Norwegian University of
Science and Technology

Monitoring patient's condition based on breath detection

Kamil Stanislaw Adamczyk

Master of Science in Computer Science
Submission date: June 2016
Supervisor: Asbjørn Thomassen, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

Abstract

A constant increase in elderly population in developed countries closely followed by a continuous reduction of costs, inter alia in healthcare, calls for inventing efficient methods in eldercare. For that purpose, this study is devoted to an introductory elaboration of a computer-based monitoring system that extracts patient information based on respiration analysis. By combining the field of respiratory medicine with machine learning, an empirical study has been conducted in accordance to prior methodical review of the state-of-the-art research. Consequently, Sleep Apnea-Hypopnea Syndrome is explored in the context of automated event classification by applying Artificial Neural Network and Support Vector Machine classifiers. Experiments have not resulted in revolutionary findings, however this thesis contributes with several valid suggestions and may be used as an introduction to the field and a foundation for further research.

Sammendrag

Som en følge av en konstant økning i andelen av eldre i utviklede land, samt en kontinuerlig kostnadsøkning i helse- og sosialtjenester, oppfordres forskningsmiljøet til å utvikle stadig mer effektive metoder innen eldreomsorg. Formålet med denne avhandlingen er derfor å utforme et automatisert overvåkningssystem som baserer seg på analyse av pasientenes åndedrett for å overvåke deres tilstand. Avhandlingen er en fusjon av lungemedisin og maskinlæring, og har resultert i gjennomføringen av en empirisk studie som følge av en omfattende litteraturgjennomgang innen dette forskningsområdet. Rapporten utforsket i den forbindelse Søvnapne-Hypopne Syndrom ved hjelp av klassifisering av søvnrelaterte hendelser. Klassifiseringen ble utført ved hjelp av et nevralt nettverk (ANN) og en støttevektormaskin (SVM). Eksperimentene indikerte ikke signifikante forbedringer, men denne avhandlingen bidrar med nyttige innspill angående databasert pasientovervåkning. I følge resultatene utviste begge klassifiseringsmetodene tilsvarende ytelse. Likevel utgjør avhandlingen en god introduksjon til forskingsfeltet og et grunnlag for videre forskning.

Preface

This Master of Science thesis was conducted at Norwegian University of Science and Technology (NTNU) in Trondheim.

I would like to thank my supervisor assistant professor Asbjørn Thomassen (Department of Computer and Information Science, IDI) for the guidance.

Kamil S. Adamczyk

Trondheim, June 10, 2016

Table of Contents

List of figures

List of tables

Abbreviations

AASM	American Academy of Sleep Medicine
ANN	Artificial Neural Network
AUC	Area Under the Curve
DWT	Discrete Wavelet Transform
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
EOG	Electrooculography
FFT	Fast Fourier Transform
FPR	False Positive Rate
PSG	Polysomnogram
RBF	Radial Basis Function
ROC	Receiver operating characteristic
SAHS	Sleep Apnea-Hypopnea Syndrome
SVM	Support Vector Machine
TPR	True Positive Rate

1 Introduction

The following is an introductory chapter presenting the essence of this research. A brief description of the background and motivational factors is given in the first section, which furthermore will be extended in the next chapter. Subsequently, the goal and research questions are defined, followed by an explanation of research method and structure of this paper.

1.1 Background and Motivation

Computer-based systems are becoming a reality in nearly all areas replacing processes hitherto executed by humans. Some of them – such as medicine – requires a major effort put into research due to tremendous emphasis on reliability. Furthermore, social factors such as the increase in the elderly population closely and a continuous cost reduction in healthcare motivates to devote this study to partly elaborate an automated monitoring system. Moreover, a closer consideration of sleep disorders has revealed the importance of Sleep Apnea-Hypopnea Syndrome, which has a prevalence between 3 and 7 per cent among the adult population. Present practice of diagnosis is cumbersome and leaves a lot to be desired resulting in patients' unawareness of the disease's presence. As a consequence, an ambient, movable and cheap monitoring system is desired, hence some steps towards it will be taken by this study. Among a numerous of available approaches, it has been decided that an interesting aspect related to extraction of respiratory information will be addressed by current research. For that purpose, an entirely automated recognition of SAHS-events will be examined utilizing the two most frequently used classification algorithms inferred from the structured literature review of the field.

1.2 Goal and Research Questions

Goal Elaborate an intelligent monitoring system for detection of Sleep Apnea-Hypopnea Syndrome (SAHS) based on respiratory information analysis.

In this thesis, there will be taken steps towards making a machine learning based monitoring system that extracts useful information from breath. Although there are many interesting features that one might consider, the focus will be on detection of Sleep Apnea-Hypopnea Syndrome.

Research question 1 Which of the following machine learning algorithms; Artificial Neural Network or Support Vector Machine, performs better on classification of SAHS?

Research question 2 Which signal pre-processing technique will result in better performance when considering the first research question?

1.3 Research Method

The methodology applied for this research is a practical approach consisting of an empirical study. The best way of addressing the research questions stated above is to conduct the experiments. The software for data pre-processing has been developed for this purpose, making it possible to perform desired tests. By designing the study in such a manner, various algorithms with equivalent setups can be compared against each other using the same dataset. This will in turn lead to more reliable and valid results.

1.4 Structure

This document consists of five chapters followed by a bibliography and an appendix. Lists of content, figures, tables and abbreviations can be found in the front matter. Each chapter starts with a short introductory paragraph describing its content. The first chapter introduce readers to the content of this thesis. Chapter two provides the fundamental theory needed to understand the field of study. Subsequently, the literature review protocol is specified followed by the motivation for research. The third chapter contains an overview of a model of the proposed system, a detailed explanation of the database, and a description of pre-processing software developed in order to conduct the tests. The experimental plan with the obtained results are described in the following chapter. In the last chapter the evaluation and discussion of results is given in addition to contributions made to the field, several suggestions about further work and a brief summary of this study.

2 Background Theory and Motivation

This chapter provides the fundamental theory needed to entirely understand this field of study. Background theory is followed by a section specifying the structured literature review protocol used in the first phase of research, where the field was explored and state-of-the-art knowledge gained. This laid the foundation for the motivation described in the last subchapter.

2.1 Background theory

Essential knowledge from the field of respiratory medicine and machine learning is provided in this section. A description of present techniques for respiration rate monitoring and specification of equivalent events is given, followed by an explanation of adequate topics within machine learning and signal processing.

2.1.1 Respiration rate monitoring

Respiratory monitoring can typically be classified into two groups based on whether the measuring instruments makes physical contact with the patient or not. Contact based instruments are still the most used in clinical environments, due to their reliability and maturity. Unfortunately, wearing sensors implies decrease of the patient's comfort, thus other approaches are being explored. Despite the major interest of noncontact methods for respiration rate monitoring, the safety aspect is yet not entirely solved. The main concerns are associated with data acquisition obstructions and noise, which may lead to inaccurate or even absent measurements.

An overview of currently used monitoring techniques can be found in (F.Q. AL-Khalidi, 2011), where the authors present the state of art within the field. Together with (C. Brüser, 2015) who discusses complementary unobtrusive techniques, they constitute an overview of up-to-date respiration rate monitoring methods. A comprehensive list of those techniques with corresponding parameters can be found in Table 2.1 and Table 2.2. It is worth to mention that they can be combined into multimodal systems, which may result in uncertainty reduction and safety enhancement.

Contact based respiration monitoring	Parameters measured
Acoustic	Breathing sound
Airflow	Inhaled / exhaled air volume, temperature, pressure
Chest and abdominal movement detection	Band stretching
Transcutaneous CO_2 monitoring	Skin temperature
Oximetry probe	Blood saturation
ECG-Derived	Fluctuation in ECG

Table 2.1 Contact based respiration monitoring methods.

Noncontact based respiration monitoring	Parameters measured
Radar	Chest movement – Doppler (microwaves)
Optical based	Visible range image, depth image
Thermal imaging	Thermographic image, temperature
Ballisto- and seismocardiography	Mattress pressure, deformation, vibrations
Laser	Laser dot position, vibration
Electrical/Magnetic impedance	Change of organ volume and electric properties
Photopletysmographic Imaging	Skin color variation caused by blood volume changes

Table 2.2 Noncontact based respiration monitoring methods.

2.1.2 Respiratory events

Respiratory events are sleep disorders or patterns associated with abnormal breathing or periodical absence of breath. There are several types of sleep disorders differentiated by the scale of duration and the reason of occurrence. A present standard defining the rules of scoring respiratory events has been commissioned in 2004 by American Academy of Sleep Medicine (AASM) based on the original Rechtschaffen and Kales (R&K) manual from 1968 (M. H. Silber, 2007). In the following, the description of respiratory events is given according to (Thorpy, 2012) and the definitions in (R. B. Berry, 2012).

Sleep disorders are divided into two main groups; central and obstructive. Central sleep disorders are characterized by breath cessation caused by respiratory effort reduction or its absence. Respiration failure is caused by a temporal brain dysfunction that results in the absence of the signal that was supposed to be sent to the muscles responsible for inhaling. On the other hand, obstructive disorder is a breath cessation caused by an obstruction in the upper airway, usually as an effect of laxity of the throat muscles. Additionally, a combination of both aforementioned groups is referred to as mixed sleep disorder. Depending on the extent of respiration disruption, sleep disorders are further divided into apnea and hypopnea events. According to the AASM-definition of an apnea, the amplitude is reduced by at least 90 % implying a significant or complete breath interruption. A hypopnea is rather an overly shallow breathing where, by definition, the amplitude is reduced by at least 30 %. In both cases the event must last for 10 seconds or longer in order to be accepted.

Furthermore, due to the occurrence of sleep patterns in the utilized database (3.3), an explanation is hereby given for the theory completeness. Cheyne-Stokes is a breathing pattern of at least three consecutive cycles consisting of a progressively increasing then gradually decreasing signal, followed by an apnea or hypopnea. Periodic breathing is a breathing pattern of at least three consecutive cycles consisting of breathing signal, followed by an apnea or hypopnea.

Present standard procedure for diagnosis of sleep disorders is called polysomnography. It is an overnight sleep study conducted in a clinical environment that records multiple physiological parameters. Monitored body functions may vary yet a typical PSG includes; activity of the brain [EEG], eyes [EOG], muscles [EMG], hearth [ECG], and respiratory measurements in form of airflow, effort and oxygen saturation.

2.1.3 Machine learning

Supervised learning

Machine learning technique that learns a model based on labeled training data entries consisting of an input in the form of a feature vector with a corresponding output value. Mainly used for tasks as classification and regression.

Unsupervised learning

Machine learning technique that learns a model based on unlabeled training data entries, used for tasks such as clustering and dimensionality reduction.

Artificial Neural Network (ANN)

Biologically inspired machine learning algorithm that attempts to model the human brain, applicable to various tasks such as classification, regression, clustering and association (S. B. Maind, 2014). The model is built of units called perceptrons – equivalent to biological neurons – consisting of an input vector **x** = [x_1, …, x_n], a weight vector **w** = [w_{1j}, …, w_{nj}] and an activation function (see Figure 1). The output of a perceptron is decided by a threshold function that process the weighted sum of the inputs. A feedforward ANN is a network of perceptrons arranged in layers, where each perceptron is connected to all perceptrons that belong to the subsequent layer as shown in Figure 2. In that manner, the network becomes acyclic with one-directional flow of information. Backpropagation is a method for supervised learning in a feedforward ANN with the objective of modifying weights in order to minimize the value of a cost function, by calculating the gradient with respect to the weight vectors.

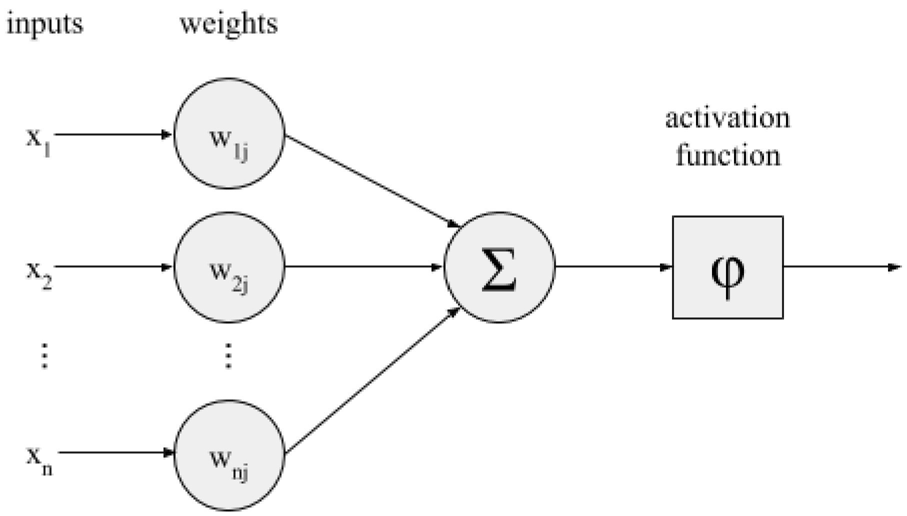

Figure 1 Logical representation of a perceptron with the input vector x, the weight vector w.

