with their handy descriptions. The distance-based features are categorized in the non-linear group.

Table 6. Summary of the conventional features extracted from PSG recordings.

Ref.	Signal	Description	T*	TF*	F*	E*	NL*
F1		Arithmetic Mean	●				
F2		Maximum	●				
F3		Minimum	●				
F4		Standard Deviation	●				
F5		Variation	●				
F6		Skewness	●				
F7		Kurtosis	●				
F8		Median	●				
F9		Petrosian Fractal Dimension					●
F10		Rényi Entropy				●	
F11		Spectral Entropy				●	
F12		Permutation Entropy				●	
F13		Approximation Entropy				●	
F14		Hjorth Parameter (Activity)	●				
F15	EEG	Hjorth Parameter (Mobility)	●				
F16		Hjorth Parameter (Complexity)	●				
F17		Mean Curve Length					●
F18		Zero-Crossing Number	●				
F19		Mean Energy					●
F20		Mean Teager Energy					●
F21		Hurst Exponent					●
F22		Mean Quadratic Value of WP Coefficients in Delta Band		●			
F23		Mean Quadratic Value of WP Coefficients in Theta Band		●			
F24		Mean Quadratic Value of WP Coefficients in Alpha Band		●			
F25		Mean Quadratic Value of WP Coefficients in Spindle Band		●			
F26		Mean Quadratic Value of WP Coefficients in Beta1 Band		●			

F27		Mean Quadratic Value of WP Coefficients in Beta2 Band		●			
F28		Mean Quadratic Value of WP Coefficients in All Frequency Bands		●			
F29		F24/(F22+F23)		●			
F30		F22/(F24+F23)		●			
F31		F23/(F22+F24)		●			
F32		F24/F23		●			
F33		F22/F23		●			
F34		Mean of the Absolute Values of WP Coefficients in All Bands		●			
F35		Standard Deviation of WP Coefficients in All Bands		●			
F36	EMG	Spectral Power			●		
F37	EMG	Maximum of the Spectral Power Distribution			●		
F38	EMG	Mean of the Spectral Power Distribution			●		
F39	EMG	Standard Deviation of the Spectral Power Distribution			●		
F40	EMG	Temporal Energy					●
F41	EMG	Ratio of the Temporal Energy of Current Epoch to The Energy of Previous Epoch					●
F42	EMG	Ratio of the Temporal Energy of Current Epoch to the Energy of Next Epoch					●
F43	EOG	Mean	●				
F44	EOG	Energy					●
F45	EOG	Maximum	●				
F46	EOG	Standard Deviation	●				
F47	EOG	Skewness	●				
F48	EOG	Kurtosis	●				

* T: Temporal, TF: Time-Frequency, F: Frequency, E: Entropy, NL: Non-Linear

Table 7. Summary of distance-based features extracted from PSG recordings.

Ref.	Signal	Description
F49	EEG	Itakura Distance of AR Coefficients
F50	EEG	Itakura Distance of Spectral Coefficients
F51	EEG	Itakura-Saito Distance of AR Coefficients
F52	EEG	Itakura-Saito Distance of Spectral Coefficients
F53	EMG	Itakura Distance of AR Coefficients
F54	EMG	Itakura Distance of Spectral Coefficients
F55	EMG	Itakura-Saito Distance of AR Coefficients
F56	EMG	Itakura-Saito Distance of Spectral Coefficients
F57	EOG	Itakura Distance of AR Coefficients
F58	EOG	Itakura Distance of Spectral Coefficients
F59	EOG	Itakura-Saito Distance of AR Coefficients
F60	EOG	Itakura-Saito Distance of Spectral Coefficients
F61	ECG	Itakura Distance of AR Coefficients

F62		Itakura Distance of Spectral Coefficients
F63		Itakura-Saito Distance of AR Coefficients
F64		Itakura-Saito Distance of Spectral Coefficients
F65	EEG & EOG	Itakura Distance of AR Coefficients,
F66		Itakura Distance of Spectral Coefficients
F67		Itakura-Saito Distance of AR Coefficients
F68		Itakura-Saito Distance of Spectral Coefficients
F69	EEG & EMG	Itakura Distance of AR Coefficients
F70		Itakura Distance of Spectral Coefficients
F71		Itakura-Saito Distance of AR Coefficients
F72		Itakura-Saito Distance of Spectral Coefficients
F73	EEG	COSH Distance of AR Coefficients
F74		COSH Distance of Spectral Coefficients
F75	EMG	COSH Distance of AR Coefficients
F76		COSH Distance of Spectral Coefficients
F77	EOG	COSH Distance of AR Coefficients
F78		COSH Distance of Spectral Coefficients
F79	ECG	COSH Distance of AR Coefficients
F80		COSH Distance of Spectral Coefficients

4-5 Feature Normalization

The extracted features from PSG signals are in different ranges, and this variety can bias the results of the following steps. Feature normalization methods are usually utilized for avoiding this bias. In this thesis, two different types of normalization methods were used: standardization (or Z-score normalization) and Min-Max. The effect of each method in feature ranking and classification was evaluated. In standardization, the features were rescaled so that they have zero mean and unit variance. In Min-Max, features were scaled to the fixed range of [0 1]. This rescaling is necessary for many machine learning algorithms.

4-6 Feature Selection

In this section the goal is to use conventional feature selection methods to prepare the features set for classifiers. In order to select a subset of features containing most of the original feature set information, we proposed a two-step feature selection method. In the first step we reduced the similarity between pairs of the features. In the second step, different feature ranking methods were applied.

4-6-1 Similarity Reduction

To remove the features with high levels of similarity, a feature selection method was proposed. The existence of similar features negatively affect the stability [190] of the feature ranking results; therefore, applying feature selection can improve the overall performance of the proposed algorithm [191]. After the L1-norm between each pair of feature vectors was calculated, a similarity threshold was defined. The feature pair, whose L1-norm was lower than the threshold level, was considered strongly similar. This way, features were clustered into groups of similar features, and one feature per cluster was selected as representative. The representative feature was the one with the lowest computational complexity. Alternatively, it is possible to use Principal Component Analysis (PCA) for finding the most dissimilar features. However, there is a main reason why we did not use PCA. In PCA for finding a non-redundant feature set would lead to keeping and calculating all the features in the classification and practical application steps, whereas by using the similarity threshold, the most redundant features can be detected and omitted from the feature set in the application step.

4-6-2 Feature Ranking

For analysing the contribution and evaluating the potency of the different kind of features, feature ranking techniques were adopted. In particular, we used ReliefF, minimum Redundancy Maximum Relevance (mRMR-MID and mRMR-MIQ), Fisher score, Chi-square and Information Gain (IG) techniques. Next, each technique will be briefly described:

4-6-2-1 ReliefF

Originally proposed by Kira and Rendell in 1992 [192], Relief [193] is an instance-based method for estimating a feature discrimination power. In this method, for a randomly selected sample, the $2k$ nearest neighbours are considered: k neighbours from the same class (*hits*) and from a different class (*misses*). Then, the distance of the random sample from the hits and misses is calculated. A quality (discrimination power) coefficient is updated according to this distance, i.e., the feature with lower distance from hits will have higher quality. ReliefF is an extension of the Relief method that removes the two-class problems restriction and reduces its sensitivity to noisy and incomplete data.

4-6-2-2 minimum Redundancy-Maximum Relevance (mRMR)

This is a feature selection method that selects a subset of features by maximizing the relevance of each feature to the target class and minimizes the redundancy between the selected features. It was mainly proposed by Peng et al. [194] for dealing with the redundancy problem. The redundancy and relevance are calculated using mutual information, whereas the objective function is defined by either the difference between redundancy and relevance (mRMR-MID) or the ratio between relevance and redundancy (mRMR-MIQ).

4-6-2-3 Fisher Score

This method is one of the most efficient, being widely used for feature ranking. Its main idea is to find a group of features with maximum distance between the data points from different classes and minimum distance between the data points of the same class in the feature space [195]. Since the Fisher score is calculated individually for each feature, the selected feature set can be redundant.

4-6-2-4 Chi-square Test

This is a statistical test to measure the independency of events. In feature selection, it is used to evaluate whether or not the occurrence of a specific value of a feature and a specific class are independent. Despite the fact that Chi-square was proposed exclusively for categorical data, this method was later extended to the continuous case [196]. For calculating the Chi-square statistics of each feature, the range of the numerical feature should be discretized into intervals. The features are ranked according to Chi-square statistics without taking into account the interactions between features like Fisher score.

4-6-2-5 Information Gain (IG)

This method proposed by Ross Quinlan [197] is a widely used feature-ranking algorithm. It works based on a DT generated from the training set. To select the effective feature in each node of the tree the IG measure is used. In other words, IG measures how much information each specific feature provides with respect to each class. Therefore, considering the notion of the DT, IG depends on how much information was available before knowing the feature and on how much would be available after. A common measure for the information is Shannon entropy, although any measure that allows for evaluating the information content of a feature will be applicable.

4-7 Classification

The process of labelling the data into relevant classes is called classification. The first step in the classification process is the identification of the features or characteristics that will enable the highest discrimination between the different groups of data. A classification model is developed in such a way that it provides the structure for how the classification processes' actions will be realized. Ideally, this model should be chosen to optimize the performance of the classification system, although it may need to be revised as the classifier design progresses. A classifier is then implemented and "trained" to recognize the chosen features in the data, or to determine the best input-to-output mapping. Once the system has trained and learned, it is ready to classify specific inputs. Then, the system can be tested and evaluated with such metrics as speed of computation and accuracy of classification [118]. For sleep stage classification variety of the classifiers have been used. In this thesis we will compare conventional to the recently emerged methods. Five different classifiers were chosen as follows: *k*-NN, ANN, DSVM, and deep learning (SSAE and CNN). DSVM classifier has been comprehensively discussed in the subsection 3-3-1, therefore it will not be explained in this section.

4-7-1 *k*-NN Classifier

k-NN is a fairly simple and straightforward classifier to implement, which makes it very appealing. Briefly, it has two parameters that need to be adjusted, which can directly affect the classification performance. The first is the number of neighbours while the second is the dissimilarity measure that indicates the neighbouring relationships [198].

Euclidean distance was used as the distance measure for the *k*-NN classifier in this work. In each experiment, considering the number of features, the classification accuracy for the 1, 2, …20 neighbourhood was calculated, and the one leading to maximum accuracy was selected as the optimum neighbourhood number.

4-7-2 Shallow ANN Classifier

In the state of art section of this thesis we discussed about ANN. Considering the slowness of learning procedure of ANN, it is one of the wildly applied classifiers to biosignals.

A standard FFNN consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment while other neurons get activated through weighted connections from previously active neurons. Some neurons may influence the environment by triggering actions.

The work flow for the neural network design process has seven primary steps as follows [199]:

i. Collect data;

ii. Create the network;

iii. Configure the network;

iv. Initialize the weights and biases;

v. Train the network;

vi. Validate the network;

vii. Use the network to test;

A standard three-layer FFNN is illustrated in Fig. 17, which consists of an input layer, a hidden layer and an output layer [199], [200].

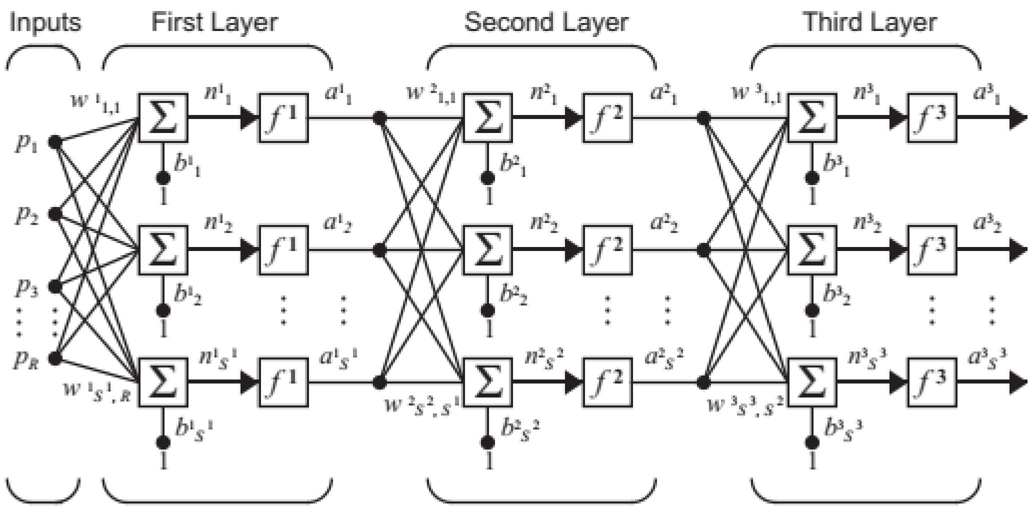

Figure 17. Three layers neural network [199].

In the following output of each layer is given by:

$$\mathbf{a}^1 = \mathbf{f}^1\left(\mathbf{W}^1\mathbf{p} + \mathbf{b}^1\right) \tag{7}$$

$$\mathbf{a}^2 = \mathbf{f}^2\left(\mathbf{W}^2\mathbf{a}^1 + \mathbf{b}^2\right) \tag{8}$$

$$\begin{aligned} \mathbf{a}^3 &= \mathbf{f}^3\left(\mathbf{W}^3\mathbf{a}^2 + \mathbf{b}^3\right) \\ &= \mathbf{f}^3\left(\mathbf{W}^3\mathbf{f}^2\left(\mathbf{W}^2\mathbf{f}^1\left(\mathbf{W}^1\mathbf{p} + \mathbf{b}^1\right) + \mathbf{b}^3\right) + \mathbf{b}^3\right) \end{aligned} \tag{9}$$

Figure 17 shows an output layer (layer 3) and a hidden layer (layers 2). The input layer is composed of input vectors [p_1, $p_{2, \ldots,}$ p_R]. A layer whose output is the network output is called the output layer [$a^3{}_1$, $a^3{}_{2, \ldots,}$ $a^3\mathrm{S}^3$]. The other layers are called hidden layers with outputs of [$a^1{}_1$, $a^1{}_{2, \ldots,}$ $a^1\mathrm{S}^1$] and

[$a^2{}_1$, $a^2{}_2$, ..., $a^2{}_{S^2}$]. All inputs in each layer goes through the transfer function (**f**), which produces the scalar neuron output vector **a**.

4-7-3 Stacked Sparse Auto Encoder (SSAE)

An autoencoder is a special type of neural network whose output values are equal to the inputs. Figure 18 presents an autoencoder structure. An autoencoder typically consists of an encoder and a decoder and it is trained in an unsupervised manner using backpropagation. During training, a cost function that measures the error between input and output of the autoencoder is optimized. In other words, the autoencoder tries to learn the identity function. By applying special constraints on the network such as the number of hidden units, an autoencoder can learn new representation or coding of the data [201].

Suppose the input vector to the autoencoder is a set of un-labelled data $\mathbf{x} \in \mathbb{R}^{D_x}$. This vector is encoded to another vector $\mathbf{z} \in \mathbb{R}^{D_1}$ in the hidden layer as follows:

$$\mathbf{z} = h^1\left(\mathbf{W}^1\mathbf{x} + \mathbf{b}^1\right) \tag{10}$$

where h^1 is the transfer function of the encoder, $\mathbf{W}^1$ is the weight matrix and $\mathbf{b}^1$ is the bias vector of the encoder. Then, the autoencoder tries to decode this new representation back to the original input vector as follows:

$$\hat{\mathbf{x}} = h^2\mathbf{z} = h^2\left(\mathbf{W}^1\mathbf{x} + \mathbf{b}^1\right) \tag{11}$$

where h^2 is the transfer function of the decoder, $\mathbf{W}^2$ is the weight matrix and $\mathbf{b}^2$ is the bias vector of the decoder. Sparse autoencoder is a specific type of autoencoder in which, in order to encourage the sparsity of the output of the hidden layer, a constraint is imposed on the number of active hidden neurons. The cost function of sparse autoencoder is slightly different from the original autoencoder as follows:

$$E = \underbrace{\frac{1}{N}\sum\sum\left(\mathbf{x}-\hat{\mathbf{x}}\right)^2}_{\text{mean squared error}} + \underbrace{\lambda\,\Omega_{weights}}_{\text{weight regularization}} + \underbrace{\beta\,\Omega_{sparsity}}_{\text{sparsity regularization}} \tag{12}$$

where N is the length of the input vector, λ is the weight regularization parameter and β is the sparsity regularization parameter [202].

A SSAE is a neural network with several sparse autoencoders. In this architecture, the output of each autoencoder is fully connected to the inputs of the next autoencoder. Greedy layer-wise training strategy is usually used for training SSAE. After the training of each layer is complete, a fine tuning is usually performed for enhancing the learned weights using backpropagation algorithm. Fine tuning can greatly improve the performance of the stacked autoencoder [201]. Figure 19 shows the training steps of a stacked autoencoder with two layers. Training of this stacked autoencoder has three steps:

i. **Step1:** initial pretraining of layer 1;

ii. **Step 2:** optimize the weights of the second layer using the weights of the first layer;

iii. **Step 3:** model fine-tuning by connecting all layers together;

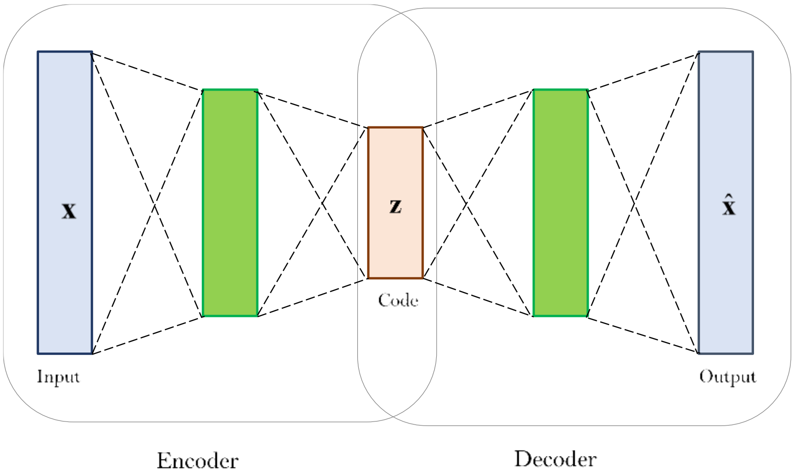

Figure 18. Structure of an autoencoder with 3 fully-connected layers.

