Automated Detection of Central Apnea in Preterm Infants

Marina-Marinela Nano
AUTOMATED DETECTION OF CENTRAL APNEA IN PRETERM INFANTS

by

Marina-Marinela Nano

in partial fulfillment of the requirements for the degree of

Master of Science
in Electrical Engineering

at the Delft University of Technology,
to be defended publicly on Tuesday October 27, 2015 at 2:00 PM.

Supervisor: Dr. ir. Richard Heusdens, TU Delft
Thesis committee: Dr. ir. Richard Heusdens, TU Delft
Dr. ir. Richard Hendriks, TU Delft
Prof. dr. ir. Geert Leus, TU Delft
Dr. Peter Andriessen, Máxima Medisch Centrum
Dr. Carola van Pul, Máxima Medisch Centrum

This thesis is confidential and cannot be made public until September 31, 2015.

An electronic version of this thesis is available at http://repository.tudelft.nl/.
In 2010, an estimated 14.9 million babies were born preterm, which amounted to 11.1% of all livebirths worldwide, ranging from about 5% in several European countries to 18% in some African countries. The rate of preterm births has increased remarkably. Prematurity of birth can predispose neonates to undesirable cessations of breathing, a condition termed as Apnea of Prematurity. The prevalence of this condition poses problems, because when untreated or inadequately treated Apnea of Prematurity, may impair development. This thesis investigates the automated central apnea detection in preterm infants based on raw waveform analysis of one-lead ECG and chest impedance signals. For this purpose, 18 novel features and 34 features of existing research that characterize different aspects of chest impedance and ECG signals were extracted for automated apnea classification. Features aim to extract information regarding respiratory and cardiac regularity, estimated from chest impedance and ECG signals. These features are indicators of some properties of cardio-respiratory physiology, which is not independent of the presence of apnea and thus can be in turn used to classify apnea. The objective is to find the most discriminative subset of features from one-lead ECG and chest impedance signals that can be used by a machine-learning approach to study and accurately detect central apnea. This was achieved by applying feature selection algorithms in order to remove redundant or irrelevant features without incurring much loss of information. In this thesis, nine hours of continuously recorded data of ten very low-birth-weight infants (birth weight < 1,500 gr) undergoing continuous cardiopulmonary monitoring in the NICU at Maxima Medisch Centrum from 2008 were included in the analysis. The dataset was annotated by two neonatologists. Results from this work indicate that the analysis of chest impedance and ECG signals with a support vector machine can automatically detect Apnea of Prematurity.
ACKNOWLEDGMENTS

Though only my name appears on the cover of this thesis, a great many people have contributed to its production.

I want to express my sincere gratitude to my supervisors dr. Peter Andriessen and dr. Carola van Pul, from Máxima Medisch Centrum, for the trust, the insightful discussion, for your support during the whole period of this study, and especially for your patience and guidance during the writing process. I would like to acknowledge dr. S Bambang Oetomo for the time and effort for annotating the dataset and enriching my knowledge in the area of apnea in prematurity.

I would like to express my deepest gratitude to my supervisor, Associate Professor dr. Richard Heusdens, for supporting me during these past two years. His motivation, encouragement and excellent guidance provided me with an excellent atmosphere for doing research. Richard has been always very supportive and has given me the freedom to pursue various ideas without objection. I would like to thank the rest of my thesis committee. Assistant Professor Dr. Richard Hendriks for his co-supervision and encouragements. Professor dr. Geert Leus for taking the time to evaluate my thesis.

I would also like to thank Rohan Joshi for his help during the past nine months. I am also thankful to him for reading my report, commenting on my views and helping me understand and enrich my ideas.

Most importantly, none of this would have been possible without the support and patience of my family and friends. I have to give a special mention for the support given by my friend Vana Panagiotou throughout the last two years.

Finally, I appreciate the financial and moral support from Hatzikyriakeio Child Care Institution during the past seven years of my studies. Without their help, none of my educational achievements could have been possible.

Marina-Marinela Nano
Delft, September 2015
CONTENTS

List of Figures ix

List of Tables xv

1 Introduction 1

2 Background of central apnea and apnea detection techniques 5
   2.1 Clinical presentation of Central Apnea (CA) . . . . . . . . . . . . . . . . . . . . . 5
   2.2 Apnea, Bradycardia and Desaturation (ABD) . . . . . . . . . . . . . . . . . . . . . 7
   2.3 Review of techniques for detecting apnea . . . . . . . . . . . . . . . . . . . . . . 9
   2.4 Difficulties of detecting central apnea . . . . . . . . . . . . . . . . . . . . . . . 11

3 Methodology 12
   3.1 Overview of physiological signals. . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
   3.2 Description of the dataset . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
   3.3 Description of the Annotation System Software . . . . . . . . . . . . . . . . . . . 14
   3.4 Feature Generation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
      3.4.1 Signal Processing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
      3.4.2 Existing features . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
         3.4.2.1 Respiratory features . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
         3.4.2.2 Cardiac features . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
      3.4.3 New features . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
         3.4.3.1 Respiratory features . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
         3.4.3.2 Cardio-respiratory features . . . . . . . . . . . . . . . . . . . . . . . 29
   3.5 Feature Visualization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
   3.6 Classification technique . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
      3.6.1 Support Vector Machine . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
      3.6.2 Feature selection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6.3</td>
<td>Applied approach</td>
<td>37</td>
</tr>
<tr>
<td>3.7</td>
<td>Evaluation</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>Results</td>
<td>42</td>
</tr>
<tr>
<td>4.1</td>
<td>General classification of dataset by annotators</td>
<td>42</td>
</tr>
<tr>
<td>4.2</td>
<td>Based on respiratory features</td>
<td>43</td>
</tr>
<tr>
<td>4.3</td>
<td>Based on cardiac features</td>
<td>48</td>
</tr>
<tr>
<td>4.4</td>
<td>Based on respiratory and cardiac features</td>
<td>49</td>
</tr>
<tr>
<td>4.5</td>
<td>Impact of epoch duration and percentage of apnea per epoch</td>
<td>51</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Respiratory features</td>
<td>51</td>
</tr>
<tr>
<td>4.6</td>
<td>Comparison</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>Discussion and Conclusions</td>
<td>57</td>
</tr>
<tr>
<td>A</td>
<td>Appendix</td>
<td>60</td>
</tr>
<tr>
<td>A.1</td>
<td>Feature visualization</td>
<td>60</td>
</tr>
<tr>
<td>A.2</td>
<td>Classification results based on cardiac features</td>
<td>78</td>
</tr>
<tr>
<td>A.3</td>
<td>Impact of epoch duration and percentage of apnea per epoch based on cardiac features</td>
<td>82</td>
</tr>
</tbody>
</table>

**Bibliography** | 84
LIST OF FIGURES

1.1 Three minute segments of chest impedance signal .......................... 2

2.1 Relation of apnea, bradycardia and oxygen desaturation. Bradycardia and oxygen desaturation are immediate effects of apnea. However, these conditions also inhibit respiratory drive, which may either cause or worsen apnea. Reproduced from Martin et al. [1] ................................. 8

3.1 Electrocardiogram .............................................................. 12
3.2 Photoplethysmography ....................................................... 13
3.3 Impedance pneumography .................................................. 13
3.4 Annotation system ............................................................. 15
3.5 Annotation system. Cursor mode enabled ............................... 16
3.6 Annotation system. One apnea event annotated .................... 16
3.7 Annotation system. Multiple apnea events annotated ............ 17
3.8 Figure (up): shows the raw chest impedance which includes monitor artifacts. Figure (down): shows the pre-processed (orange) after removing monitor artifact .......................................................... 18
3.9 Example of subject-dependence removal from RR interval series ...... 19
3.10 Chest impedance and the locations of the detected peaks and troughs .... 25
3.11 Sampling of chest impedance by using R-peaks as time unit .......... 29
3.12 Relation between RR Interval and IBI during apnea episodes and normal breathing .......................................................... 30
3.13 Regularity of breathing cycle (R30). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. .......................................................... 32
3.14 Discriminative power of all respiration features (existing and new) computed by using ASMD. ................................................. 33
3.15 Discriminative power of all cardiac features (existing and new) computed by using ASMD. .................................................. 33
3.16 Description of the basic idea of SVMs in a two-dimensional space, where the two classes “-1” and “+1” can be separated by the optimal hyperplane (with support vectors) using two features $x_1$ and $x_2$. .......................... 34

3.17 Overview of the apnea detection system ................................................. 38

3.18 An annotated segment of apnea can occur in the beginning, middle, end of the epoch as well as it might be equated to the whole epoch. .......................... 38

3.19 Subject-Specific and Subject-Independent approaches .............................. 39

4.1 ROC and PR curves for both Subject-Independent and Subject-Specific for CFS algorithm .......................................................... 44

4.2 ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm .......................................................... 46

4.3 ROC and PR curves for both Subject-Independent and Subject-Specific for CFS algorithm .......................................................... 48

4.4 ROC and PR curves for both Subject-Independent and Subject-Specific for CFS algorithm .......................................................... 50

4.5 Apnea detection performance of subject-independent approach for varying epoch duration .......................................................... 52

4.6 Apnea detection performance of subject-independent approach for varying apnea percentage per epoch ............................................ 52

4.7 ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm .......................................................... 53

4.8 ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm .......................................................... 55

4.9 Histograms of apnea and non-apnea classes and mean value of apnea epoch and of nine epochs preceding and following apnea .................. 56

A.1 Spectral power of respiratory frequency (R1). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$. .................................................. 60

A.2 Spectral power in VLF (R2). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$. .................................................. 60

A.3 Spectral power in LF (R3). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$. .................................................. 61

A.4 Spectral power in HF (R4). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$. .................................................. 61

A.5 Ratio of LF and HF bands (R5). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$. .................................................. 61
A.6 Respiratory frequency estimated in the frequency domain (R6). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ................................. 62

A.7 Respiratory frequency estimated in the time domain (R7). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 62

A.8 Breath-by-breath correlation (R8). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 62

A.9 Envelope Power (R9). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 63

A.10 Breath Length Variation (R10). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 63

A.11 Sample entropy of respiratory effort (R11). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 63

A.12 Standardized median of respiratory peaks (R12). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 64

A.13 Standardized median of respiratory troughs (R13). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 64

A.14 Respiratory peak regularity measured by sample entropy (R14). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 64

A.15 Respiratory trough regularity measured by sample entropy (R15). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.001 \). ....................................................... 65

A.16 Median of peak-to-trough amplitude sequence (R16). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.0001 \). ....................................................... 65

A.17 Ratio of respiratory peak and trough regularity measured by sample entropy (R17). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.001 \). ....................................................... 65

A.18 Respiratory regularity of breath by breath duration (R18). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at \( p < 0.001 \). ....................................................... 66
A.19 Correlation of peak-trough breath duration (R19). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001. ................................................................. 66

A.20 Regularity of inhalation-exhalation cycle (R20). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001. ................................................................. 66

A.21 Correlation of inhalation-exhalation cycle (R21). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001. ................................................................. 67

A.22 Median of inhalation time period (R22). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 67

A.23 Ratio of median value of inhalation and exhalation time periods (R23). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 68

A.24 IQR of depth of breathing based on peaks distribution R(24). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 68

A.25 IQR of depth of breathing based on troughs distribution (R25). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 68

A.26 IQR of breathing cycle based on peaks distribution (R26). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 69

A.27 IQR of breathing cycle based on troughs distribution (R27). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001. ................................................................. 69

A.28 IQR of inhalation cycle distribution (R28). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 69

A.29 IQR of exhalation cycle distribution (R29). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 70

A.30 Regularity of inhalation-exhalation depth (R31). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001. ................................................................. 70