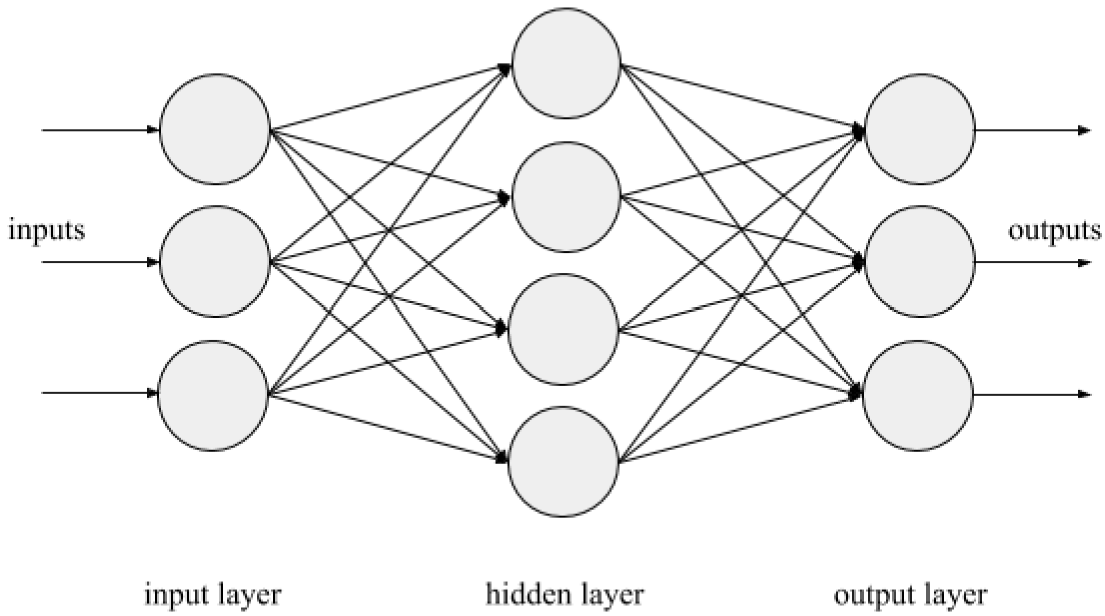

Figure 2 A feedforward Artificial Neural Network with one hidden layer.

Support Vector Machine (SVM)

Supervised learning algorithm applicable for classification and regression that separates hyperplanes by maximizing the distance between the hyperplane boundary and the closest samples as shown in Figure 3 (V. Vapnik, 1995). Originally intended for linearly separable data, however extendable to nonlinear problems by a technique called a kernel trick, which transforms the input into a feature space of higher dimensionality. A Gaussian radial basis function (RBF) kernel is the most common among kernel functions that may be utilized for this kernel trick.

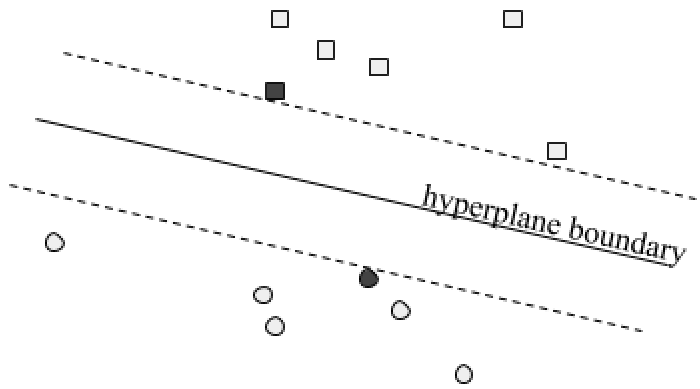

Figure 3 A hyperplane boundary separating two classes with equivalent margin indicated as the area between dashed lines.

Classifier evaluation

An analysis of a classifier involves a consideration of various metrics calculated on the basis of obtained classification results. A confusion matrix is often used for the purpose of a systematic performance visualization. A table of confusion including equivalent metrics used in the evaluation of the conducted tests is provided in Figure 4. Additionally, metric definitions with brief intuitive descriptions are provided in Table 2.3.

		Predicted condition		
	Total population	Predicted positive	Predicted negative	
True condition	Positive	True positive	False negative	True positive rate (TPR), **Recall**
	Negative	False positive	True negative	False positive rate (FPR), Fall-out
	Accuracy	**Precision**		

Figure 4 A table of confusion presenting the relationship between important evaluation metrics.

Metric	Formula	Description
Accuracy (ACC)	$\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	The amount of correctly classified entries
Recall	$\frac{\Sigma \text{ True positive}}{\Sigma \text{ True positive} + \Sigma \text{ False negative}}$	The amount of true positives that were found
Precision	$\frac{\Sigma \text{ True positive}}{\Sigma \text{ True positive} + \Sigma \text{ False positive}}$	The amount of found entries that were correct
F-measure	$2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$	Harmonic mean of recall and precision
ROC area (AUC)	$\int \text{TPR(T)FPR(T)dT}$	Chance that a positive entry is ranked higher than a negative one

Table 2.3 Definition and description of important classifier evaluation metrics.

2.1.4 Signal processing

Discrete Wavelet Transform (DWT)

A wavelet transform of discrete signal *x* that preserves both frequency and temporal information. A DWT(*x*) is calculated by simultaneously passing the signal through high- and low-pass filters, denoted as h[n] and g[n] respectively. Each iteration results in detailed coefficients of equivalent level obtained by the high-pass filter, and approximation coefficients given by the low-pass filter that are used for further decomposition. The filters remove half of the frequencies of *x*, thus due to the Nyquist's rule the output is downsampled by 2. Figure 5 illustrates a level 2 discrete wavelet transform.

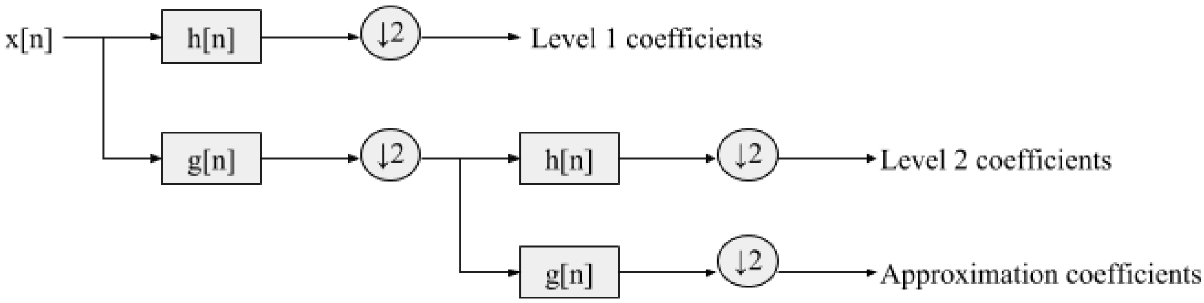

Figure 5 A diagram of discrete wavelet transform of level 2

Fast Fourier Transform (FFT)

A discrete Fourier transform that decomposes a signal *x* into its corresponding representation in a frequency domain. The transform results in coefficients of sinusoids ordered by the frequency. Given the sequence of N complex numbers, the definition is as follows:

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{\frac{-2\pi ikn}{N}} \quad , k \in \mathbb{Z}$$

2.2 Structured literature review protocol

The first phase of the research comprised the literature review. For that purpose a structured protocol is specified in order to execute this task in an efficient and comprehensive manner. Search engines, keywords, and the evaluation and inclusion criteria used during the literature review are defined in this section.

2.2.1 Search engines

It is important to use trustworthy and reliable sources during exploration of scientific papers. In order to ensure a most comprehensive search, a set of academically recognized search engines have been used. A complete list of engines can be found in Table 2.4. The structured procedure of using those engines is to traverse through the entire list given in Table 2.4 and search for the keywords specified in subsection 2.2.2. Conducting the search in such a way may cause overlapping results, since articles are often available at several locations simultaneously.

ID	Search engine	Link
L1	ACM Digital Library	http://dl.acm.org/
L2	IEEE Xplore Digital Library	http://ieeexplore.ieee.org/Xplore/home.jsp
L3	ISI Web of Knowledge	http://apps.webofknowledge.com/
L4	ScienceDirect	http://www.sciencedirect.com/
L5	CiteSeerX	http://citeseerx.ist.psu.edu/
L6	Springer Link	http://link.springer.com/
L7	Academia	https://www.academia.edu/
L8	Hindawi	http://www.hindawi.com/
L9	National Centre for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
L10	Wiley Online Library	http://onlinelibrary.wiley.com/

Table 2.4 A comprehensive list of search engines used during the literature review.

2.2.2 Keywords

Once the search engines are specified, it is time to perform the most crucial part of the literature review, namely to find the appropriate keywords for searching. This task is essential in order to obtain a broad knowledge in a field of interest and it should be well-thought-out. This study concatenates several disciplines – computer science, signal processing, biotechnology and medicine – that all have been considered both in a separate and combined manner. Furthermore, it is important to be aware of the significance of using alternative words such as synonyms and scientific terms. Taking that into consideration, a list of keywords used for searching during the literature review have been created, and a subset with the most significant entries can be found in Table 2.5.

ID	Keyword	Alternative (synonym, variation, scientific term)
W1	Breath	Breathing, respiration, respiratory, airway
W2	Monitoring	Detection, observation, diagnosis
W3	Noninvasive	Non-invasive, unobtrusive, unconstrained, ambient, non-intrusive, video-based, Kinect
W4	Rate	Rhythm, response
W5	Disorder	Event, sleep disorder
W6	Apnea	Sleep Apnea-Hypopnea Syndrome, SAHS, hypopnea, sleep apnea
W7	Automated	Computer-assisted, computer-aided
W8	Anomaly	Abnormal
W9	Polysomnography	PSG, sleep study
W10	Patient	Ribcage, chest, abdomen, abdominal
W11	Learning	Machine learning
W12	Classification	Artificial neural network, ANN, support vector machine, SVM
W13	Feature	Feature extraction, informative feature, wavelet, energy based, fast Fourier transform, FFT

Table 2.5 List of significant keywords used for searching during the literature review.

2.2.3 Inclusion criteria

Searching in search engines specified in subsection 2.2.1 by using keywords listed in subsection 2.2.2 resulted in an enormous amount of scientific papers. Reading extensively through all of them would be very inefficient – if possible at all – thus, the inclusion criteria are introduced in order to filter out an satisfactory articles. The inclusion criteria ordered by the importance are:

1. Title
2. Publication year
3. Abstract
4. Introduction
5. List of references

Titles contain significant information and are often a good indicator of the articles' content. Considering the fact that disciplines of this study are expanding rapidly, it urges for using the publication year as the second criterion. It is often preferred to use up-to-date research, besides relevant older publications will often be found as references in more recent work. The succeeding task is to read through the abstract and introduction, which hopefully will provide an overview of the content. Finally, it might be a good practice to check the list of references that will ensure the credibility.

A review protocol is defined as follows; start with the inclusion criterion of highest importance and move to the lower level only when the current criterion is fulfilled. The article should be added to the list of relevant scientific papers for in-depth examination when all inclusion criteria are satisfied. An example-record of a research paper that successfully has passed all criteria can be found in Table 2.6.

Search engine ID	L6
Keyword ID	W1, W2, W3
Title	Vision-based patient monitoring: a comprehensive review of algorithms and technologies
Publication year	2015
Link	http://link.springer.com/article/10.1007/s12652-015-0328-1
Notes	Good collection of currently used methods, different approaches for patient monitoring

Table 2.6 An example of a record in the list of accepted research papers.

2.3 Motivation

This section presents motivating factors behind the research, including reasons for choosing this field and the importance of addressing the stated goal and research questions. Additionally, significantly important research related to current study will be reviewed. The structured literature review protocol defined in section 2.2 has been used during the knowledge acquisition phase.

Rapid growth of computer science with equivalently increased research activity in the field opens new opportunities. Some industries and professions are being exposed for partial or complete computer-based replacement. Medicine is a discipline wherein scientists are endeavoring to automate the processes executed by healthcare professionals. However, when human health and life are involved, and even a slight mistake can have colossal consequences, the expectations tend to be very high. Thus, a major effort put into research is essential, which motivates to delve into it.

The increase in the elderly population in developed countries, continuous cost reduction in healthcare, and human desire for independence motivates to create an intelligent, entirely automated, patient monitoring system (P. Rashidi, 2013). Different approaches have been proposed – both contact based and ambient, measuring single and multiple parameters – although current research tend towards noninvasive, preferably multimodal solutions (S. Sathyanarayana, 2015). Such ambient approaches emphasize the patients' comfort in addition to the desired functionality and reliability. Due to the limits of this study, which restricts both time and human resources, there will only be taken a few steps towards the aforementioned monitoring system focusing on the extraction of interesting information from breathing.

An exploration of relevant literature through searching for useful respiratory information has led to the disclosure of a very interesting and important area. Sleep Apnea-Hypopnea Syndrome (SAHS) is a medical condition causing morbidity or even mortality (G.C. Mbata, 2012) characterized by its underdiagnosis issue (V. Somers, 2008). Present practice of diagnosis method includes an overnight clinical polysomnography (see section 2.1.2) followed by manual analysis of results by a physician. Here, however, a problem arises. Symptoms must have been perceived in order to get a medical referral, which unfortunately is not an easy task mainly caused by patient's unawareness. Furthermore, polysomnogram-analysis is a demanding task that entails major cost in both time and effort by the clinician, consequently consuming human resources and increasing medical centers' economic cost (D. Alvarez-Estevez, 2015). This motivates to elaborate an automated scoring method that can substitute the manual procedure, simultaneously making the analysis more efficient.

The essential part of making a computer-aided scoring system, for the purpose of finding and labeling SAHS-events in polysomnograms, is to find proper classification algorithms. Different approaches have been proposed as the result of recent research in this area. Despite the huge progress, there has not yet been proposed any solution that satisfies the reliability required in medical tools. A relatively recent review written by (D. Alvarez-Estevez, 2015) has been found very useful as the starting point for further study. It contains a structured and comprehensive review of currently applied methods for computer-assisted diagnosis of SAHS with equivalent comparison of results. However, the main issue with reviews is that the experiments are conducted with different datasets in various environments, thus a direct comparison is inadequate. Although, it provides an indication of the performance and, more importantly, allows to get a quantitative overview of frequently used methods. Two algorithms, Artificial Neural Network (ANN) and Support Vector Machine (SVM), have distinguished themselves among the approaches proposed for classification of SAHS-events. The research questions of this study have consequently been defined, and can be found in section 1.2.