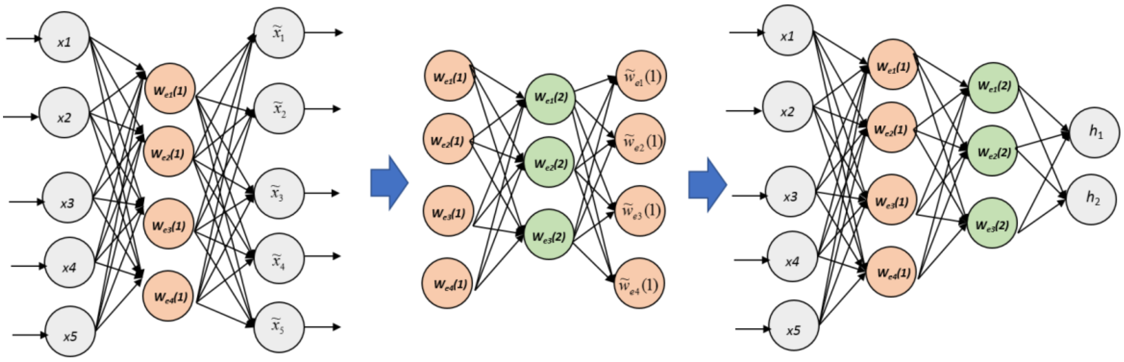

Figure 19. Training of two layers stacked autoencoder [203].

The Softmax function is also known as the normalized exponential and can be considered the multi-class generalization of the logistic sigmoid function [204]. The Softmax function is the output unit activation function that should come after the last fully connected layer for multi-class classification problems. After SSAE, a Softmax function is used as classifier and stacked to the network as the output layer.

4-7-4 Convolutional Neural Network (CNN)

A CNN is a multilayer perceptron designed specifically to recognize two-dimensional shapes with a high degree of invariance to translation, scaling, skewing, and other forms of distortion. Learning section of this classifier is done in supervised method which includes the following structure [205], [206]:

i. Feature extraction;

ii. Feature mapping;

iii. Subsampling;

The weights in all layers of a CNN are learned through training. Also, the network learns to extract its own features automatically.

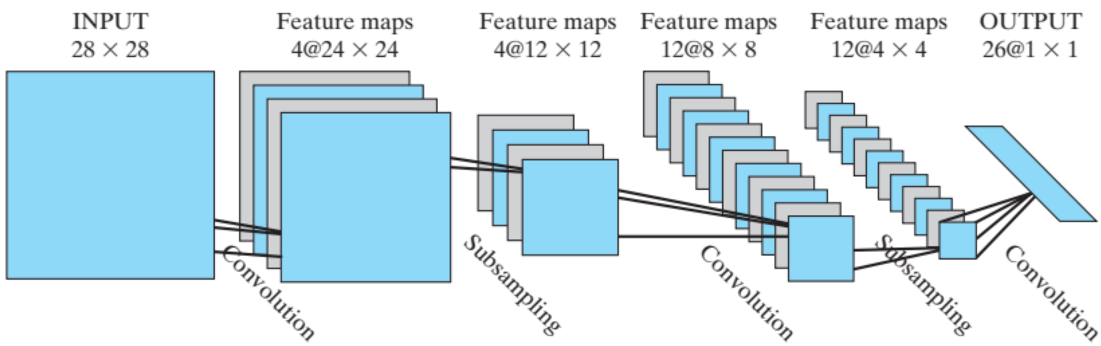

Figure 20. Convolutional network for image processing such as handwriting recognition.

Figure 20 illustrates the architectural design of a CNN made up of an input layer, four hidden layers, and an output layer. This network is designed to perform image processing. The input layer consists of 28×28 sensor neurones, receives the images of different characters that have been approximately centred and normalized in size.

However, a major breakthrough in the field occurred in 2012 when the Deep CNN AlexNet won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) by a big margin to all other competing models [179].

Therefore, an overall architecture of this CNN will be described in this part of thesis.

As depicted in figure 21, the AlexNet contains eight layers with weights; the first five are convolutional and the remaining three are fully-connected. The output of the last fully-connected layer (eighth layer) is fed to a 1000-way Softmax which produces a distribution over the 1000 class labels. AlexNet maximizes the multinomial logistic regression objective, which is equivalent to maximizing the average across training cases of the log-probability of the correct label under the prediction distribution. The kernels of the second, fourth, and fifth convolutional layers are connected only to those kernel maps in the previous layer. The kernels of the third convolutional layer are connected to all kernel maps in the second layer. The neurons in the fully-connected layers are connected to all neurons in the previous layer. Response-normalization layers follow the first and second convolutional layers. In the Rectified Linear Units (ReLU) non-linearity is applied to the output of every convolutional and fully-connected layer. The first convolutional layer filters the 224×224×3 input image with 96 kernels of size 11×11×3 with a stride of 4 pixels. The second convolutional layer takes as input the output of the first convolutional layer and filters it with 256 kernels of size 5×5×48. The third, fourth, and fifth convolutional layers are connected to one another without any intervening pooling or normalization layers. The third convolutional layer has 384 kernels of size 3×3×256 connected to the outputs of the second convolutional layer. The fourth convolutional layer has 384 kernels of size 3×3×192, and the fifth convolutional layer has 256 kernels of size 3×3×192. The fully-connected layers have 4096 neurons each [207].

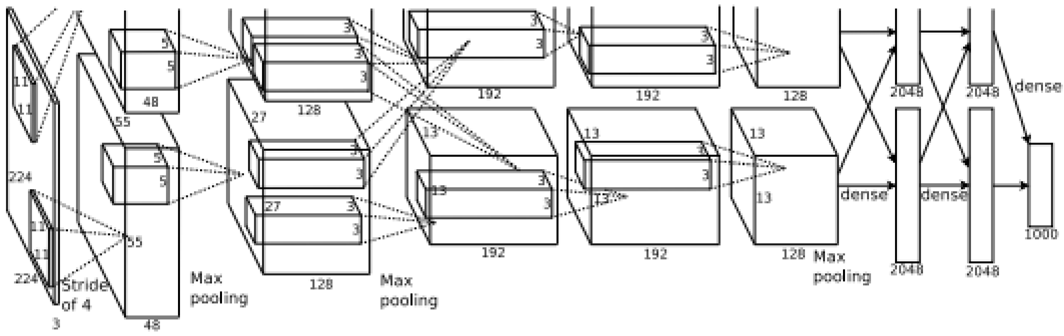

Figure 21. An illustration of the AlexNet architecture, explicitly showing the layers [207].

4-8 Brief Summary

Section 4 described the methods applied in this thesis in two main research topics which are denoising and classification. We considered FFNN as a shallow neural network and taking about the deep learning methods, SSAE and CNN are explained.

In the next chapter, we will explain our methodology in each section of this thesis including denoising and classification.

Chapter 5

5 Methodology and Results

In this section, we will discuss the proposed methodologies in each main topic of this thesis. First methodology is based on the description of the proposed denoising algorithm and the corresponding details. As second methodology, the details of the proposed methods for sleep stage classification will be discussed.

In the following a list of our main contributions can be found:

- Artefact detection and cancellation for EEG signal with application in sleep stage classification.
- Adaptive denoising EEG from artefacts with origin of EOG and EMG signals.

- Comparing different kinds of classifiers in sleep stage classification.
- Analysing the performance of the DSVM classifier.
- Creating 2D signals from conventional 1D sleep signals and classify them using deep learning methods.

5-1 Removing Artefact of EOG and EMG from EEG for Sleep Stage Classification

In this sub-section, a new algorithm will be proposed for artefact removing of EEG with application in sleep stage classification. Rather than other works that used artificial noise, in this study we used real EEG data contaminated with EOG and EMG for evaluating the proposed artefact removal algorithm's efficiency. The classification accuracy was the evaluation criterion. The artefact detection was performed by thresholding the EEG-EOG and EEG-EMG cross correlation coefficients. Then, the segments considered contaminated were denoised by NLMS adaptive filtering. Using a single EEG channel, four sleep stages consisting of Awake, S1 + REM, S2 and S3 were classified. A WP based feature set together with MLFN was deployed for sleep stage classification purposes. Simulation results showed that artefact removed EEG allowed a classification accuracy improvement of around 14%.

5-1-1 Methodology

Figure 22 shows an overview of the sleep stage classification with proposed sleep EEG artefact removal scheme.

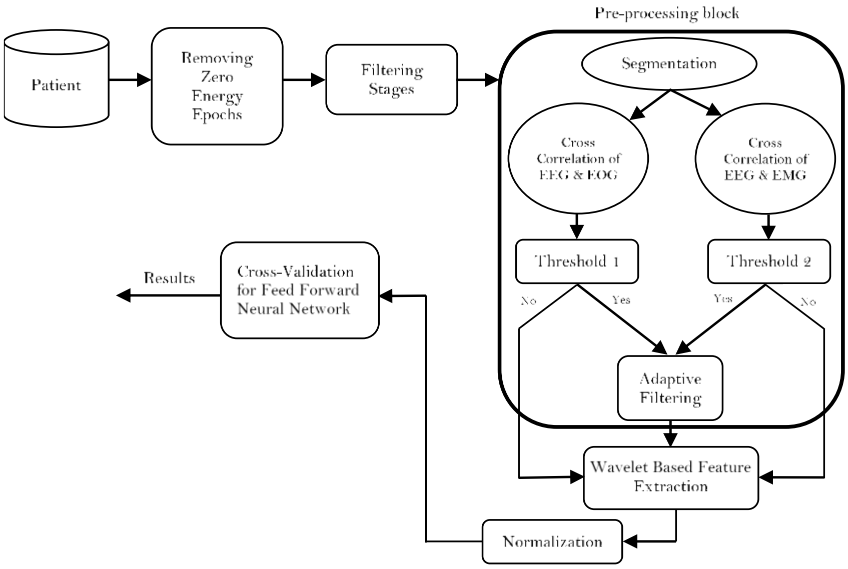

Figure 22. Flowchart of the proposed algorithm for denoising the EEG.

In this study, PSG records of the Sleep EDF Database were used which was explained in the chapter 4. All PSGs were divided into 30-seconds epochs and annotated according to the R&K manual. Pz-Oz EEG channel, horizontal EOG and submental chin EMG recordings of all the subjects were used.

Visual inspection of the data revealed that occasionally there exists one or more consecutive epochs in which the energy of the signal was zero probably due to hardware failure. Therefore, all pre-processing steps in chapter 4 including database pruning and synchronization and filtering and windowing were applied.

In this study, WP based decomposition and reconstruction methods as a filter for removing unwanted frequency band of physiological data was applied [208].

Conventionally, it is assumed that the measured EEG is a linear combination of cerebral activity with one or more kind of artefacts.

Therefore, in this study, for detecting the EOG and EMG contamination, the filtered EEG, EOG and EMG recordings were divided into 1000-sample segments and then the cross correlation of each EEG segment was calculated with the corresponding EOG and EMG segment. If the absolute value of the EEG-EOG cross correlation coefficients or EEG-EMG cross correlation coefficients was more than threshold 1 or threshold 2 respectively, the corresponding segment would be fed to an artefact removal block which was based on NLMS adaptive filtering. We fed to NLMS a recorded reference of the artefact (in our case horizontal EOG and submental chin EMG) to adjust a vector of weights that models the contamination according to an optimization algorithm.

If the thresholding conditions for cross correlation coefficients were not satisfied, the relevant EEG segment would be copied to the output without any change.

In order to perform sleep stage classification, the output of the pre-processing block was used for feature extraction. There are four types of main brain waves that can be distinguished by their frequency range. These frequency bands are called Delta (0-3.99 Hz), Theta (4-7.99 Hz), Alpha (8-13 Hz) and Beta (>13 Hz) [1]. EEG is analysed in the frequency domain. Moreover, EEG is a non-stationary signal and simultaneous time-frequency analysis can be quite useful. In this study, a WP tree with 7 decomposition levels and Daubechies order 2 (db2) mother wavelet was used for feature extraction. Different frequency bands of EEG including Delta, Theta, Alpha, spindle, Beta1 and Beta 2 were extracted according to the scheme proposed in [41]. The following statistical features were calculated for each epoch using the WP coefficients:

i. Energy of the WP coefficients for each frequency band (F1-F6);

ii. Total Energy (F7);

iii. Mean of the absolute values of WP coefficients for all frequency bands (F8);

iv. Standard deviation of WP coefficients for all frequency bands (F9);

v. Energy ratio of various frequency bands (F10 to F14);

F10 is the ratio between the energy in the Alpha band and the sum of the energy in the Delta and Theta bands. F11 is the ratio between the energy in the Delta band and the sum of the energy in the Alpha and Theta bands. F12 is the ratio between the energy in the Theta band and the sum of the energy in the Alpha and Delta bands. F13 is the ratio between the energy in the Alpha band and the energy in the Theta band and F14 is the ratio between the energy in the Delta band and the energy in the Theta band.

Since the range of values of raw EEG varies broadly, to avoid that features with larger numeric values dominate those with smaller numeric values affecting the accuracy of the classification technique, standardization method from chapter 4 was applied. Therefore, each feature x_{ij} was independently normalized by applying the following equation:

$$x'_{ij} = \frac{x_{ij} - \bar{\mathbf{x}}_i}{\sigma_{\mathbf{x}_i}} \tag{13}$$

where $\bar{\mathbf{x}}_i$ and $\sigma_{\mathbf{x}_i}$ are the mean and the standard deviation of each independent feature vector $\mathbf{x}_i$.

In this study MLFN was used for the classification of sleep stages. The two-layer feed forward backpropagation neural network consisting of 14 input neurons, 12 hidden neurons and 4 output neurons for discrimination between the four sleep stages Wake, REM+S1, S2 and S3 was applied to the extracted feature set. A sigmoid transfer function in the hidden layer and a linear transfer function in the output layer were selected. Levenberg-Marquardt training algorithm was chosen.

5-1-2 Results

The performance of the proposed method was assessed using the six subjects selected from the dataset mentioned in section 5-1-1. In the artefact detection stage, a threshold of 0.5 (*Threshold 1*) for EEG-EOG cross correlation coefficients and 0.25 (*Threshold 2*) for EEG-EMG cross correlation coefficients were used. These thresholds were selected empirically considering highest classification accuracy. Three different result validation approaches including subjective and objective methods were applied.

The cross-correlation coefficients for EEG-EOG and EEG-EMG which were detected by thresholding before and after applying the artefact removal algorithm are shown in figure 23. A significant reduction in the correlation coefficients is noticeable after the artefact removal procedure.

Figure 24 and 25 illustrate the cancellation of EOG and EMG artefacts from contaminated EEG segments. It can be seen that the artefacts can be correctly eliminated without distorting the original EEG.

After the completion of the artefact removal stage, the data was fed to the feature extraction algorithm. Unlike the more conventional approaches in the literature, which import all the existing stages to the neural network, we used a quantity of training data to be selected out of each patient for training the MLFN. This method is suitable for large databases helping on the reduction of the computational complexity of the classifier training stage.

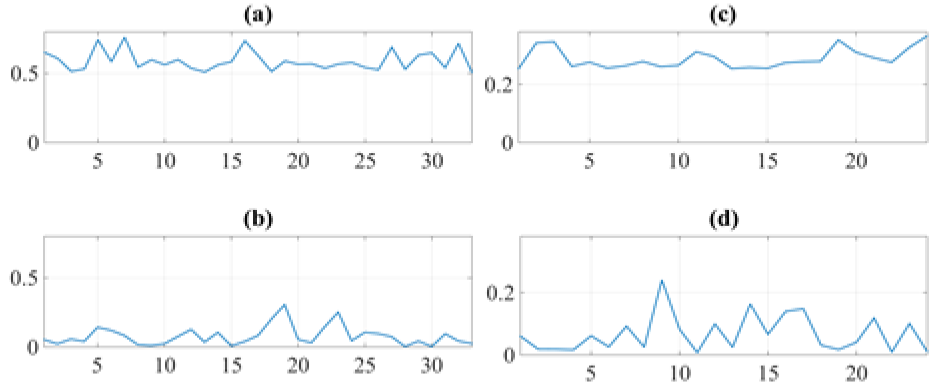

Figure 23. Cross correlation coefficient's absolute values, (a) EEG-EOG before artefact removal, (b) EEG-EOG after artefact removal, (c) EEG-EMG before artefact removal, (d) EEG-EMG after artefact removal algorithm.

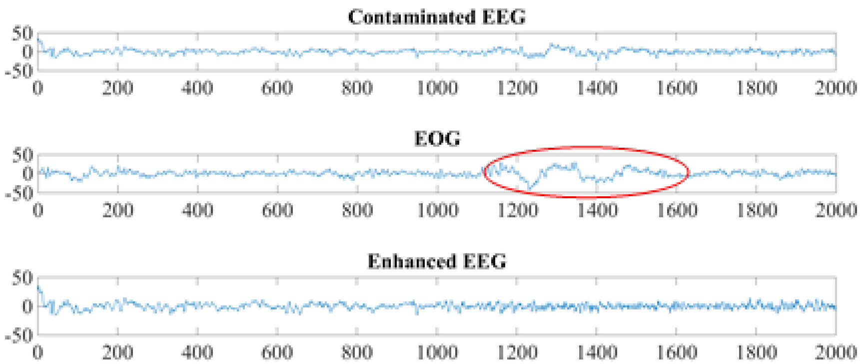

Figure 24. EOG artefact cancelation from contaminated EEG.

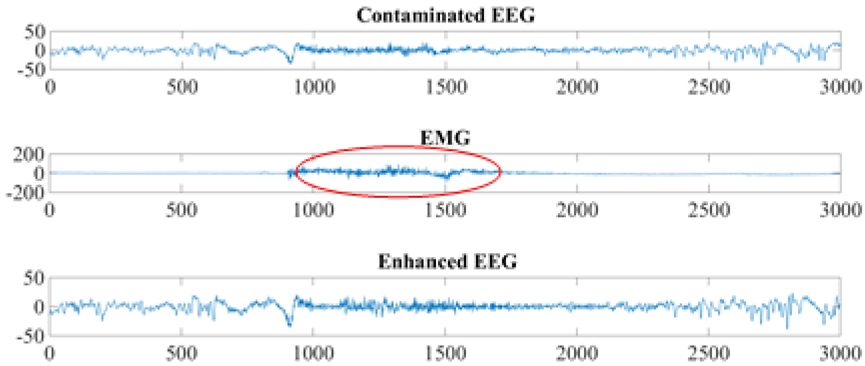

Figure 25. EMG artefact cancelation from contaminated EEG.