A.31 RR LF band (C1). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001. ................................................................. 70
A.32 RR HF band (C2). The significance of difference (5%) was found between the
two classes using an unpaired Mann–Whitney test at p < 0.001. .......... 71
A.33 RR standard deviation (C3). The significance of difference (5%) was found
between the two classes using an unpaired Mann–Whitney test at p < 0.001. . 71
A.34 RR respiratory frequency (C4). The significance of difference (5%) was found
between the two classes using an unpaired Mann–Whitney test at p < 0.001. . 71
A.35 RR respiratory power (C5). The significance of difference (5%) was found be-
tween the two classes using an unpaired Mann–Whitney test at p < 0.001. . 72
A.36 LF/HF Ratio (C6). The significance of difference (5%) was found between the
two classes using an unpaired Mann–Whitney test at p < 0.001. .......... 72
A.37 Longest Shortest RR difference (C7). The significance of difference (5%) was
found between the two classes using an unpaired Mann–Whitney test at p <
0.001. ................................................................. 72
A.38 Detrended RR mean (C8). The significance of difference (5%) was found be-
tween the two classes using an unpaired Mann–Whitney test at p < 0.001. . 73
A.39 RR mean (C9). The significance of difference (5%) was found between the two
classes using an unpaired Mann–Whitney test at p < 0.001. ................. 73
A.40 Sample Entropy of RR (C10). The significance of difference (5%) was found
between the two classes using an unpaired Mann–Whitney test at p < 0.001. . 73
A.41 Mean absolute deviation of RR interval (C11). The significance of difference
(5%) was found between the two classes using an unpaired Mann–Whitney
test at p < 0.001. ................................................................. 74
A.42 SDSD measure (C12). The significance of difference (5%) was found between
the two classes using an unpaired Mann–Whitney test at p < 0.001. ......... 74
A.43 RMSSD measure (C13). The significance of difference (5%) was found be-
tween the two classes using an unpaired Mann–Whitney test at p < 0.001. . 75
A.44 Inter-quartile range (C14). The significance of difference (5%) was found be-
tween the two classes using an unpaired Mann–Whitney test at p < 0.001. . 75
A.45 NN50 first (C15). The significance of difference (5%) was found between the
two classes using an unpaired Mann–Whitney test at p < 0.001. ......... 75
A.46 pNN50 first (C16). The significance of difference (5%) was found between the
two classes using an unpaired Mann–Whitney test at p < 0.001. .............. 76
A.47 NN50 second (C17). The significance of difference (5%) was found between
the two classes using an unpaired Mann–Whitney test at p < 0.001. .......... 76
A.48 pNN50 second (C18). The significance of difference (5%) was found between
the two classes using an unpaired Mann–Whitney test at p < 0.001. ........... 76
A.49 Synchronized correlation of RR interval and respiration (C19). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. ........................................ 77

A.50 Correlation of RR and IBI (C20). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. ........................................ 77

A.51 Correlation of RR and inhalation-exhalation time period (C21). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. 77

A.52 ROC and PR curves for both Subject-Independent and Subject-Specific for CFS algorithm ........................................ 78

A.53 ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm ........................................ 81

A.54 Apnea detection performance of subject-independent approach for varying epoch duration ........................................ 82

A.55 Apnea detection performance of subject-independent approach for varying apnea percentage per epoch ........................................ 83
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Factors related to the occurrence of apnea, and some possible physiological mechanisms involved in apnea. Reproduced and modified from Lopes [2]</td>
<td>6</td>
</tr>
<tr>
<td>3.1</td>
<td>Summary of subject demographics and some statistics (N = 10)</td>
<td>14</td>
</tr>
<tr>
<td>3.2</td>
<td>List of existing respiratory features</td>
<td>20</td>
</tr>
<tr>
<td>3.3</td>
<td>List of existing respiratory features</td>
<td>22</td>
</tr>
<tr>
<td>3.4</td>
<td>List of proposed respiration features</td>
<td>24</td>
</tr>
<tr>
<td>3.5</td>
<td>List of proposed cardio-respiratory features</td>
<td>30</td>
</tr>
<tr>
<td>3.6</td>
<td>Confusion matrix</td>
<td>40</td>
</tr>
<tr>
<td>3.7</td>
<td>Confusion matrix of the first algorithm</td>
<td>40</td>
</tr>
<tr>
<td>3.8</td>
<td>Confusion matrix of the second algorithm</td>
<td>41</td>
</tr>
<tr>
<td>4.1</td>
<td>Statistics of occurrence of apnea per subject for both Annotators</td>
<td>42</td>
</tr>
<tr>
<td>4.2</td>
<td>Statistics of mean agreement between the two Annotators</td>
<td>43</td>
</tr>
<tr>
<td>4.3</td>
<td>Classification results using respiratory features and CFS for the first annotator</td>
<td>44</td>
</tr>
<tr>
<td>4.4</td>
<td>Classification results using respiratory features and CFS for the second annotator</td>
<td>45</td>
</tr>
<tr>
<td>4.5</td>
<td>Classification results using respiratory features and CFS for the both annotators</td>
<td>45</td>
</tr>
<tr>
<td>4.6</td>
<td>Classification results using respiratory features and mRMR for the first annotator</td>
<td>45</td>
</tr>
<tr>
<td>4.7</td>
<td>Classification results using respiratory features and mRMR for the second annotator</td>
<td>46</td>
</tr>
<tr>
<td>4.8</td>
<td>Classification results using respiratory features and CFS for the both annotators</td>
<td>47</td>
</tr>
<tr>
<td>4.9</td>
<td>All new respiratory features selected at least once</td>
<td>47</td>
</tr>
<tr>
<td>4.10</td>
<td>Classification results using cardiac features and CFS algorithm for the first annotator</td>
<td>49</td>
</tr>
</tbody>
</table>
### List of Tables

4.11 Classification results using cardiac features and CFS algorithm for the second annotator .................................................. 49
4.12 Classification results using respiratory and cardiac features and CFS for the first annotator .............................................. 50
4.13 Classification results using respiratory and cardiac features and CFS for the second annotator ........................................ 51
4.14 Classification results using respiratory and cardiac features and CFS for the both annotators ........................................... 51
4.15 Classification results using respiratory features and CFS for the first annotator ............................................................. 54
4.16 Classification results using respiratory features from chest impedance after applying Lee's et al [3] method .......................... 54
4.17 Classification results using cardiac features and CFS algorithm for the first annotator ..................................................... 55
4.18 Classification results using cardiac features and chest impedance after applying Lee's et al [3] method ............................ 56
A.1 Classification results using cardiac features and CFS algorithm for the first annotator .............................................. 79
A.2 Classification results using cardiac features and CFS algorithm for the second annotator .............................................. 79
A.3 Classification results using cardiac features and CFS algorithm for both annotators .................................................. 80
A.4 Classification results using cardiac features and mRMR algorithm for the first annotator ............................................. 80
A.5 Classification results using cardiac features and mRMR algorithm for the second annotator ......................................... 80
A.6 Classification results using cardiac features and mRMR algorithm for both annotators ............................................... 81
A.7 All new cardiac features selected at least once .................................................. 81
A.8 Classification results using respiratory features for epoch duration of 30 seconds .................................................. 82
A.9 Classification results using respiratory features for apnea percentage of 45% per epoch ........................................ 83
1

INTRODUCTION

Background

The normal duration of pregnancy, also known as gestational age, is between 37-42 weeks. A baby born during this period is considered to be born at term. If, on the other hand, a baby is born before 37 weeks or after 42 weeks, it is classified as premature and post term respectively [4]. In 2010, an estimated 14.9 million babies were born preterm, 11.1% of all livebirths worldwide, ranging from about 5% in several European countries to 18% in some African countries [5]. The rate of preterm birth has increased in many locations, predominantly because of increasing indicated preterm births and preterm delivery of artificially conceived multiple pregnancies [6]. Common reasons for preterm births include pre-eclampsia or eclampsia, and intrauterine growth restriction. Births that follow spontaneous preterm labour and preterm premature rupture of the membranes—together called spontaneous preterm births—are regarded as a syndrome resulting from multiple causes, including infection or inflammation, vascular disease, and uterine overdistension [6]. Preterm birth is a major determinant of neonatal mortality and morbidity and has long-term adverse consequences for health [7] [8]. In the first weeks, the complications of premature birth may include, respiratory distress syndrome (surfactant deficiency, leading to alveolar collapse), chronic lung disease or bronchopulmonary dysplasia (damage of the lung parenchym caused by immaturity or artificial ventilation) and cerebral hemorrhage. In addition to the aforementioned disease conditions, prematurity of birth can also predispose neonates to undesirable cessations of breathing, a condition termed as Apnea of Prematurity (AOP) [9]. The prevalence of this problem is very high since over than fifty percent of premature infants experience apnea and the existence of apnea is nearly universal in infants who are born less than 1 kg [10].

Definition of Apnea

The literature defines clinically significant apnea in infants as breathing pauses that last for >20 seconds or for > 10 seconds if associated with bradycardia or oxygen desaturation (see Figure 1.1), but there is no consensus about the duration of apnea, the degree of
change in oxygen saturation, or severity of bradycardia that should be considered pathologic [10]. Zhao et al. [11] and Paolillo et al. [12] attempt to be more precise defining apnea as a cessation of breathing for more than 15–20 seconds, or accompanied by oxygen desaturation ($SpO_2 \leq 80$ for $\geq 4$ seconds) and bradycardia (heart rate $< 2/3$ of baseline for $\geq 4$ seconds). It has to be underlined that this definition may vary depending on the infant's symptomatology. Moreover, there is no unanimity about the duration of apnea that should be considered pathologic, and there is no agreement regarding the percentage of change in oxygen saturation or severity of bradycardia that constitutes a significantly important apnea event [10]. The lack of consensus on the definition of apnea, highlights the difficulty in understanding this problem. This in turn hampers efforts to detect apnea and report its prevalence.

**Classification of Apnea**

Apnea is traditionally classified as central (10%–25% of cases), obstructive (10%–25%), or mixed (50%–75%) based on the presence or absence of upper airway obstruction [13]. In central apnea, a cessation of inspiratory efforts with no evidence of obstruction is observed. In obstructive apnea, there is no airflow, even if the infant tries to breathe resulting in chest wall motion without airflow throughout the entire apneic episode. The obstruction can be due to a combination of passive pharyngeal collapse and either active or passive laryngeal closure [14]. Mixed apnea consists of obstructed respiratory efforts usually following central apneic episodes. However, in some cases, airway obstruction precedes central pauses. Usually respiratory efforts against a closed airway prolong apneas and worsen oxygenation and bradycardia [15].

**Causes of Apnea**

AOP likely reflects a “physiological” rather than a “pathological” immature state of res-
piratory control since it usually resolves by term gestation [16]. Regular rhythmic breathing necessitates a patent airway, central respiratory drive (originating from respiratory centers in the brainstem modulated by input from peripheral neural and chemical receptors) and adequate functioning of the muscles of respiration [17]. Chemical feedback cooperates in the maintenance of ideal chemical concentrations such as oxygen and carbon dioxide (central and peripheral chemoreceptors) [17]. AOP is thought to be secondary to immaturity of brainstem centers that regulate breathing. This immaturity in regulation of breathing is also manifested by immaturity in the respiratory responses to hypoxia, hypercapnia and an exaggerated inhibitory response to stimulation of airway receptors [9].

Obstructive apnea, on the other hand, may be associated with poor pharyngeal tone leading to pharyngeal collapse with the negative airway pressures generated during inspiration, or it may result from incoordination of the upper airway muscles involved in maintaining airway patency due to immaturity of the neuronal signaling mechanisms [18].

**Consequences of Apnea**

Although AOP is, by definition, a self-limiting disease, untreated or inadequately treated AOP may result in short and/or long-term consequences. In premature infants, desaturation and bradycardic episodes often occur along with apnea and can be accompanied by a rise in stroke volume [11]. However, prolonged apnea and bradycardia can decrease the systemic blood pressure and lead to cerebral hypoperfusion, increasing the risk of impaired neurological development. The long-term consequences of apnea are not well known [11], since it is difficult to prove a link between apnea and poor neurodevelopmental outcomes, due to the possible coexistence of neurological injury in premature infants. However, there is evidence that an increased number of AOP days were associated with neurodevelopmental impairment such as cerebral palsy and blindness at three years of age [19]. AOP is often the rate-limiting process in NICU discharge. In one survey, 74% of neonatal specialists required an apnea-free period of five to seven days before discharging a baby from NICU [20].

**Current detection of Apnea—Reason of research**

Preterm newborns or critically ill newborns are cared for in an intensive care setting using various medical devices including patient monitors, ventilators and infusion pumps for treating and monitoring their vital functions. Patient monitoring is used to continuously acquire the ECG, respiration rate (using impedance pneumography), oxygen saturation (SpO2), invasive arterial blood pressure (when available) and temperature of all babies. Algorithms, built into the patient monitors are intended to detect all clinically significant deteriorations, including apnea.