Addressing the research questions require an in-depth knowledge of machine learning and signal processing, thus further literature review has been proceeded. Articles of major significance related to topics of classification are hereby discussed. The following articles (M. Emin Tagluk, 2010) (M. Tagluk, 2010) (N. Sezgin, 2009) compose an implicit series of research conducted by an almost unchanged group of scientists. The introduced approaches, with slight variations, propose the use of ANNs for classification of SAHS-events with Discrete Wavelet Transform as a suggested signal pre-processing and feature extraction method. Although those articles appear as independent, the experimental reproducibility has been impeded due to an incomplete description, forcing readers to consider all of them to get a sufficient explanation. The idea of taking advantage of ANNs together with wavelet transform in the context of SAHS classification has been confirmed by other studies such as (O. Fontenla-Romero, 2005) and (B. Guijarro-Berdiñas, 2012). The latter research attempts to combine ANNs with SVM in order to improve the classification accuracy and precision, constructing a multimodal expert system. Pure use of SVM approach has been proposed by (Y. Maali, 2012) with preceding extraction of statistical information from a wavelet transform. All aforementioned approaches – henceforth referenced to as *focus collection* – have produced satisfactory results, but have been conducted with different data making direct comparison inadequate. Taking that into consideration has motivated current study to perform an equitable comparison of used algorithms by reproducing the approaches and conducting experiments with the same dataset.

3 Model

The focus of this chapter is directed towards the model of classification process. Prior to that, however, an introductory description of an automated monitoring system is provided. Finally, a specification of the utilized database is given as the last part of this chapter.

3.1 Monitoring system

This section concentrates on computer-aided monitoring system deduced by the goal of current study. The intention is to acquire an overall overview of the system architecture that will constitute the context. Thus, some ideas and thoughts regarding possible functionality are being presented.

Elaboration of a system for medical purpose demands careful consideration of the requirements. Reliability and decision correctness should ideally be – or considering a more realistic scenario – approach 100 per cent. Additionally, real time performance is desired in order to enable quick reaction to anomalies. Anomalies in this context are events caused by abnormal breathing that are to be detected by the system.

The proposed patient monitoring platform consists of three main modules; data acquisition component, learning unit, and anomaly handler. The first module is intended to gather the information about the patient that in turn constitutes the input data for the learning unit. Various channels can be utilized for this purpose according to desired parameters to be monitored. This research, however, exploits channels associated with respiration. Considering an ideal scenario all channels are noncontact-based surrounding the patient in an unobtrusive way. Preferring ambient sensors, as opposed to contact-based, ensures that the patient's comfort remains intact making the system imperceptible. Unfortunately, present technology does not allow to fulfill that ideal scenario without compromising other requirements, thus in some cases contact-based sensors are necessary. Some ambient breath monitoring approaches that can be proposed are those tracking chest and abdomen movement, taking advantage of thermal features, or utilizing traits of pressure and sound. Adequate devices may be video cameras capturing visible and/or invisible light, thermal cameras, audio recorders and pressure mats. However, as is the case for the majority of noncontact-based devices, multimodality is preferable due to exposure to noise caused by environment distortion.

The learning unit is responsible for machine learning and anomaly recognition tasks. Plenty of useful information can be extracted from data acquired by the first module according to the sensors that have been utilized. With proper processing such respiratory data can be used for detection of sleep disorders or falls. Additionally, general information about the subject such as age can be estimated, followed by vital signs, sleep stage, mood and emotion recognition. Anomaly recognition is done by applying machine learning methods for classification and regression. As mentioned above, real time performance is desirable, thus analyzing a continuous signal is an apt choice. Nevertheless, classification performed on discrete chunks of data proves to be relevant for instance during scoring of SAHS-events. Thus, both unsupervised and supervised learning may be considered depending on the current task and the amount of disposable data.

The last module is an anomaly handler that aims to perform actions in case of an aberration occurrence. The list of possible actions is long, thus just a few examples will be presented. The obvious choice is to start the alarm procedure by notifying responsible personnel such as a nurse. Another possibility is to add functionality to tune the appropriate parameters in the corresponding medical treatment devices. The latter suggestion would result in an entirely automated system but implies an enormous liability and the need of total reliability and trust of the system. Proposition to a less responsible yet valid function of this module would be to play some kind of relaxing music or show a pleasant image, in order to improve the patient's mood or evoke positive emotions. Furthermore, all learned data can be stored and used for further diagnosis or treatment.

3.2 Classification model

This section describes the model including the software that has been developed for the purpose of this study. The implementation and design decisions will hereby be discussed.

Recall that the classification of SAHS-events is the motivation main behind this study and consequently also the main focus of the experiments. As discussed in section 2.3, there have been chosen six studies (focus-collection articles) that produce the best results on SAHS classification task, and equivalent approaches has been reproduced in this study. In order to conduct a comparison of most the used classification algorithms in the SAHS context, a testing platform was needed. A model depicting the classification process can be found in Figure 6.

The reader is highly encouraged to get familiar with its content as it will be of help to understand the rest of this thesis.

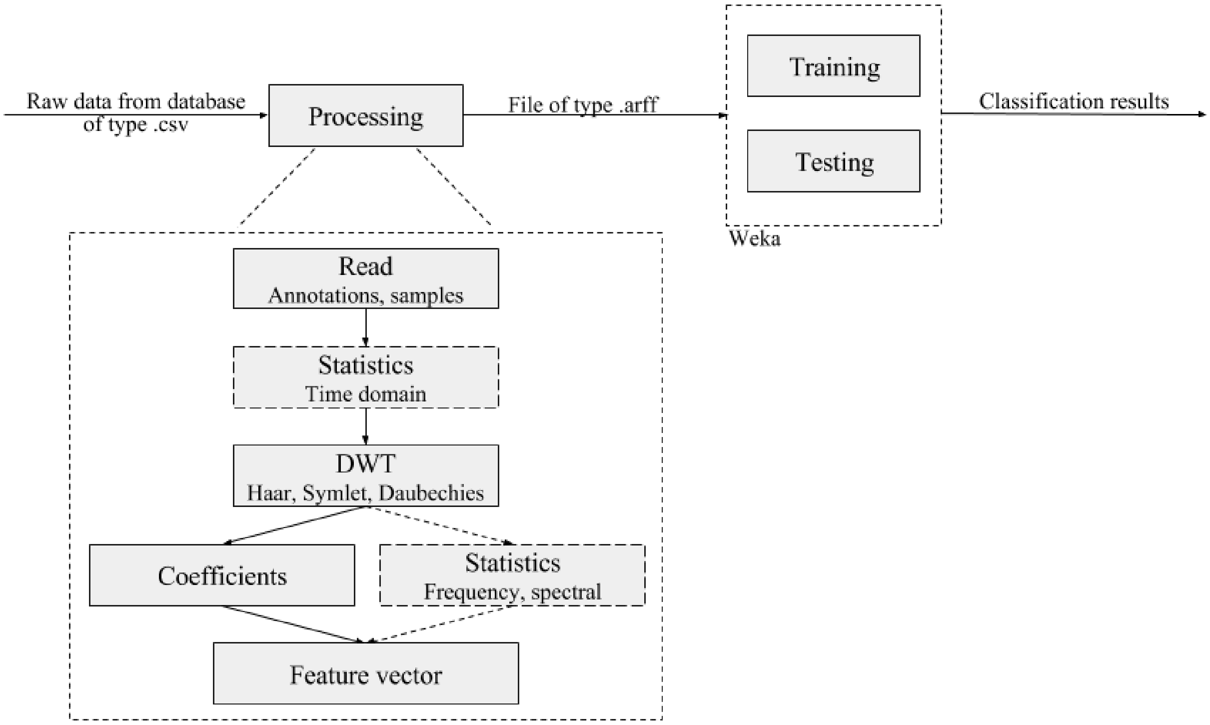

Figure 6 Model of classification process consisting of raw data input, processing unit, learning unit and classification output.

The classification process starts with acquisition of raw patient data. As discussed in section 3.1, such data can be acquired in a contact-based or unobtrusive manner, utilizing various types of sensors. Obtaining sufficient amount of data by carrying it out as a part of this thesis would require additional resources and may lead to an uncomplete study, which has been main motivation to decline this option. Thus, the raw data used for the experiments has been acquired from a freely accessible database that is described in the following section (3.3).

When the raw data is acquired, the subsequent step is to process the input data and make it suitable for classification step. Thus, a design and development of an adequate processing software has been required. Java programming language with corresponding libraries listed in Table 4.1 has been used for this purpose. The first stage of processing consist of reading raw signal samples and annotations into the program. Respiratory annotations are given as external text files – each corresponding to one patient. As described in section (3.3) signal samples have been downloaded in the comma-separated values-format (.csv) with four input channels in addition to a timestamp. During the second stage, the focus is on extracting significant features from the signal that will constitute a feature vector – an entry in the training data. More thorough explanation; every scope of samples corresponding to a single respiratory annotation is grouped into a signal chunk that subsequently is transformed into a frequency domain. There are two main arguments why taking direct use of raw signal in the time domain is a bad idea. The first concerns the nature of SAHS-event of which the length may vary, while a classifier requires a fixed number of inputs. Furthermore, considering a sampling rate of 128Hz and the definition of SAHS-events which states its length to last for more than 10 seconds, the smallest chunk contains 1280 samples. Such a high number of inputs would cause overfitting making the classifier to perform very poorly. In order to allow a wider spectrum of experiments, current implementation supports multiple approaches of signal chunk transform and feature vector composition. According to focus-collection the discrete wavelet transform (see subsection 2.1.4 for description of DWT) is recommended as the pre-processing technique of raw input signal. The reasoning behind that postulate is that respiratory signals may change much over time – a characteristic of non-stationary signal – making it desired to recognize variations by preserving both frequency and temporal information. Furthermore, the support for three wavelet transform methods – Haar, Symlet and Daubechies – has been added to current implementation. Wavelet transform results in a list of coefficients ordered by a descending detail level. Two approaches – coefficient- and statistics-based – has been independently tested while creating a feature vector according to focus-collection studies. The most popular method is to form the feature vector by taking the mean value of all coefficients corresponding to each detail level of wavelet transform. In such a manner, given a certain level of detail, the feature vector will as desired have a constant length regardless of SAHS-event duration. Second method utilizes both original time domain signal as well as frequency domain by analyzing its statistical characteristics. As proposed by (D. Alvares, 2012) statistical features includes arithmetic mean, variance, skewness and kurtosis for both time and frequency domain, in addition to median frequency, total spectral power and peak amplitude. Once the feature vector is composed, equivalent annotation class is added as the last element, subsequently adding a new line to training file in the attribute-relation file format (.arff). The last stage of processing is to combine data in a desired way by joining several training files. This procedure enables to create training files that are based on multiple patients simultaneously providing a better learning basis. A sequence diagram of processing module described above can be found in Figure 7.

So far in the classification process, the acquired patient-data has been preprocessed and is now ready for the main stage, classification. For that purpose, the open-source software Weka 3.8, which provides a collection of machine learning algorithms has been used. Weka is using previously mentioned (.arff) as the main input file format, thus all training files are created accordingly. Due to the objective of this study, classifiers being utilized are MultilayerPerceptron and LibSVM, which corresponds to a feedforward Artificial Neural Network with backpropagation and Support Vector Machine classification algorithms (see subsection 2.1.3 for detailed explanation). After choosing an appropriate machine learning algorithm with equivalent settings, the classifier is trained and the obtained model is saved in (.model)-format for eventual future reuse. Testing is conducted by performing classification with previously saved model on data acquired from a new patient. Classifier output contains valuable information that is used for evaluation and comparison of performance. Adequate parameters will be discussed in the following chapters, but firstly it is highly recommended to get familiar with subsection 2.1.3 where an explanation of classifier metrics is given.

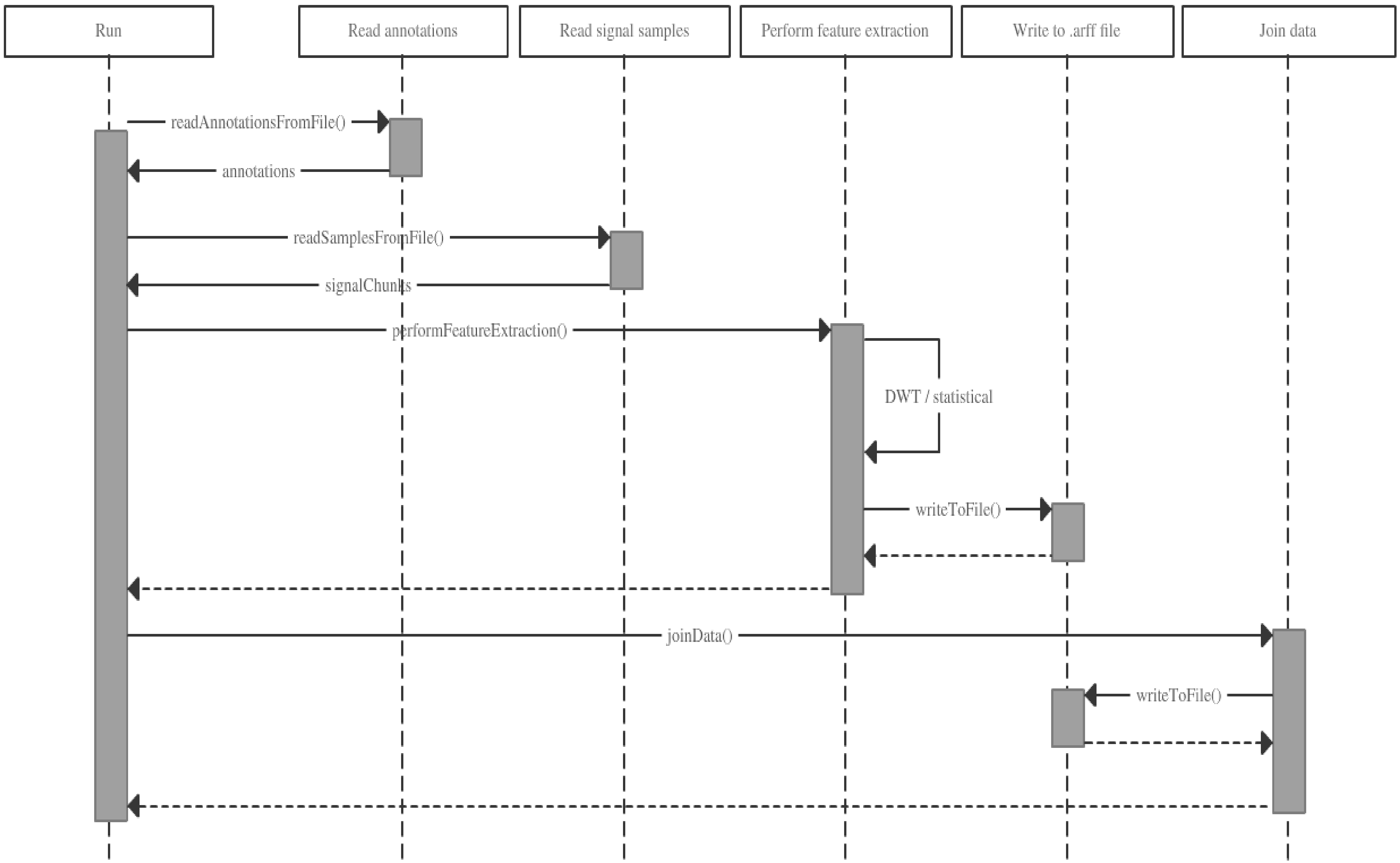

Figure 7 Sequence diagram of signal processing stage.