To assess the effectiveness of our artefact removal algorithm, we studied the sleep stage classification accuracy for raw (after removing zero energy epochs), filtered and artefact removed data. Table 8 shows the results of statistical analysis for comparison of each stage and overall accuracy for all the above-mentioned data. The results were validated using repeated random sub-sampling method, also known as Monte Carlo cross-validation technique. It can be observed that there is an improvement in the performance of the classifier after filtering the data, but the best performance is achieved by applying the proposed artefact removal algorithm.

Table 8. Results of the statistical analysis for comparison of each stage and overall accuracy.

	Wake (%)	REM + S1 (%)	S2 (%)	S3 (%)	Overall (%)
Raw	77.56	87.08	74.67	78.11	63.70
Filtered	79.44	78.75	83.26	90.74	70.60
Proposed method	87.08	87.25	87.38	90.93	77.80

5-2 Performance Analysis of Conventional Classifiers in Sleep Stage Classification

In this sub-section, the performance of conventional classifiers such as *k*-NN, MLFN and DSVM are evaluated. This evaluation was done by applying them to a variety of feature extraction and selection methods. Also, both available databases were utilised and in the following we will have two studies in order to discuss methodologies and results related to each of these datasets. A comprehensive feature set was extracted from these recordings. The extracted features are among the most common and effective features used in sleep stage classification from temporal, spectral, entropy-based and nonlinear categories. The classification methods were evaluated and compared using three criteria: classification accuracy, sensitivity, and specificity.

5-2-1 Methodology for First Study

In this study, the database downloaded from Physionet website was processed. Figure 26 shows the block diagram of the proposed algorithm for comparing the classifiers for sleep stage classification.

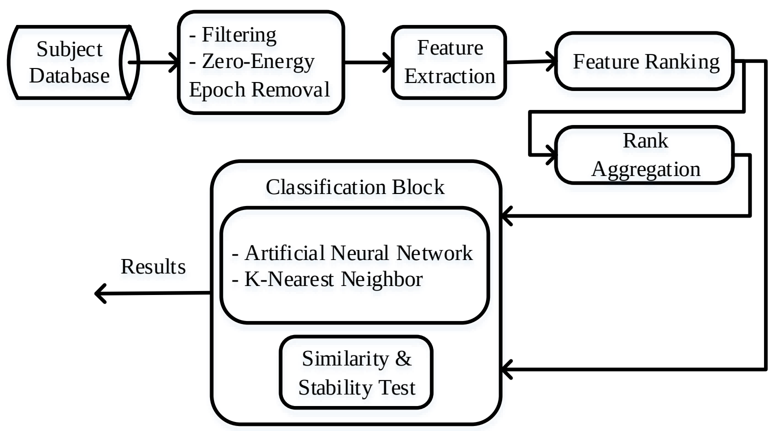

Figure 26. Block diagram for comparison of two classification methods for first study.

The data used in this study was obtained from Physionet Sleep-EDF Expanded Dataset [180]. Pz-Oz channel EEG together with submental chin EMG and horizontal EOG each sampled at 100 Hz were used in the evaluations. Table 9 illustrates the number of stages available per subject.

Table 9. Summary of the data provided by six selected subjects in Sleep-EDF Expanded Dataset.

	Wake	**REM**	**S1**	**S2**	**S3**
Subject #1	146	122	101	527	136
Subject #2	41	159	71	351	284
Subject #3	85	226	120	392	180
Subject #4	40	143	47	266	152
Subject #5	149	80	102	428	218
Subject #6	131	142	135	378	198

All pre-processing steps in chapter 4, including database pruning, synchronization, filtering and windowing were applied to this database as well.

In order to explore the information contained in PSG recordings, a set of features were extracted from EEG, H_EOG and submental chin EMG of each subject. This feature set includes 49 features that can be categorized into time, frequency, joint time-frequency domain, entropy-based and nonlinear types. Table 10 summarises information about the different features used in this study and their brief description.

The physiological differences from subject to subject and equipment related variations have considerable impact on the features extracted from the PSG recordings. Moreover, since there are usually a wide variety of feature types extracted for characterizing sleep stages, the amplitude and unit of features will also vary.

Table 10. Summary of the features extracted from PSG recordings in the Physionet Sleep-EDF Expanded Database.

Signal	Category	Feature Name
EEG	Time Domain (F1 to F12)	Statistical Features (Minimum Value, Maximum Value, Arithmetic Mean, Standard Deviation, Variance, Skewness, Kurtosis, Median), Zero-crossing Rate, Hjorth Parameters (Activity, Mobility and Complexity) [6].
	Time-Frequency Domain (F13 to F26)	Features Extracted from WP Coefficients including Energy of α, δ, $\beta 1$, $\beta 2$, θ and Spindle bands, Total Energy of all bands, Energy ratio of ($\frac{\alpha}{\delta+\theta}$, $\frac{\delta}{\alpha+\theta}$, $\frac{\theta}{\alpha+\delta}$, $\frac{\delta}{\theta}$, $\frac{\alpha}{\theta}$), Statistical Features (mean and standard deviation of coefficients in all the bands).
	Entropy (F27 to F30)	Spectral Entropy, Rényi Entropy, Approximate Entropy, Permutation Entropy [6].
	Non-linear (F31 to F36)	Petrosian Fractal Dimension, Teager Energy, Energy, Mean Curve Length, Hurst Exponent [6], ISD.
EOG	Time Domain (F37 to F41)	Mean, Maximum, Standard Deviation, Skewness, Kurtosis [209].
	Non-linear (F42)	Energy [209].
EMG	Frequency Domain (F43 to F46)	Total Power in the EMG Frequency Spectrum, Statistical Features of EMG Frequency Spectrum (Maximum, Mean, Standard Deviation) [209].
	Non-linear (F47 to F49)	Energy, Ratio of the EMG Signal Energy for the Current Epoch and Previous Epoch, Ratio of the EMG Signal Energy for the Current Epoch and Next Epoch [209].

The features may also get extreme values, i.e. extremely low or extremely high values. Data post-processing is an important step in this respect. The aim of feature post-processing is to enable classification algorithms to uniformly handle the features with different units and ranges as well as reducing the influence of extreme values. Feature post-processing can be a scaling operation (normalization/standardization) or a feature transformation operation. In this study, standardization method applied to the extracted features.

To select a subset of features containing most of the original feature set information, we used seven different feature selection methods: ReliefF, MRMR-MID, MRMR-MIQ, Fisher Score, Chi-Square (Chi2), IG and Conditional Mutual Information Maximization (CMIM). We have also implemented two different rank aggregation methods, Borda and Robust Rank Aggregation (RRA), to evaluate their ability to produce better feature rankings compared to conventional feature ranking methods [210].

We selected two simple and widely used classifiers: k-NN and MLFN to discriminate five sleep stages W, S1, S2, S3 and REM. By selecting $k=1$, nearest neighbour was utilized. The k-NN classifier is the simplest nonparametric classifier and assigns a pattern to a specific class based on its nearest neighbour's class. In spite of its simplicity, in [211] it has been proved that, if the utilized database is fairly large, the error bound for nearest neighbour rule is quite tight, i.e. equal or less than twice the Bayes error. Also, neural networks are known to be very powerful computing models that can learn from training examples. Neural networks have been successfully used in a broad range of data mining applications including classification [212]. Accuracy as main criterion is considered for evaluating and comparing the different classification methods. The performance of the sleep stage classification is evaluated using repeated random sub-sampling validation. To measure the classification accuracy, the overall accuracy value is calculated as follows [213]:

$$\text{Accuracy} = \frac{\text{No. of true detections}}{\text{Total no. of epochs}} \tag{14}$$

Six subjects were selected from the Physionet database for evaluating and comparing the feature ranking and the rank aggregation methods. For filtering EEG and EOG signals, Daubechies order 20 (db20) was used as the mother wavelet. The filtered data was segmented into 30-second epochs. From each epoch, a feature vector containing 49 features was extracted. After feature standardization, the feature vectors were fed into seven feature selection methods. Then, in order to aggregate the results, the outputs of these seven feature ranking methods were used by Borda and RRA, producing two additional ranked lists of features.

For sleep stage classification, the parameters of the classifiers are set as follows. The Euclidean distance was chosen as the distance metric for the k-NN classifier. For the three-layer neural network classifier 12 hidden neurons and a sigmoid transfer function were selected in our simulations. The Levenberg-Marquardt training algorithm was adopted for minimizing the cost function because of its fast and stable convergence. In contrast with conventional approaches in the literature, which imports all the existing epochs to the classifier, we used a quantity of epochs selected out of each subject. In this method, selected epochs from each subject have two characteristics. Firstly, the number of epochs was the same for all the subjects. Second, the number of epochs for each stage was dependent on the number of occurrences of that stage for each subject. This method is suitable for large databases helping on the computational complexity reduction of the classifier training stage.

In our simulations, 50 subsets were generated out of the original training set by bootstrapping.

5-2-2 Results for First Study

In order to estimate the generalization ability of the classifier, repeated random sub-sampling validation with 200 runs was applied. Figure 27 depicts the classification accuracy of *k*-NN and MLFN classifiers for different feature selection methods. Results show that MRMR-MID achieves the highest classification performance. On the other hand, the performance of the aggregation methods was in the average level, although they are known to generate more stable results and better accuracy.

As figure 27 shows, starting with one feature, each additional feature typically leads to an increment in the classification accuracy. However, at some point, the increment on the classification accuracy for each additional feature is not significant leading to an elbow in the graph. Inspired by the "elbow" point in the cost-benefit curves, in this study we used the Kneedle algorithm proposed in [214] for determining the optimal feature number which provides a satisfactory trade-off between the selected number of features and the classification accuracy. Table 11 illustrates the top 10 features selected by each method. In general, the accuracy of MLFN was higher than *k*-NN for all the ranking methods. The highest accuracy was 77%, achieved by RRA and MLFN.

Table 11. Top 10 features selected by each method and the optimum number selected by Kneedle algorithm (Corresponding Accuracy).

	ReliefF	**Fisher**	**Chi2**	**IG**	**CMIM**	**MRMR-MID**	**MRMR-MIQ**	**Borda**	**RRA**
Top 10 Features	F28	**F36**	F35	F9	F15	F35	F35	**F36**	**F36**
	F36	F35	F9	F35	**F36**	F39	F42	F35	F35
	F7	F31	F11	F11	F9	**F36**	F15	F9	F9
	F49	F9	F31	F31	F8	F22	**F36**	F31	F31
	F41	F29	**F36**	**F36**	F1	F15	F22	F22	F27
	F27	F11	F27	F4	F34	F31	F23	F27	F22
	F20	F25	F26	F27	F35	F29	F31	F29	F17
	F23	F27	F4	F26	F28	F23	F38	F11	F29
	F6	F12	F25	F25	F6	F9	F29	F15	F11
	F22	F22	F14	F29	F48	F38	F9	F20	F20
MLFN	7 (0.75)	5 (0.76)	7 (0.76)	7 (0.76)	**3** (0.74)	5 (0.76)	5 (0.76)	5 (0.76)	7 (0.77)
***k*-NN**	7 (0.69)	5 (0.71)	9 (0.73)	9 (0.73)	**3** (0.68)	7 (0.75)	11 (0.75)	9 (0.74)	7 (0.73)

5-2-3 Methodology for Second Study

In the second study the dataset from ISRUC-Sleep was processed to evaluate the performance of conventional classifiers. Figure 28 shows the study framework used in this study. In the following sub-sections, each part will be described in detail.

All the 80 features from tables 6 and 7 were used which can be categorized into temporal, time-frequency domain, entropy-based, non-linear and distance-based features.

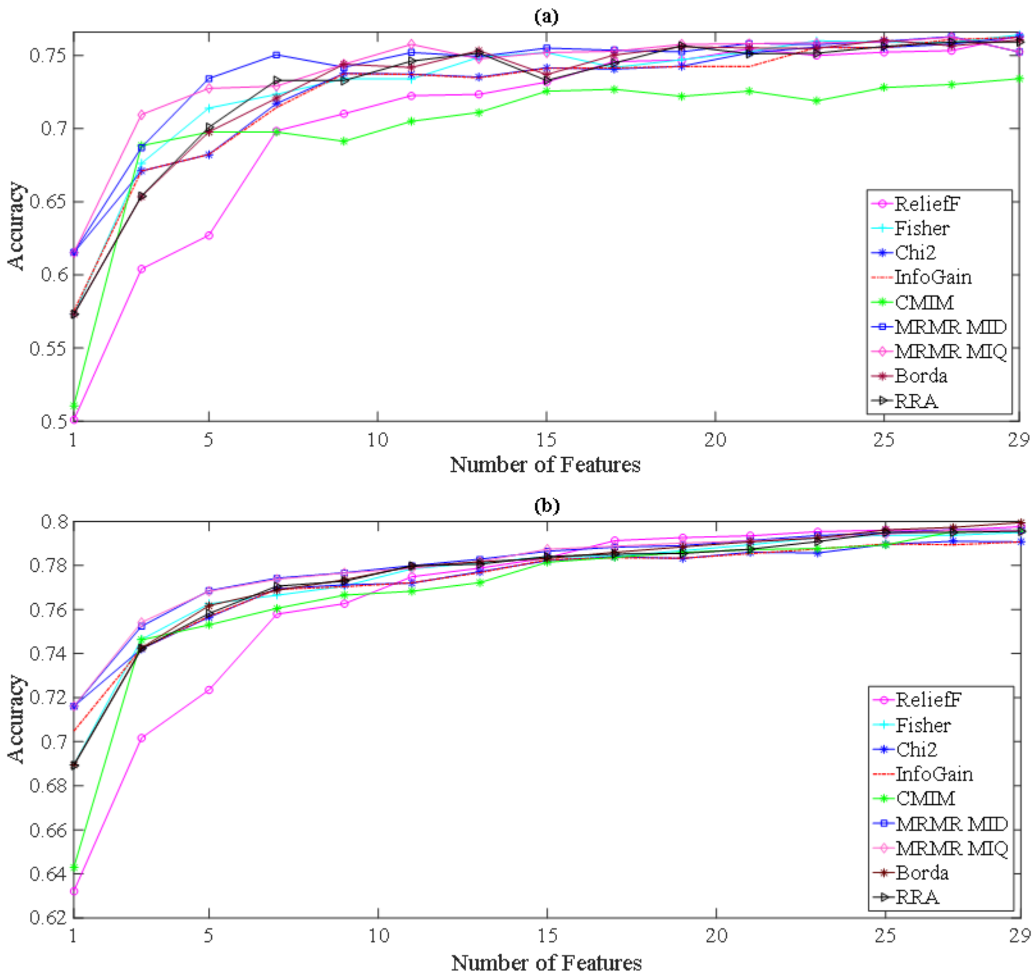

Figure 27. Classification accuracy for different feature selection methods, (a) *k*-NN classifier, (b) MLFN classifier.

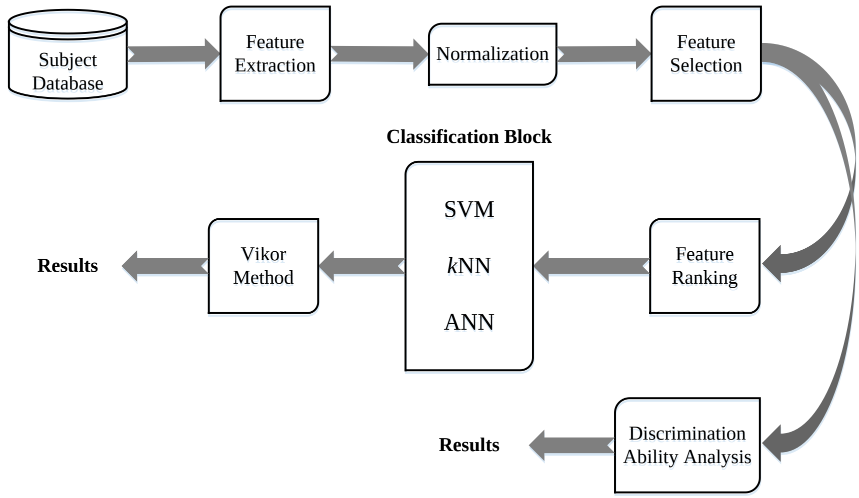

Figure 28. Sleep study framework for second study.

The extracted features from PSG signals are in different ranges, and this variety can bias the results of the subsequent steps. In this study, two different types of normalization methods from subsection 4-5 were used: standardization (or Z-score normalization) and Min-Max. The effect of each method in classification was evaluated. In standardization, the features are rescaled so that they have zero mean and unit variance. In normalization, features are scaled to the fixed range of [0 1]. This rescaling is necessary for many machine learning algorithms.

To remove the features with high levels of similarity, a feature selection method which was mentioned in sub-section 4-6-1, was proposed in this study.

We used all the six feature ranking methods mentioned in section 4-6-2. Each of these methods was applied on the conventional, distance-based and total feature sets (combined conventional and distance-based feature sets), and in all, 18 ranked lists of features were achieved.

For sleep stage classification three different classifiers were used: k-NN, MLFN and DSVM. The main reason for choosing these three different classifiers was that we did not want to restrict the significance of the comparison to one specific family of classifiers. On the other hand, we aimed to choose a variety of classifiers including the simplest, most used and the one that usually shows the best performance. Euclidean distance was used as the distance measure for the k-NN classifier. In each experiment, the classification accuracy for the $1, 2, \ldots 20$ neighbourhood was calculated. The one leading to maximum accuracy was selected as the optimum neighbourhood number. For the MLFN classifier, a three-layered MLFN with 20 hidden neurons for the conventional and total feature sets and 12 hidden neurons for the distance-based feature set were used.

DSVM was used instead of conventional multi-SVMs. The reason for choosing DSVM was that it outperforms conventional multi-SVMs (OAO and OAA) while utilizing lower number of SVM in the structure [116], [117], [215], [216]. RBF was selected as the kernel function, and sigma was set to 3.0 for the conventional and total feature sets and 1.1 for the distance-based feature set.