Apnea is defined as cessation of breathing as mentioned above. During cessation of breathing there is no airflow. However, airflow measurements have a more obtrusive character and therefore impedance pneumography is used. Impedance pneumography is most often used in apnea monitoring units in neonates. The monitor detects small changes in electrical impedance as lung volume changes from air entering and leaving the lungs during respiration [21]. The monitor passes a safe, low-amplitude, high-frequency (20-120
KHz) current through the chest by means of two electrodes placed on the left and right side of the chest. Impedance variations in the thorax result in voltage variation in the monitor’s detect circuit; these variations are interpreted as breaths. The same electrodes are also used to monitor the electrical activity heart and determine heart rate [21]. Apnea monitoring also includes indirect methods of apnea detection such as monitoring other physiological parameters linked to presence of absence of adequate respiration (e.g. heart rate and/or oxygen saturation).

Frequently, monitors fail to alarm during apnea episodes because they sense artifacts as respiration. Typical artifacts include vibration from nearby instruments, heart activity and patient movement [21]. It is also likely that some monitors may misinterpret diaphragmatic activity (e.g. gasping during upper-airway obstruction) as a breath. Studies have shown that bedside detection of AOP today exhibits poor sensitivity and specificity rates as the prevalence of false alarms is high while the AOP is under-detected [22]. Moreover, nursing documentation does not provide accurate manual monitoring of apnea either. Although bedside monitors have better sensitivity and specificity than nursing documentation, future research should be directed to improve the specificity of bedside monitoring [22].

**Purpose of Thesis**

The aim of this research project is to create an improved algorithm for automated central apnea detection in preterm infants based on raw waveform analysis of one-lead ECG and chest impedance signals.

**Thesis Outline**

In Chapter 2, background of apnea and apnea detection methods are presented. First, clinical presentation of apnea is introduced. Then, the complex relation of apnea, bradycardia and desaturation is discussed. Existing techniques and challenges of detecting apnea are reviewed.

In Chapter 3 the applied methodology is described in details. First an overview of the physiological signals used in this work is presented and then the description of the annotation follows. In sections, feature generation and visualization, existing and new respiratory and cardiac features are illustrated. In addition, the classification technique that was followed, as well as applied approaches and evaluation metrics are discussed.

In Chapter 4 statistics between the two annotators and classification results are presented. Classification is applied based on respiratory, cardiac and combination of cardiac and respiratory features. Furthermore, CFS and mRMR feature selection algorithms are compared as well as annotations cases (annotations based on first annotator, second annotator and both annotators). In addition, we will investigate the impact of duration of the epoch and the percentage of apnea included in an apneic epoch on the apnea detection performance of the algorithm. Last, we will compare the performance of our system with the approach that Lee et al [3] proposed.

Chapter 5 presents discussion and conclusions.
BACKGROUND OF CENTRAL APNEA AND APNEA DETECTION TECHNIQUES

2.1. CLINICAL PRESENTATION OF CENTRAL APNEA (CA)

Breathing is an essential, involuntary and dynamic process that is modulated by multiple central and peripheral inputs so that oxygen and metabolic demands of cells and tissues can be met. During early development in the uterus, the structure and function of all components (sensors, controls and effectors) of the integrated respiratory network are going through significant modification such that such that ventilation evolves from sporadic fetal breathing to more sustained breathing seen in infants born at term gestation [23]. However, the fetus does not rely on ventilation to oxygenate tissues. Therefore, it is not necessary for breathing to be sustained since it can be modulated by arterial oxygen tension and blood glucose levels [24]. Upon preterm birth, the brain is not mature and appears to influence adversely the rhythmic firing of brainstem neurons and thus contributes to the development of apneic episodes.

Central apneic episodes occur when there is a lack of respiratory effort due to either a cessation of output from the central respiratory centers or the inability of the efferent peripheral nerves and respiratory muscles required to receive or process the signals from the brain. This can be due to immaturity of the system, as seen in certain premature infants, who have a decreased response to hypercapnia (increased carbon dioxide levels). However, this is not the only cause. The mechanisms through which different factors are related to the occurrence of apnea are not entirely established. Table 2.1 presents a relationship between some of the possible physiological mechanisms involved that are reported in the literature as to the occurrence of apnea [2].

For example, anemia is associated with apnea, because lowered oxygen-carrying capacity of red blood cells lead to hypoxia, resulting in respiratory depression [25]. Other studies however suggest that the role of gastroesophageal reflux and anemia remains controversial [11].

Central Apnea is treated by stimulating the central nervous system by using drugs like
Table 2.1: Factors related to the occurrence of apnea, and some possible physiological mechanisms involved in apnea. Reproduced and modified from Lopes [2]

<table>
<thead>
<tr>
<th>Factor</th>
<th>Possible mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDA</td>
<td>Leads to hypoxia and/or muscle fatigue</td>
</tr>
<tr>
<td>Anemia</td>
<td>Leads to hypoxia</td>
</tr>
<tr>
<td>Hypoglycemia</td>
<td>Affects brain stem and muscles due to central, lack of substrate to respiratory muscles</td>
</tr>
<tr>
<td>Hypocalcemia</td>
<td>Central, lack of substrate to respiratory muscles</td>
</tr>
<tr>
<td>Infection</td>
<td>Leads to hypoxemia? Increased metabolic rate, central inhibition</td>
</tr>
<tr>
<td>Prematurity</td>
<td>CNS immaturity, muscle fatigue</td>
</tr>
<tr>
<td>Thermal instability</td>
<td>Leads to inhibitory afferent stimulation</td>
</tr>
<tr>
<td>Intracranial pathology</td>
<td>Direct effect on central respiratory centers</td>
</tr>
<tr>
<td>Gastroesophageal reflux</td>
<td>Inhibitory reflex influencing upper airways</td>
</tr>
<tr>
<td>Drugs</td>
<td>CNS function depression</td>
</tr>
</tbody>
</table>

methylxanthine. Methylxanthine compounds such as caffeine, theophylline, and aminophylline are administered to premature infants as respiratory stimulants to decrease central apnea [26]. These drugs are powerful central nervous system stimulants and likely reduce apnea by multiple physiological and pharmacological mechanisms [11] [26] [27].

Systematic reviews of caffeine therapy in AOP have shown that both caffeine and theophylline are effective in reducing apnea within 2 to 7 days of starting treatment. Caffeine is safer and has a wider therapeutic range than theophylline [11] [27] [28]. Methylxanthines do carry some risks of adverse events. Toxic levels may produce tachycardia, cardiac dysrhythmias, and feeding intolerance and at very high doses may accelerate seizures. Mild diuresis and delayed gastric emptying can also be seen in very low birth weight infants. Methylxanthines also increase energy expenditure, possibly leading to diminished growth in premature infants, requiring supplementation of calories. [11].

As mentioned previously, apnea is defined by the cessation of respiratory airflow. The duration of cessation necessary to be qualified as a true apneic event has changed dramatically over the last few decades: two minutes in 1956 [29], one minute in 1959 [30], 30 seconds in 1970 [31], and 20 seconds or shorter if associated with bradycardia or oxygen
2.2. APNEA, BRADYCARDIA AND DESATURATION (ABD)

While AOP is related to immaturity, the reasons behind the causes and occurrence of apnea in premature infants are not entirely clear or understood. Some definitions described in Chapter 1 include bradycardia and oxygen desaturation for defining AOP; however, the relation between these three conditions is not clear and quite complex, as can be seen in the following Figure 2.1.

It has been widely assumed that apnea of prematurity is caused by immaturity of brainstem respiratory rhythm generation that is in proximity to sites of central \( \text{CO}_2 \) (carbon dioxide) chemosensitivity regulating respiratory activity and ventilatory responses to carbon dioxide [1]. Whereas adults increase their ventilation through an increase in both tidal volume and frequency, preterm infants do not increase frequency in response to carbon dioxide increase which implies that the hypercapnic ventilatory responses are impaired in preterm infants [1].

Martin et al. [1] also state that a decrease in \( \text{PaO}_2 \) (partial oxygen pressure) and thus in oxygen saturation is the typical response to apnea in preterm infants. Furthermore, they support that the decrease in oxygenation is related directly to the duration of apnea and the initial level of \( \text{PaO}_2 \). Mohr et al. [33] showed that, despite the slow drop of \( \text{SpO}_2 \) in very long apneas, the available measure of deficit in oxygen transport is approximately proportional to the duration of the cessation of breathing. And as shown in Figure 2.1 oxygen desaturation (hypoxia) is though to cause bradycardia. On the other hand, Martin et al. [1] also claim that bradycardia may also follow apnea without a measurable fall in oxygen saturation, suggesting a strong mediated phenomenon that is not necessarily triggered by hypoxemia in some infants.

Analyzing the relationship between apneic events, bradycardia and hypoxemia; another study [34] found that i) 86% of bradycardias were accompanied with a fall of oxygen saturation to \( \leq 80\% \), ii) 83% of bradycardias were accompanied with an apneic event of four seconds or longer and iii) 79% of bradycardias were accompanied with both apneic event and desaturation (ABD events). It was observed that bradycardia began 4.8 seconds (median) after the onset of apnea event or 4.2 seconds after (median) the onset of oxygen desaturation [35].

There is no doubt that bradycardia is related to both oxygen supply and apnea, due to different neural inputs that signal heartbeats. Bradycardia is mediated by a chemoreceptor reflex caused by hypoxemia (which is common in neonates) [34]. In a situation of oxygen deficit, peripheral chemoreceptors signal the heart to slow down. It is the lack of oxygen, and not the apnea itself, responsible for this reaction [34]. However, further studies indi-
2. BACKGROUND OF CENTRAL APNEA AND APNEA DETECTION TECHNIQUES

Figure 2.1: Relation of apnea, bradycardia and oxygen desaturation. Bradycardia and oxygen desaturation are immediate effects of apnea. However, these conditions also inhibit respiratory drive, which may either cause or worsen apnea. Reproduced from Martin et al. [1]

cate that the absence of lung inflation could also cause bradycardia. Thus, apnea alone may result in bradycardia [34]. Poets [34] claims that it is not just the chemoreceptor input that influences bradycardia. The conclusion was that bradycardia results from the lack of oxygen in the blood, but is worsened by the apnea itself [34].

The other condition associated with an ABD event is arterial oxygen saturation, or \( SaO_2 \). This is the ratio of oxygenated hemoglobin (hemoglobin that is carrying oxygen molecules) over the total amount of hemoglobin in a given volume. Oxygen saturation reflects how much oxygen the blood transports throughout the body, and is related to the partial pressure of oxygen in the lungs. The normal level of oxygen saturation in the blood is around 95%.

Tourneux et al. [36] confirmed the inverse relationship between functional residual capacity (FRC) and the likelihood of apnea-related desaturation. Lung volume is an important determinant of the speed with which desaturation develops during voluntary breathholding, and pre-apneic lung volume was found to have a strong influence on the hypoxemia that occurs during sleep apnea in adults [34]. The FRC serves as a buffer to stabilize oxygenation during brief periods of apnea. Poets et al. [37] found an inverse correlation between FRC and the speed with which \( SpO_2 \) fell during desaturation for neonates, i.e., the lower the lung volume, the more rapid the fall in \( SpO_2 \) during apnea.
2.3. REVIEW OF TECHNIQUES FOR DETECTING APNEA

Poets [34] emphasizes the importance of sufficient lung volume and the impact that a loss in lung volume has by providing evidence which underline the close relationship. Studies showed that a loss in lung volume was crucial for the development of hypoxemic episodes in association with abdominal muscle contractions in extremely premature infants, being involved in 80% of desaturations to <75% SpO2 in these infants [34]. Furthermore, these hypoxemic episodes were associated with a mean decrease in resting lung volume by 69% of tidal volume.

In conclusion, the physiology of AOP is not well understood. In general, bradycardia results from the lack of oxygen in the blood. However, is worsened by the apnea itself. Moreover, apnea (when absence of lung inflation occur) may lead to bradycardia.

2.3. REVIEW OF TECHNIQUES FOR DETECTING APNEA

Apnea is a very common problem and constitutes a major reason for prolonged stay in neonatal intensive care units (NICUs). In 2006, the National Institutes of Health and the Food and Drug Administration convened a consensus group to define the issues associated with AOP. The goal of this group was to define future research goals for improving diagnosis, understanding etiology, and developing more effective treatments. The group identified the following issues requiring attention [10]:

- Lack of standardization for definition, diagnosis, and treatment of AOP
- Unproven benefit of intervention
- Lack of real-time data documenting AOP events
- Unevaluated sustained treatment improvement at 7 days or later
- Failure to address confounding conditions
- Unsubstantiated AOP–gastroesophageal reflux disease relationship more than 50 ms
- Undetermined role of AOP affecting long-term neurodevelopmental outcomes

Obviously, detecting and analyzing apnea plays a central role [10].