3.3 Database

In order to find proper data that satisfy the requirements, a certain number of inclusion criteria were needed. Thus, while choosing database the main concern was on the presence of respiratory rate signal, respiratory event annotations, the duration, environment and the number of subjects. This section contains a detailed description of the database used in this study.

For the purpose of this thesis an adequate database have been found using the free access collection of recorded physiologic signals provided by PhysioNet[1]. The recordings were collected and assembled at St. Vincent's University Hospital Sleep Disorders Clinic by (W. McNicholas, L. Doherty, S. Ryan, J. Garvey, P. Boyle, E. Chua). The database contains overnight polysomnograms, sets of vital parameters recorded during sleep, from 25 adult subjects with suspected sleep-disordered breathing. Complete list of input signals included in PSGs with corresponding explanation can be found in Table 3.1.

Input channel	**Explanation**
Electroencephalography [EEG]	Electrical activity of the brain, measured in areas (C3-A2), (C4-A1)
Left Electrooculography [EOG]	Left eye movement
Right Electrooculography [EOG]	Right eye movement
Submental Electromyography [EMG]	Electrical activity produced by muscles in submental space (located in the midline under the chin)
Electrocardiography [ECG]	Electrical activity of the hearth
Oro-nasal airflow	Air-pressure measured by thermistor
Ribcage movements	Strain gauge (uncalibrated with abdomen)
Abdomen movements	Strain gauge (uncalibrated with ribcage)
Peripheral oxygen saturation [SpO_2]	Concentration of oxygen (O_2) in the blood measured by finger pulse oximeter
Snoring	Sound measured by tracheal microphone
Body position	

Table 3.1 Input channels included in polysomnograms of this database.

[1] https://physionet.org/

This database contains two types of annotations – sleep stages and respiratory events – annotated by an experienced sleep technologist. Sleep stages are divided into eight categories; wake, REM, stage 1, stage 2, stage 3, stage 4, artifact and indeterminate. Considering the fact that the focus of this thesis is on respiration rate rather than on sleep monitoring, details of sleep stages are omitted. As described in theory subsection 2.1.2 we can differentiate several respiratory events. Those annotated in the database are divided into three groups and can occur simultaneously. First group is a set of sleep disorders containing obstructive sleep apnea (APNEA_O), central sleep apnea (APNEA_C), mixed sleep apnea (APNEA_M), obstructive sleep hypopnea (HYP_O), central sleep hypopnea (HYP_C) and mixed sleep hypopnea (HYP_M). The total number of respiratory event occurrences related to first group is listed in Table 3.2. The second group are patterns including periodic breathing (PB) and Cheyne-Stokes (CS), while the third group indicates presence of heartbeat arrhythmia. Capital letters in parenthesis corresponds to description of given events used in annotation-files.

Event	**APNEA_O**	**APNEA_C**	**APNEA_M**	**HYP_O**	**HYP_C**	**HYP_M**
Number of occurrences	167	266	106	1446	974	109

Table 3.2 Total number of all annotated sleep disorder events in the database.

This database also contain metadata that may be of interest when considering additional features. A set of parameters such as height, weight, gender, BMI and age of each subject is listed at the database's main page[2]. Furthermore, respiratory annotation files corresponding to each patient are provided in text format (.txt) and have been downloaded directly from the same location. Raw data samples have been acquired through Cygwin Terminal[3] that allows Windows-users take advantage of Linux-commands. Due to memory concerns, only the desired channels – abdomen, ribcage, airflow and oxygen saturation (SpO2) – have been included in the comma-separated values-formatted (.csv) files. Data samples have been downloaded by the command given below – in this case samples of subject 25 saved locally as "025_samples.csv". A list of options including corresponding descriptions of the command is provided in Appendix B: PhysioNet.

```
rdsamp -r ucssb/ucssb025.rec -c -H -v -pd -s abdo ribcage Flow SpO2
> 025_samples.csv
```

[2] https://physionet.org/physiobank/database/ucddb/
[3] https://www.cygwin.com/

4 Experiments and Results

The objective of the chapter is to address the research questions defined in section 1.2, by applying experimental methodology. The experimental plan is drafted in the first section followed by a detailed description of the test environment including the corresponding parameters. The results were obtained by conducting experiments in accordance to predefined plan, and are presented in the final section of this chapter.

4.1 Experimental plan

This section presents a plan designed for the empirical part of this thesis. Adopting this strategy during the study ensures deliberated experiments conducted in a systematic way. Referring to the research questions (1.2), the experimental plan should concern comparison of classification algorithms and equivalent pre-processing techniques. The compared algorithms are Artificial Neural Network and Support Vector Machine. Several aspects need to be considered in order to ensure the completeness of tests.

Firstly, it is important to look into the training file and ponder how to compose it. One option is to decide on a certain amount of minutes, M, and use annotations, A, corresponding to every subject within the scope of M. Choosing this alternative would result in a diversified training file of length 25*A, but it may lead to insufficient training data due to lack of event occurrence guarantee during M. Furthermore, an SAHS-event classifier should perform in a generalized manner i.e. training must be accomplished beforehand making testing entirely automated, thus this option is undesirable. In consequence, a preferred solution of constructing the training file is to take advantage of all annotations corresponding to a certain number of subjects, X. It is also interesting to find out, to what degree learning will be affected by various values of X. Thus, experiments will be conducted utilizing two training files; one consisting of single subject (x1) and another joining five subjects (x5), constituting approximately 10 and 50 per cent of all events respectively.

Second factor to be considered is whether feature vectors should be taken the absolute value of. There is discrepancy between focus-collection studies regarding this topic, thus both possibilities will be tested. This will allow to find out if there is any advantage of applying absolute value on training entries in the SAHS-event classification context.

The subsequent task is to examine the correlation between the input channel and the classification performance. The objective is to determine if the current method is transferable to other channels, e.g. video-based input. Furthermore, it is interesting to examine how classification may be affected by taking advantage of multichannel. For that purpose, three single input channels will be tested in sequence. Depending on the outcome, the two obtaining best results will subsequently be combined and used for further testing.

Addressing the second research question, data pre-processing methods will be examined. Four techniques are being proposed by focus-collection studies. Three of them varies on type of wavelet transform function, while the last one utilizes statistical analysis during feature vector composition. Moreover, authors of (D. Alvares, 2012) suggests that frequency-based statistics should be extracted from signal processed by Fast Fourier Transform, rather than Discrete Wavelet Transform as argued by remaining focus-collection studies. In order to verify that assertion, both techniques will be used for statistical analysis. Thus, the experiments should be conducted with all five pre-processing procedures, which will provide the extent of the impact on classifier performance.

The final aspect to be verified is to determine training parameters that will result in the best classification outcome. Although it is significant for the end results, the task tends to be time consuming. Thus, all experiments defined above will be performed with default training parameters. In such a way, results are obtained efficiently and can be compared relatively to each other preserving ipso facto validity. However, after the first phase of experiments is done, the factors resulting in best classification outcome should be used for further investigation of training parameters. Additionally, experimental efficiency can be improved by exploiting holdout sets. Thus, cross-validation will be used during the first phase of testing providing results for temporary comparison. Utilization of the trained models for testing in a more realistic environment will be performed during the second phase, where sleep disorder-events corresponding to other patients will be classified.

Ergo, experiments should be conducted in a systematic manner according to Phase I and II protocols defined correspondingly in Figure 8 and Figure 9.

```
For each Pre-processing technique {Haar, Symlet 7, Daubechies 4,
statistics(DWT), statistics(FFT)}
      For each Input channel {abdomen, ribcage, airflow,
      multichannel*}
            For each Absolute value of feature vector {true, false}
                  For each Training file composite {x1, x5}
                        Run ANN
                        Run SVM
```

Figure 8 Pseudocode specifying experimental protocol of Phase I.

```
Given pre-processing factors resulting in best outcome of Phase I
For each Classification parameters setup {default, adjusted}
      Run ANN
      Run SVM
```

Figure 9 Pseudocode specifying experimental protocol of Phase II.

4.2 Experimental setup

In order to deliver a reproducible study, the description of tools used in context of experiments is given in this section.

The experiments have been conducted using a tailor made software together with machine learning freeware as described in section 3.2. The program that has been developed for the purpose of this research is written in Java SE Runtime Environment 8 utilizing the additional libraries listed in Table 4.1. The programming environment used for the implementation, compilation and running of this program is Eclipse IDE (4.5.1). Furthermore, classification has been performed using Weka 3.8 Explorer, which is provided by the Machine Learning Group at the University of Waikato. Specifications of the computer used as testing environment for all experiments can be found Table 4.2.

Library	Version	Description
JWave	160109	Wavelet transforms
Joda-Time	2.9.2	Date and time
Apache Commons Math	3.6.1	Mathematics and statistics
Weka	3.8	Machine learning

Table 4.1 External libraries used in the implementation.

Feature	Specification
Operating system	Windows 8.1 Pro, 64-bit
Processor	Intel® Core™ i7-4770 CPU, 3.40GHz
RAM (random-access memory)	16 GB

Table 4.2 Test environment specifications.

Referring to experimental plan defined in the previous section certain variables need to be determined. The first aspect to decide is which of the available subjects should be included in the classifier training. Careful revision of the respiratory event occurrences led to choosing patient number 025 due to abundant and varying events. In order to get a broad representation of events when combining five patients (x5), the following subjects has been chosen; 003, 006, 010, 025 and 027. Secondly, considering the input channels to be used, three signals has been chosen for testing – abdomen, ribcage and airflow. The reason for omitting oxygen saturation is that it only can be acquired in a contact-based manner, making the transition to ambient solutions impossible. The last factor is to determine pre-processing techniques, among which the following five has been chosen; Haar, Symlet 7, Daubechies 4, and statistical based on DWT and FFT. Applying to the first three methods, as described in section 3.2, the feature vector is composed of the mean values of corresponding wavelet transform levels. The latter techniques utilizes statistical information such as; first to fourth-order statistical moments in both time and frequency domain, median frequency, total spectral power and peak amplitude. A transparent summary of parameters mentioned in current paragraph can be found in Table 4.3.

Factor	Parameter
Training file	x1: 025 x5: merge{003, 006, 010, 025, 027}
Input channel	abdomen, ribcage, airflow
Pre-processing technique	Discrete Wavelet Transform: - Haar - Symlet 7 - Daubechies 4 Statistics: - time domain: arithmetic mean (M1t), variance (M2t), skewness (M3t), kurtosis (M4t) - frequency domain: arithmetic mean (M1f), variance (M2f), skewness (M3f), kurtosis (M4f), median frequency (MF) - spectral features: total spectral power (P_T), peak amplitude (PA)
Cross-validation	10-fold
Multilayer Perceptron (ANN)	Default: - Hidden layers: 1 (7 nodes) [Weka: a (attributes + classes)/2] - Learning rate: 0.3 - Momentum coefficient: 0.2 - Training time: 500 Adjusted: - Hidden layers: 2 (15 nodes each) - Learning rate: 0.05 - Momentum coefficient: 0.95 - Training time: 50000
SVM	Default: - Kernel type: Radial basis function (RBF) - C (cost): 1.0

SVM (cont.)	- Gamma: 0.0 Adjusted: - Kernel type: Radial basis function (RBF) - C (cost): 1.5 - Gamma: 0.005

Table 4.3 List of experimental parameters.

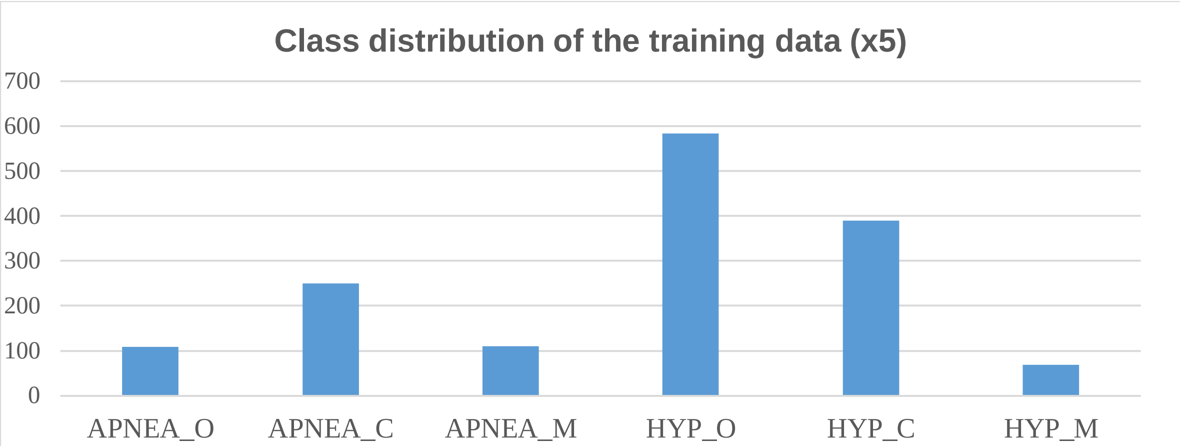

Figure 10 The distribution of classes in training file composed of five merged subjects (x5).