For each ranked list of features, created by one of the ranking methods, and each specific classifier, the classification accuracy was calculated for the top 25 features. Since it is always desirable to achieve the maximum accuracy with the minimum complexity, to find the optimum number of features, a Multi-Criteria Decision Making (MCDM) method called Vikor was used [217]. The Vikor method was originally developed for MCDM problems with contrasting and conflicting criteria. In our case, the accuracy and number of features were two conflicting criteria. This method ranks and selects a set of alternative solutions for the problem at hand, helping decision makers to reach a final decision. The various J alternative solutions are denoted as $a_1, a_2, \ldots, a_J$. Suppose that there are n criteria. f_{ij} is the value of the i^{th} criterion for j^{th} solution, a_j. The compromise ranking is performed by comparing its closeness to the ideal solutions of the criteria (utopian solution F^*). The distance measure of the Vikor method is developed from the L_p-metric as:

$$L_{p,j} = \left\{ \sum_{i=1}^{n} \left[w_i \left(f_i^* - f_{ij} \right) / \left(f_i^* - f_i^- \right) \right]^p \right\}^{\frac{1}{p}}, \qquad (15)$$
$$1 \le p \le \infty; \quad j = 1, 2, \ldots, J,$$

where f_i^* and f_i^- are the best and the worst solutions of the i^{th} criterion. After determining the best and the worst solutions for all criteria, the Vikor algorithm has the following steps:

i. Compute the values S_j and R_j, j=1, 2, …, J as:

$$S_j = \sum_{i=1}^{n} w_i \left(f_i^* - f_{ij} \right) / \left(f_i^* - f_i^- \right), \qquad (16)$$

$$R_j = \max_i \left[w_i \left(f_i^* - f_{ij} \right) / \left(f_i^* - f_i^- \right) \right], \qquad (17)$$

where w_i is the weight of i^{th} criterion expressing its importance.

ii. Compute the values Q_j as:

$$
\begin{aligned}
Q_j &= \upsilon\left(S_j - S^*\right)/\left(S^- - S^*\right) \\
&\quad + (1-\upsilon)\left(R_j - R^*\right)/\left(R^- - R^*\right) \\
&\quad \text{where} \\
S^* &= \min_j S_j, \quad S^- = \max_j S_j, \\
R^* &= \min_j R_j, \quad R^- = \max_j R_j,
\end{aligned}
\tag{18}
$$

where υ is the maximum group utility, here $\upsilon = 0.5$.

iii. Sort the values of *S*, *R* and *Q* in decreasing order, obtaining three ranked lists.

iv. The alternative that minimizes *Q* is selected as the compromise solution if two conditions of "acceptable advantage" and "acceptable stability in decision making" are satisfied. For more information about these conditions, refer to [71].

Feature similarity was evaluated using the method described before for both conventional and distance-based feature sets. The threshold value for L1-norm between each pair of feature vectors was empirically set to $1e^{-15}$.

5-2-4 Results for Second Study

For conventional and distance-based feature sets, the similar groups were detected and are listed in table 12.

As shown in table 12, several similar cases were found using this measure. For example, the Hjorth activity parameter is the same as the variation. Furthermore, the COSH distance is the symmetric version of the Itakura-Saito distance.

Table 12. Similar feature groups from the conventional and distance-based feature sets.

Conventional Feature Set	**Group 1**			**Group 2**		
	F36, F38 and F40			F6 and F14		
Distance-based Feature Set	**Group 1**	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**
	F52 and F74	F55 and F75	F56 and F76	F60 and F78	F63 and F79	F64 and F80

From each group of similar features, the feature with the lowest computational complexity was selected as group representative. Therefore, F14, F38 and F40 were removed from the conventional feature set. F74, F75, F76, F78, F79 and F80 were also removed from the distance-based feature set. After removing the redundant features, 45 features remained in the conventional feature set and 26 features remained in the distance-based feature set.

To assess the usefulness of pruning feature sets, the sleep stage classification accuracy before and after feature selection was evaluated using the conventional, distance-based, and total feature sets. The results obtained by the *k*-NN classifier with Euclidean distance are shown in table 13. The optimum number of neighbours for each case was found (shown in brackets in table 13) by evaluating the performance of the classifier for different numbers of neighbours. According to the results, removing similar features led to an average improvement of 0.61% for all cases. The maximum improvement (2.07%) was observed in the pruning of the conventional feature set using the standardization method. Additionally, it is noticeable that the accuracy of the classification with the Min-Max method is in all cases higher than the one with the standardization method. This emphasizes the importance of selecting a proper feature normalization method before classification.

Table 13. Classification accuracy for the original, pruned and both together (total) feature sets using the *k*-NN classifier.

Normalization \ Features	Distance-Based	Pruned Distance-Based	Conventional	Pruned Conventional	Total
STD	60.88 (15)	61.03 (5)	70.90 (15)	72.97 (26)	73.26 (12)
Min-Max	62.30 (10)	62.37 (5)	73.94 (8)	74.10 (8)	74.42 (6)

Tables 14-22 demonstrate the 5-stage (Wake, REM, S1, S2 and S3) classification accuracy results along with the optimum number of features selected by the Vikor method for the three feature sets and the three classifiers. The reliability of the results was validated using 10 times repeated 10-fold cross validation method on the whole data from 10 healthy subjects. Simulations were performed using a PC with 3.40 GHz Intel® Core™ i7-3770 CPU, 8 GB of RAM, Windows 10 (64 bits), and MATLAB R2015b. For each ranked list of features, created by one of the ranking methods, and each classifier, the overall classification accuracy, sensitivity and specificity were calculated for the top 25 features.

The analysis of the results reveals that, starting with one feature, each additional feature typically leads to an increment in the classification accuracy. However, at some point, the increment on the classification accuracy for each additional feature is not significant. Inspired by MCDM problems, in this paper the Vikor method was applied to the classification results for determining the optimal feature number that provides a satisfactory trade-off between the selected number of features and the classification accuracy. This method is one of the most common MCDM techniques with straightforward calculations. Accuracy and number of features were two conflicting criteria with the corresponding weights of 0.7

(w_1) and 0.3 (w_2), respectively, meaning that, in our sleep stage classification system, classification accuracy had priority over complexity. Figure 29 shows a sample of the Vikor method results for the features scaled by standardization method, ranked with ReliefF and classified by *k*-NN classifier. The utopian solution, shown with a black star, represents the ideal solution in which the accuracy is maximum, and the number of features is minimum. The point selected by the Vikor method in each case is the closest point of the Pareto front (the set of solutions) to the utopian solution considering the weights of the two criteria.

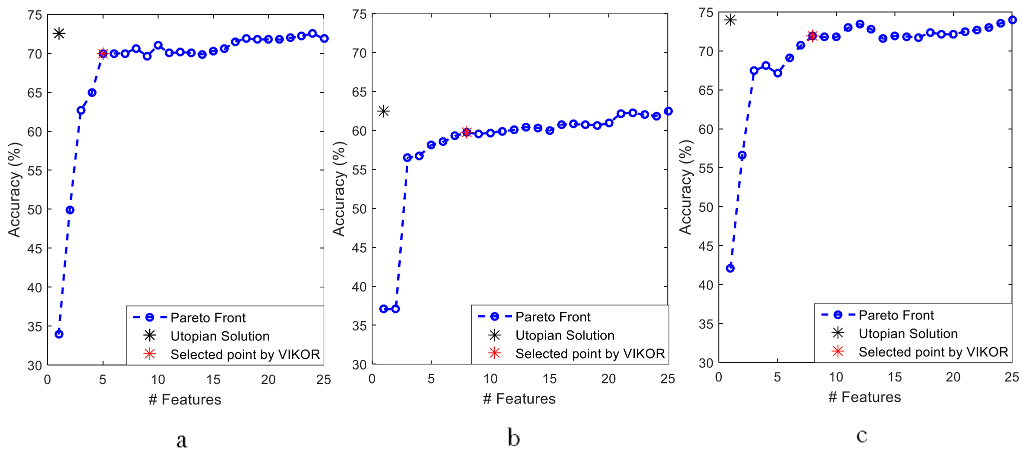

Figure 29. Optimum number of features selected by the VIKOR method for the (a) conventional, (b) distance-based, and (c) total feature sets.

In the following, assessment of the results related to the *k*-NN classifier (tables 14-16) will be presented.

i. The maximum enhancement in classification accuracy after adding the distance-based features to the conventional feature set occurred in mRMR-MID with Min-Max.

ii. For all three feature sets, the maximum accuracy, regardless of the feature normalization method, was achieved by mRMR-MID or mRMR-MIQ methods. Seven or eight (in one case) features were selected by the Vikor method to achieve this accuracy. The Itakura distance of EEG-EOG spectral coefficients, Itakura-Saito distance of EEG-EOG spectral coefficients, and Itakura distance of EMG AR coefficients are among these features.

iii. For all three feature sets, the minimum accuracy, regardless of the feature normalization method, was achieved by the Chi-square method.

iv. For most of the ranking methods, adding distance-based features to the conventional feature set improved the sensitivity and specificity of the classification.

Table 14. *k*-NN classifier results for the conventional feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
#Features	5	6	10	7	9	8	8	8	7	8	8	8
# Neighbours	18	16	20	11	20	20	12	12	12	20	16	8
Sensitivity	72.8	72.9	75.6	72.5	73.7	71	71.5	72.7	71.3	74.6	73.1	72.9
Specificity	93.4	93.2	94	93.4	93.5	92.6	93.1	93.2	92.9	93.8	93.3	93.4
Accuracy	70	70.9	72.1	71.3	72.9	70.8	69.7	71.6	69	71.9	69.2	72.7

Table 15.. *k*-NN classifier results for the distance-based feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	8	11	6	6	6	5	11	12	10	8	10	8
# Neighbours	19	6	16	9	17	9	10	11	18	12	17	12
Sensitivity	64.3	61.7	62.5	65.6	64	63	64.3	63.3	63.9	60	64.7	61.5
Specificity	91.2	90.6	90.4	91.1	90.6	91	91.5	90.8	91	89.9	91	90.3
Accuracy	59.7	59	61.5	60.6	61.9	60	62	60	61	56.3	61.1	56.6

Table 16. *k*-NN classifier results for the total feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	8	8	8	7	7	7	7	6	7	8	8	10
# Neighbours	14	10	11	6	17	12	10	11	11	10	19	10
Sensitivity	75.1	73.3	74.1	77.4	75.1	75.3	76.5	73.4	72.3	70.6	74	75.4
Specificity	93.8	93.7	93.6	94.2	93.9	93.6	94.2	93.5	93	92.4	93.8	94.3
Accuracy	72	71	73.2	73	72.2	72.3	71.1	71	71	70	71	70.3

Next, assessment of results related to the MLFN classifier (tables 17-19) will be discussed.

i. The maximum enhancement in classification accuracy after adding the distance-based features to the conventional feature set occurred in mRMR-MIQ with standardization.

ii. For all three feature sets, the maximum accuracy, regardless of feature normalization method, was achieved by the mRMR-MID or mRMR-MIQ method. Up to 11 features were selected by the Vikor method to achieve this accuracy. The Itakura distance of the EEG-EOG spectral coefficients, Itakura-Saito distance of the EEG-EOG spectral coefficients, and Itakura distance of the EMG AR coefficients are among these features.

iii. Compared to the results of the *k*-NN classifier, the overall accuracy, sensitivity and specificity of the ANN classifier is higher for three feature sets.

Table 17. MLFN classifier results for the conventional feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	9	10	11	11	9	11	8	10	9	9	11	8
Sensitivity	72.6	77.7	75.9	78.3	74.9	76	73.9	74.6	73.4	76.9	73.6	75.4
Specificity	93.7	94.4	94	94.6	93.7	94	93.5	93.6	93.3	94.2	93.4	93.9
Accuracy	79	80	80	80.6	79	79.8	79.8	79.2	78.5	79.7	78.7	79.6

Table 18. MLFN classifier results for the distance-based feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	9	7	7	7	13	7	11	13	15	15	15	15
Sensitivity	62.1	59.9	63.3	61.1	64.8	61.3	63.4	63.6	66.1	64	65.1	63
Specificity	90.5	90	90.9	90.2	91.1	90.3	90.8	90.9	91.5	90.5	91.2	90.7
Accuracy	74.3	72.1	75.2	74	75.6	74	75	74.2	75	73.1	75	73.1

Table 19. MLFN classifier results for the total feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	8	9	9	10	8	11	9	9	9	14	8	10
Sensitivity	75.1	75.4	76.5	76.7	76.7	78.8	74	74.8	73.3	76.3	74	74.2
Specificity	93.8	93.8	94.1	94.3	94.2	94.7	93.5	93.7	93.3	94.1	93.5	93.5
Accuracy	79.5	79.2	80.2	79.9	80.2	80.4	79.2	79.1	79.2	79.5	79.2	78.5

Next, assessment of results related to the DSVM classifier (tables 20-22) will be discussed.

i. The maximum enhancement in classification accuracy after adding the distance-based features to the conventional feature set occurred in mRMR-MIQ with Min-Max.

ii. For all three feature sets, the maximum accuracy, regardless of the feature normalization method, was achieved by the mRMR-MID or mRMR-MIQ methods. Up to 13 features were selected by the Vikor method to achieve this accuracy. The Itakura distance of the EEG-EOG spectral coefficients, Itakura-Saito distance of the EEG-EOG spectral coefficients, and Itakura distance of the EMG AR coefficients are among these features.

iii. Considering that the overall performance of the DSVM classifier, including accuracy, sensitivity and specificity, is the highest among the classifiers used in this paper, it can be concluded that DSVM outperforms *k*-NN and MLFN classifiers in sleep stage classification.

Observing the results for all the classifiers, we can conclude that the accuracy obtained by Min-Max is higher than standardization in most cases. Furthermore, the presence of the distance-based features among selected features by the Vikor method shows their positive contribution to sleep stage classification.

Table 20. DSVM classifier results for the conventional feature set.

	ReliefF		**mRMR-MID**		**mRMR-MIQ**		**Fisher**		**Chi-square**		**IG**	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	10	10	10	11	9	9	8	11	8	9	8	8
Sensitivity	79.2	74.4	80.1	78.5	79	76.3	77.2	76.6	73.2	78.4	76.3	75.7
Specificity	95.3	94.2	95.7	94.9	95.6	94.6	95.2	94.6	94.7	95.4	94.9	94.7
Accuracy	83.7	84.5	84.0	84.7	84.0	83.8	81.5	81.7	81.0	81.9	81.0	81.8

Table 21. DSVM classifier results for the distance-based feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	9	11	7	6	8	6	11	11	9	15	15	15
Sensitivity	61.1	60.6	70.1	63.6	70.3	60.7	64.1	58.3	62.3	62.9	68.5	64.4
Specificity	91.1	90.9	93.4	92.1	93.4	91.1	91.8	90.7	91.7	91.5	92.8	92.5
Accuracy	78.1	77.2	79.7	79.3	79.8	77.8	79.2	78.1	77.8	78.7	79.4	79.2

Table 22. DSVM classifier results for the total feature set.

	ReliefF		mRMR-MID		mRMR-MIQ		Fisher		Chi-square		IG	
	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max	STD	Min-Max
# Features	11	9	8	13	8	11	9	14	9	14	9	15
Sensitivity	79.3	76	81.6	79.8	80.6	80.5	75.1	76.3	75.3	73.8	77.5	76.5
Specificity	95.5	94.9	96.5	96.3	96.1	96	94.6	95.3	94.6	94.3	94.9	94.8
Accuracy	84.8	82.0	84.4	85.5	84.7	85.3	81.3	81.9	80.8	81.6	80.8	81.7

5-3 Performance Analysis of Deep Learning methods for 1D Sleep Stage Classification

In this section, a deep learning-based dimension reduction and classification method was proposed for automatic sleep stage classification. In order to enhance the feature vector, before feeding it to the deep network, a feature selection method (described in sub-section 4-6-1), was applied for removing the features with minimum information. Two-layer SSAE together with Softmax classifier was selected as the deep network model. The performance of the proposed method was compared with Softmax and k-nearest neighbour classifiers. Simulation results show that proposed deep learning structure outperformed others in terms of classification accuracy.

5-3-1 Methodology

Figure 30 shows an overview of the proposed sleep stage classification algorithm with the proposed feature transformation scheme.

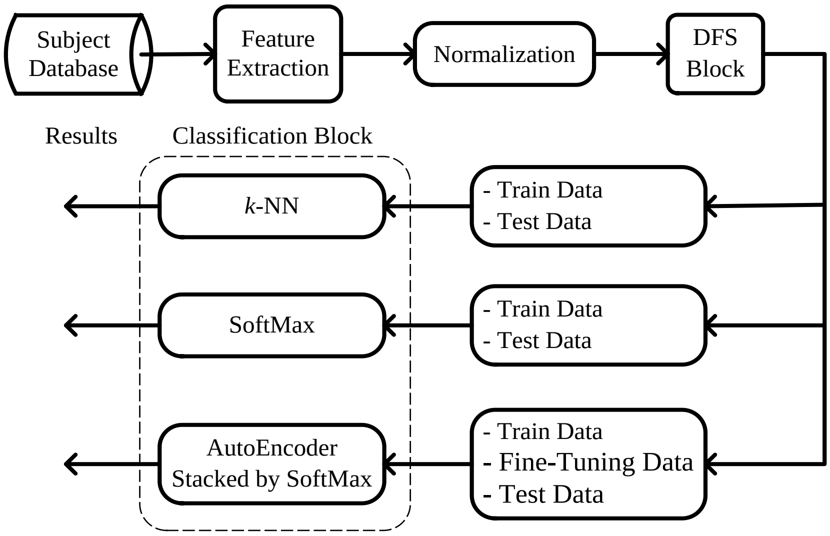

Figure 30. Flowchart of the proposed algorithm for 1D sleep stage classification.

In this study, we used the publicly available dataset, ISRUC-Sleep [85] which details are presented in section 4-1. For the evaluation of our proposed method, we used C3-A2 EEG channel, right EOG and chin EMG channels. In this study, we used all the data from 10 healthy subjects from ISRUC-Sleep dataset. The number of epochs for these 10 subjects is 954, 941, 824, 794, 944, 853, 814, 1000, 969, and 796. Totally we have 8889 epochs from this database. To avoid overfitting, we used all of them. All the pre-processing methods in sub-section 4-2 were applied to this dataset to prepare it for further processing.

All signals used in this study, were divided into 30-second epochs. A set of features were extracted from each epoch of EEG, EOG and EMG recordings of each subject. This feature set includes 49 features that can be considered as time, frequency, joint time-frequency domain, entropy-based and nonlinear types. Table 10 in sub-section 5-2-1 shows a summary of these features.

In order to standardize the range of features, a Min-Max normalization method was applied. Details about this normalization method can be found in sub-section 4-5.