In general apnea detection is based on analysis of the respiratory peaks. Lee's et al. [3] developed an algorithm for detecting central apnea in neonates by filtering the cardiac artifact from the chest impedance. During an apnea, the absence of breaths allows the heartbeat to dominate the chest impedance. The heart rate slows giving in this way more time for the heart to fill with blood, resulting in larger fluctuations in impedance. Without impedance fluctuations caused by breathing (approximately 1 ohm), smaller impedance changes (0.5 ohm) due to blood pumping through the heart may be misinterpreted as breaths. Usually heart rate of neonates is between 120 to 160 bpm. During an apnea event, the heart rate slows down and can move into respiratory frequency ranges. Lee's et al. [3] synchronized chest impedance to heartbeats. A band-stop Butterworth filter removed the cardiac artifact of the chest impedance, next resampled was performed using the QRS signal as a clock. A high-pass filter removed artifact due to patient movement. The chest
impedance signal was then renormalized using an envelope function [3]. Finally, Lee et al. [3] calculated the probability of apnea from the renormalized signal’s standard deviation. High fluctuations, i.e. large standard deviations, indicate normal breathing, and thus have low apnea probability. Minimal fluctuations occur during an apnea, resulting in high apnea probability. After applying the filtering algorithm to a randomly selected sample of 237 apnea episodes of coexisting bradycardia and $O_2$ desaturation (HR< 100bmp and $SpO_2$ <85%) selected from five infants, three clinicians at the University of Virginia unanimously agreed on 234 apnea episodes [3]. Out of these apnea events, 212 were automatically detected with Lee’s et al. algorithm [3], resulting in a 91% accuracy rate. In addition Lee et al. [3] found an 11% false-positive rate and a 37% false-negative rate based on random sampling [3].

In 2014 Vergales et al. [38] compared nursing records of AOP to apnea events detected by Lee’s et al. algorithm [3]. They analyzed 19.9 patient years of signal data for 276 very low-birth-weight infants (birth weight <1,500 gr). Their algorithm defined apnea as > 10 seconds if combined with bradycardia (HR< 100bmp) and desaturation ($SpO_2$ <85%). They concluded that the current gold standard of documentation of AOP, the nursing record, does not document the majority of prolonged central apnea events detected by their computer algorithm. Of the 5,275 algorithm-detected prolonged apnea events > 30 seconds, only 26% had nurse-recorded documentation within 1 hour. Monitor alarms sounded in only 74% of events of algorithm-detected prolonged apnea events > 10 seconds. Specifically, of the 3,019 nurse-recorded events, only 68% had any algorithm-detected ABD event. Thus, implying under-detection of apnea, emphasizing the need for improved algorithms.

On the other hand, Altuve et al [39] propose an apnea-bradycardia detector who takes into account not only the instantaneous values of beat-to-beat interval (RR interval) time series, like current monitor do, but also the their temporal evolution (e. g. their dynamics). In current NICU monitors, AB is detected by processing the cardiac cycle length (RR interval) extracted from the electrocardiogram (ECG) and by comparing its instantaneous values to a fixed or relative threshold [39]. This conventional analysis (based only on the amplitude analysis of RR interval time series ) is not so effective to accurately detect apnea-bradycardia events since there is a significant intra- and inter-patient variability of heart rate [21]. The detector is based on a set of hidden semi-Markov models, representing the temporal evolution of RR time series. A preprocessing step, including quantization and delayed version of the observation vector, is also proposed to maximize detection performance. This approach was quantitatively evaluated through simulated and real signals, the latter extracted from ECG of thirty two preterm infants monitored from neonatal intensive care units (NICU) with frequent apnea-bradycardia episodes [39]. Their best detector showed an improvement on average of around 15% in sensitivity and 7% in specificity compared to two conventional detectors (fixed or relative threshold detections) used in NICU. A reduced detection delay of approximately 2 seconds was also observed with respect to conventional detectors [39].
2.4. DIFFICULTIES OF DETECTING CENTRAL APNEA

The complicated physiology that determines apnea along with the inter and intra-patient variability in respiration and heart rate makes the detection of apnea challenging. While long apnea episodes are typically followed by bradycardia and desaturation, this is not always the case. At times, bradycardia or desaturation may in fact trigger an apnea. Furthermore, short apneas can occur without any other reflection on heart rate or oxygen saturation. To complicate things even more, heart rate and respiration rate are measured from the same sensors (a set of three chest leads). So possibly heart rate interference may occur during apnea leading to misinterpretation of heart rate as breathing.

Measuring respiration rate from chest impedance is quite difficult especially if the baby’s weight is very low which is common on very preterm babies. Small babies have a large surface area compared to their volume, and little body fat. Detecting of respiration rate even more difficult since the lungs are premature and the air volume inhaled and exhaled is relatively small to detect.
3.1. OVERVIEW OF PHYSIOLOGICAL SIGNALS

The aim of this thesis is to create an algorithm for automated apnea detection for preterm infants based on a set of non-invasive physiological signals. The signals that are commonly measured in a patient monitor are:

- the electrocardiogram (ECG)
- the chest impedance pneumography (CI)
- the photoplethysmography signal (PPG)

![Three ECG leads attached on a preterm baby](image1)

![Normal ECG signal](image2)

(a) Three ECG leads attached on a preterm baby  
(b) Normal ECG signal

Figure 3.1: Electrocardiogram

An ECG signal (see Figure 3.1) is a recording of the electrical activity produced by the heart. The heart is a four-chambered pump that provides the driving force for the circulation of blood. With each heart beat the synchronized depolarization and repolarization spreading through the heart causes currents in the extracellular fluid that establish field potentials over the whole body. ECG can provide very valuable information about apnea
events and has been broadly studied for the detection of apnea. One of the most important signals which can be obtained from an ECG is the beat-by-beat series of the heart rate.

PPG is an optic method measuring light absorption upon passage of oxygenated blood. It provides a measure of the volume of tissue blood where the pulsatile component of the heartbeat is measured and the peripheral circulation is evaluated [40]. This measurement is related to arterial vasoconstriction or vasodilatation generated by the autonomic nervous system and is modulated by the heart cycle. This noninvasive technique attaches a sensor to a finger or toe. The sensor where it emits red and infrared light. Oxygenated hemoglobin absorbs more infrared light, while regular hemoglobin absorbs more red light. Based on the absorption ratio of the two different wavelengths, the fractional oxygen saturation in the blood is calculated.

![Pulse oximeter](image1.png)  
(a) Pulse oximeter  

![Fingertip Photoplethysmogram waveform](image2.png)  
(b) Fingertip Photoplethysmogram waveform  

Figure 3.2: Photoplethysmography

![Three ECG leads attached on a preterm baby](image3.png)  
(a) Three ECG leads attached on a preterm baby  

![Impedance pneumography waveform](image4.png)  
(b) Impedance pneumography waveform  

Figure 3.3: Impedance pneumography

Impedance pneumography is a noninvasive technology available for measuring respiratory function for decades and is commonly used in clinical research. Impedance pneumography employs low amplitude, high frequency alternating current (AC) between two surface electrodes to record volume changes at the rib cage (RC) during a respiratory cycle. Based on Ohm’s Law, the voltage drop is linearly related to the impedance between the electrodes and increases during inspiration and decreases during expiration.
3. Methodology

3.2. Description of the dataset

Data of ECG, chest impedance and oxygen saturation were collected from NICU bedside monitors (GE Medical models Solar 8000M and I). ECG is tracked using three ECG leads collected at 240 Hz, with electrodes on both sides of the heart to record its electrical activity. Heart pulses are detected, and a heart rate is calculated. Heart rate computed from the monitor was used during annotation from neonatologists (see section 3.3). Chest impedance waveforms are collected with the same leads at 60 Hz by passing a high frequency signal through the leads. Impedance changes with breaths due to the fluctuations of air in the lungs. Oxygen saturation signal (0.5Hz), measured by PPG was used in this project.

Nine hours of continuously reordered data of ten very low-birth-weight (VLBW) infants (birth weight < 1,500 gr) undergoing continuous cardiopulmonary monitoring in the NICU at Maxima Medisch Centrum from 2008 were included in the analysis. In all ten infants apnea events were present. Patient demographics are statistics are presented in Table 3.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean ±SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>7 males and 3 females</td>
<td></td>
</tr>
<tr>
<td>GA (weeks)</td>
<td>28.5 ±1.58</td>
<td>27 - 31</td>
</tr>
<tr>
<td>BW (gr)</td>
<td>1041.5 ±203.37</td>
<td>580 - 1250</td>
</tr>
</tbody>
</table>

3.3. Description of the Annotation System Software

The aim of this research is the development of an algorithm for automatic detection of central apnea. The validation of this algorithm is based on the comparison of its results with reference annotations and therefore the development of user-friendly annotation software is required. This section presents an application, developed in Matlab®, for the annotation of respiratory signals.

The annotation process is time-consuming and demands concentration and rigor, as there is high inter-and-intra variation in the chest impedance of preterm infants. It should also be noted that this tool is intended to be used mainly by healthcare professionals, who tend to have overloaded agendas and no programming skills. For these reasons, the main requirement of the software is user-friendliness: the annotation system must be simple, quick and flexible. Additionally, the application format was chosen because of its resemblance to the monitor used at the NICU, so that the healthcare professionals will be familiar and comfortable as much as possible. The software interface comprises five physiological signals (see Figure 3.4):

- ECG
3.3. DESCRIPTION OF THE ANNOTATION SYSTEM SOFTWARE

- Oxygen saturation (SpO$_2$)
- Heart Rate (HR)
- Respiration rate (RR)
- Respiration signal

The red dashed line is the threshold used at the NICU monitors for bradycardia (≤ 80 bpm) and desaturation (≤ 80%).

The physiological signals are presented in three minutes segments. ECG corresponds to the last ten seconds of the three minutes segment. The user can annotate apnea episodes (on the chest impedance) based on the chest impedance but also based on other vital signs that are associated with apnea such as oxygen saturation and heart rate. It should be mentioned that heart rate and respiration rate parameters are exported from the monitor and not re-calculated for performing annotation.

By clicking on the Annotate-button, cursor mode is enabled on the current three minutes (see Figure 3.5). The user may click the beginning and end of apnea segment on the chest impedance and the apnea episode will be marked with a red line (see Figure 3.6). It should be noted that a three minute segment may consist of multiple apnea events, as many as the user deems appropriate (see Figure 3.7).

One of the most important features of this application is the possibility of modifying the respiratory annotations performed in previous three minute segments. That is an important feature since it is possible that the user may reconsider his judgment on the existence of an apnea event. By moving forward (next-button) and backward (back-button)
the user may click the show-button, upon which the red line segments will appear on chest impedance if apnea episodes were previously annotated. Then the user can modify his annotation.

Nine hour data from ten preterm infants were annotated by two clinical experts (neonatologists).
3.4. Feature Generation

3.4.1. Signal Processing

The two signals, ECG and chest impedance are filtered to eliminate any artifact (outside the breathing and heart rate range) or unwanted components of the signal.

The raw chest impedance signal of all subjects was preprocessed before feature extraction. First, monitor artifacts were removed (see Figure 3.8). Then, raw chest impedance signals were filtered with a 10th order Butterworth low-pass filter with a cut-off frequency of 1.67 Hz for the purpose of eliminating high frequency noise. Low frequency noise (patient movement/baseline wander) was removed by using 4th order Butterworth high-pass filter with a cut-off frequency of 0.1 Hz. These cut-off frequencies assume an upper bound of 100 breaths per minute and a lower bound of 6 breaths per minute.

For locating peaks and troughs of chest impedance, the turning points were simply identified based on sign change of signal slope. The falsely detected ‘dubious’ peaks and troughs were corrected when [41]:

- too short intervals between peak and trough pairs occurred, i.e. the sum of two successive intervals is less than the median of all intervals over the entire recording
- two small amplitudes occurred, i.e. the peak-to-trough difference is smaller than 15% of the median of the entire chest impedance

These methods were verified by comparing automatically detected results with manually annotated peaks and troughs from thoracic inductance plethysmography. An accuracy
of 98% was achieved. This algorithm was applied for peak-trough detection on chest impedance after validation performed based on empirical observation of all data.