The Phase II-experiments have been conducted with the pre-processing parameters specified in Table 4.4 that resulted in best outcome during the first phase. Furthermore, adjusted classifier parameters listed in Table 4.3 for correspondingly ANN and SVM, have been found based on focus-collection studies, where all setups have been tested and the very best were chosen.

Factor	**Parameter**
Training file	merge{003, 006, 010, 025, 027}
Testing file	merge{002, 005, 007, 008, 009, 011, 012, 014, 015, 017, 019, 020, 021, 022, 023, 024, 026}
Absolute value of feature vector	false
Pre-processing technique	Daubechies
Input channel	Ribcage

Table 4.4 List of pre-processing parameters used during Phase II. Note: values based on results obtained in Phase I.

4.3 Experimental results

This section presents the results obtained from the experiments conducted according to experimental plan (4.1) and setup (4.2). The tests have been divided into two phases, thus this section contains two subsections describing each stage respectively.

4.3.1 Phase I

The results gained during the first stage of the experiments are hereby presented respectively to the order introduced in experimental plan (4.1). The first interesting aspect is how classification is affected by number of entries in the training file. For that purpose, two types of training files have been used throughout all tests, containing SAHS-events of correspondingly one and five subjects. The appropriate measure chosen to draw out equivalent information was the arithmetic mean of correctly classified instances. Table 4.5 shows the results corresponding to both cases. Subsequently, the same calculation is performed considering whether the absolute value has been utilized. The average of both true- and absolute valued training entries that have been classified correctly is shown in Table 4.6.

One subject (x1)	Five subjects combined (x5)
37.67 %	44.44 %

Table 4.5 Arithmetic mean of correctly classified instances given training files consisting of single and multiple subjects.

True value	Absolute value
41.56 %	40.56 %

Table 4.6 Arithmetic mean of correctly classified instances given training files composed of true and absolute values of entries.

Numerous tests have been run in order to examine how the input channels impact classifiers' performance. Due to varying pre-processing techniques that can affect the results itself, feature vectors based on both wavelet coefficients and statistical analysis have consistently been used for every input channel. Results in form of a column chart are shown in Figure 11, depicting the average rate of correctly classified respiratory events in accordance to input channels. Computing the arithmetic mean of each channel cumulatively for all methods results in following values; abdomen: 40.72 %, ribcage: 42.41 %, and airflow: 37.18 %.

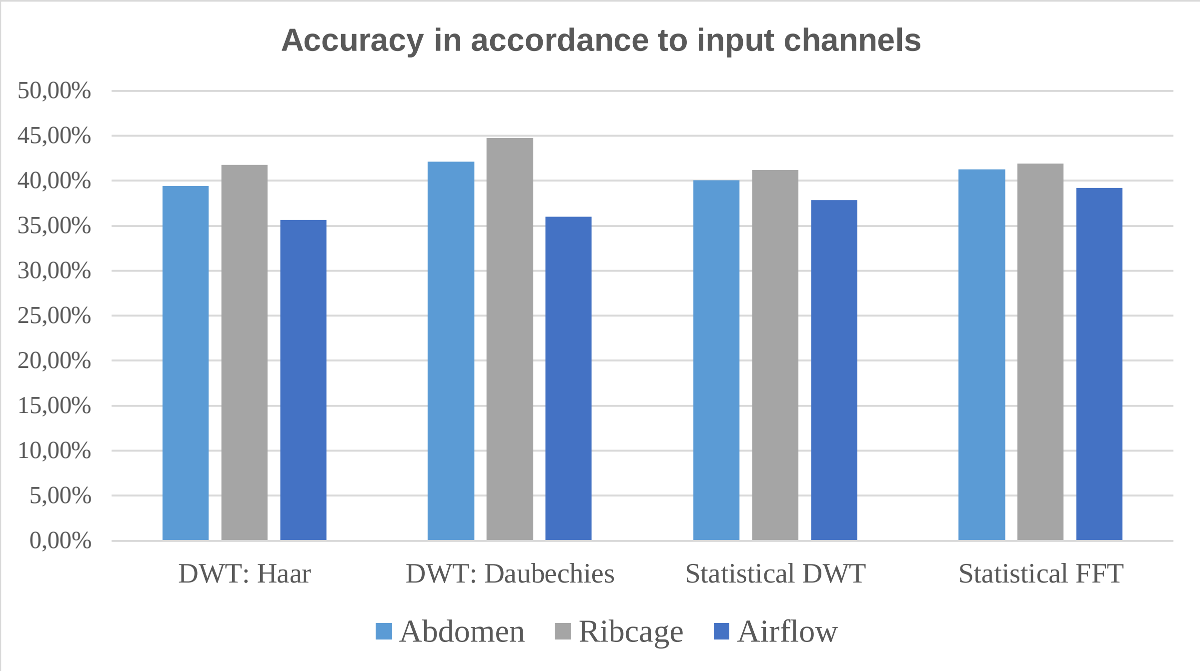

Figure 11 Arithmetic mean of correctly classified instances in accordance to input channels (abdomen, ribcage, and airflow), given four different pre-processing techniques (Haar- and Daubechies wavelet transform, statistical analysis based on Daubechies transform and statistical analysis based on Fourier transform).

Once the idea of how various input channels affect the classification output is recognized, a decision regarding the composition of multichannel can be made. Although the discussion and reasoning will be provided in the next chapter, it was decided to join the abdomen and ribcage signals to constitute a multichannel. Setting single- and multichannel to a constant value enable the comparison of wavelet transform-based pre-processing techniques. Figure 12 displays the average of properly classified events according to wavelet transforms given input channels. Mean values were calculated cumulatively for both single and multiple channel corresponding to each wavelet technique are as follows; Haar: 41.30 %, Daubechies: 43.79 %, and Symlet: 43.53 %. Furthermore, it is of interest to collate pre-processing methods that composes feature vectors based on wavelet coefficients and statistical analysis parameters. Thus, Figure 13 presents averaged classification output, based on results of all three single input channels, correspondingly taking advantage of the aforementioned pre-processing techniques. Daubechies is chosen as the wavelet transform function due to the best results so far. In addition to average of pre-processing methods, an explicit division into classification algorithms is shown.

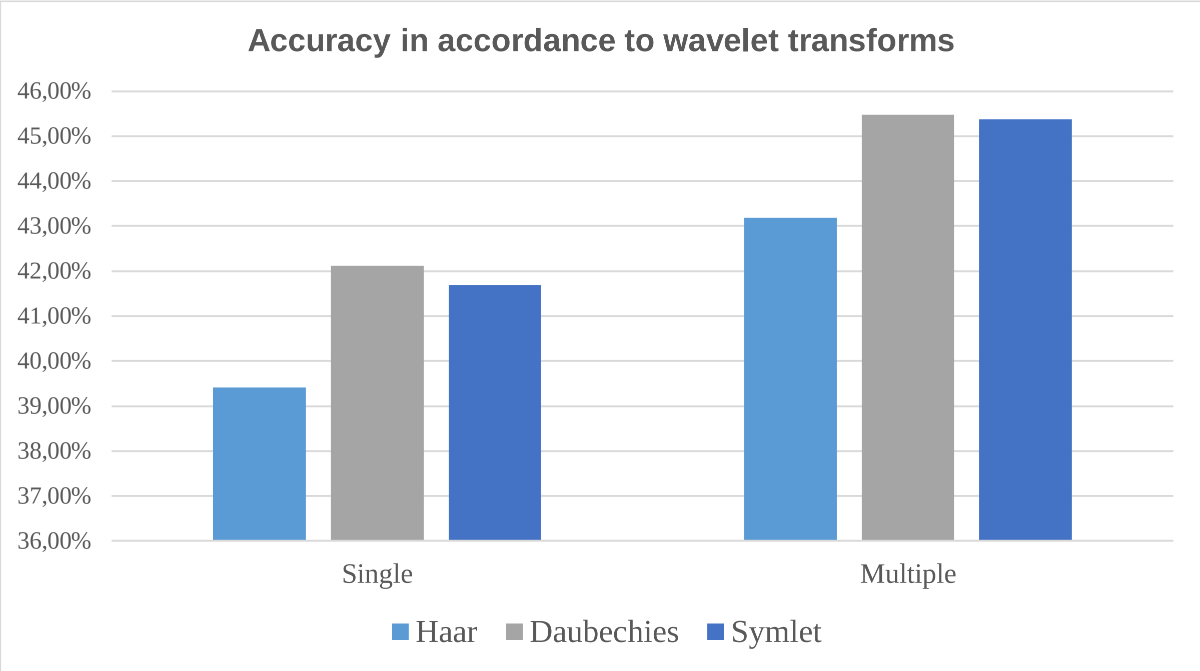

Figure 12 Arithmetic mean of correctly classified instances in accordance to wavelet transforms (Haar, Daubechies, and Symlet), given single (abdomen) and multiple (abdomen + ribcage) input channels.

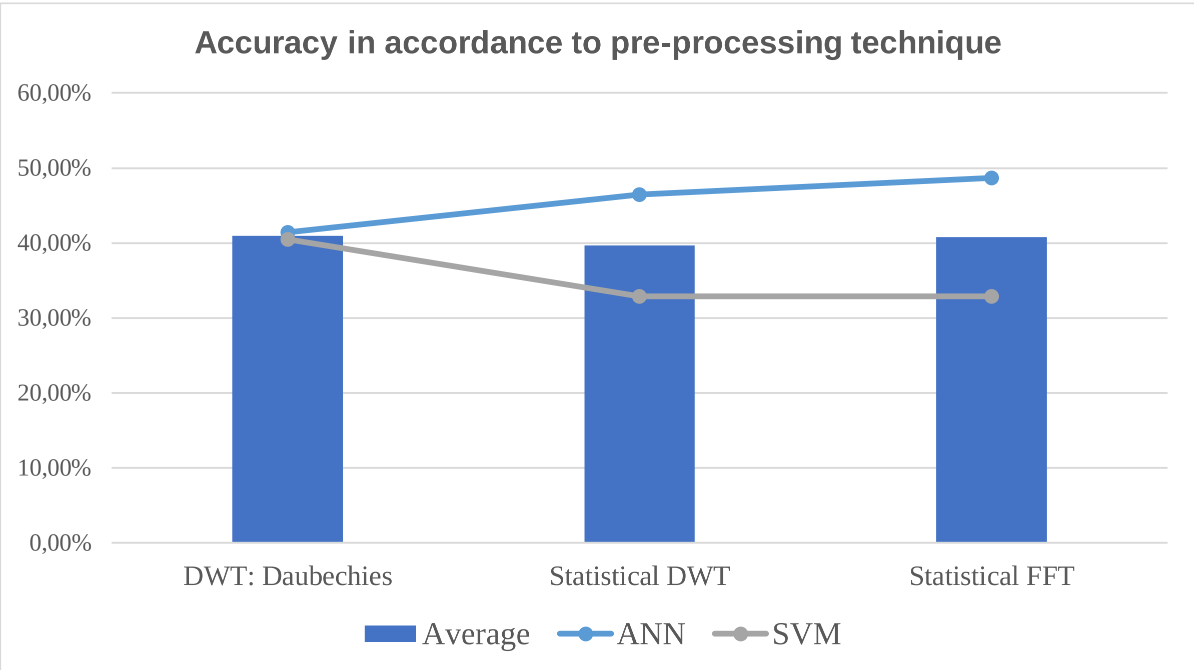

Figure 13 Arithmetic mean of correctly classified instances in accordance to pre-processing techniques (Daubechies wavelet transform, statistical analysis of Daubechies wavelet transform, and statistical analysis of Fourier transform), given all three input channels (abdomen, ribcage, and airflow).

At last, the arithmetic mean of the classifiers’ accuracy for all experiments that have been conducted so far is given in Table 4.7.

Multilayer Perceptron (ANN)	**Support Vector Machine**
43.39 %	38.73 %

Table 4.7 Arithmetic mean of correctly classified instances by ANN- and SVM-classifiers based on all experiments of the first phase.

4.3.2 Phase II

The focus of this subsection is directed towards the results obtained during the second phase of the experiments. As opposed to the first stage, the emphasis in this section was on actual performance of classifiers in a realistic environment rather than pre-processing parameters. From inductive reasoning, it has been decided that the parameters resulting in best outcome of Phase I-experiments (specified in Table 4.4), will be utilized as the only configuration of training entries for the following tests. The succeeding results are obtained by classifiers that have been trained on five merged subjects, and tested on the remaining patients. Phase II-experiments has been repeated twice with correspondingly default and tuned classification parameters.

Results from the first run of Phase II with default classification parameters are presented below. The classification accuracy with equivalent amount of both correctly and incorrectly classified respiratory events is shown in Table 4.8. Additionally, Table 4.9 and Table 4.10 displays significant learning metrics of correspondingly ANN- and SVM-classifier, followed by their confusion matrices (Table 4.12 and Table 4.13).

	Correctly classified	**Incorrectly classified**
ANN	41.11 % (631 instances)	58.89 % (904 instances)
SVM	41.89 % (643 instances)	58.11 % (892 instances)

Table 4.8 The performance of classifiers (ANN and SVM, default parameters) in terms of accuracy and an explicit number of classified instances.

	Recall	Precision	F-measure	AUC
APNEA_O	0	0	0	0.511
APNEA_C	0.733	0.118	0.203	0.788
APNEA_M	0	0	0	0.606
HYP_O	0.612	0.531	0.569	0.584
HYP_C	0.225	0.456	0.302	0.579
HYP_M	0.057	0.125	0.078	0.532

Table 4.9 Significant learning metrics (recall, precision, f-measure, and ROC area) of the ANN-classifier (default parameters).

	Recall	Precision	F-measure	AUC
APNEA_O	0	0	0	0.5
APNEA_C	0.6	0.113	0.19	0.704
APNEA_M	0	0	0	0.498
HYP_O	0.604	0.526	0.562	0.547
HYP_C	0.274	0.444	0.339	0.533
HYP_M	0	0	0	0.5

Table 4.10 Significant learning metrics (recall, precision, f-measure, and ROC area) of the SVM-classifier (default parameters).