There are many potential advantages in removing the features before final modelling and classification. Fewer features mean lower computational complexity. Also, some features may reduce the performance by their corrupt distributions. Consider a feature that a single value for all the samples. According to [218], this feature is called "zero-variance predictor". Even if it has little effect on the next step, this feature should be discarded from feature set, because it has no information. Similarly, some features may have few unique values that occur with low frequency. These features are called "near-zero variance predictors". Kuhn et al. [218] defines two criteria for detecting near-zero variance features as follows:

i. The ratio of unique values to the number of samples is low, for example 10%.

ii. The ratio of the frequency of the most dominant value to the frequency of the second dominant value is high, for example 20.

Using these two criteria, we applied Discriminative Feature Selection (DFS) to remove the features that didn't have enough discriminative power. As a result, the dimension of feature set was reduced to 37.

The EEG, EOG and EMG signals of each subject in the dataset were divided into 30-second epochs. After feature extraction and normalization, the feature sets were fed to DFS block to eliminate the near-zero variance features. According to the criteria mentioned before, 12 features were recognized as near-zero variance feature and removed from our sleep data model, as follows: maximum value (F1), minimum value (F2), variation (F5), median (F8), Petrosian fractal dimension (F31), permutation entropy (F30), Hjorth parameter (Activity) (F10), zero crossing number (F9), total power in the EMG frequency spectrum (F43), mean of power in the EMG frequency spectrum (F45), absolute energy of the time domain EMG signal (F47), maximum value of time domain EOG signal (F38). After the feature vector was set, data were divided into two parts, training, testing, using 10-fold cross validation method. For fine tuning step of SSAE, part of training data was utilized. Our deep learning consists of three layers, a two-layer SSAE and a Softmax layer. The number of hidden units for the first and second layer of SSAE was 20 and 12, respectively.

For finding the best hyper-parameters for the autoencoders, we tried several models by adjusting sparsity regularization parameter, weight regularization parameter and number of iterations. We used autoencoders with logistic sigmoid activation function for both layers.

The performance of the proposed algorithm was compared with two other classifiers: Softmax and *k*-NN classifier. The number of neighbours was set to 18 and Euclidean distance was used as a measure of distance for *k*-NN.

5-3-2 Results

To evaluate our system's performance, we used classification accuracy as the evaluation criterion. Table 23 shows the individual sleep stage and overall classification accuracy extracted from the confusion matrix for the three different classifiers. The boldface-numbers indicate the best performance. It is noticeable that SSAE method outperforms the other two classifiers in terms of overall accuracy. Also, for the individual sleep stages, in most cases SSAE discriminates the stages better. In addition to the higher performance, SSAE provides a considerable reduction in the dimension of the feature vector. Considering that the second layer of SSAE had 12 hidden units, SSAE succeeded to decrease the dimension from 37 to 12, which means 67% reduction. Therefore, SSAE is a powerful tool to generate more descriptive features from the original feature vector.

Table 23. Results of the statistical analysis for comparison of each stage and overall accuracy.

Classifiers	Wake (%)	REM (%)	S1 (%)	S2 (%)	S3 (%)	Overall Accuracy (%)
Softmax	80	61.66	65	**90**	78.33	74.9
***k*-NN**	85	66.66	61.66	70	83.33	73.33
SSAE	**91**	**77**	**69**	87	**87**	**82.2**

In order to confirm the advantage of DFS block, the performance of SSAE-based sleep stage classification with and without this step was investigated. Without using DFS block, 49 original features were fed to SSAE. The classification accuracy achieved in this way was 74.1% which is almost 8% less than the accuracy with DFS block.

5-4 Performance Analysis of Deep Learning methods for 2D Sleep Stage Classification

Most of the existing methods for automatic sleep stage classification are relying on hand-crafted features. In this study, the goal is to develop a deep learning-based method that automatically exploits time-frequency spectrum of EEG signal, removing the need for manual feature extraction. Using CWT, we extracted the time-frequency spectrogram for the EEG signal of 10 healthy subjects and converted to RGB images. The images were classified using transfer learning of AlexNet, a pre-trained CNN. Evaluation results showed that our method can achieve state of the art accuracy, while having higher overall sensitivity.

5-4-1 Methodology

In this study, we benchmarked our work with EEG recordings of a publicly available database. We used the open-access comprehensive ISRUC-Sleep dataset [85]. All the pre-processing methods in sub-section 4-2 were applied to this dataset to prepare it for further processing.

EEG is a nonstationary signal and temporal or spectral representation by itself cannot reflect the information hidden in EEG. CWT [219] is a proper method to analyse the non-stationary signals such as EEG. As the window of CWT gets shrunk and dilated, it maps the variations of the signal into a time-frequency spectrogram. In this study, each epoch of EEG data is presented by time-frequency spectrogram image. Epochs are concatenated in a way that there is no overlap between them. There are a set of parameters that affect the CWT spectrogram, including the mother wavelet and number of frequency bins. After extracting the spectrogram for each epoch, we converted it into an RGB image by using a colour map (or colour table) to map pixel values into the actual colour values.

Our approach was based on transfer learning, which is developed based on the fact that human brain is able to discover the underlying structure in previously learned knowledge and transfer this knowledge to new tasks [220]. Transfer learning is a machine learning method where the learning of a new task is improved through the use of the learning for a previous task as a starting point. The main advantage of transfer learning is the reduction in the number of training samples necessary to achieve a desired performance on correlated problems.

We chose AlexNet [207] which is a CNN trained on a subset of ImageNet datasets [221]. CNNs are deep feed-forward neural networks with trainable weights. These networks are mainly used for visual tasks since 1980s [222], minimizing the need for pre-processing related to feature extraction. AlexNet was trained on 1.2 million images and can classify them into 1000 categories. This network comprises 25 layers where eight of them have learnable weights (5 convolutional layers and 3 fully connected layers). The input is a 227 by 227 pixels RGB image.

Images were created for C3-A2 sensor EEG signal for healthy subjects as explained in sub-section 4-1. To reduce the computational complexity while preserving the information in the EEG, a decimation with factor 2 was applied to the signals. Then, using DWT [219], the signals were filtered with lower cut-off frequency of 0.1 and higher cut-off frequency of 40. CWT mother wavelet was set to cmor15-1. For the time-frequency representation of each epoch, the number of frequency bins was chosen to be 2048 with the aim of properly capturing sleep dynamics. The time-frequency spectrograms were created in the range of 0.3 Hz and 35 Hz, according to the guidelines of AASM. The spectrograms were converted to RGB images using Jet colour map with 256 colours. The mapping of spectrogram to images was done through linear mapping. Although the dynamic range of sleep spectrograms was relatively high, using logarithmic/exponential mapping would either highlight the unwanted noise in the signals or fade out the desired high activity areas. While the generated image resolution was 3000 (points in time) in 2048 (points in frequency), the final image resolution was set to AlexNet input size (227 in 227) through image resizing. The last three layers of AlexNet are by default configured for 1000 classes. We transferred these layers to fit the five-stage sleep classification problem by replacing these layers with a fully connected Softmax layer. Simulations were done using a PC with 3.40 GHz Intel® Core™ i7-3770 CPU, 8 GB of RAM, Windows 10 (64 bits), and MATLAB R2015b. Several sets of training parameters were tested to determine the optimal set of parameters. AlexNet was trained using Stochastic Gradient Descent (SGD) with learning rate of 10^{-4} on mini-batches of 64 samples. To assess the generalization ability, the proposed model was validated by 10-fold cross validation.

5-4-2 Results

Figure 31 shows sample RGB images of each sleep stage, as typical inputs to the pretrained CNN. Each individual image depicts the signal energy distribution during an epoch, i.e. 30s of EEG signal (horizontal axis) and frequency interval of 0.3 to 35 Hz (vertical axis). These images can be easily interpreted using AASM guidelines. The spectrogram for wake stage (figure 31.a) shows high Alpha band activity in the frequency interval of 8 to 13Hz. Figure 31.b, corresponding to S1, shows activity in the range of 4-7Hz with slow background activity (in frequencies less than 2 Hz).

As the subject moves to S2, low amplitude activity in the range of 12-14 Hz, together with low frequency K complexes, are noticeable. During S3, slow wave activity in frequencies less than 4 Hz is typical. Mixed frequency, low amplitude activity, resembling S1 is present in REM spectrogram.

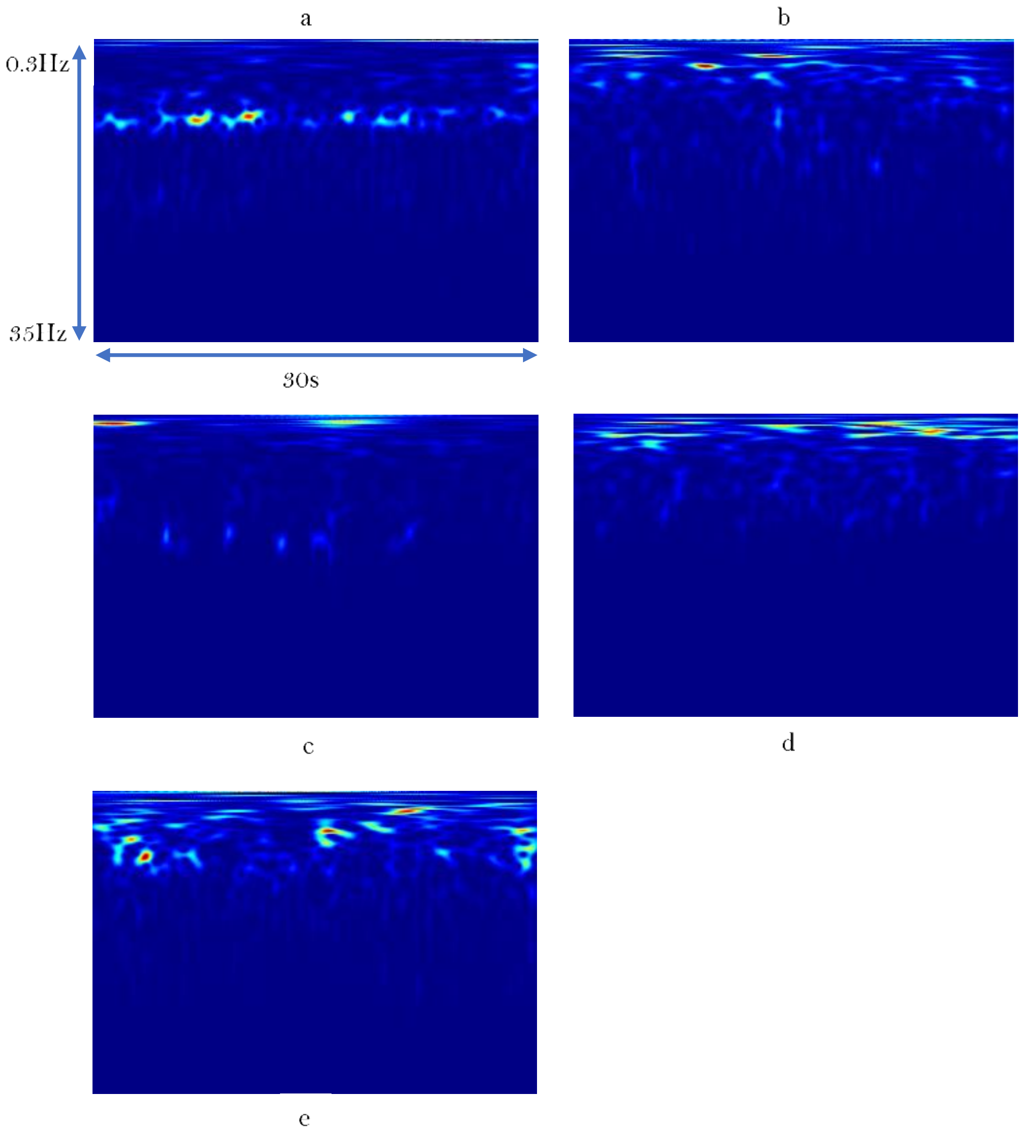

Figure 31. Sample RGB image for five sleep stages created from time-frequency spectrogram (a) Wake, (b) S1, (c) S2, (d) S3, and (e) REM.

The proposed method performance was evaluated using per stage and overall performance scores, including per stage sensitivity, specificity, accuracy and overall accuracy as shown in Table 24 The most correctly classified stage is wake (96.3%) followed by S3 (89.2%), S2 (84.1%) and REM (82.4%). As typical for automatic sleep stage classification systems, the lowest classification accuracy is for S1. The main reason for this is that EEG signals have similar patterns in S1, REM and even S2. On the other hand, the distribution of sleep stages along a sleep study is not balanced. Usually S1 is rare compared to other stages and the classification system has difficulty in learn enough and predict it correctly. The sensitivity results confirm the above-mentioned similarity of EEG patterns in S1, S2, and REM. The specificity or true negative results for all stages are higher than 90%. Our results can be compared to the state of the art works presented in [179], although our system was trained with a lower number of samples. In contrast to our work, they achieved the highest per stage accuracy in S3 (94%). The accuracy for wake stage was 87% compared to 96.3% in our work. The system in [179] outperformed our system in classifying S1 stage (75% vs. 64%). On the other hand, overall sensitivity of our system is higher. Both systems reached state of the art overall accuracy [173], [175].

Table 24. Sensitivity, specificity, accuracy per stage, and overall accuracy for proposed method. All numbers are in percentage.

	Sensitivity	**Specificity**	**Accuracy (per stage)**
Wake	92	97	96.3
S1	64	94	67.6
S2	79	91	84.1
S3	85	95	89.2
REM	78	96	82.4
Overall Accuracy			**84**

Chapter 6

6 Discussion and Conclusion

There are many solutions for the denoising and classification of sleep PSG signals. In this thesis, we proposed three methods and explored their utility and benefits in the study of sleep. As the first method, in the pre-processing part, we proposed an artefact detection algorithm based on the adaptive filtering technique. Two different types of artefacts (body and chin movement from EMG and eye movement from EOG) were studied; the performance of the denoising algorithm was quantified by the classification accuracy. The main advantage of the developed algorithm is the cancelation of the noise instead of removing it. This methodology avoids data loss by keeping the noisy epochs.

As the second novel method in this thesis, we proposed studying the conventional classifiers performance on the comprehensive features set used in sleep stage classification. The main goal of this work was to have an objective look at the conventional classifier's performance.

The emphasis of this thesis was to explore the potential of deep learning in sleep stage classification. For this aim, we proposed two new methodologies for 1D and 2D sleep stage classification. One for feature transformation (dimension reduction) and classification in 1D environment and the other one for feature extraction and classification in 2D environment.

In the following, the first sub-sections will discuss the data we used in this thesis. The second sub-section will present the benefits of denoising EEG signals. The third sub-section will describe the cons and pros of conventional classifiers results and handcrafted features. In fourth sub-section, the most important topic of this thesis, deep learning for classification and feature extraction/transformation, will be presented.

6-1 Datasets

In this sub-section we will discuss cons and pros of the datasets used in this thesis.

Sleep-EDF Database [Expanded]:

As mentioned in chapter 4, this database includes two groups of recordings, namely SC and ST. In our simulations, we didn't use SC files, since EMG data for these recordings was a zero-amplitude or no data recording.

Through careful analysis ST recordings, a number of issues were detected that made some of the recordings unsuitable for being used in the evaluations. These issues are as follows:

- Lack of S4 (according to R&K guidelines),
- Artefacts such as severe movement or sensor misconnection,
- Unsynchronized EEG data and hypnogram,
- Lack of stage 3 epochs,
- Severely corrupted EEG data.

As a result, six recordings were selected out of twenty-two and the corresponding hypnograms were converted from R&K to AASM. Eliminating 16 out of twenty-two recordings led to major data reduction which can limit the power of the final classifying system.

Pz-Oz channel EEG together with submental chin EMG and horizontal EOG each sampled at 100 Hz were used in the evaluations. Table 9 illustrates the number of stages available per subject.

ISRUC_Sleep Database:

For our evaluations, we used PSG recordings from healthy subjects. Nine male and one female subjects aged between 30 and 58 participated in the recordings. Each recording contains signals from 19 channels. The data include six EEG channels: F3-A2, C3-A2, O1-A2, F4-A1, C4-A1, and O2-A1 from which we selected the C3-A2 EEG channel. The C3-A2 channel is the commonly used EEG channel in sleep stage classification (refer to chapter 3) and is among the recommended channels by AASM. In addition to one EEG signal, we used the signals from right EOG and chin EMG, and ECG channels of all ten subjects.

The recordings in this dataset didn't have the issues mentioned for previous database. However, the main limitation of ISRUC_sleep database was the interscorer variability especially for S1. The studied sleep stage classification system was evaluated based on a hypnogram created from the consensus of two experts on visual sleep scoring. There were some cases of disagreement, especially on S1 which made the confirmation of the obtained results difficult. Moreover, the database was pre-processed, and raw data was not available for possible changes in the pre-processing step.

One common issue with open access sleep databases is the problem of unbalanced data, meaning that sleep stages are not equally present in the database. Training a classifier with unbalanced data leads to unequal learning of different stages, thus stage-wise accuracy for S1 is usually lower than for the other classes. This problem becomes even more highlighted with the fact that EEG has similar wave patterns for S1 and REM. Therefore, it is important to use a feature extraction and classification scheme that can handle the domain knowledge and is not dependent on the availability of labelled data. Deep learning techniques are suitable choices considering these criteria.

6-2 Denoising EEG Signal

To the best of our knowledge there is a deficiency of EEG artefact removing studies in the sleep stage classification context that can remove the EOG and EMG artefacts from EEG without rejecting epochs and at the same time evaluates the performance of the classifier for the denoised data. Sub-section 5-1 is a contribution in this regard. Our findings showed that the proposed method for artefact cancelation is reliable for sleep stage classification giving an accuracy improvement. According to the obtained results, artefact removed-EEG allowed a classification accuracy improvement of around 14% compared to raw EEG.

Conventional classifiers are not able to handle domain knowledge. In other words, they cannot extract relevant information from raw data. Therefore, feature extraction is an essential prerequisite for them. Considering that the final classification accuracy is in direct relationship with feature vector quality and enhancement of raw signal quality can improve the quality of features, proper removal of EMG and EOG artefacts is of paramount importance. The same is not valid for deep learning-based sleep stage classification. In these systems, the leakage of information from EOG and EMG channels to EEG can be a positive contribution to improve the classification accuracy since this leakage can be used for efficient unsupervised feature extraction.