![Respiratory effort signal](image)

There are a number of noise sources that can cause interference in the ECG. These include mains interference (50 Hz), muscle artifact and electrode contact noise as well as patient movement [42]. The ECG signal is band-pass filtered, the high-pass cut-off frequency being chosen as 0.67 Hz to eliminate baseline wander. The low-pass cut-off is 40 Hz to eliminate mains noise and any other high-frequency noise whilst preserving characteristics of the ECG of interest, for example the QRS amplitude. Pre-processing is typical for monitoring and extracting information from the signal, however some information of the original signal is lost. This loss does not have an impact on the analysis of this project.

For R-peak detection in each subject’s ECG signal, a R-peak detection algorithm (used in fetal monitoring [43]) was used in order to achieve higher robustness and accuracy compared to Pan and Tompkins’ QRS detector [44]. The R-peak locations are used to derive RR-based features which may directly provide information about apnea. In order to remove subject-dependence from the features, a normalization step on the RR interval series (see figure 3.8 where yellow line is the measured RR interval series) was carried out. For each subject, a normalized RR series (see figure 3.8 where blue line is the normalized RR interval series) was calculated by dividing by the mean RR interval (producing an RR sequence with a unity mean)[45]. The RR interval series exhibit significant variation over the entire 9 hour data. An interesting marker of changes in sleep state may be the relative changes in the RR
interval series rather than the absolute value. Relative changes were quantified in the RR series by detrending the series with a 15-min. The detrended RR (see figure 3.8 red line) is simply the current interval length minus the average length over the previous 15 min [45].

(a) Raw, normalized and detrended RR interval series  
(b) Zoomed version of (a): Normalized RR interval series  
(c) Zoomed version of (a): Detrended RR interval series

Figure 3.9: Example of subject-dependence removal from RR interval series

### 3.4.2. EXISTING FEATURES

The aim of this work is to detect neonatal apnea using ECG and chest impedance signal. Since apnea detection and sleep staging in adults employ ECG and chest impedance respectively, this work exploits this similarity by selecting 33 features from existing research. Notably, apnea detection in adults does not employ chest impedance, since it is rendered redundant by employing a continuous airflow sensor.

### 3.4.2.1. RESPIRATORY FEATURES

Table 3.2 lists and describes all the existing respiratory features. Features aim to extract information related to regularity of respiration, respiratory frequency, spectral power and
Table 3.2: List of existing respiratory features

<table>
<thead>
<tr>
<th>Feature Index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Spectral power of respiratory frequency</td>
</tr>
<tr>
<td>R2</td>
<td>Spectral power in very low frequency (VLF) band (0.01–0.04 Hz)</td>
</tr>
<tr>
<td>R3</td>
<td>Spectral power in low frequency (LF) band (0.04–0.15 Hz)</td>
</tr>
<tr>
<td>R4</td>
<td>Spectral power in high frequency (HF) band (0.4–1.5 Hz)</td>
</tr>
<tr>
<td>R5</td>
<td>Ratio of spectral power between LF and HF bands</td>
</tr>
<tr>
<td>R6</td>
<td>Respiratory frequency estimated in the frequency domain</td>
</tr>
<tr>
<td>R7</td>
<td>Respiratory frequency estimated in the time domain</td>
</tr>
<tr>
<td>R8</td>
<td>Breath-by-breath correlation</td>
</tr>
<tr>
<td>R9</td>
<td>Envelope Power</td>
</tr>
<tr>
<td>R10</td>
<td>Breath Length Variation</td>
</tr>
<tr>
<td>R11</td>
<td>Sample entropy of respiratory effort</td>
</tr>
<tr>
<td>R12</td>
<td>Standardized median of respiratory peaks</td>
</tr>
<tr>
<td>R13</td>
<td>Standardized median of respiratory troughs</td>
</tr>
<tr>
<td>R14</td>
<td>Respiratory peak regularity measured by sample entropy</td>
</tr>
<tr>
<td>R15</td>
<td>Respiratory trough regularity measured by sample entropy</td>
</tr>
<tr>
<td>R16</td>
<td>Median of peak-to-trough amplitude sequence</td>
</tr>
</tbody>
</table>

Note: the references for the existing features are 1–10 from Redmond and Heneghan 2006 [45] related to neonatal bounds [46], Redmond et al. [47], 11 [48], 12 - 16 Long et al. [49]

depth. First the respiratory effort spectrum (the square of the discrete Fourier transform (DFT) of the chest impedance for that epoch, windowed with a Hanning window) was computed [45] [47]. From the spectrum the logarithm of the power in the three bands —VLF(0.01–0.04 Hz) (R2), LF (0.04–0.15 Hz) (R3), and HF (0.4–1.5 Hz) (R4)— was calculated. In addition, the ratio of the logarithm of the spectral between LF and HF was computed (R5). The definition of these bands is taken directly from the corresponding definitions for ECG signals for neonates [46]. Furthermore the respiratory frequency in frequency domain was estimated as the frequency of peak power in the range of 0.04 Hz – 1.5 Hz (R6), and also the logarithm of the power at that frequency (R1). The respiratory frequency in time domain was estimated as the inverse of mean time between adjacent peaks and between adjacent troughs (R7) [45] [47]. Breath-by-breath correlation is computed as the maximum correlation between adjacent breath cycle time (R8) [45]. Envelope power is defined as the mean value of standard deviation of the peak and trough values for the epoch divided by the standard deviation of the chest chest impedance for the epoch (R9). The mean value of the standard deviation of the time between peak locations and the standard deviation of the time between trough locations is denoted as breath length variation (R10) [45] [47]. The non-linear sample entropy measure examines the regularity of a time series and has been broadly used in quantifying regularity of biomedical time series [48] [50] [51] [52]. Sample entropy was implemented with the PhysioNet toolkit sampen [53]. More details about sample entropy we be presented in section 3.4.3.1.
The next five features (feature R12-R16) attempt to exploit respiratory information based on the depth of breathing. Depth-based features were extracted from the peak and trough sequences (i.e., upper and lower envelopes) of the chest impedance [49]. Let consider \( p = p_1, p_2, ..., p_n \) and \( t = t_1, t_2, ..., t_n \) the peak and trough amplitude sequences from an epoch. Then the standardized median of the peaks (and troughs) was computed by dividing the median by their interquartile range (IQR, the difference between the 3rd and the 1st quartile), such that

\[
P_{sdm} = \frac{\text{median}(p_1, p_2, ..., p_n)}{\text{IQR}(p_1, p_2, ..., p_n)} \quad (3.1)
\]

\[
T_{sdm} = \frac{\text{median}(t_1, t_2, ..., t_n)}{\text{IQR}(t_1, t_2, ..., t_n)} \quad (3.2)
\]

These two features consider the mean respiratory depth and its variability at the same time in terms of inhalation (for peaks) and exhalation (for troughs). Feature 12 and 13 are measuring the regularity of the upper and lower envelopes [49]. Let’s consider a time series with \( n \) data points \( u = u_1, u_2, ..., u_n \), let \( v(i) = u_i, u_{i+1}, ..., u_{i+m-1} \) (\( 1 \leq i \leq n - m + 1 \)) where the epoch length \( m \) is an positive integer and \( m < n \). Then for each \( i \), we have \( B_{i,m}(r) = (n - m + 1)^{-1} \eta(r) \), in which \( \eta(r) \) is the number \( j \) such that \( d_m[v(i), v(j)] \leq r(1 \leq j \leq n - m, j \neq i) \) where the distance metric \( d_m \) between two subsequences \( v(i) \) and \( v(j) \) is given by

\[
d_m[v(i), v(j)] = \text{max} \ |u_{i+l} - u_{j+l}| \text{ for all } l = 0, 1, ..., m - 1.
\]

For a higher dimension \( m + 1 \), we have \( A_{i,m}(r) \). Then the sample entropy of the time series \( u \) is defined by

\[
SE = -\ln \left[ \frac{A^m(r)}{B^m(r)} \right] \quad (3.3)
\]

where

\[
A^m(r) = \frac{1}{n - m} \sum_{i=1}^{n-m} A_{i,m}(r), \quad (3.4)
\]

\[
B^m(r) = \frac{1}{n - m} \sum_{i=1}^{n-m} B_{i,m}(r), \quad (3.5)
\]

Similarly, the sample entropy measures of the peak and trough sequences are

\[
P_{se} = -\ln \left[ \frac{A^m_{\text{peak}}(r)}{B^m_{\text{peak}}(r)} \right], \quad (3.6)
\]

\[
T_{se} = -\ln \left[ \frac{A^m_{\text{trough}}(r)}{B^m_{\text{trough}}(r)} \right], \quad (3.7)
\]
in which \( r \) is 0.2 standard deviation of the peak or the trough sequence and \( m \) is 2. \( R \) and \( m \) values were experimentally chosen to maximize the discriminative power of the two features. Finally, the median of peak-to-trough differences expresses the range of inhale and exhale depths[49]. It was computed as

\[
PT_{\text{diff}} = \text{median} \left[ (p_1 - t_1), (p_2 - t_2), \ldots, (p_n - t_n) \right] \tag{3.8}
\]

### 3.4.2.2. Cardiac Features

<table>
<thead>
<tr>
<th>Feature Index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>RR LF band</td>
</tr>
<tr>
<td>C2</td>
<td>RR HF band</td>
</tr>
<tr>
<td>C3</td>
<td>RR standard deviation</td>
</tr>
<tr>
<td>C4</td>
<td>RR respiratory frequency</td>
</tr>
<tr>
<td>C5</td>
<td>RR respiratory power</td>
</tr>
<tr>
<td>C6</td>
<td>LF/HF Ratio</td>
</tr>
<tr>
<td>C7</td>
<td>Longest Shortest RR difference</td>
</tr>
<tr>
<td>C8</td>
<td>Detrended RR mean</td>
</tr>
<tr>
<td>C9</td>
<td>RR mean</td>
</tr>
<tr>
<td>C10</td>
<td>Sample Entropy of RR</td>
</tr>
<tr>
<td>C11</td>
<td>Mean absolute deviation of normalized RR interval</td>
</tr>
<tr>
<td>C12</td>
<td>SDSD measure</td>
</tr>
<tr>
<td>C13</td>
<td>RMSSD measure</td>
</tr>
<tr>
<td>C14</td>
<td>Inter-quartile range</td>
</tr>
<tr>
<td>C15</td>
<td>( \text{NN}50 \text{first} )</td>
</tr>
<tr>
<td>C16</td>
<td>( \text{NN}50 \text{second} )</td>
</tr>
<tr>
<td>C17</td>
<td>( \text{pNN}50 \text{first} )</td>
</tr>
<tr>
<td>C18</td>
<td>( \text{pNN}50 \text{second} )</td>
</tr>
</tbody>
</table>

This section presents a pool of 18 heart rate variability (HRV) features that will be used for detecting AOP (see Table 3.3). Spectral representations of the RR interval series have been widely used previously for a variety of applications [45] [54] [55] [51] [56]. To calculate a power spectral density estimate, the normalized RR intervals corresponding to the specific epoch are windowed (using a Hanning window). The square of its DFT is taken as a single periodogram estimate of the interval-based power spectral density. RR interval series are unevenly spaced. So before computing power spectral density, RR interval series of the epoch were resampled at 4 Hz to obtain an equidistantly spaced time series [46] [57].
From this spectral estimate five features were calculated [45]: i) the logarithm of the normalized LF (power in the 0.04 – 0.15 Hz band) (C1), ii) the logarithm of the normalized HF (power in the 0.4 – 1.5 Hz band) (C2), where normalization is achieved by dividing by the total power in the VLF, LF, and HF bands (0.01 Hz – 1.5 Hz), iii) the LF/HF power ratio (C6), iv) the mean respiratory frequency (C4), which is defined by finding the frequency of maximum power in the HF band and v) the logarithm of the power at the mean respiratory frequency (C5). The definition of these bands is taken directly from the corresponding definitions for ECG signals for neonates [46].

In addition to the RR spectral features a range of temporal RR features for each epoch was computed. These features include [45]: i) the mean normalized RR interval (C9) [54] [55], ii) the standard deviation of normalized RR interval (C3)[54], iii) the difference between the longest and shortest normalized RR interval in the epoch (C7) [55], iv) the mean value of the RR detrend interval in the epoch (C8) and v) the Sample Entropy of the normalized RR interval (C10) [51]. The difference between longest and shortest RR interval within the epoch is an attempt to quantify some of the dynamic behavior within the epoch [45]. RR detrended interval value is a measure of the present relative to the previous 15 minutes of normalized RR interval value. So by measuring the mean RR detrended interval of an epoch, we attempt to explore whether or not the present mean heart RR interval differs from the mean RR interval value over the last 15 minutes. This allows the caption of hysteretic cardiac variations (usually bradycardia) than usually last longer than the apnea episodes.