	Recall	Precision	F-measure	AUC
P-value	0.377	0.279	0.537	0.016

Table 4.11 Paired t-test for ANN- and SVM-classifier metrics given in Table 4.9 and Table 4.10. Calculated with alpha value equal to 0.05.

	APNEA_O	APNEA_C	APNEA_M	HYP_O	HYP_C	HYP_M
APNEA_O	0	27	0	49	14	1
APNEA_C	0	44	0	10	3	3
APNEA_M	0	8	0	11	6	1
HYP_O	0	154	2	454	126	6
HYP_C	0	138	2	307	131	3
HYP_M	0	2	0	24	7	2

Table 4.12 Confusion matrix of the ANN-classifier (default parameters). Vertical axes: true condition. Horizontal axes: predicted condition.

	APNEA_O	APNEA_C	APNEA_M	HYP_O	HYP_C	HYP_M
APNEA_O	0	18	1	46	26	0
APNEA_C	0	36	0	18	6	0
APNEA_M	0	7	0	14	5	0
HYP_O	0	138	2	448	154	0
HYP_C	0	118	4	300	159	0
HYP_M	0	1	0	26	8	0

Table 4.13 Confusion matrix of the SVM-classifier (default parameters). Vertical axes: true condition. Horizontal axes: predicted condition.

The results from the second run of Phase II, ordered as above, with adjusted classification parameters are presented below. The classification accuracy is shown in Table 4.14, the significant learning metrics of ANN- and SVM-classifiers are correspondingly displayed in Table 4.15 and Table 4.16, and finally the confusion matrices can be found in Table 4.17 and Table 4.18 respectively.

	Correctly classified	**Incorrectly classified**
ANN	46.38 % (712 instances)	53.62 % (823 instances)
SVM	49.77 % (764 instances)	50.23 % (771 instances)

Table 4.14 The performance of classifiers (ANN and SVM, adjusted parameters) in terms of accuracy and an explicit number of classified instances.

	Recall	**Precision**	**F-measure**	**ROC area**
APNEA_O	0	0	0	0.486
APNEA_C	0.15	0.103	0.122	0.653
APNEA_M	0.038	0.029	0.033	0.569
HYP_O	0.836	0.51	0.633	0.58
HYP_C	0.141	0.436	0.213	0.514
HYP_M	0	0	0	0.5

Table 4.15 Significant learning metrics (recall, precision, f-measure, and ROC area) of the ANN-classifier (adjusted parameters).

	Recall	**Precision**	**F-measure**	**ROC Area**
APNEA_O	0	0	0	0.5
APNEA_C	0.033	0.667	0.063	0.516
APNEA_M	0	0	0	0.5
HYP_O	0.918	0.504	0.65	0.536
HYP_C	0.139	0.45	0.213	0.518
HYP_M	0	0	0	0.5

Table 4.16 Significant learning metrics (recall, precision, f-measure, and ROC area) of the SVM-classifier (adjusted parameters).

	APNEA_O	APNEA_C	APNEA_M	HYP_O	HYP_C	HYP_M
APNEA_O	0	7	1	72	11	0
APNEA_C	1	9	2	37	11	0
APNEA_M	0	0	1	20	4	1
HYP_O	1	26	17	620	76	2
HYP_C	0	42	13	440	82	4
HYP_M	0	3	1	27	4	0

Table 4.17 Confusion matrix of the ANN-classifier (adjusted parameters). Vertical axes: true condition. Horizontal axes: predicted condition.

	APNEA_O	APNEA_C	APNEA_M	HYP_O	HYP_C	HYP_M
APNEA_O	0	0	0	77	14	0
APNEA_C	0	2	0	46	12	0
APNEA_M	0	0	0	20	6	0
HYP_O	0	0	0	681	61	0
HYP_C	0	1	0	499	81	0
HYP_M	0	0	0	29	6	0

Table 4.18 Confusion matrix of the SVM-classifier (adjusted parameters). Vertical axes: true condition. Horizontal axes: predicted condition.

5 Evaluation and Conclusion

An evaluation and discussion of the obtained results is hereby given in this chapter. Moreover, key contributions made to the field are presented, followed by several suggestions on how current study may be extended in the future. Finally, a compact summary of this thesis can be found at the end.

5.1 Evaluation

The objective of this study, as defined by the research questions (1.2), is to compare the most frequently used machine learning algorithms in the context of SAHS-event classification. For that purpose, due to empirical characteristics of current research, numerous of experiments have been conducted according to the experimental plan defined in the previous chapter. The obtained results are evaluated in present section.

The experiments have been designed to address several aspects that ensued from research questions. The first issue that needed to be determined was whether to use the absolute value of the feature vectors when composing the input files for the classifiers. For that purpose, all experiments of Phase I have been run twice, with both true- and absolute values. Evaluating Table 4.6, which contains the mean of correctly classified respiratory events given the presence of absolute value, leads to a reasoning that it does not improve the classification. Due to an accuracy difference of 1 % in favor of true values, it has been resolved that usinf the absolute value of the feature vectors negatively affects the output of SAHS-event classification based on utilized classifiers. As a consequence, the experiments of the second phase have been run on true-valued feature vectors.

Secondly, the extent of impact of experimental parameters on classifiers' performance is considered based on the results. The classifiers have been trained on data consisting of respiratory events of correspondingly one and five merged subjects, constituting approximately 10 % and 50 % of all available events in the database. According to Table 4.5, a significant variation in classification accuracy can be revealed caused by the amount of entries in the training file. Increasing the number of entries from 10 to 50 per cent causes the rate of correctly classified events being raised by 7%. Due to current findings it has been decided to utilize half of all available respiratory events during Phase II.

An analysis of results allows to determine which of the pre-processing techniques should be selected in order to improve classification performance, which directly refers to the second research question. Five methods have been proposed according to experimental plan (4.1) of which three are functions of wavelet transform. Studying Figure 12 induces that irrespective of utilization of single or multiple input channels, the Daubechies transform results on average in slightly better classification outcome than Symlet or Haar. This has consequently led to a decision of using Daubechies as the default wavelet transform function during Phase II tests. Furthermore, both the best- and worst-performing wavelet transform function are collated with the remaining pre-processing techniques in Figure 11, where the accuracy in accordance to each input channel is depicted. A closer analysis leads to the inference that the Daubechies wavelet transform should be the favored pre-processing method when exploiting abdomen and ribcage channels. As it turns out in case of airflow channel, variables obtained by statistical analysis of raw signal utilizing Fourier transform results in a better outcome. Additionally, it is interesting to pay attention to Figure 13 while debating on pre-processing procedures, which shows how the accuracy of each classifier is being affected. By averaging classifiers' performance, all three methods result in comparable outcome. However, considering each by itself shows clearly that statistical-based classification significantly improves the ANN-classifier, while the performance of the SVM-classifier decreases. Moreover, the Daubechies transform can be interpreted as the most general pre-processing technique due to approximately equal result of both classifiers.

Results provide a valuable insight into the input channels with equivalent information on how those correlates with the performance of classifiers. Considering Figure 11 an obvious difference in classification accuracy in accordance to the utilized channel is noticeable regardless of pre-processing technique. Current findings can be construed in a way that the performance of classifiers is dependent on input channels. This indicates that methods cannot be directly transferred to other channels without previous consideration of other possibilities. Despite the differing accuracy due to various channels, it is noteworthy that ribcage results in best performance irrespective of pre-processing technique. Taking that into consideration followed to a decision of choosing ribcage as the default input channel for the second phase of experiments. Furthermore, a multichannel composed by merging the two best-performing channels – ribcage and abdomen – has been applied for testing. An average of obtained accuracies regarding single- and multiple channel of various wavelet functions is depicted in

Figure 12. A significant improvement is observable for all transforms where multichannel was applied.

So far the evaluated results have been exclusively based on the accuracy. The rates of correctly classified instances during each phase are given in Table 4.7, Table 4.8 and Table 4.14, with the latter two belonging to Phase II. A quick comparison of the content leads to the conclusion that on average classifiers performs better – raising the accuracy by 7 % – given the adjusted parameters of the second phase. Furthermore, one may argue that according to the aforementioned tables, the SVM-classifier is a relative winner by approaching the rate of correctly classified respiratory events of 50 %. However, the accuracy itself may not be a sufficient measure when comparing two classifiers. For example, in case of unequally distributed training data, the classifier may always prefer the most represented class resulting in high accuracy, simultaneously omitting all remaining classes. This issue is known as the accuracy paradox that results in poor predictive power, consequently making the classifier useless in the given domain. For that reason, other metrics derived from the confusion matrix should be considered.

Phase II has been divided into two runs with correspondingly default and adjusted classification parameters, assuming that the latter would result in significantly better output. However, according to obtained results this assumption turns out to be wrong. Although, as mentioned in the previous paragraph the accuracy of the second run has increased significantly, a closer analysis of additional metrics proves the opposite. Taking into consideration two pairs of tables by collating equivalently Table 4.9 with Table 4.15 and Table 4.10 with Table 4.16, shows that averaging recall, precision, F-measure and ROC area results in higher values during the first run. In the case of an ANN-classifier, all four measures indicates better performance obtained by experiments conducted with default classification parameters, while for SVM-classifier it is true in 3 out of 4 metrics. Due to unsatisfactory results of the second run and a desire to preserve consistency in the current study, the results of first the run are used as a basis for further evaluation of classifiers.

A closer study of confusion matrices of ANN- and SVM-classifier displayed respectively in Table 4.12 and Table 4.13 reveals some important characteristics. Both matrices appear fairly similar and are most conspicuous by the overall poor predictive power. In the best case scenario, the only fields unequal to zero should occur along the diagonal, which is highlighted in green color scale. However, the obtained results deviates from that ideal scenario by numerous of wrongly classified instances. Furthermore, an indication of prioritizing overrepresented classes by both classifiers is noticeable. In case of the SVM-classifier, both *APNEA_O* and *HYP_M* classes are being completely omitted, while *HYP_O* is repeatedly preferred. Many useful metrics can be calculated based on the content of confusion matrices. Four chosen metrics of ANN- and SVM-classifier – recall, precision, F-measure, and ROC area – are presented in Table 4.9 and Table 4.10, respectively. Taking the recall into consideration shows that half of the classes are approaching the value of 0, consequently making the corresponding events to stay undetected. Furthermore, the same pattern is observable for precision that may indicate a large number of false positives. Moreover, the ROC area is almost exclusively higher than 50 % for all classes, indicating that the classifiers will rank a randomly chosen positive instance higher than a negative one. Based on a closer study of all metrics, the classifiers do a weak job separating classes that is an undesired behavior leading to a poor predictive power. However, it may be interesting to compare the metrics of the ANN- and SVM-classifier against each other. An appropriate statistical method for such collation is a paired t-test, which have been calculated for each field with the gained results shown in Table 4.11. Given the null hypothesis H_0 stating that both classifiers perform equally, the P-value is less than α only for the case of ROC area. This allows to reject H_0 and suggest that ANN-classifier performs better in terms of AUC. Due to high P-values of the remaining fields, there is not enough evidence to reject the equivalent hypothesis.

5.2 Discussion

The results have raised several questions that are discussed in this section. The main focus is on what could have been done differently in order to obtain more satisfactory results. For that reason, various parts of the classification process are addressed together with the limitations.

The most crucial part of the classification process is the training phase, where classifiers learns the model based on training entries. This stage has a direct impact on the classifier performance and consequently the classification output. Regarding the utilized algorithms, the training may be time consuming which is acceptable for the current application, while classification itself may be done in real time. Given the obtained results, the performance of the tested classifiers with equivalent predictive power leaves much to be desired. However, comparing the outcome to the probability of guessing a certain class – more precisely by studying if the values of the

ROC area exceeds 0.5 – allows to conclude that while the performance is poor, it is not tragic. As mentioned in the previous section, in order to avoid the accuracy paradox, it is important that the training entries are equally distributed among all classes. However, regarding Figure 10 some classes are overrepresented leading to uneven class prioritization. Thus, one possible improvement would be to level out the number of training entries for each class, by either removing the excess or adding the absent. Considering the fact that the more relevant[4] training entries the better, the latter suggestion should be preferred. As a consequence, one may significantly increase the percentage of all available events to the advantage of training file. Due to the limitation in quantity of raw data with equivalent SAHS-event annotations, this would lead to a decrease of test entries. Furthermore, another issue associated with the training phase is the need for finding appropriate classification parameters, as they may have a tremendous effect on the classifier performance. A variety of different setups have been verified for both classifiers during current research, leading to unsatisfactory results. However, it does not mean that such a setup is nonexistent, but rather imply a need of broader exploration.

In connection of the intent to improve the classification outcome it may be worth to consider the characteristics of SAHS-events. It is significant to recall from subsection 2.1.2 that sleep events may occur as a result of various physiological or psychological dysfunctions. Humans breathe by both thorax and abdomen, but the extent of using each of them tends to diverse among individuals. As a consequence, in case of contact-base monitoring it is important to carefully consider the placement of sensors, or the area of interest in case of noncontact-based solutions. Utilizing only a single sensor increases the risk of omitting events observable in other body parts. Furthermore, the movement of corresponding muscles is possible without any actual inhaling or exhaling. Thus, the use of multimodality is preferred in order to prevent the system to be exposed to ambiguity and to secure exhaustive event recognition. Relating this issue to the conducted experiments of this study confirms that, as stated in the previous section, taking advantage of multichannel improves the classification output. Moreover, identical events may appear differently in the acquired data due to breathing disparity among subjects. A suggested solution would be to increase the amount of training entries significantly by adding new subjects, which presumably would lead to a better classification model. All aforementioned factors together with the wide definition of SAHS-events makes the learning and classification tasks very demanding, which led to unsatisfactory results. Applying the changes mentioned

[4] Not outliers nor causing overfitting.

above would most probably result in a better outcome, however due to the limited scope of this study – in terms of resources and time – they remain as suggestions for future work.