This study is supported and validated by the following publication:

- A. A. Gharbali, J. M. Fonseca, S. Najdi, and T. Y. Rezaii, "Automatic EOG and EMG Artifact Removal Method for Sleep Stage Classification," *IFIP Advances in Information and Communication Technology*, *Springer International Publishing*, 2016, pp. 142–150.

6-3 Conventional Classifiers

As described in sub-section 5-2-1, we analysed the performance of two common classifiers, namely *k*-NN and MLFN in sleep stage classification. Also, this study was a contribution regarding the mutual relation between classifiers and feature set. A comprehensive set of the most common features in sleep stage classification together with several feature ranking techniques were used. Table 11 illustrates the top 10 features selected by each method. Itakura Spectral Distance (ISD) (F36) always appears in the top 10 for all the methods. Although different feature ranking methods have their own specific criteria for ranking the features, observing ISD in the top 10 list, means that ISD is a preferable feature for all feature selection methods.

In addition to ISD, there are some other features that can be considered most preferable. EEG ZCR (F9) is a simple, yet effective feature that is listed in top 10 by all the methods except ReliefF. Following ZCR, Petrosian fractal dimension (F31), Hurst exponent (F35), WP feature (F22), approximate entropy (F29), spectral entropy (F27), and Hjorth mobility parameter (F11) were selected by at least five ranking methods to be included in top 10 list.

The optimum number of features for each method, which is selected by the Kneedle algorithm, is also shown in table 11. For MLFN and *k*-NN classifiers, a slight difference exists in the optimum number. Considering the maximum accuracy that the methods reach in their optimum points, the MRMR-MID method using *k*-NN classifier outperforms all the others with seven selected features. Also, both MRMR methods using MLFN classifier outperform all the other methods with five features.

The CMIM method reaches its best accuracy with the first 3 features on both the classifiers. Considering figure 27, its accuracy is equal or less than the MRMR-MID method's accuracy at that point. Unanticipatedly, none of the aggregation methods outperformed the rest of the feature ranking methods. One possible reason is that the aggregation methods, especially Borda, are affected by the performance of all the methods from best to worst.

It should be considered that the results presented in this study are obtained through using Physionet Sleep-EDF Expanded Database which is already used in several previous sleep studies [40], [41], [72], [86] and can be supposed as verified enough to be used in such a comparative study.

This study is supported and validated by the following publications:

- S. Najdi, A. A. Gharbali, and J. M. Fonseca, "Feature ranking and rank aggregation for automatic sleep stage classification: a comparative study," *Biomedical Engineering Online*, vol. 16, no. S1, p. 78, Aug. 2017.

- S. Najdi, A. A. Gharbali, and J. M. Fonseca, "A Comparison of Feature Ranking and Rank Aggregation Techniques in Automatic Sleep Stage Classification Based on Polysomnographic Signals," in *4th International Conference, IWBBIO*, 2016, pp. 230–241.

In sub-section 5-2-3, the goal was to evaluate conventional classifiers performance in sleep stage classification with a new feature set. The new feature set consisted of conventional and distance-based feature set. The distance-based feature set included 32 features that were extracted by calculating the distance between AR and spectral coefficients of EEG, EOG, EMG, and ECG signals. The distance measures used were Itakura, Itakura-Saito and COSH, all common in speech signal processing. Extensive assessments were performed to reveal the weaknesses and strengths of these features' classification ability and classifiers' performance.

Similar features were removed from the feature sets by thresholding L1-norm between feature vectors. This step was advantageous because removing these features reduces the final feature vector dimensionality and enhances the stability of the feature-ranking results. Moreover, according to the results of table 13, this step led to an improvement in the classification accuracy. This improvement was expected since the existence of redundant features has no positive effects on the classification results and increases the computational complexity of the whole system.

To find the most useful features for describing PSG signals for classifiers, six feature-ranking methods, namely ReliefF, mRMR-MID, mRMR-MIQ, Chi-square, Fisher, and IG, were applied to the three feature sets (conventional, distance-based and total).

To further investigate the contribution of distance-based features, three different classifiers, *k*-NN, MLFN and DSVM, were used. Previous studies [20], [71] showed that combining different types of features, i.e. temporal, spectral, time-frequency domain and nonlinear, would lead to a satisfactory level of classification accuracy using a lower number of features. In this study, it has been shown that using distance-based features together with conventional ones can further improve the performance of the sleep scoring system. This improvement is noticeable in the results of all three classifiers. According to the results of the Vikor method in tables 14-22, carefully selected measures from the total feature set are sufficient to reach an 85% average accuracy.

Regarding the compatibility of feature ranking and classifier, all classifiers achieved the highest accuracy with either mRMR-MID or mRMR-MIQ. In particular, mRMR-MID with Min-Max normalization gave the highest accuracy with 13 features in which the EEG-EOG Itakura distance of spectral coefficients and EMG Itakura distance of AR coefficients are selected from the distance-based feature set (table 22). This result is consistent with our previous studies in which we used the Physionet sleep database [180] to evaluate the applicability of rank aggregation to the sleep scoring problem. Moreover, simulation results showed that DSVM outperformed the other two classifiers for all the feature ranking and normalization methods.

6-3-1 GPU vs. CPU Performance Analysis for MLFN

Training the classifier with sufficient amount of the data is crucial for obtaining a robust and reliable classifier. One of the main concerns in next generation sleep monitoring systems is the ability of hardware to manage that huge amount of data. GPUs or graphics processing units are used for mathematically-intensive tasks to assist CPU and improve the overall performance of computer or electronic device.

In our study for evaluation of the conventional classifiers, considering the computational complexity of MLFN, the GPU unit was utilized to accelerate the simulation process with significant differences when compared to the CPU. The MATLAB parallel computing toolbox was used for distributing the processes among several sessions in the computer. A maximum of eight local workers could be provided by the parallel computing toolbox on a single workstation [217]. Nevertheless, the number of usable workers for a process is dependent on the number of cores of the processor in the computer. In our case, a core-i7 processor provided four workers considering the version of MATLAB. A NVIDIA GeForce GT 640 card was used as GPU. The elapsed time for 20 neurons of ANN using CPU was 4426.57 seconds while using the GPU it was 1116.51 seconds. Therefore, it can be determined that, for this type of computation, GPU can be at least 4-times faster than CPU.

This study is supported and validated by the following publication:

- A. A. Gharbali, S. Najdi, and J. M. Fonseca, "Investigating the contribution of distance-based features to automatic sleep stage classification," *Computers in Biology and Medicine*, vol. 96, pp. 8–23, May 2018.

6-4 Deep Learning Classifiers

Although feature transformation based on deep learning has been already used in several machine learning applications, the advantages and potentials of applying these methods in sleep stage classification problems have not been explored yet. The study in sub-section 5-3 is a contribution in this regard. A new method was proposed for dimension reduction and classification based on SSAEs. The results showed that SSAE can be considered as an appropriate tool for decreasing the complexity of sleep scoring issues and classifying compared to *k*-NN and Softmax classifiers. DFS block showed in this study promising results not only in reducing computational complexity but also, on increasing the accuracy. SSAE outperformed other classifiers on all the stages except S2, which leads SSAE to have better overall accuracy comparing others.

- S. Najdi, A. A. Gharbali, and J. M. Fonseca, "Feature Transformation Based on Stacked Sparse Autoencoders for Sleep Stage Classification," *in Technological Innovation for Smart Systems*, 2017, pp. 191–200.

The promising results of this study imply that deep learning techniques can be proper alternatives to the conventional classifiers. The main disadvantage of conventional classifiers is that they are closely dependent on a set of hand-crafted feature set extracted from PSG signals. The quality of this feature set is determined by the expert's creativity and prior knowledge of the problem. Moreover, extracting these features requires extensive human labour and their modelling capabilities are limited. Therefore, it is highly desirable to make the feature learning and classification stages automatic and less dependent on expert's knowledge.

Most of the existing methods for automatic sleep stage classification rely on hand-crafted features and there are limited number of studies that apply deep learning techniques for unsupervised feature learning and classification. On the other hand, in image processing areas, deep learning techniques already achieved significant results and their usefulness and efficiency are confirmed in several applications. Motivated by these results, a new sleep scoring algorithm based on image classification was proposed.

The idea of using 2D images instead of conventional hand-crafted 1D features is the application of deep learning-based methods (CNN) which is an emerging trend in sleep stage classification. It was always a concern for researchers to provide an efficient feature set for classifiers to increase their performance. Therefore, several feature extraction and selection methods were proposed to reach this goal. The main advantage of this method was elimination of the need for manual feature extraction and selection, while taking advantage of the advancements in deep image classification domain. High resolution time-frequency spectrograms of sleep epochs were extracted using CWT and converted to RGB images. The extracted images were intuitive and interpretable according to AASM guidelines. These images were fed to a pre-trained CNN, AlexNet. We took advantage of transfer learning which allows rapid progress and facilitates the modelling for the second task. Classification results showed that although our system was trained with a lower number of samples compared to similar studies, it was able to achieve state of the art accuracy and higher overall sensitivity.

It should be noted that because of hardware limitations we couldn't use all the epochs of ten healthy subject's data for training CNN. Classification accuracy could be further improved with more training data. Nevertheless, this study was one of the first studies in the sleep stage classification area where the need for feature extraction and selection stages was eliminated by using a deep network.

This study is supported and validated by the following publication:

- A. A. Gharbali, S. Najdi, and J. M. Fonseca, "Transfer Learning of Spectrogram Image for Automatic Sleep Stage Classification," *Springer International Publishing*, 2018, pp. 522–528.

6-6 Future Work

Due to the database available at the moment of this research, EOG and EMG channels were utilized for cancellation of their contamination in EEG. The performance of the proposed method can be improved in the future using other PSG channels like ECG. On the other hand, the proposed artefact detection scheme was only able to detect linear artefact by evaluating cross correlation. Extending it to nonlinear artefact removal can extend the applicability of the proposed framework.

Moreover, one of the main criteria to evaluate the performance of the denoising algorithms is SNR. Considering the lack of clean EEG, SNR evaluation was not applicable in our study. As a future work, to have a clean EEG instead of using conventional sleep records, it is suggested to get the Electrocochleography (ECoG) database. In contrast with the non-invasive techniques (EEG), ECoG is an invasive technique which provides brain signals that have an exceptionally high SNR, less susceptibility to artefacts than EEG and a high spatial and temporal resolution. Then, it is possible to manually add the signals from EOG, EMG and ECG for denoising purposes.

For evaluating the performance of conventional classifiers, in sub-section 5-2-1, we tried to involve a comprehensive list of features and even newly introduced ones in the sleep study area. Also, a group of conventional feature selection techniques which performs selection independent of classifiers were used. We applied these techniques to the available datasets at the moment and tried to pre-process them in the same way to remain objective. Nevertheless, generalizing these results to all future sleep studies requires further study and analysis by using other sleep databases as well.

Also, in this study for evaluating the generalization ability of classifiers we used repeated random subsampling validation. In [19], it is mentioned that due to the data subdivision dependency resulted from validation methods that are based on random subsampling, patient cross validation was preferred. Therefore, future steps will involve verifying the results with different databases, applying and comparing more rank aggregation methods and using patient cross validation and comparing the results with common validation methods.

Despite the advantages of the study in sub-section 5-2-3, there were some limitations. For example, the proposed system was designed for classification of data acquired from healthy subjects. Generalization of the results to unhealthy or elderly subjects would require modifications in the classification rules. Future work can include the analysis of other databases as well as different classifiers for further investigation. Furthermore, comparing the performance of handcrafted features with automatically extracted features by deep networks can be useful for further development of deep learning-based methods.

For 1D deep learning classification study future works can be as follows:

- Comparing the performance of other conventional classifiers such as SVM and RF with SSAE in sleep stage classification;
- Comparing other deep learning classifiers performance with the SSAE;
- Compare SSAE's performance with other classifiers using different sleep databases;
- In order to have a better comparison it is better to add other criteria rather than accuracy such as sensitivity and specificity;

Transfer learning was used as deep learning approach for classification of sleep epoch images. The reason that higher accuracy was not achieved in our study for image classification is the deficiency of raw data in the available datasets. Deep learning techniques are data intensive and we used a dataset that included just ten healthy subjects. This number is quite small compared to the amount of data required to correctly train a deep architecture like CNN. Future work can include providing larger datasets to approach an optimal accuracy, further optimization of parameters and testing of different pre-trained CNNs.

References

[1] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, R. M. Lioyd, C. L. Marcus, and B. V. Vaughn, *AASM - Manual for the Scoring of Sleep and Associated Events version 2.1*. 2014.

[2] D. Migotina, "Symbolic Representation of The Sleep Electroencephalogram Application in The Analysis of Sleep," Instituto Superior Técnico, 2014.

[3] M. G. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, and A. Walters, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep," *Sleep Med.*, vol. 2, no. 6, pp. 537–553, Nov. 2001.

[4] A. A. Gharbali, J. M. Fonseca, S. Najdi, and T. Y. Rezaii, "Automatic EOG and EMG Artifact Removal Method for Sleep Stage Classification," in *IFIP Advances in Information and Communication Technology*, Springer International Publishing, 2016, pp. 142–150.

[5] J. Kim, "A Comparative Study on Classification Methods of Sleep Stages by Using EEG," *J. Korea Multimed. Soc.*, vol. 17, no. 2, pp. 113–123, Feb. 2014.

[6] B. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms," *J. Med. Syst.*, vol. 38, no. 3, p. 18, Mar. 2014.

[7] M. Peker, "A new approach for automatic sleep scoring: Combining Taguchi based complex-valued neural network and complex wavelet transform," *Comput. Methods Programs Biomed.*, vol. 129, pp. 203–216, Jun. 2016.

[8] A. Subasi, M. K. Kiymik, M. Akin, and O. Erogul, “Automatic recognition of vigilance state by using a wavelet-based artificial neural network,” *Neural Comput. Appl.*, vol. 14, no. 1, pp. 45–55, Mar. 2005.

[9] M. E. Tagluk, N. Sezgin, and M. Akin, “Estimation of Sleep Stages by an Artificial Neural Network Employing EEG, EMG and EOG,” *J. Med. Syst.*, vol. 34, no. 4, pp. 717–725, Aug. 2010.

[10] A. R. Hassan and M. I. H. Bhuiyan, “An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting,” *Neurocomputing*, vol. 219, pp. 76–87, 2017.

[11] A. R. Hassan and M. I. H. Bhuiyan, “Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting,” *Comput. Methods Programs Biomed.*, vol. 140, pp. 201–210, 2017.

[12] M. Diykh and Y. Li, “Complex networks approach for EEG signal sleep stages classification,” *Expert Syst. Appl.*, vol. 63, pp. 241–248, 2016.

[13] M. Diykh, Y. Li, and P. Wen, “EEG sleep stages classification based on time domain features and structural graph similarity,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, 2016.

[14] S. Mahvash Mohammadi, S. Kouchaki, M. Ghavami, and S. Sanei, “Improving time–frequency domain sleep EEG classification via singular spectrum analysis,” *J. Neurosci. Methods*, vol. 273, pp. 96–106, 2016.

[15] N. Burioka, M. Miyata, G. Cornélissen, F. Halberg, T. Takeshima, D. T. Kaplan, H. Suyama, M. Endo, Y. Maegaki, T. Nomura, Y. Tomita, K. Nakashima, and E. Shimizu, “Approximate entropy in the electroencephalogram during wake and sleep.,” *Clin. EEG Neurosci.*, vol. 36, no. 1, pp. 21–24, 2005.

[16] M. Obayya and F. E. Z. Abou-Chadi, “Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm,” in *2014 31st National Radio Science Conference (NRSC)*, 2014, pp. 265–272.

[17] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, Y.-H. Pan, and Y.-H. Wang, "Automatic Stage Scoring of Single-Channel Sleep EEG by Using Multiscale Entropy and Autoregressive Models," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1649–1657, Jun. 2012.

[18] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, and H. Dickhaus, "Classification of Sleep Stages Using Multi-wavelet Time Frequency Entropy and LDA," *Methods Inf. Med.*, vol. 49, no. 3, pp. 230–237, Jan. 2010.

[19] L. J. Herrera, C. M. Fernandes, A. M. Mora, D. Migotina, R. Largo, A. Guillen, and A. C. Rosa, "Combination of Heterogeneous EEG Feature Extraction Methods and Stacked Sequential Learning for Sleep Stage Classification," *Int. J. Neural Syst.*, vol. 23, no. 03, p. 1350012, Jun. 2013.

[20] M. Radha, G. Garcia-Molina, M. Poel, and G. Tononi, "Comparison of feature and classifier algorithms for online automatic sleep staging based on a single EEG signal," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 1876–1880.

[21] H. G. Jo, J. Y. Park, C. K. Lee, S. K. An, and S. K. Yoo, "Genetic fuzzy classifier for sleep stage identification," *Comput. Biol. Med.*, vol. 40, no. 7, pp. 629–634, 2010.

[22] L. J. Herrera, a. M. Mora, and C. M. Fernandes, "Symbolic Representation of the EEG for Sleep Stage Classification," in *11th International Conference on Intelligent Systems Design and Applications*, 2011, pp. 253–258.

[23] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Comput. Biol. Med.*, vol. 42, no. 12, pp. 1186–1195, Dec. 2012.

[24] S. Güneş, K. Polat, and Ş. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7922–7928, Dec. 2010.

[25] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated

sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, Oct. 2012.

[26] Yi Li, Fan Yingle, Li Gu, and Tong Qinye, "Sleep stage classification based on EEG Hilbert-Huang transform," in *2009 4th IEEE Conference on Industrial Electronics and Applications*, 2009, pp. 3676–3681.

[27] T. H. Sanders, M. McCurry, and M. a Clements, "Sleep stage classification with cross frequency coupling," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, vol. 2014, pp. 4579–4582.

[28] R. Boostani, F. Karimzadeh, and M. Nami, "A comparative review on sleep stage classification methods in patients and healthy individuals," *Comput. Methods Programs Biomed.*, vol. 140, pp. 77–91, 2017.

[29] A. R. Hassan and M. I. H. Bhuiyan, "A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features," *J. Neurosci. Methods*, vol. 271, pp. 107–118, 2016.

[30] P. Tian, J. Hu, J. Qi, X. Ye, D. Che, Y. Ding, and Y. Peng, "A hierarchical classification method for automatic sleep scoring using multiscale entropy features and proportion information of sleep architecture," *Biocybern. Biomed. Eng.*, vol. 37, no. 2, pp. 263–271, 2017.