Other efficient cardiac features used for apnea detection include [54] [55] [58] [56]:

- Mean absolute deviation of normalized RR interval, defined as mean of absolute values obtained by the subtraction of the mean normalized RR interval values from all the normalized RR interval values in an epoch (C11)
- The SDSD measures, defined as standard deviation of the differences between adjacent normalized RR interval of the epoch (C12)
- The RMSSD measures, defined as the square root of the mean of the sum of the squares of differences between adjacent normalized RR interval of the epoch (C13)
- Inter-quartile range, defined as the difference between the 75th and 25th percentiles of the RR normalized interval distribution (C14)
- NN50first, number of pairs of successive normalized RR intervals where the first RR interval exceeds the second RR interval by more than 50 ms (C15)
- NN50second, number of pairs of successive normalized RR interval where the second RR interval exceeds the first RR interval by more than 50 ms (C16)
- pNN50first, defined as NN50first divided by the total number of normalized RR intervals (C17)
- pNN50second, defined as NN50second divided by the total number of normalized RR intervals (C18)
3.4.3. New features

3.4.3.1. Respiratory features

Table 3.4: List of proposed respiration features

<table>
<thead>
<tr>
<th>Feature Index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R17</td>
<td>Ratio of respiratory peak and trough regularity measured by sample entropy</td>
</tr>
<tr>
<td>R18</td>
<td>Respiratory regularity of breath by breath duration</td>
</tr>
<tr>
<td>R19</td>
<td>Correlation of peak-trough breath duration</td>
</tr>
<tr>
<td>R20</td>
<td>Regularity of inhalation-exhalation cycle</td>
</tr>
<tr>
<td>R21</td>
<td>Correlation inhalation-exhalation cycle</td>
</tr>
<tr>
<td>R22</td>
<td>Median of inhalation time period</td>
</tr>
<tr>
<td>R23</td>
<td>Ratio of median value inhalation and exhalation time periods</td>
</tr>
<tr>
<td>R24</td>
<td>IQR of depth of breathing based on peaks distribution</td>
</tr>
<tr>
<td>R25</td>
<td>IQR of depth of breathing based on troughs distribution</td>
</tr>
<tr>
<td>R26</td>
<td>IQR of breathing cycle based on peaks distribution</td>
</tr>
<tr>
<td>R27</td>
<td>IQR of breathing cycle based on troughs distribution</td>
</tr>
<tr>
<td>R28</td>
<td>IQR of inhalation cycle distribution</td>
</tr>
<tr>
<td>R29</td>
<td>IQR of exhalation cycle distribution</td>
</tr>
<tr>
<td>R30</td>
<td>Regularity of breathing cycle</td>
</tr>
<tr>
<td>R31</td>
<td>Regularity of inhalation-exhalation depth</td>
</tr>
</tbody>
</table>

A total of 15 new features are proposed in this section. The ratio of Respiratory peak regularity measured by sample entropy \( (P_{se} \text{ see equation 3.6}) \) and Respiratory trough regularity measured by sample entropy \( (T_{se} \text{ see equation 3.7}) \) introduced on Long et al. [49] was computed (R17).

The locations of peaks and troughs should include the information with respect to regularity of breathing. Let consider \( Locs_p = loc_{p_1}, loc_{p_2}, ..., loc_{p_n} \) and \( loc_t = loc_{t_1}, loc_{t_2}, ..., loc_{t_n} \) the peak and trough location sequences from an epoch. Then the differences of the peaks and differences of troughs locations (Inter-breath intervals) were computed as:

\[
Peak_{diff} = [(loc_{p_1} - loc_{p_2}), (loc_{p_2} - loc_{p_3}), ..., (loc_{p_{n-1}} - loc_{p_n})] \tag{3.9}
\]

\[
Trough_{diff} = [(loc_{t_1} - loc_{t_2}), (loc_{t_2} - loc_{t_3}), ..., (loc_{t_{n-1}} - loc_{t_n})] \tag{3.10}
\]
where \((\text{loc}_{p_1} - \text{loc}_{p_2})\) is denoted as \(\text{diff}_{\text{Locs}_{p_1}}\) and \((\text{loc}_{t_1} - \text{loc}_{t_2})\) is denoted as \(\text{diff}_{\text{Locs}_{t_1}}\) respectively. Then the differences of breaths durations based on peaks and troughs were calculated as:

\[
\text{Breath}_{\text{diff}_p} = \left[ (\text{diff}_{\text{Locs}_{p_1}} - \text{diff}_{\text{Locs}_{p_2}}), (\text{diff}_{\text{Locs}_{p_3}} - \text{diff}_{\text{Locs}_{p_4}}), \ldots \right] \quad (3.11)
\]

\[
\text{Breath}_{\text{diff}_t} = \left[ (\text{diff}_{\text{Locs}_{t_1}} - \text{diff}_{\text{Locs}_{t_2}}), (\text{diff}_{\text{Locs}_{t_3}} - \text{diff}_{\text{Locs}_{t_4}}), \ldots \right] \quad (3.12)
\]

\[
\text{Breath}_{\text{diff}} = \left[ \text{Breath}_{\text{diff}_p}, \text{Breath}_{\text{diff}_t} \right] \quad (3.13)
\]

Then the standard deviation of equation 3.13 is denoted as "Respiratory regularity of breath by breath duration" (R18). In addition, the maximum correlation between equation 3.11 and 3.12 is denoted as "Correlation of peak-trough breath duration" (R19) (see Figure 3.10).

![Figure 3.10: Chest impedance and the locations of the detected peaks and troughs](image)

As illustrated in Figure 3.10, a breathing cycle is the period between two consecutive troughs (or breaths) and thereby the inhalation period is the time period starting from a trough up to the next consecutive peak. On the other hand exhalation period is the time period starting from a peak and ending to the next consecutive trough. Let consider again \(\text{Locs}_{p_1}, \text{loc}_{p_2}, \ldots, \text{loc}_{p_n}\) and \(\text{loc}_{t_1}, \text{loc}_{t_2}, \ldots, \text{loc}_{t_n}\) the peak and trough location sequences from an epoch. Then the inhalation and exhalation cycles within an epoch are expressed as:

\[
\text{Inhal}_{\text{cycle}} = \left[ (\text{loc}_{t_1} - \text{loc}_{p_2}), (\text{loc}_{t_2} - \text{loc}_{p_3}), \ldots, (\text{loc}_{t_{n-1}} - \text{loc}_{p_n}) \right] \quad (3.14)
\]
\[ Exhal_{cycle} = \left[ (loc_{p_1} - loc_{t_1}), (loc_{p_2} - loc_{t_2}), \ldots, (loc_{p_n} - loc_{t_n}) \right] \] (3.15)

where \((loc_{t_1} - loc_{p_2})\) is denoted as \(ih_1\) is the first inhalation cycle in the epoch where there are in total \(K\) consecutive inhalation cycles \((k = 1, 2, \ldots, K)\) and \((loc_{p_1} - loc_{t_1})\) is denoted as \(ex_1\) is the first exhalation cycle in the epoch where there are in total \(L\) consecutive exhalation cycles \((l = 1, 2, \ldots, L)\). Then the differences of inhalation and exhalation cycles are expressed as:

\[ Inhal_{Cycle}_{diff} = [(in_{1} - in_{2}), (in_{2} - in_{3}), \ldots, (in_{K-1} - in_{K})] \] (3.16)

\[ Exhal_{Cycle}_{diff} = [(ex_{1} - ex_{2}), (ex_{2} - ex_{3}), \ldots, (ex_{L-1} - ex_{L})] \] (3.17)

\[ InhalExhal_{vec} = Inhal_{Cycle}_{diff} \ Exhal_{Cycle}_{diff} \] (3.18)

Then the standard deviation of equation 3.18 is denoted as "Regularity of inhalation-exhalation cycle" (R20). In addition, the maximum correlation between equation 3.16 and 3.17 is denoted as "Correlation inhalation-exhalation cycle" (R21). Additionally, the median of inhalation (eq 3.14) time period (R22) and ratio of median values of inhalation and exhalation time periods (R23) were computed.

Two data sets (for example apnea and non-apnea epochs) can have the same mean value of a feature but they can be entirely different. Thus to describe data, one needs to know the extent of variability. This is given by the measures of dispersion. Inter-Quartile-Range (IQR) is a commonly used measure of dispersion. IQR is the difference between 75th and 25th percentiles of a distribution. IQR was used to derive discriminative (between apnea and normal breathing epoch) information from chest impedance. The features that were computed are listed below:

- IQR of depth of breathing based on peaks distribution (R24) (where \(p = p_1, p_2, \ldots, p_n\) the peak amplitude sequences from an epoch)
- IQR of depth of breathing based on troughs distribution (R25) (where \(t = t_1, t_2, \ldots, t_n\) trough amplitude sequences from an epoch)
- IQR breathing cycle based on peaks distribution (R26) (see eq. 3.9)
- IQR breathing cycle based on troughs distribution (R27) (see eq. 3.10)
- IQR inhalation cycle distribution (R28) (see eq. 3.14)
- IQR exhalation cycle distribution (R29) (see eq. 3.15)
Sample entropy provides information regarding how time series signals change with time, by comparing each time series signal with a lagged form of itself. Sample Entropy is less sensitive to noise and can be applied for short-length time series data [48]. Additionally, it is resistant to short strong transient interferences (outliers) such as spikes. These characteristics make Sample Entropy an appealing tool for nonlinear analysis of physiological signals. In spite of these advantages over other non-linear estimators, the Sample Entropy is not widely used. Sample Entropy measure has been used to evaluate the signal complexity of the cyclic behavior of HRV in obstructive sleep apnea syndrome [59] and for analysis of neonatal heart rate variability [51]. In this study, Sample Entropy is investigated for the first time as a feature extracted in the automatic detection of AOP.

For calculating the Sample Entropy, the embedding dimension \( m \) and vector comparison distance \( r \) must be specified. It is common to set the embedding dimension parameter \( m \) to be \( m = 1, 2 \) or \( 3 \) and to set the vector comparison distance (tolerance) \( r \) to be some percentage (usually 0.1-0.25) of the standard deviation of the time series so as not to depend on the absolute amplitude of the signal [48]. In our study, \( r \) of 0.20 of the standard deviation of the sequence and \( m \) of 2 were experimentally chosen to maximize the discriminative power of the features. Sample Entropy is the negative logarithm of the conditional probability that two sequences similar for \( m \) points remain similar at the next point, where self-matches are not included in calculating the probability. Thus, a larger value often corresponds to more irregularity or complexity in the time series data.

Let consider a breath cycle as the time from one peak to next adjacent peak. In most cases the breaths will be of different lengths, so in this case the shorter length is zero padded to make it of equal length. Suppose that \( br_n = [\text{loc}_{p_n} , \text{loc}_{p_n} + 1 : \text{loc}_{p_{n+1}}] \) is the \( n \)th breath cycle in the epoch where there are in total \( K \) consecutive breathing cycles \((n = 1, 2, ..., N)\). So the vector that includes all breath cycles will look like:

\[
BreathCycle_{vec} = [br_1 \ br_2 \ .... \ br_N]
\] (3.19)

the Sample Entropy measure of \( BreathCycle_{vec} \) is determined as shown in the following steps [48] [60]:

1. Given \( N \) data points from the time series (equation 3.19) which is rewritten as \( br(n) = [br(1), br(2), ..., br(N)] \) take \( m \) vectors \( Br_m(1), ..., Br_m(N-m+1) \) is defined as \( Br_m(i) = [br(i), br(i+1), ..., br(i+m-1)] \) for \( 1 \leq i \leq N-m+1 \). These vectors stand for \( m \) consecutive \( br \) values, starting at the \( ith \) sample.

2. Let \( r \) denote the noise filter level which is defined as:

\[
r = g \cdot std \quad \text{for} \quad g = 0.1, 0.2, ..., 0.5
\] (3.20)

where \( std \) represents the standard deviation of the data sequence \( Br \).