According to the results obtained from experiments conducted during this study, there does not exists any strong evidences to prefer one classifier above another. Although the accuracy of SVM-classifier is ipso facto higher, a closer consideration of confusion matrices and inferred metrics – recall, precision, F-measure and AUC – indicates an even and comparable predictive power on the tested data. There are however noticeable differences in terms of classifiers' performance in accordance to other factors, such as choice of pre-processing or input channel. This phenomenon could be exploited by taking advantage of multiple algorithms simultaneously, thus building a more intelligent decision platform as proposed by (B. Guijarro-Berdiñas, 2012). A recommended starting point given the results from the first phase would be to pre-process the ANN-classifier with statistical analysis, leaving wavelet transform as the more suitable option for the SVM-classifier. Furthermore, based on the experimental outcome, it would be suggested to process ribcage channel with ANN, simultaneously dissuading to classify data from airflow channel by SVM.

5.3 Contributions

Several contributions have been made to the field through this study that are discussed in this section. Some concrete steps in the form of an empirical study has been taken as a result of addressing a broader goal defined in section 1.2. Firstly, due to an adequate coverage of background theory together with an extensive up-to-date literature review, this research may be used as an introduction to the field of Sleep Apnea- Hypopnea Syndrome and the equivalent task of automated sleep event-classification. Additionally, an overview of a monitoring system based on breath-data with corresponding design thoughts have been presented and can serve as a starting point for further development. Finally, a tailor-made software has been developed for the purpose of this research, where two most frequently used algorithms in the context of SAHS-event classification – Artificial Neural Network and Support Vector Machine – has been tested on the same dataset. The results themselves cannot be counted as a significant contribution to the field as they happen to be unsatisfactory and only concern two classification algorithms, however the program with attached database may be useful for further research.

5.4 Future work

The current research could be extended in two separate directions that are presented in the following. The first one involves a direct continuation of the ongoing work by comparing another classification algorithms in the same environment utilizing the same dataset, consequently following to a more thorough results and a broader scope of study. Several changes have been suggested in the discussion section (5.2) with the purpose of improving the prediction power of the already used classifiers. On the other hand, considering the main objective of this study, the obvious choice would be to point future work in the direction of further elaboration of a breath-based monitoring system. That could involve, depending on the available resources, either the focus on other parts of the intended system or a design of a whole platform. Some adequate possibilities that may be worth consideration are; detection of sleep stages, prediction of subject information such as age and gender, tracking the vital signs, recognition of emotions such as pain, depression and anger, detection of fall, and support for patient-caregiver communication and alarm modules. Moreover, as mentioned in section 3.1, a variety of noncontact-based sensors could be utilized, such as video cameras capturing visible and/or invisible light, thermal cameras, audio recorders and pressure mats. There exists numerous of studies exploring each of the aforementioned issues that combined may form a desired monitoring system.

5.5 Conclusion

This section provides a summary that concludes the empirical study described in this paper. An increase in elderly population together with a rapidly growing field of computer science were the main motivating factors to proceed and delve into the issue of automated patient monitoring. It was decided to address a very interesting aspect related to extraction of the information based on breathing. A broad goal with following research questions (1.2) was consequently defined pointing the study in the direction of sleep disorder classification. Some general reflections regarding the intended monitoring system have been presented, however considering the scope of this study, only a few steps towards the goal have in fact been taken. In order to acquire an adequate knowledge, a comprehensive literature review has been conducted according to a previously prepared structured review protocol (2.2). In the absence of appropriate equipment, laboratory and test subjects, it has been decided to obtain necessary data from a free access database (3.3). Furthermore, a specific software has been developed for the purpose of pre-processing of raw data samples that creates an appropriate input for the classifiers. As a result of the literature review, it has been determined to compare two most frequently used classification algorithms in the context of Sleep Apnea-Hypopnea Syndrome, utilizing open source software Weka. A variety of experiments have been conducted in accordance to experimental plan (4.1) testing the performance of Artificial Neural Network and Support Vector Machine through the classification of SAHS-events. Despite a precise abidance with the guidelines inferred from the focus-collection papers, the obtained outcome differed significantly implying unsatisfactory results. Consequently, several improvements with proposed directions of further research have been suggested (5.2, 5.4) leaving the topic open with none grand conclusions drawn.

Bibliography

B. Guijarro-Berdiñas, E. H.-P. D. P.-B., 2012. A mixture of experts for classifying sleep apneas. *Expert Systems with Applications,* 39(8).

C. Brüser, C. A. T. W. M. W. S. L., 2015. Ambient and unobtrusive cardiorespiratory monitoring techniques. *IEEE Reviews in Biomedical Engineering.*

D. Alvares, R. H. J. M. F. d. C., 2012. Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis. *Medical Engineering and Physics,* 34(8).

D. Alvarez-Estevez, V. M.-B., 2015. Computer-Assisted Diagnosis of the Sleep Apnea-Hypopnea Syndrome : A Review. *Hindawi Publishing Corporation, Sleep Disorders.*

F.Q. AL-Khalidi, R. S. D. B. H. E. S. T., 2011. Respiration Rate Monitoring Methods: A Review. *Wiley-Liss, Inc..*

G.C. Mbata, J. C., 2012. Obstructive Sleep Apnea Hypopnea Syndrome. *Annals of Medical and Health Sciences Research.*

M. Emin Tagluk, M. A. N. S., 2010. Classıfication of sleep apnea by using wavelet transform and artificial neural networks. *Expert Systems with Applications,* 37(2).

M. H. Silber, S. A.-I. M. H. B. S. C. M. M. G.-D. M. S. K. S. K. M. K. T. P. M. P. C. I., 2007. The Visual Scoring of Sleep in Adults. *Journal of Clinical Sleep Medicine,* 3(2).

M. Tagluk, N. S., 2010. Classification of sleep apnea through sub-band energy of abdominal effort signal using wavelets + neural networks. *Journal of Medical Systems,* 34(6).

N. Sezgin, M. E. T., 2009. Energy based feature extraction for classification of sleep apnea syndrome. *Computers in Biology and Medicine,* 39(11).

O. Fontenla-Romero, B. G.-B. A. A.-B. V. M.-B., 2005. A new method for sleep apnea classification using wavelets and feedforward neural networks. *Artificial Intelligence in Medicine,* 34(1).

P. Rashidi, A. M., 2013. A survey on ambient-assisted living tools for older adults. *IEEE Journal of Biomedical and Health Informatics,* 17(3).

R. B. Berry, R. B. D. J. G. D. G. C. I. V. K. K. C. L. M. R. M. S. P. S. F. Q. S. R. K. P. S. S. L. D. W. M. M. T., 2012. Rules for Scoring Respiratory Events in Sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *Journal of Clinical Sleep Medicine.*

S. B. Maind, P. W., 2014. Research Paper on Basic of Artificial Neural Network. *International Journal on Recent and Innovation Trends in Computing and Communication,* 2(1).

S. Sathyanarayana, R. K. S. S. S. S. T., 2015. Vision-based patient monitoring: a comprehensive review of algorithms and technologies. *Journal of Ambient Intelligence and Humanized Computing.*

Thorpy, M. J., 2012. Classification of Sleep Disorders. *Neurotherapeutics.*

V. Somers, D. W. R. A. W. A. F. C. A. C. S. D. J. F. C. H. L. O. T. P. R. R. M. W. T. Y., 2008. Sleep Apnea and Cardiovascular Disease. *Circulation*, 118(10).

V. Vapnik, C. C., 1995. Support-Vector Networks. *AT&T Bell Labs., Machine Learning, Kluwer Academic Publishers*.

W. McNicholas, L. Doherty, S. Ryan, J. Garvey, P. Boyle, E. Chua, n.d. *Vincent's University Hospital Sleep Disorders Clinic - Sleep apnea database.* [Online] Available at: https://physionet.org/physiobank/database/ucddb/#acknowledgments

Y. Maali, A. A.-J. L. L., 2012. Self-advising SVM for sleep apnea classification. *CEUR Workshop Proceedings,* Volum 944.

Appendix A: Results

Phase I

Experiment ID	**Correctly classified instances**
	Accuracy [%] (cross validation)
Haar - Single input channel	
ann0_b1_cA_Haar_abs0	35,8
ann0_b1_cA_Haar_abs1	31,64
svm0_b1_cA_Haar_abs0	42,96
svm0_b1_cA_Haar_abs1	37,18
ann0_b5_cA_Haar_abs0	41,83
ann0_b5_cA_Haar_abs1	40,64
svm0_b5_cA_Haar_abs0	42,82
svm0_b5_cA_Haar_abs1	42,49
ann0_b1_cR_Haar_abs0	36,03
ann0_b1_cR_Haar_abs1	36,03
svm0_b1_cR_Haar_abs0	43,19
svm0_b1_cR_Haar_abs1	38,34
ann0_b5_cR_Haar_abs0	47,32
ann0_b5_cR_Haar_abs1	42,75
svm0_b5_cR_Haar_abs0	47,12
svm0_b5_cR_Haar_abs1	43,28
ann0_b1_cF_Haar_abs0	38,11
ann0_b1_cF_Haar_abs1	34,64
svm0_b1_cF_Haar_abs0	27,25
svm0_b1_cF_Haar_abs1	25,87
ann0_b5_cF_Haar_abs0	40,64
ann0_b5_cF_Haar_abs1	41,43
svm0_b5_cF_Haar_abs0	38,72
svm0_b5_cF_Haar_abs1	38,32
Haar - Multiple input channel	
ann0_b1_cAR_Haar_abs0	36,03

ann0_b1_cAR_Haar_abs1	34,87
svm0_b1_cAR_Haar_abs0	44,57
svm0_b1_cAR_Haar_abs1	39,03
ann0_b5_cAR_Haar_abs0	48,05
ann0_b5_cAR_Haar_abs1	44,93
svm0_b5_cAR_Haar_abs0	51,16
svm0_b5_cAR_Haar_abs1	46,79
Daubechies4 - Single input channel	
ann0_b1_cA_D4_abs0	37,88
ann0_b1_cA_D4_abs1	39,26
svm0_b1_cA_D4_abs0	45,5
svm0_b1_cA_D4_abs1	39,49
ann0_b5_cA_D4_abs0	43,48
ann0_b5_cA_D4_abs1	43,02
svm0_b5_cA_D4_abs0	44,14
svm0_b5_cA_D4_abs1	44,08
ann0_b1_cR_D4_abs0	43,19
ann0_b1_cR_D4_abs1	41,57
svm0_b1_cR_D4_abs0	44,8
svm0_b1_cR_D4_abs1	43,19
ann0_b5_cR_D4_abs0	46,13
ann0_b5_cR_D4_abs1	46,26
svm0_b5_cR_D4_abs0	46,39
svm0_b5_cR_D4_abs1	46,39
ann0_b1_cF_D4_abs0	33,26
ann0_b1_cF_D4_abs1	39,72
svm0_b1_cF_D4_abs0	27,02
svm0_b1_cF_D4_abs1	27,71
ann0_b5_cF_D4_abs0	41,5
ann0_b5_cF_D4_abs1	42,03
svm0_b5_cF_D4_abs0	38,72
svm0_b5_cF_D4_abs1	38,25
Daubechies4 - Multiple input channel	

ann0_b1_cAR_D4_abs0	42,57
ann0_b1_cAR_D4_abs1	37,88
svm0_b1_cAR_D4_abs0	46,19
svm0_b1_cAR_D4_abs1	43,42
ann0_b5_cAR_D4_abs0	47,92
ann0_b5_cAR_D4_abs1	46,59
svm0_b5_cAR_D4_abs0	50,43
svm0_b5_cAR_D4_abs1	48,78
Symlet7 - Single input channel	
ann0_b1_cA_S7_abs0	39,95
ann0_b1_cA_S7_abs1	36,72
svm0_b1_cA_S7_abs0	45,03
svm0_b1_cA_S7_abs1	37,64
ann0_b5_cA_S7_abs0	42,89
ann0_b5_cA_S7_abs1	42,89
svm0_b5_cA_S7_abs0	44,8
svm0_b5_cA_S7_abs1	43,61
Symlet7 - Multiple input channel	
ann0_b1_cAR_S7_abs0	37,64
ann0_b1_cAR_S7_abs1	35,57
svm0_b1_cAR_S7_abs0	46,19
svm0_b1_cAR_S7_abs1	44,57
ann0_b5_cAR_S7_abs0	47,19
ann0_b5_cAR_S7_abs1	48,71
svm0_b5_cAR_S7_abs0	51,69
svm0_b5_cAR_S7_abs1	51,37
Single input stats FFT	
ann0_b1_cA_stats_abs0	45,03
ann0_b1_cA_stats_abs1	42,03
svm0_b1_cA_stats_abs0	26,33
svm0_b1_cA_stats_abs1	35,56
ann0_b5_cA_stats_abs0	52,28
ann0_b5_cA_stats_abs1	52,35

svm0_b5_cA_stats_abs0	38,25
svm0_b5_cA_stats_abs1	38,19
ann0_b1_cR_stats_abs0	49,19
ann0_b1_cR_stats_abs1	48,96
svm0_b1_cR_stats_abs0	26,33
svm0_b1_cR_stats_abs1	26,33
ann0_b5_cR_stats_abs0	54,4
ann0_b5_cR_stats_abs1	53,08
svm0_b5_cR_stats_abs0	38,39
svm0_b5_cR_stats_abs1	38,72
ann0_b1_cF_stats_abs0	44,8
ann0_b1_cF_stats_abs1	44,57
svm0_b1_cF_stats_abs0	24,71
svm0_b1_cF_stats_abs1	24,71
ann0_b5_cF_stats_abs0	49,11
ann0_b5_cF_stats_abs1	48,58
svm0_b5_cF_stats_abs0	38,65
svm0_b5_cF_stats_abs1	38,65
Single input stats DWT	|Best DWT: Daubechies
ann0_b1_cA_stats_abs0	45,5
ann0_b1_cA_stats_abs1	43,65
svm0_b1_cA_stats_abs0	29,56
svm0_b1_cA_stats_abs1	30,72
ann0_b5_cA_stats_abs0	48,84
ann0_b5_cA_stats_abs1	46,86
svm0_b5_cA_stats_abs0	37,99
svm0_b5_cA_stats_abs1	37,59
ann0_b1_cR_stats_abs0	46,88
ann0_b1_cR_stats_abs1	49,88
svm0_b1_cR_stats_abs0	27,02
svm0_b1_cR_stats_abs1	27,25
ann0_b5_cR_stats_abs0	51,22
ann0_b5_cR_stats_abs1	51,1

svm0_b5_cR_stats_abs0	38,32
svm0_b5_cR_stats_abs1	37,99
ann0_b1_cF_stats_abs0	41,11
ann0_b1_cF_stats_abs1	42,03
svm0_b1_cF_stats_abs0	25,87
svm0_b1_cF_stats_abs1	25,64
ann0_b5_cF_stats_abs0	45,2
ann0_b5_cF_stats_abs1	45,73
svm0_b5_cF_stats_abs0	38,58
svm0_b5_cF_stats_abs1	38,58

Table A 1 Phase I results in terms of accuracy.