[31] A. Ouanes and L. Rejeb, "A Hybrid Approach for Sleep Stages Classification," *Proc. 2016 Genet. Evol. Comput. Conf. - GECCO '16*, pp. 493–500, 2016.

[32] M. Peker, "An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms," *Neurocomputing*, vol. 207, pp. 165–177, 2015.

[33] A. R. Hassan and M. I. Hassan Bhuiyan, "Automatic sleep scoring using statistical features in the EMD domain and ensemble methods," *Biocybern. Biomed. Eng.*, vol. 36, no. 1, pp. 248–255, 2016.

[34] Ö. F. Alçin, S. Siuly, V. Bajaj, Y. Guo, A. Şengu¨r, and Y. Zhang, "Multi-category EEG signal classification developing time-frequency texture features based Fisher Vector encoding method," *Neurocomputing*, vol. 218, pp. 251–258, Dec. 2016.

[35] T. L. T. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, "Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain," *Med. Biol. Eng. Comput.*, vol. 55, no. 2, pp. 343–352, 2017.

[36] K. Aboalayon, M. Faezipour, W. Almuhammadi, and S. Moslehpour, "Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation," *Entropy*, vol. 18, no. 9, p. 272, 2016.

[37] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," *Neurocomputing*, vol. 104, no. c, pp. 105–114, Mar. 2013.

[38] K. a. I. Aboalayon, H. T. Ocbagabir, and M. Faezipour, "Efficient sleep stage classification based on EEG signals," in *IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014*, 2014, pp. 1–6.

[39] T. Kayikcioglu, M. Maleki, and K. Eroglu, "Fast and accurate PLS-based classification of EEG sleep using single channel data," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7825–7830, Nov. 2015.

[40] V. Bajaj and R. B. Pachori, "Automatic classification of sleep stages based on the time-frequency image of EEG signals," *Comput. Methods Programs Biomed.*, vol. 112, no. 3, pp. 320–328, Dec. 2013.

[41] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients," *2008 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2008, pp. 1151–1154, Aug. 2008.

[42] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and

bootstrap aggregating," *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, Feb. 2016.

[43] R. Acharya U., O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan, "Non-linear analysis of EEG signals at various sleep stages," *Comput. Methods Programs Biomed.*, vol. 80, no. 1, pp. 37–45, Oct. 2005.

[44] J. Fell, J. Röschke, K. Mann, and C. Schäffner, "Discrimination of sleep stages: A comparison between spectral and nonlinear EEG measures," *Electroencephalography and Clinical Neurophysiology*, vol. 98, no. 5. pp. 401–410, 1996.

[45] R. Kaplan, Y. Wang, K. Loparo, M. Kelly, and R. Bootzin, "Performance evaluation of an automated single-channel sleep–wake detection algorithm," *Nat. Sci. Sleep*, vol. 6, p. 113, Oct. 2014.

[46] A. Pasieczna and J. Korczak, "Classification Algorithms in Sleep Detection—A Comparative Study," 2016, vol. 8, pp. 113–120.

[47] M. Elmessidi, S. T. Ben Hamida, B. Ahmed, and T. Penzel, "Accurate automatic identification of slow wave sleep using a single electro-oculogram channel," *Middle East Conf. Biomed. Eng. MECBME*, pp. 232–235, 2014.

[48] J. Virkkala, J. Toppila, P. Maasilta, and A. Bachour, "Electro-oculography-based detection of sleep-wake in sleep apnea patients," *Sleep Breath.*, vol. 19, no. 3, pp. 785–789, Sep. 2015.

[49] S. M. Isa, I. Wasito, A. M. Arymurthy, and A. Noviyanto, "Kernel Dimensionality Reduction on Sleep Stage Classification using ECG Signal," *Int. J. Comput. Sci. Issues*, vol. 8, no. 1, pp. 1178–1181, 2011.

[50] A. Noviyanto and A. M. Arymurthy, "Sleep stages classification based on temporal pattern recognition in neural network approach," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–6.

[51] M. Adnane, Z. Jiang, and Z. Yan, "Sleep–wake stages classification and sleep

efficiency estimation using single-lead electrocardiogram," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 1401–1413, 2012.

[52] Q. K. Le, Q. D. K. Truong, and V. T. Vo, "A tool for analysis and classification of sleep stages," *2011 Int. Conf. Adv. Technol. Commun. (ATC 2011)*, no. Atc, pp. 307–310, Aug. 2011.

[53] S. Khalighi, T. Sousa, and U. Nunes, "Adaptive Automatic Sleep Stage Classification under Covariate Shift," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 2259–2262.

[54] S. Khalighi, T. Sousa, G. Pires, and U. Nunes, "Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels," *Expert Syst. Appl.*, vol. 40, no. 17, pp. 7046–7059, 2013.

[55] L. Zoubek, S. Charbonnier, S. Lesecq, A. Buguet, and F. Chapotot, "Feature Selection for Sleep/Wake Stages Classification Using Data Driven Methods," *Biomed. Signal Process. Control*, vol. 2, no. 3, pp. 171–179, Jul. 2007.

[56] T. Lajnef, S. Chaibi, P. Ruby, P.-E. Aguera, J.-B. Eichenlaub, M. Samet, A. Kachouri, and K. Jerbi, "Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines," *J. Neurosci. Methods*, vol. 250, pp. 94–105, 2015.

[57] T. Sousa, A. Cruz, S. Khalighi, G. Pires, and U. Nunes, "A two-step automatic sleep stage classification method with dubious range detection," *Comput. Biol. Med.*, vol. 59, pp. 42–53, Apr. 2015.

[58] S. Khalighi, T. Sousa, D. Oliveira, G. Pires, and U. Nunes, "Efficient feature selection for sleep staging based on maximal overlap discrete wavelet transform and SVM," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, no. July, pp. 3306–3309.

[59] F. Ebrahimi, M. Mikaili, E. Estrada, and H. Nazeran, "Assessment of Itakura Distance as a Valuable Feature for Computer-aided Classification of Sleep

Stages," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, vol. 2007, pp. 3300–3303.

[60] L. Chen, Y. Zhao, J. Zhang, and J. Zou, "Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7344–7355, Nov. 2015.

[61] E. Estrada, P. Nava, H. Nazeran, K. Behbehani, J. Burk, and E. Lucas, "Itakura Distance: A Useful Similarity Measure between EEG and EOG Signals in Computer-aided Classification of Sleep Stages.," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2, pp. 1189–1192, 2005.

[62] T. Willemen, D. Van Deun, V. Verhaert, M. Vandekerckhove, V. Exadaktylos, J. Verbraecken, S. Van Huffel, B. Haex, and J. Vander Sloten, "An Evaluation of Cardiorespiratory and Movement Features With Respect to Sleep-Stage Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 2, pp. 661–669, Mar. 2014.

[63] V. C. Figueroa Helland, a. Gapelyuk, a. Suhrbier, M. Riedl, T. Penzel, J. Kurths, and N. Wessel, "Investigation of an Automatic Sleep Stage Classification by Means of Multiscorer Hypnogram," *Methods Inf. Med.*, vol. 49, no. 5, pp. 467–472, 2010.

[64] P. Fonseca, X. Long, M. Radha, R. Haakma, R. M. Aarts, and J. Rolink, "Sleep stage classification with ECG and respiratory effort," *Physiol. Meas.*, vol. 36, no. 10, pp. 2027–2040, Oct. 2015.

[65] F. Chapotot and G. Becq, "Automated Sleep-Wake Staging Combining Robust Feature Extraction, Artificial Neural Network Classification, and Flexible Decision Rules," *Int. J. Adapt. Control Signal Process.*, vol. 24, no. 5, pp. 409–423, 2009.

[66] G. Becq, S. Charbonnier, F. Chapotot, A. Buguet, L. Bourdon, and P. Baconnier, "Comparison Between Five Classifiers for Automatic Scoring of Human Sleep

Recordings," in *Classification and Clustering for Knowledge Discovery*, vol. 127, 2005, pp. 113–127.

[67] S. F. Liang, C. E. Kuo, F. Z. Shaw, Y. H. Chen, C. H. Hsu, and J. Y. Chen, "Combination of expert knowledge and a genetic fuzzy inference system for automatic sleep staging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 10, pp. 2108–2118, 2016.

[68] J. Foussier, P. Fonseca, X. Long, and S. Leonhardt, "Automatic Feature Selection for Sleep/Wake Classification with Small Data Sets," *6th Int. Conf. Bioinforma. Model. Methods Algorithms*, pp. 1–7, 2013.

[69] M. Čić, J. Šoda, and M. Bonković, "Automatic classification of infant sleep based on instantaneous frequencies in a single-channel EEG signal," *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2110–2117, Dec. 2013.

[70] I. Zhovna and I. D. Shallom, "Automatic detection and classification of sleep stages by multichannel EEG signal modeling," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, vol. 2008, no. c, pp. 2665–2668.

[71] A. Krakovská and K. Mezeiová, "Automatic sleep scoring: A search for an optimal combination of measures," *Artif. Intell. Med.*, vol. 53, no. 1, pp. 25–33, Sep. 2011.

[72] J. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán, and D. Peluffo, "Automatic Sleep Stages Classification Using EEG Entropy Features and Unsupervised Pattern Analysis Techniques," *Entropy*, vol. 16, no. 12, pp. 6573–6589, Dec. 2014.

[73] L. G. Doroshenkov, V. A. Konyshev, and S. V. Selishchev, "Classification of human sleep stages based on EEG processing using hidden Markov models," *Biomed. Eng. (NY).*, vol. 41, no. 1, pp. 25–28, Jan. 2007.

[74] E. Estrada, H. Nazeran, F. Ebrahimi, and M. Mikaeili, "EEG signal features for computer-aided sleep stage detection," in *2009 4th International IEEE/EMBS*

Conference on Neural Engineering, 2009, pp. 669–672.

[75] J. Shi, X. Liu, Y. Li, Q. Zhang, Y. Li, and S. Ying, "Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning," *J. Neurosci. Methods*, vol. 254, pp. 94–101, Oct. 2015.

[76] H. Simões, G. Pires, U. Nunes, and V. Silva, "Feature Extraction and Selection for Automatic Sleep Staging Using EEG," in *ICINCO*, 2010, pp. 128–133.

[77] Z. Liu, J. Sun, Y. Zhang, and P. Rolfe, "Sleep staging from the EEG signal using multi-domain feature extraction," *Biomed. Signal Process. Control*, vol. 30, pp. 86–97, 2016.

[78] J. Virkkala, J. Hasan, A. Värri, S.-L. Himanen, and K. Müller, "Automatic sleep stage classification using two-channel electro-oculography," *J. Neurosci. Methods*, vol. 166, no. 1, pp. 109–115, Oct. 2007.

[79] S. Liang, C. Kuo, Y. Lee, W. Lin, Y. Liu, P. Chen, F. Cherng, and F. Shaw, "Development of an EOG-Based Automatic Sleep-Monitoring Eye Mask," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 11, pp. 2977–2985, Nov. 2015.

[80] K. Šušmáková and A. Krakovská, "Discrimination ability of individual measures used in sleep stages classification," *Artif. Intell. Med.*, vol. 44, no. 3, pp. 261–277, Nov. 2008.

[81] M. Ronzhina, O. Janoušek, J. Kolářová, M. Nováková, P. Honzík, and I. Provazník, "Sleep scoring using artificial neural networks," *Sleep Med. Rev.*, vol. 16, no. 3, pp. 251–263, Jun. 2012.

[82] J. P. Lanquart, M. Dumont, and P. Linkowski, "QRS artifact elimination on full night sleep EEG," *Med. Eng. Phys.*, vol. 28, no. 2, pp. 156–165, 2006.

[83] M. B. Hamaneh, N. Chitravas, K. Kaiboriboon, S. D. Lhatoo, and K. A. Loparo, "Automated Removal of EKG Artifact From EEG Data Using Independent Component Analysis and Continuous Wavelet Transformation," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1634–1641, Jun. 2014.

[84] W. Zhou and J. Gotman, "Removal of EMG and ECG artifacts from EEG based on wavelet transform and ICA.," *26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 1, no. 5, pp. 392–395, 2004.

[85] S. Khalighi, T. Sousa, J. M. Santos, and U. Nunes, "ISRUC-Sleep: A comprehensive public dataset for sleep researchers," *Comput. Methods Programs Biomed.*, vol. 124, pp. 180–192, Feb. 2016.

[86] J. Zhou, X. Wu, and W. Zeng, "Automatic detection of sleep apnea based on EEG detrended fluctuation analysis and support vector machine," *J. Clin. Monit. Comput.*, vol. 29, no. 6, pp. 767–772, Dec. 2015.

[87] S.-H. Oh, Y.-R. Lee, and H.-N. Kim, "A Novel EEG Feature Extraction Method Using Hjorth Parameter," *Int. J. Electron. Electr. Eng.*, vol. 2, no. 2, pp. 106–110, 2014.

[88] P. Piñero, P. Garcia, L. Arco, A. Álvarez, M. M. García, and R. Bonal, "Sleep stage classification using fuzzy sets and machine learning techniques," *Neurocomputing*, vol. 58–60, pp. 1137–1143, Jun. 2004.

[89] S. Özşen, "Classification of sleep stages using class-dependent sequential feature selection and artificial neural network," *Neural Comput. Appl.*, vol. 23, no. 5, pp. 1239–1250, Oct. 2013.

[90] J. Zhang, Y. Wu, J. Bai, and F. Chen, "Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers," *Trans. Inst. Meas. Control*, vol. 38, no. 4, pp. 435–451, Apr. 2016.

[91] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal-state-of-the-art and guidelines," *J. Neural Eng.*, vol. 12, no. 3, p. 031001, Jun. 2015.

[92] S. Devuyst, T. Dutoit, T. Ravet, P. Stenuit, M. Kerkhofs, and E. Stanus, "Automatic Processing of EEG-EOG-EMG Artifacts in Sleep Stage Classification," in *IFMBE Proceedings*, 2009, vol. 23, pp. 146–150.

[93] N. Mammone and F. Morabito, "Enhanced Automatic Wavelet Independent

Component Analysis for Electroencephalographic Artifact Removal," *Entropy*, vol. 16, no. 12, pp. 6553–6572, Dec. 2014.

[94] H. Nolan, R. Whelan, and R. B. Reilly, "FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection," *J. Neurosci. Methods*, vol. 192, no. 1, pp. 152–162, Sep. 2010.

[95] M. K. Islam, A. Rastegarnia, and Z. Yang, "Methods for artifact detection and removal from scalp EEG: A review," *Neurophysiol. Clin. Neurophysiol.*, vol. 46, no. 4–5, pp. 287–305, Nov. 2016.

[96] R. J. Croft, J. S. Chandler, R. J. Barry, N. R. Cooper, and A. R. Clarke, "EOG correction: A comparison of four methods," *Psychophysiology*, vol. 42, no. 1, pp. 16–24, Jan. 2005.

[97] K. T. Sweeney, T. E. Ward, and S. F. McLoone, "Artifact removal in physiological signals--practices and possibilities.," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 488–500, May 2012.

[98] S. Romero, M. A. Mañanas, and M. J. Barbanoj, "Ocular Reduction in EEG Signals Based on Adaptive Filtering, Regression and Blind Source Separation," *Ann. Biomed. Eng.*, vol. 37, no. 1, pp. 176–191, Nov. 2008.

[99] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, no. 4, pp. 626–638, Apr. 1996.

[100] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, and I. Merlet, "Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 127, 2012.

[101] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, "EMG and EOG artifacts in brain computer interface systems: A survey.," *Clin. Neurophysiol.*, vol. 118, no. 3, pp. 480–94, Mar. 2007.

[102] T. Wiener and T. Newton, *Fundamentals of Adaptive Filtering 2.1*. J. Wiley & Sons, 2013.

[103] J. Dhiman, S. Ahmad, and K. Gulia, "Comparison between Adaptive filter Algorithms (LMS, NLMS and RLS)," *Int. J. Sci. Eng. Technol. Res.*, vol. 2, no. 5, pp. 2278–7798, 2013.

[104] A. GuruvaReddy and S. Narava, "Artifact Removal from EEG Signals," *Int. J. Comput. Appl.*, vol. 77, no. 13, pp. 17–19, Sep. 2013.

[105] P. S. Kumar, R. Arumuganathan, K. Sivakumar, and C. Vimal, "Removal of artifacts from EEG signals using adaptive filter through wavelet transform," in *2008 9th International Conference on Signal Processing*, 2008, pp. 2138–2141.

[106] M. K. Ahirwal, A. Kumar, G. K. Singh, and N. D. Londhe, "Performance prediction of adaptive filters for EEG signal," *IET Sci. Meas. Technol.*, vol. 11, no. 5, pp. 525–531, Aug. 2017.

[107] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Wiley, 2013.

[108] M. K. Ahirwal, A. Kumar, and G. K. Singh, "EEG/ERP Adaptive Noise Canceller Design with Controlled Search Space (CSS) Approach in Cuckoo and Other Optimization Algorithms," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 10, no. 6, pp. 1491–1504, Nov. 2013.

[109] E. Estrada, H. Nazeran, G. Sierra, F. Ebrahimi, and S. K. Setarehdan, "Wavelet-based EEG denoising for automatic sleep stage classification," in *CONIELECOMP 2011, 21st International Conference on Electrical Communications and Computers*, 2011, pp. 295–298.

[110] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Jan. 2002.

[111] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42,

no. 8, pp. 1778–1790, Aug. 2004.

[112] Y. Liu, K. Wen, Q. Gao, X. Gao, and F. Nie, "SVM based multi-label learning with missing labels for image annotation," *Pattern Recognit.*, vol. 78, pp. 307–317, Jun. 2018.

[113] M. Dadgostar, S. K. Setarehdan, S. Shahzadi, and A. Akin, "Classification of Schizophernia Using SVM via fNIRS," *Biomed. Eng. Appl. Basis Commun.*, vol. 30, no. 02, p. 1850008, Apr. 2018.

[114] K. Benabdeslem and Y. Bennani, "Dendogram based SVM for multi-class classification," in *28th International Conference on Information Technology Interfaces, 2006.*, 2006, pp. 173–178.

[115] F. Takahashi and S. Abe, "Decision-tree-based multiclass support vector machines," in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02.*, 2002, vol. 3, pp. 1418–1422.

[116] G. Madzarov, D. Gjorgjevikj, and I. Chorbev, "A Multi-class SVM Classifier Utilizing Binary Decision Tree," *Informatica*, vol. 33, no. 2. 2009.