3. The distance between vectors \( Br_m(i) \) and \( Br_m(j) \), \( d[Br_m(i), Br_m(j)] \) is defined as the maximum absolute difference between their scalar components:
3. Methodology

\[ d \left[ Br_m(i), Br_m(j) \right] = \max_{k=0, \ldots, m-1} (|br(i+k) - br(j+k)|) \quad (3.21) \]

4. For a given \( Br_m(i) \) count the number of \( j \) \( (1 \leq i \leq N - m \ (j \neq i)) \) such that \( d \left[ Br_m(i), Br_m(j) \right] \leq r \). This is number is represented as \( B_i \). Then, for \( 1 \leq i \leq N - m \),

\[ B_i^m(r) = \frac{1}{N-m-1} B_i \quad (3.22) \]

Here, note that only the first \( N - m \) vectors of length \( m \) are considered in order to ensure that for \( 1 \leq i \leq N - m \), the vector \( Br_{m+1}(i) \) is also defined.

5. Define \( B^m(r) \) as:

\[ B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \quad (3.23) \]

Then the dimension is incremented to \( m + 1 \) and \( A_i \) is computed as the number of \( Br_{m+1}(i) \) within \( r \) of \( Br_{m+1}(j) \), for \( j \) \( (1 \leq j \leq N - m \ (j \neq i)) \). Then \( A_i^m(r) \) is defined as:

\[ A_i^m(r) = \frac{1}{N-m-1} A_i \quad (3.24) \]

and \( A^m(r) \) is defined as:

\[ A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \quad (3.25) \]

Thus, \( B^m(r) \) represents the probability that two sequences will match for \( m \) points, whereas \( A^m(r) \) represents the probability that two sequences will match for \( m + 1 \) points.

6. Finally, the sample entropy of \( br(n) = \{br(1), br(2), \ldots, br(N)\} \) vector is defined as:

\[ SampEn_{br} = -\ln \left[ \frac{A^m_{br}(r)}{B^m_{br}(r)} \right], \quad (3.26) \]

This feature is denoted as "Regularity of breathing cycle" (R30).

For measuring the regularity of inhalation to exhalation depth we compute the Sample Entropy of vector \( [(\text{loc}_{p_1} - \text{loc}_{t_1}), (\text{loc}_{t_1} - \text{loc}_{p_2}), (\text{loc}_{p_2} - \text{loc}_{t_2}), \ldots, (\text{loc}_{p_n} - \text{loc}_{t_n})] \) which is defined as:

\[ SampEn_{InEx} = -\ln \left[ \frac{A^m_{InEx}(r)}{B^m_{InEx}(r)} \right], \quad (3.27) \]

This feature is denoted as "Regularity of inhalation-exhalation depth" (R31).

All the new features that are proposed on this section are listed on Table 3.4.
3.4.3.2. Cardio-respiratory features

A pool of three new features are proposed in this section. Cardioventilatory coupling (CVC) has been extensively studied [61] [62] [63] and is defined as the initiation of inspiration by baroreceptor afferent nerve activity, such that inspiratory onset occurs at a constant time interval after a heart beat [64] [65]. We attempt to explore the correlation/synchronization between respiration and heart rate. For this purpose several features have been developed. In this study, cardioventilatory correlation is investigated for the first time as a feature extracted in the automatic detection of AOP.

Since there is a relation between heart beat and respiration rate, we down-sampled the chest impedance by using the ECG signal as a clock. More specifically, only the chest impedance value that corresponds to each R-peak detected within an epoch was considered resulting in the down-sampled version of chest impedance (see Figure 3.11). Then the maximum correlation between this signal and the RR normalized interval was calculated. This feature is denoted as "Synchronized correlation of RR interval and respiration" (C19).

In addition we wanted to check the cohesion between RR intervals and Inter-breath Interval (IBI). During normal breathing the respiration rate would contain small variation and the same appears to be the case for RR interval, while during apnea episodes the breathing pattern become more irregular usually accompanied with bradycardia (seen clearly on Figure 3.12). This behavior was tried to be captured by computing the maximum correlation of normalized RR interval and IBI based on peaks (C20) (see equation 3.9). Additionally, the maximum correlation between RR interval and inhalation-exhalation time periods was computed (C21).
3. Methodology

(a) Respiratory effort  
(b) ECG  
(c) IBI  
(d) RR Interval

Figure 3.12: Relation between RR Interval and IBI during apnea episodes and normal breathing

Table 3.5: List of proposed cardio-respiratory features

<table>
<thead>
<tr>
<th>Feature Index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C19</td>
<td>Synchronized correlation of RR interval and respiration</td>
</tr>
<tr>
<td>C20</td>
<td>Correlation of RR interval and IBI</td>
</tr>
<tr>
<td>C21</td>
<td>Correlation of RR interval and inhalation-exhalation time period</td>
</tr>
</tbody>
</table>

3.5. Feature Visualization

Feature visualization is important since it casts data into a format that can be grasped and understood much more quickly and easily than the raw numbers and equations alone. In this way the discriminative power of features becomes easily recognized and provides a more intuitive understanding of the features described on section 3.4. For this purpose box plots and histograms were used. Moreover, for each feature, the mean values of each feature of each apnea epoch and mean values of nine epochs preceding and following an apnea epoch as well as standard deviation of the mean were computed. This analysis was performed by pooling epochs over all subjects and is denoted as analysis of the apnea and non-apnea epochs.
Box plots and histograms are useful for comparing distributions. As shown in Figure 3.13, the new respiratory features were found to significantly differ between apnea (Figure 3.13 c: red line) and non-apnea (Figure 3.13 c: blue line) epochs for most features, as we can see from the histogram. This means that the information regarding respiratory depth and regularity estimated from chest impedance signal, which are indicators of some properties of respiratory physiology, is not independent of presence of apnea and thus it can be in turn used to classify apnea. It is worth noting that a comparison of existing and new respiratory features will enhance even more the discriminative power of the new respiratory features that are proposed in this work. Feature visualization of existing respiratory and cardiac features can be found in Appendix A.1.

The absolute standardized mean difference (ASMD) was used to measure the discriminative power of a single feature (existing and new). Given a feature $f$, it is computed as the absolute mean difference of the feature values between apnea and non-apnea epochs divided by the standard deviation of the values over all epochs from all subjects:

$$\text{ASMD}_f = \frac{|\mu_{\text{apnea}}^f - \mu_{\text{non-apnea}}^f|}{\sigma_f}$$

where $\mu_{\text{apnea}}^f$ and $\mu_{\text{non-apnea}}^f$ express the sample mean of apnea and non-apnea epochs, respectively, and $\sigma_f$ is the sample standard deviation. A higher discriminative power in separating the two classes translates to a larger ASMD value. ASMD was computed for both respiratory and cardiac features (see Figures 3.14 and 3.15). As it can be seen from Figures 3.14 the discriminative power of the new proposed features (Figure 3.14 features included in the red box) is high. Comparing new and existing respiratory features it is clear that the discriminative power of the new features in general reaches or in some cases exceeds the discriminative power of the four best representative of existing respiratory features (R7,R9,R10,R11). For cardiac features the discriminative power is among the highest, with features C19, C20 and C21 (Figure 3.15 features included in the red box) exceeding the discriminative power of all existing cardiac features, since the new cardio-respiratory features involve information from chest impedance signal. As seen from Figure 3.14 respiration features have higher discriminative power than cardiac features. The cause of this observation will be discussed in chapter 4.
Figure 3.13: Regularity of breathing cycle (R30). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. 

(a) Box plot of apnea and non-apnea (normal breathing) classes

(b) Analysis of the apnea and non-apnea (normal breathing) epochs

(c) Histogram of apnea and non-apnea (normal breathing) classes
3.5. **Feature Visualization**

Figure 3.14: Discriminative power of all respiration features (existing and new) computed by using ASMD.

Figure 3.15: Discriminative power of all cardiac features (existing and new) computed by using ASMD.
3.6. **CLASSIFICATION TECHNIQUE**

3.6.1. **SUPPORT VECTOR MACHINE**

SVM-based classifiers have been widely used for biological sequence analysis such as Sleep Apnea detection in adults [56] [66] [67] [68]. SVM is a machine learning method for obtaining the optimal boundary of two data sets which are non-linearly mapped in a very high dimension feature space; independently on the probabilistic distributions of training vectors in the sets. The fundamental idea of SVM is, separating the binary labeled training data with a hyperplane that has maximum distance from them, known as maximum margin hyperplane. Figure 2 illustrates the basic idea of SVM for a binary-class problem. The circles and squares represent samples of two classes “-1” and “+1”, respectively; x1 and x2 represent two features. SVM handles the class separation by using a kernel function to map the data into a different space with a hyperplane. The theory of SVMs can be briefly described as follows [69].

Consider a training set \((x_1, y_1), ..., (x_L, y_L)\), where \(y\) is either +1 or -1, and \(x_i\) is the input vector corresponding to the \(i\)th sample and labeled by \(y_i\), depending on its class. Each input \(x_i\) is projected to a higher-dimensional inner product space, called a feature space, by \(z = \phi(x)\), via a nonlinear mapping \(\phi : \mathbb{R}^n \rightarrow \mathbb{R}^d\). Considering the case when the data are linearly separable in \(\mathbb{R}_d\), there exists a vector \(w \in \mathbb{R}_d\) and a scalar \(b\) (hyperplane bias) that define the separating hyperplane as:

\[
\mathbf{w} \cdot \mathbf{x} + b = 0 \quad \text{such that} \quad y_i(w \cdot x_i) + b \geq 1 \quad \text{for} \quad i = 1, ..., L
\]  

(3.29)
Distance of each training data \( x_i \) from the hyperplane is given by:

\[
d_i = \frac{\mathbf{w} \cdot x_i + b}{||\mathbf{w}||}
\] (3.30)

Combining inequality 3.29 and 3.30, for all \( x_i \) result in:

\[
y_i d_i \geq \frac{1}{||\mathbf{w}||}
\] (3.31)

Therefore, \( \frac{1}{||\mathbf{w}||} \) is the lower bound on the distance between the training data \( x_i \) and the separating hyperplane.

The maximum margin hyperplane can be considered as the solution of the problem of maximizing the \( \frac{1}{||\mathbf{w}||} \) subject to the constraint 3.29, or equivalently by solving the following problem:

\[
\text{minimize } z = \frac{1}{2} \mathbf{w} \cdot \mathbf{w}
\] (3.32)

subject to the constraints  \( y_i (\mathbf{w} \cdot x_i) + b \geq 1 \) for \( i = 1, ..., L \) (3.33)

Constructing an optimal hyperplane is therefore a quadratic programming problem that can be solved by constructing a Lagrangian resulting in the following dual problem:

\[
\text{maximize } W(a) = \sum_{i=1}^{L} a_i - \frac{1}{2} \sum_{i,j=1}^{L} a_i a_j y_i y_j (z_i \cdot z_j)
\] (3.34)

subject to the constraints \( 0 \leq a_i \leq C \) \( i = 1, ..., L \) (3.35)

and the constraint \( \sum_{i=1}^{L} y_i a_i = 0 \) (3.36)

Let’s denote \((\alpha_1, \alpha_2, ..., \alpha_L)\) are the L nonnegative Lagrange multipliers associated with the constraints 3.29. Only a few \( \alpha_i \) will be greater than zero. The corresponding \( x_i \) are the support vectors which lie along the margins of the decision boundary. The regularization parameter \( C \) represents the tradeoff between the maximum margin and the minimum classification error.

By using kernels it is possible to make all the necessary operations in the input space by using \( (z_i \cdot z_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle = k(x_i, x_j) \) as \( k(x_i, x_j) \) an inner product in the feature space. The decision function can be written in terms of these kernels, as follows:

\[
y = \text{sign} \left( \sum_{\alpha_i > 0} y_i a_i k(x_i, x_j) + b \right)
\] (3.37)
In this work, we used the radial basis function as a kernel function:

$$k(x_i, x_j) = \exp \left( \frac{||x_i - x_j||^2}{2\sigma^2} \right)$$ (3.38)

Linear classifiers cannot deal with non-linearly repeatable data and/or noisy data. One solution for this problem would be to use a net of simple linear classifiers (neurons): a Neural Network. However they suffer from problems such as: local minima; many parameters; heuristics to train. Another solution is to map data into a richer feature space (as mentioned above) including non-linear features and then use a linear classifier. For these reasons, a radial basis function was used as a kernel function leading to a non-linear classifier.

3.6.2. Feature Selection

Variable and feature selection have become the focus of much research in areas of application for which datasets with many variables are available. The objective of variable selection is three-fold: improving the prediction performance of the predictors, providing faster and more cost-effective predictors, and providing a better understanding of the underlying process that generated the data. The central premise when using a feature selection technique is that the data contains many features that are either redundant or irrelevant, and can thus be removed without incurring much loss of information [70]. Redundant or irrelevant features are two distinct notions, since one relevant feature may be redundant in the presence of another relevant feature with which it is strongly correlated [71]. In this work Minimal-redundancy-maximal-relevance (mRMR) and Correlation-based Feature Selection (CFS) algorithms will be used.