	Abdomen avg [%]	**Ribcage avg [%]**	**Flow avg [%]**
|Haar	39,42	41,7575	35,6225
|Daubechies	42,10625	44,74	36,02625
|stats FFT	41,2525	41,925	39,2225
|stats DWT	40,08875	41,2075	37,8425
|all	40,716875	42,4075	37,1784375

Table A 2 The accuracy in the context of input channels and pre-processing techniques.

	DWT avg |A	**DWT avg |AR**	**DWT avg |A & AR**
|Haar	39,42	43,17875	41,299375
|Daubechies	42,10625	45,4725	43,789375
|Symlet	41,69125	45,36625	43,52875

Table A 3 The accuracy in the context of pre-processing techniques and the use of multichannel.

	Daubechies avg	**Stats DWT avg**	**Stats FFT avg**
|ANN	41,44166667	46,5	48,69833333
|SVM	40,47333333	32,92583333	32,90166667
|all	40,9575	39,71291667	40,8

Table A 4 The accuracy in the context of the classifiers and the pre-processing techniques.

abs0 avg	abs1 avg	b1 avg	b5 avg	ann0 avg	svm0 avg
41,5571875	40,5578125	37,674375	44,440625	43,3896875	38,7253125

Table A 5 The overall accuracy in the context of use of absolute value, training data composition and the classifiers.

Phase II

Test subject ID_algorithm	Correctly classified instances
	Accuracy [%]
ANN: Default algorithm setup	
2_ann0	45,97
5_ann0	47,37
7_ann0	43,84
8_ann0	15,79
9_ann0	18,42
11_ann0	37,14
12_ann0	49,01
14_ann0	46,15
15_ann0	37,84
17_ann0	47,06
19_ann0	31,73
20_ann0	38,36
21_ann0	53,75
22_ann0	22,22
23_ann0	48,69
24_ann0	25,32
26_ann0	46,43
SVM: Default algorithm setup	
2_svm0	46,77
5_svm0	49,12
7_svm0	47,95
8_svm0	52,63

9_svm0	17,11
11_svm0	42,86
12_svm0	47,02
14_svm0	43,96
15_svm0	48,65
17_svm0	42,65
19_svm0	37,5
20_svm0	43,84
21_svm0	55
22_svm0	25,93
23_svm0	42,41
24_svm0	27,27
26_svm0	48,81

Table A 6 Phase II results in terms of accuracy.

Run 1 ANN: Default setup

=== Evaluation on test set ===

=== Summary ===

Correctly Classified Instances 631 41.1075 %

Incorrectly Classified Instances 904 58.8925 %

Kappa statistic 0.0942

Mean absolute error 0.2126

Root mean squared error 0.338

Relative absolute error 90.7158 %

Root relative squared error 101.807 %

Total Number of Instances 1535

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0	0	0	0	0.511	APNEA_O
0.733	0.223	0.118	0.733	0.203	0.788	APNEA_C
0	0.003	0	0	0	0.606	APNEA_M
0.612	0.506	0.531	0.612	0.569	0.584	HYP_O
0.225	0.164	0.456	0.225	0.302	0.579	HYP_C
0.057	0.009	0.125	0.057	0.078	0.532	HYP_M
0.411	0.315	0.437	0.411	0.399	0.585	Weighted Avg.

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
0	27	0	49	14	1	| a = APNEA_O
0	44	0	10	3	3	| b = APNEA_C
0	8	0	11	6	1	| c = APNEA_M
0	154	2	454	126	6	| d = HYP_O
0	138	2	307	131	3	| e = HYP_C
0	2	0	24	7	2	| f = HYP_M

Run 1 SVM: Default setup

=== Evaluation on test set ===

=== Summary ===

Correctly Classified Instances 643 41.8893 %

Incorrectly Classified Instances 892 58.1107 %

Kappa statistic 0.0852

Mean absolute error 0.1937

Root mean squared error 0.4401

Relative absolute error 82.6592 %

Root relative squared error 132.58 %

Total Number of Instances 1535

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0	0	0	0	0.5	APNEA_O
0.6	0.191	0.113	0.6	0.19	0.704	APNEA_C
0	0.005	0	0	0	0.498	APNEA_M
0.604	0.509	0.526	0.604	0.562	0.547	HYP_O
0.274	0.209	0.444	0.274	0.339	0.533	HYP_C
0	0	0	0	0	0.5	HYP_M
0.419	0.333	0.427	0.419	0.407	0.543	Weighted Avg.

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
0	18	1	46	26	0	| a = APNEA_O
0	36	0	18	6	0	| b = APNEA_C
0	7	0	14	5	0	| c = APNEA_M
0	138	2	448	154	0	| d = HYP_O
0	118	4	300	159	0	| e = HYP_C
0	1	0	26	8	0	| f = HYP_M

Run 2 ANN: Adjusted setup

MultilayerPerceptron -L 0.05 -M 0.95 -N 50000 -V 0 -S 0 -E 10 -H "15, 15" -G -B -C -R

LibSVM wrapper, original code by Yasser EL-Manzalawy (= WLSVM)

Time taken to build model: 222.18 seconds

=== Evaluation on test set ===

=== Summary ===

Correctly Classified Instances 712 46.3844 %

Incorrectly Classified Instances 823 53.6156 %

Kappa statistic 0.0559

Mean absolute error 0.2159

Root mean squared error 0.3707

Relative absolute error 92.1349 %

Root relative squared error 111.6765 %

Total Number of Instances 1535

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0	0.001	0	0	0	0.486	APNEA_O
0.15	0.053	0.103	0.15	0.122	0.653	APNEA_C
0.038	0.023	0.029	0.038	0.033	0.569	APNEA_M
0.836	0.752	0.51	0.836	0.633	0.58	HYP_O
0.141	0.111	0.436	0.141	0.213	0.514	HYP_C
0	0.005	0	0	0	0.5	HYP_M
0.464	0.408	0.416	0.464	0.392	0.55	Weighted Avg.

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
0	7	1	72	11	0	| a = APNEA_O
1	9	2	37	11	0	| b = APNEA_C
0	0	1	20	4	1	| c = APNEA_M
1	26	17	620	76	2	| d = HYP_O
0	42	13	440	82	4	| e = HYP_C
0	3	1	27	4	0	| f = HYP_M

Run 2 SVM: Adjusted setup

LibSVM -S 0 -K 2 -D 3 -G 0.005 -R 0.0 -N 0.5 -M 40.0 -C 1.5 -E 0.001 -P 0.1 -seed 1

=== Classifier model (full training set) ===

Time taken to build model: 0.33 seconds

=== Evaluation on test set ===

=== Summary ===

Correctly Classified Instances 764 49.772 %

Incorrectly Classified Instances 771 50.228 %

Kappa statistic 0.0519

Mean absolute error 0.1674

Root mean squared error 0.4092

Relative absolute error 71.4465 %

Root relative squared error 123.2601 %

Total Number of Instances 1535

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.033	0.001	0.667	0.033	0.063	0.516	APNEA_O
0	0	0	0	0	0.5	APNEA_C
0.918	0.846	0.504	0.918	0.65	0.536	APNEA_M
0.139	0.104	0.45	0.139	0.213	0.518	HYP_O
0	0	0	0	0	0.5	HYP_C
0.498	0.448	0.44	0.498	0.397	0.525	HYP_M
0.033	0.001	0.667	0.033	0.063	0.516	Weighted Avg.

=== Confusion Matrix ===

a	b	c	d	e	f	<-- classified as
0	0	0	77	14	0	| a = APNEA_O
0	2	0	46	12	0	| b = APNEA_C
0	0	0	20	6	0	| c = APNEA_M
0	0	0	681	61	0	| d = HYP_O
0	1	0	499	81	0	| e = HYP_C
0	0	0	29	6	0	| f = HYP_M

Paired t-test

t-Test: Paired Two Sample for Means

	Recall ANN	*Recall SVM*
Mean	0,271166667	0,246333333
Variance	0,104903767	0,087162267
Observations	6	6
Pearson Correlation	0,983688956	
Hypothesized Mean Difference	0	
df	5	
t Stat	0,969016793	
P(T<=t) one-tail	0,188520908	
t Critical one-tail	2,015048373	
P(T<=t) two-tail	**0,377041816**	
t Critical two-tail	2,570581836	

t-Test: Paired Two Sample for Means

	Precision ANN	*Precision SVM*
Mean	0,205	0,1805
Variance	0,0534592	0,0582199
Observations	6	6
Pearson Correlation	0,97901027	
Hypothesized Mean Difference	0	
df	5	
t Stat	1,214047106	
P(T<=t) one-tail	0,139469475	
t Critical one-tail	2,015048373	
P(T<=t) two-tail	**0,27893895**	
t Critical two-tail	2,570581836	

t-Test: Paired Two Sample for Means

	F-measure ANN	*F-measure SVM*
Mean	0,192	0,181833333
Variance	0,0482148	0,053696967
Observations	6	6
Pearson Correlation	0,987592808	
Hypothesized Mean Difference	0	
df	5	
t Stat	0,663161015	
P(T<=t) one-tail	0,268286688	
t Critical one-tail	2,015048373	
P(T<=t) two-tail	**0,536573375**	
t Critical two-tail	2,570581836	

t-Test: Paired Two Sample for Means

	ROC area ANN	*ROC area SVM*
Mean	0,6	0,547
Variance	0,0097244	0,0063328
Observations	6	6
Pearson Correlation	0,940303855	
Hypothesized Mean Difference	0	
df	5	
t Stat	3,60174937	
P(T<=t) one-tail	0,007757315	
t Critical one-tail	2,015048373	
P(T<=t) two-tail	**0,01551463**	
t Critical two-tail	2,570581836	

Appendix B: PhysioNet

Name

rdsamp - read WFDB signal files

Synopsis

rdsamp -r *record* [*options* ...]

Description

rdsamp reads signal files for the specified *record* and writes the samples as decimal numbers on the standard output. If no *options* are provided, **rdsamp** starts at the beginning of the record and prints all samples. By default, each line of output contains the sample number and samples from each signal, beginning with channel 0, separated by tabs.

Options include:

-c

Produce output in CSV (comma-separated value) format (default: write output in tab-separated columns).

-f *time*

Begin at the specified *time*. By default, **rdsamp** starts at the beginning of the record.

-h

Print a usage summary.

-H

Read the signal files in high-resolution mode (default: standard mode). These modes are identical for ordinary records. For multifrequency records, the standard decimation of oversampled signals to the frame rate is suppressed in high-resolution mode (rather, all other signals are resampled at the highest sampling frequency).

-l *interval*

Limit the amount of output to the specified time *interval* (in standard time format; default: no limit). If both **-l** and **-t** are used, **rdsamp** stops at the earlier of the two limits.

-p

Print times in seconds and milliseconds, and values in physical units. By default, **rdsamp** prints times in sample intervals and values in A/D units.

-P

Same as **-p**, but yields higher precision in the sample values (8 decimal places rather than 3).

A single character can be attached to either **-p** or **-P** to choose the

format for the printed times in the first column of output. The choices are:

-pd (or **-Pd**)

Print time of day and date if known, as [hh:mm:ss DD/MM/YYYY]. The base time and date must appear in the header file for the record; otherwise, this format is equivalent to "e" format (below).

-pe (or **-Pe**)

Print the elapsed time from the beginning of the record, as hh:mm:ss.

-ph (or **-Ph**)

Print the elapsed time in hours.

-pm (or **-Pm**)

Print the elapsed time in minutes.

-ps (or **-Ps**)

Print the elapsed time in seconds. This is the default format when using **-p** or **-P**.

-pS (or **-PS**)

Print the elapsed time in sample intervals.

-s *signal-list*

Print only the signals named in the *signal-list* (one or more input signal numbers or names, separated by spaces; default: print all signals). This option may be used to re-order or duplicate signals.

-S *signal*

Search for the first valid sample of the specified *signal* (a signal name or number) at or following the time specified with **-f** (or the beginning of the record if the **-f** option is not present), and begin printing at that time.

-t *time*

Stop at the specified *time*. By default, **rdsamp** stops at the end of the record.

-v

Print column headings (signal names on the first line, units on the second). The names of some signals are too wide to fit in the columns; such names are shortened by omitting the initial characters (since names of related signals often differ only at the end, this helps to make the columns identifiable). Names of units are shortened when necessary by omitting the final characters, since the initial characters are usually most important for distinguishing different units.

-X

Produce output in WFDB-XML format (same as the CSV format produced using the **-c** option, but wrapped within an XML header and trailer). This format is recognized and parsed automatically by **wrsamp**.

Environment

It may be necessary to set and export the shell variable **WFDB** (see **setwfdb**(1)).

Availability

This program is provided in the *app* directory of the WFDB Software Package. Run **make** in that directory to compile and install it if it have not been installed already.

The PhysioNet ATM (http://physionet.org/cgi-bin/ATM) provides web access to **rdsamp** (select **Show samples as text** from the Toolbox).

See Also

rdann(1) , **setwfdb**(1) , **wrsamp**(1)

Author

George B. Moody (george@mit.edu)

Source

http://www.physionet.org/physiotools/wfdb/app/rdsamp.c