[117] M. Bala and R. K. Agrawal, "Optimal Decision Tree Based Multi-class Support Vector Machine," *Informatica*, vol. 35, no. 2, 2011.

[118] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.

[119] S. Tufféry, *Data Mining and Statistics for Decision Making*. Chichester, UK: John Wiley & Sons, Ltd, 2011.

[120] J. M. S. Pascualvaca, C. Fernandes, A. Guillén, A. M. Mora, R. Largo, A. C. Rosa, and L. J. Herrera, "Sleep Stage Classification Using Advanced Intelligent Methods," 2013, pp. 604–612.

[121] H. Banaee, M. Ahmed, and A. Loutfi, "Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges,"

Sensors, vol. 13, no. 12, pp. 17472–17500, Dec. 2013.

[122] M. M. Baig and H. Gholamhosseini, "Smart Health Monitoring Systems: An Overview of Design and Modeling," *J. Med. Syst.*, vol. 37, no. 2, p. 9898, Apr. 2013.

[123] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[124] S. J. Redmond, P. de Chazal, C. O'Brien, S. Ryan, W. T. McNicholas, and C. Heneghan, "Sleep staging using cardiorespiratory signals," *Somnologie - Schlafforsch. und Schlafmedizin*, vol. 11, no. 4, pp. 245–256, Dec. 2007.

[125] G. J. McLachlan, *Discriminant analysis and statistical pattern recognition.* Wiley-Interscience, 2004.

[126] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, and Y.-S. Cheng, "A rule-based automatic sleep staging method," *J. Neurosci. Methods*, vol. 205, no. 1, pp. 169–176, 2012.

[127] V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang, Eds., *Handbook of Partial Least Squares.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[128] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[129] G.-B. Huang, "An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels," *Cognit. Comput.*, vol. 6, no. 3, pp. 376–390, Sep. 2014.

[130] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings*, vol. 1857, 2000, pp. 1–15.

[131] C. Avcı and A. Akbaş, "Sleep apnea classification based on respiration signals by using ensemble methods," *Biomed. Mater. Eng.*, vol. 26, no. s1, pp. S1703–S1710, Aug. 2015.

[132] Z. Yi and K. K. Tan, *Convergence Analysis of Recurrent Neural Networks*, vol. 13.

Boston, MA: Springer US, 2004.

[133] D. E. Rumelhart, J. L. McClelland, and S. D. P. R. G. University of California, *Parallel Distributed Processing*. MIT Press, 1986.

[134] T. Kohonen, *Self-Organization and Associative Memory*, vol. 8. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989.

[135] J. S. Albus, "A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC)," *J. Dyn. Syst. Meas. Control*, vol. 97, no. 3, p. 220, 1975.

[136] R. Hecht-Nielsen, *Neurocomputing*. Addison-Wesley Pub. Co, 1990.

[137] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 79, no. 8, pp. 2554–8, Apr. 1982.

[138] J. L. Elman, "Finding Structure in Time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.

[139] M. I. Jordan, "Attractor dynamics and parallelism in a connectionist sequential machine," in *Eighth Annual Conference of the Cognitive Science Society*, 1986, pp. 531–546.

[140] D. Brezak, T. Bacek, D. Majetic, J. Kasac, and B. Novakovic, "A comparison of feed-forward and recurrent neural networks in time series forecasting," in *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, 2012, pp. 1–6.

[141] M. Salmasi, H. Mahdavi-Nasab, and H. Pourghassem, "Comparison of Feed-Forward and Recurrent Neural Networks in Active Cancellation of Sound Noise," in *2011 International Conference on Multimedia and Signal Processing*, 2011, vol. 2, pp. 25–29.

[142] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi, and S. Jain, "Machine

translation using deep learning: An overview," in *2017 International Conference on Computer, Communications and Electronics (Comptelix)*, 2017, pp. 162–167.

[143] R. K. Sinha, "Artificial Neural Network and Wavelet Based Automated Detection of Sleep Spindles, REM Sleep and Wake States," *J. Med. Syst.*, vol. 32, no. 4, pp. 291–299, Aug. 2008.

[144] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, and H. Dickhaus, "Classification of sleep stages using multi-wavelet time frequency entropy and LDA," *Methods Inf. Med.*, vol. 49, pp. 230–237, 2010.

[145] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic Sleep Stage Classification Based on EEG Signals by Using Neural Networks and Wavelet Packet Coefficients," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2008, pp. 1151–1154, Jan. 2008.

[146] B. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms," *J. Med. Syst.*, vol. 38, no. 3, p. 18, 2014.

[147] Y. Kurihara and K. Watanabe, "Sleep-Stage Decision Algorithm by Using Heartbeat and Body-Movement Signals," *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans*, vol. 42, no. 6, pp. 1450–1459, Nov. 2012.

[148] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1412–1423, 2001.

[149] E. A. Vivaldi and A. Bassi, "Frequency Domain Analysis of Sleep EEG for Visualization and Automated State Detection," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 3740–3743.

[150] P. Van Hese, W. Philips, J. De Koninck, R. Van de Walle, and I. Lemahieu, "Automatic detection of sleep stages using the EEG," in *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2001, vol. 2, pp. 1944–1947.

[151] O. R. Pacheco and F. Vaz, "Integrated system for analysis and automatic classification of sleep EEG," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol.20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No.98CH36286)*, 1998, vol. 4, pp. 2062–2065.

[152] M. H. Schwaibold, T. Penzel, J. Schochlin, and A. Bolz, "Combination of AI components for biosignal processing application to sleep stage recognition," *Eng. Med. Biol. Soc. 2001. Proc. 23rd Annu. Int. Conf. IEEE*, vol. 2, pp. 1692–1694 vol.2, 2001.

[153] S. Gudmundsson, T. P. Runarsson, and S. Sigurdsson, "Automatic Sleep Staging using Support Vector Machines with Posterior Probability Estimates," in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, 2005, vol. 2, pp. 366–372.

[154] N. McGrogan, E. Braithwaite, and L. Tarassenko, "BioSleep: a comprehensive sleep analysis system," in *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2001, vol. 2, pp. 1608–1611.

[155] D. Álvarez-Estévez, J. M. Fernández-Pastoriza, E. Hernández-Pereira, and V. Moret-Bonillo, "A method for the automatic analysis of the sleep macrostructure in continuum," *Expert Syst. Appl.*, vol. 40, no. 5, pp. 1796–1803, 2013.

[156] A. Flexerand, G. Dorffner, P. Sykacekand, and I. Rezek, "An automatic, continuous and probabilistic sleep stager based on a hidden markov model," *Appl. Artif. Intell.*, vol. 16, no. 3, pp. 199–207, Mar. 2002.

[157] M. Hanaoka, M. Kobayashi, and H. Yamazaki, "Automatic sleep stage scoring based on waveform recognition method and decision-tree learning," *Syst. Comput. Japan*, vol. 33, no. 11, pp. 1–13, Oct. 2002.

[158] B. Wang, X. Wang, J. Zou, F. Kawana, and M. Nakamura, "Automatic determination of sleep stage through bio-neurological signals contaminated with artifacts by a conditional probability of the knowledge base," *Artif. Life Robot.*, vol. 12, no. 1–2, pp. 270–275, Mar. 2008.

[159] H.-J. Park, J.-S. Oh, D.-U. Jeong, and K.-S. Park, "Automated Sleep Stage Scoring Using Hybrid Rule- and Case-Based Reasoning," *Comput. Biomed. Res.*, vol. 33, no. 5, pp. 330–349, Oct. 2000.

[160] P. Anderer, G. Gruber, S. Parapatics, M. Woertz, T. Miazhynskaia, G. Klösch, B. Saletu, J. Zeitlhofer, M. J. Barbanoj, H. Danker-Hopfe, S.-L. Himanen, B. Kemp, T. Penzel, M. Grözinger, D. Kunz, P. Rappelsberger, A. Schlögl, and G. Dorffner, "An E-Health Solution for Automatic Sleep Classification according to Rechtschaffen and Kales: Validation Study of the Somnolyzer 24 × 7 Utilizing the Siesta Database," *Neuropsychobiology*, vol. 51, no. 3, pp. 115–133, May 2005.

[161] S. R. I. Gabran, S. Zhang, M. M. A. Salama, R. R. Mansour, and C. George, "Real-time automated neural-network sleep classifier using single channel EEG recording for detection of narcolepsy episodes," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, vol. 2008, pp. 1136–1139.

[162] D. Y. Li Deng, "Deep Learning: Methods and Applications," NOW Publishers, May 2014.

[163] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy Layer-Wise Training of Deep Networks," in *Advances in Neural Information Processing Systems 19 (NIPS'06)*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. MIT Press, 2007, pp. 153--160.

[164] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[165] H. Lee, C. Ekanadham, and A. Y. Ng, "Sparse deep belief net model for visual area V2," in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D.

Koller, Y. Singer, and S. T. Roweis, Eds. Curran Associates, Inc., 2008, pp. 873–880.

[166] D. Wulsin, J. Blanco, R. Mani, and B. Litt, "Semi-Supervised Anomaly Detection for EEG Waveforms Using Deep Belief Nets," in *2010 Ninth International Conference on Machine Learning and Applications*, 2010, pp. 436–441.

[167] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement.," *J. Neural Eng.*, vol. 8, no. 3, p. 036015, Jun. 2011.

[168] D. Wang and Y. Shang, "Modeling Physiological Data with Deep Belief Networks.," *Int. J. Inf. Educ. Technol.*, vol. 3, no. 5, pp. 505–511, 2013.

[169] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep Stage Classification Using Unsupervised Feature Learning," *Adv. Artif. Neural Syst.*, vol. 2012, pp. 1–9, 2012.

[170] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," Jun. 2012.

[171] E. P. Giri, M. I. Fanany, and A. M. Arymurthy, "Combining Generative and Discriminative Neural Networks for Sleep Stages Classification," Oct. 2016.

[172] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, "Mixed Neural Network Approach for Temporal Sleep Stage Classification," Oct. 2016.

[173] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic Sleep Stage Scoring Using Time-Frequency Analysis and Stacked Sparse Autoencoders," *Ann. Biomed. Eng.*, vol. 44, no. 5, pp. 1587–1597, May 2016.

[174] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 758–769, Apr. 2018.

[175] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks," Oct. 2016.

[176] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[177] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, 2017.

[178] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[179] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep Convolutional Neural Networks for Interpretable Analysis of EEG Sleep Stage Scoring," in *IEEE International Workshop on Machine Learning for Signal Processing, Sept. 25-28*, 2017.

[180] PhysioNet, "The Sleep-EDF Database [Expanded]," 2015. [Online]. Available: https://physionet.org/physiobank/database/sleep-edfx/. [Accessed: 01-Feb-2017].

[181] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[182] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.

[183] Y. Ren and Y. Wu, "Convolutional deep belief networks for feature extraction of EEG signal," in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 2850–2853.

[184] A. A. Gharbali, J. M. Fonseca, S. Najdi, and T. Y. Rezaii, "Automatic EOG and EMG Artifact Removal Method for Sleep Stage Classification," in *7th IFIP*

Advanced Doctoral Conference on Technological Innovation for Cyber-Physical Systems, 2016, pp. 142–150.

[185] S. S. Haykin, *Adaptive Filter Theory.* Prentice Hall, 2002.

[186] X. Kong, N. Thakor, and V. Goel, "Characterization of EEG signal changes via Itakura distance," in *Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society*, 1995, pp. 873–874.

[187] M. M. Deza and E. Deza, *Encyclopedia of distances.* 2009.

[188] M. Brookes, "VOICEBOX: Speech Processing Toolbox for MATLAB." 2005.

[189] S. Najdi, A. A. Gharbali, and J. M. Fonseca, "A Comparison of Feature Ranking and Rank Aggregation Techniques in Automatic Sleep Stage Classification Based on Polysomnographic Signals," in *4th International Conference, IWBBIO*, 2016, pp. 230–241.

[190] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selection algorithms: a study on high-dimensional spaces," *Knowl. Inf. Syst.*, vol. 12, no. 1, pp. 95–116, May 2007.

[191] S. Najdi, A. A. Gharbali, and J. M. Fonseca, "Feature ranking and rank aggregation for automatic sleep stage classification: a comparative study," *Biomed. Eng. Online*, vol. 16, no. S1, p. 78, Aug. 2017.

[192] K. Kira and L. A. Rendell, "The Feature Selection Problem: Traditional Methods and a New Algorithm," in *Proceedings of the 10th National Conference on Artificial Intelligence. San Jose, CA, July 12-16, 1992.*, 1992, pp. 129–134.

[193] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1–2, pp. 23–69, 2003.

[194] C. Ding and H. Peng, "Minimum Redundancy Feature Selection from Microarray Gene Expression Data," in *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, 2005,

vol. 3, no. 2, pp. 523–528.

[195] Q. Gu, Z. Li, and J. Han, "Generalized Fisher Score for Feature Selection," in *Proceedings of the International Conference on Uncertainty in Artificial Intelligence.*, 2011, vol. abs/1202.3.

[196] Huan Liu and R. Setiono, "Chi2: Feature Selection and Discretization of Numeric Attributes," in *Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence*, 1995, pp. 388–391.

[197] J. R. Quinlan, "C4.5: Programs for Machine Learning," Mar. 1993.

[198] P. A. Jaskowiak, P. A. Jaskowiak, and R. J. G. B. Campello, "Comparing Correlation Coefficients as Dissimilarity Measures for Cancer Classification in Gene Expression Data," *BRAZILIAN Symp. Bioinforma.*, pp. 1--8, 2011.

[199] M. Hudson, B. Martin, T. Hagan, and H. B. Demuth R2018a, "Neural Network Toolbox™ User's Guide."

[200] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural Network Design 2nd Edition.* .

[201] A. Ng, J. Ngiam, C. Foo, Y. Mai, and C. Suen, "UFLDL Tutorial," *http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial*, 2010.

[202] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.

[203] P. Prakash and A. S. K. Rao, *R deep learning cookbook : solve complex neural net problems with TensorFlow, H2O and MXNet.* Packt Publishing, 2017.

[204] C. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[205] S. S. Haykin and S. S. Haykin, *Neural networks and learning machines.* Prentice Hall/Pearson, 2009.

[206] Y. LeCun and Y. Bengio, "Convolutional Networks for Images, Speech, and Time-Series," *Handb. Brain Theory Neural Networks*, vol. 2d ed, 2003.

[207] A. Krizhevsky, I. Sutskever, H. Geoffrey E., and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst. 25*, pp. 1–9, 2012.

[208] A. B. Wiltschko, G. J. Gage, and J. D. Berke, "Wavelet Filtering Before Spike Detection Preserves Waveform Shape and Enhances Single-Unit Discrimination.," *J. Neurosci. Methods*, vol. 173, no. 1, pp. 34–40, Aug. 2008.

[209] M. Dursun, S. Gunes, S. Ozsen, and S. Yosunkaya, "Comparison of Artificial Immune Clustering with Fuzzy C-means Clustering in the sleep stage classification problem," in *2012 International Symposium on Innovations in Intelligent Systems and Applications*, 2012, pp. 1–4.

[210] S. Lin, "Rank aggregation methods," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 5, pp. 555–570, Sep. 2010.

[211] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification.* Wiley, 2012.

[212] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with Artificial Neural Networks:: The State of the Art," *Int. J. Forecast.*, vol. 14, no. 1, pp. 35–62, Mar. 1998.

[213] S. A. Imtiaz and E. Rodriguez-Villegas, "Recommendations for Performance Assessment of Automatic Sleep Staging Algorithms," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, vol. 2014, pp. 5044–5047.

[214] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a 'Kneedle' in a Haystack: Detecting Knee Points in System Behavior," in *2011 31st International Conference on Distributed Computing Systems Workshops*, 2011, pp. 166–171.

[215] L. Duckstein and S. Opricovic, "Multiobjective optimization in river basin development," *Water Resour. Res.*, vol. 16, no. 1, pp. 14–20, 1980.

[216] S. Opricovic and G. H. Tzeng, "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS," *Eur. J. Oper. Res.*, vol. 156, no. 2,

pp. 445–455, 2004.

[217] G. Sharma and J. Martin, "MATLAB®: A Language for Parallel Computing," *Int. J. Parallel Program.*, vol. 37, no. 1, pp. 3–36, Feb. 2009.

[218] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY: Springer New York, 2013.

[219] S. Mallat, *A Wavelet Tour of Signal Processing*. Elsevier, 2009.

[220] R. Raina, A. Y. Ng, and D. Koller, "Constructing informative priors using transfer learning," *Proc. 23rd Int. Conf. Mach. Learn.*, pp. 713–720, 2006.

[221] "ImageNet." [Online]. Available: http://www.image-net.org/. [Accessed: 23-Jan-2018].

[222] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989.

Annex List of Publications Related to the Proposed Work

Publications in International Journals

1	S. Najdi, A. A. Gharbali, and J. M. Fonseca, "Feature ranking and rank aggregation for automatic sleep stage classification: a comparative study," *Biomedical Engineering Online*, vol. 16, no. S1, p. 78, Aug. 2017.
2	A. A. Gharbali, S. Najdi, and J. M. Fonseca, "Investigating the contribution of distance-based features to automatic sleep stage classification," *Computers in Biology and Medicine*, vol. 96, pp. 8–23, May 2018.

Publications in International Conferences Proceedings

1	S. Najdi, A. A. Gharbali, and J. M. Fonseca, "A Comparison of Feature Ranking and Rank Aggregation Techniques in Automatic Sleep Stage Classification Based on Polysomnographic Signals," in *4th International Conference on Bioinformatics and Biomedical Engineering, IWBBIO*, 2016, pp. 230–241.
2	A. A. Gharbali, J. M. Fonseca, S. Najdi, and T. Y. Rezaii, "Automatic EOG and EMG Artifact Removal Method for Sleep Stage Classification," in *7th IFIP Advanced Doctoral Conference on Technological Innovation for Cyber-Physical Systems*, 2016, pp. 142–150.
3	S. Najdi, A. A. Gharbali, and J. M. Fonseca, "Feature Transformation Based on Stacked Sparse Autoencoders for Sleep Stage Classification," in *Technological Innovation for Smart Systems*, 2017, pp. 191–200.
4	A. A. Gharbali, S. Najdi, and J. M. Fonseca, "Transfer Learning of Spectrogram Image for Automatic Sleep Stage Classification," *Springer International Publishing*, 2018, pp. 522–528.