Peng et al. [72] proposed a feature selection method that can use mutual information to select features. The aim is to penalize a feature’s relevancy by its redundancy in the presence of the other selected features. The relevance (maximum relevance criterion) of a feature set S for the class c is defined by the average value of all mutual information values between the individual feature $x_i$ and the class c as follows:

$$\max D(S, c), \quad D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c)$$ (3.39)

It is likely that features selected according to Max- Relevance could have rich redundancy, i.e., the dependency among these features could be large. When two features highly depend on each other, the respective class-discriminative power would not change much if one of them were removed. Therefore, the following minimal redundancy condition can be added to select mutually exclusive features. The redundancy of all features in the set S is the average value of all mutual information values between the feature $x_i$ and the feature $x_j$: 

$$\sum_{x_i \in S, x_j \in S} I(x_i; x_j)$$
3.6. Classification Technique

\[
\min R(S), \quad R = \frac{1}{\lvert S \rvert^2} \sum_{x_i, x_j \in S} I(x_i; x_j)
\]  

(3.40)

The “minimal-redundancy-maximal-relevance” criterion is a combination of two measures given in equations 3.39 and 3.40 and is defined as follows:

\[
mRMR = \max_S \left[ \frac{1}{\lvert S \rvert} \sum_{x_i \in S} I(x_i; c) - \frac{1}{\lvert S \rvert^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \right]
\]  

(3.41)

The mRMR algorithm is an approximation of the theoretically optimal maximum-dependency feature selection algorithm that maximizes the mutual information between the joint distribution of the selected features and the classification variable. As mRMR approximates the combinatorial estimation problem with a series of much smaller problems, each of which only involves two variables, it thus uses pairwise joint probabilities which are more robust. More details about mRMR can be found in [72].

The CFS measure evaluates subsets of features on the basis of the following hypothesis: "A good feature subset is one that contains features highly correlated with (predictive of) the class, yet uncorrelated with (not predictive of) each other" [73]. Specifically, CFS is a supervised algorithm that towards finding an 'optimal' feature subset \( S \) containing features uncorrelated with each other and highly correlated with the classes. The following equation gives the merit of a feature subset \( S \) consisting of \( k \) features:

\[
\text{Merit}_{S_k} = \frac{kr_{cx}}{\sqrt{k(k+1)r_{xx}}}
\]  

(3.42)

Here, \( r_{cx} \) is the average value of all feature-class correlations, and \( r_{xx} \) is the average value of all feature-feature correlations. The CFS criterion is defined as follows:

\[
\text{CFS} = \max_{S_k} \left[ \frac{r_{cx_1} + r_{cx_2} + \cdots + r_{cx_k}}{\sqrt{k + 2(r_{x_1 x_2} + \cdots + r_{x_i x_j} + \cdots + r_{x_k x_1})}} \right]
\]  

(3.43)

The \( r_{cx_i} \) and \( r_{x_i x_j} \) variables are referred to as correlations. More details about CFS can be found in [73]. In practice, a Matlab based computational framework as part of the Siento framework [74], was used for feature selection. preprocessing

3.6.3. Applied Approach

The system presented in this work is comprised of three main components: a data extraction and preprocessing stage, a feature extraction stage and a SVM classifier. The structure of the system is shown in Figure 3.17.

Given the set of ECG and respiration signals described above, the design of an automated apnea detection system based on those signals is considered. In designing the ap-
Apnea detector, features are extracted from each 20-second epoch which are consistent with the definition of apnea. In addition, we performed a subject-specific Z-normalization for each feature. It was done per subject by subtracting the mean of feature values and dividing by their standard deviation. This allows for reducing physiological and equipment-related variations from subject to subject, thereby enhancing the discrimination between apnea and not apnea epochs.

The goal of this work is the detection of apnea continuously. Since annotation is performed continuously there are four possible options of apnea segments involved in a epoch as seen in Figure 3.18. For implementing and evaluating the automated apnea detection system, we considered an epoch to be apneic if at least 55% (any combination showed in Figure 3.18) of its duration was annotated as apnea.

A subject-independent approach was first considered where the classifier to detect apnea epochs for each patient was trained by the data from the other patients. With this approach, a leave-one-out cross validation (LOOCV) was conducted. During each iteration of the LOOCV procedure, data from nine recordings were used for training and the remaining one was used for testing (see Figure 3.19). After the cross validation, classification results obtained for each patient in each iteration's testing set were collected, and performance metrics (mean) were then computed to evaluate the classifier.

In addition, we considered a subject-specific approach, performing the apnea detection for each patient based on the classifier trained by the data from the same patient. A 10-fold cross validation (10-fold CV) was conducted to evaluate this approach. First, the
data of each patient was randomized so that both training and testing sets would contain apnea and non-apnea (normal) epochs. During each iteration of the 10-fold CV procedure, 90% of the data of each patient was used to train the classifier, while the remaining 10% was used for testing (see Figure 3.19). After that, classification results obtained for each patient in the testing set of each iteration were collected, and performance metrics (mean) were then computed.

3.7. EVALUATION

The classification performance was first evaluated using a traditional metric of overall accuracy. However, a strong imbalance between classes observed in our data made this metric less appropriate for evaluation. The Cohen’s Kappa coefficient [75] is a measure that takes into account the prior probability of a specific class occurring. This means that it offers a better understanding of the general classification performance in correctly identifying both classes. This will be illustrated with an example.

Cohen’s Kappa Coefficient [75] is a measure of interrater agreement, where the two raters in our case are the two neonatologists and the automated apnea detection system. Cohen’s Kappa Coefficient varies from one for perfect agreement to zero for a performance no better than chance. The need for such a measure is evident when for example in the dataset there are 80 non-apnea epochs and 20 apnea epochs. Therefore, even the system accurately detects all non-apnea epochs and but none of the apnea epochs we will still achieve 80% accuracy, which may appear to be quite a reasonable performance but misleading. However, in this instance kappa will be zero, which is a better measure of performance.

Let’s consider that 2 raters, A and B, are allowed to classify 20 epochs into one of two classes (apnea and non-apnea). As shown on Table 3.6, in ten time periods (upper left cell) the raters agreed that there was no apnea. In three time periods (lower right cell) did both raters agree that apnea had occurred. There was four times period (upper right cell)
in which rater 2 thought that apnea occurred, although rater 1 did not. Conversely, there was three times period (lower left cell) in which Rater 1 thought that apnea had occurred, although rater 1 did not. So two raters agree 13 times out of 20 observations giving an accuracy of 65%.

<table>
<thead>
<tr>
<th>Rater 1</th>
<th>Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-apnea</td>
<td>10</td>
</tr>
<tr>
<td>apnea</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3.6: Confusion matrix.

To compute Kappa, we first need to calculate the observed level of agreement:

\[ p_0 = \frac{10}{20} + \frac{3}{20} = 0.65 \]  

(3.44)

This value needs to be compared to the value that you would expect if the two raters were totally independent (total proportion of agreement expected by chance):

\[ p_e = \frac{14}{20} \cdot \frac{13}{20} + \frac{6}{20} \cdot \frac{7}{20} = 0.56 \]  

(3.45)

The value of Kappa is defined as

\[ \kappa = \frac{p_0 - p_e}{1 - p_e} = \frac{0.09}{0.44} = 0.2045 \]  

(3.46)

Another evaluation metric that will be used in this analysis is Area Under the Curve (AUC). AUC can be obtained either from Receiver Operator Characteristic (ROC) curve or from Precision-Recall (PR) curve. ROC are commonly used to evaluate classification performance. However, when dealing with highly imbalance dataset PR curve gives a more informative picture of an algorithm’s performance. This will be illustrated with an example.

Let’s consider a dataset that consists of 1000 epochs, where 100 of them are apneic. Let’s assume that two algorithms will be compared. The confusion matrix of the first algorithm can be seen in Table 3.7.

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>apnea</td>
<td>90</td>
</tr>
<tr>
<td>non-apnea</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 3.7: Confusion matrix of the first algorithm
False positive rate (FPR), true positive rate (TPR) or recall and precision are defined as:

\[
FPR = \frac{FP}{FP + TN} = 0.0111 \tag{3.47}
\]

\[
TPR = \frac{TP}{TP + FN} = 0.9 \tag{3.48}
\]

\[
\text{precision} = \frac{TP}{TP + FP} = 0.9 \tag{3.49}
\]

The confusion matrix of the second algorithm can be seen in Table 3.8.

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
<th>apnea</th>
<th>non-apnea</th>
</tr>
</thead>
<tbody>
<tr>
<td>apnea</td>
<td>90</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>non-apnea</td>
<td>10</td>
<td>790</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.8: Confusion matrix of the second algorithm

False positive rate (FPR), true positive rate (TPR) or recall and precision are defined as:

\[
FPR = \frac{FP}{FP + TN} = 0.1222 \tag{3.50}
\]

\[
TPR = \frac{TP}{TP + FN} = 0.9 \tag{3.51}
\]

\[
\text{precision} = \frac{TP}{TP + FP} = \frac{90}{90 + 110} = 0.45 \tag{3.52}
\]

Obviously, those are just single points in ROC and PR space, but if these differences persist across various scoring thresholds, using AUC from ROC curve, we see a very small difference between the two algorithms, whereas AUC from PR curve would show quite a large difference. To compare both methods, using AUC from ROC curve, we see that the FPR has a difference of 0.1111, which is very small. However, by using AUC from PR curve, we see that the Precision has a difference of 0.45 which is much more pronounced.
In this chapter the results of the study will be presented focusing on displaying the behavior of the automated detection algorithm with respect to the annotator’s detection. First we will present statistics between the two annotators. Classification will be applied based on respiratory, cardiac and combination of cardiac and respiratory features. CFS and mRMR feature selection algorithms will be compared as well as annotations cases (annotations based on first annotator, second annotator and both annotators). The term both annotators refers to the intersection of apnea and non-apnea epochs between the two annotators. In this analysis two approaches will be implemented, subject-independent and subject-specific. In addition, we will investigate the impact of duration of the epoch and the percentage of apnea included in an apneic epoch, on the apnea detection performance of the algorithm. Last, we will compare the performance of our system with the approach that Lee et al [3] proposed. The objective of this comparison is to investigate whether or not chest impedance after removal of cardiac interference includes more information than the classic approach of preprocessing that is used in this work.

4.1. **General classification of dataset by annotators**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>First Annotator</th>
<th>Second Annotator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (%)</td>
<td>10.87 ± 6.16</td>
<td>5.91 ± 3.27</td>
</tr>
<tr>
<td>Epochs (%)</td>
<td>28.07 ± 8.80</td>
<td>9.82 ± 4.64</td>
</tr>
</tbody>
</table>
4.2. BASED ON RESPIRATORY FEATURES

Table 4.2: Statistics of mean agreement between the two Annotators

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Time</th>
<th>Epochs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Cohen's Kappa Coefficient</td>
<td>0.482 ± 0.117</td>
<td>0.334 ± 0.102</td>
</tr>
<tr>
<td>General agreement (%)</td>
<td>92.50 ± 3.29</td>
<td>78.86 ± 5.96</td>
</tr>
<tr>
<td>Apnea agreement (1st agrees with 2nd) (%)</td>
<td>73.92 ± 12.94</td>
<td>83.85 ± 8.54</td>
</tr>
<tr>
<td>Apnea agreement (2nd agrees with 1st) (%)</td>
<td>40.79 ± 12</td>
<td>29.28 ± 9.52</td>
</tr>
</tbody>
</table>

Tables 4.1-4.2 show the statistics of the occurrence of apnea for both annotators and the statistics of agreement between the two annotators. Both statistics were computed based on time and epochs.

Annotations were performed after preprocessing chest impedance signal based on the approach that was described on section 3.4.1. In Table 4.1 we see that apnea occurrence is increased based on epochs compared to occurrence of apnea based on time. This is due to the fact that an epoch is considered apneic if at least 55% of its duration was annotated as apnea. The apnea occurrence in the dataset based on epochs was increased for about 2.58 times compared to the occurrence of apnea based on time for the first annotator and about 1.66 times for the second annotator. This means that apneic segments annotated from the second annotator, occurred mostly to be equated to the whole epoch while for the first annotator it seems that this was not the case (see Figure 3.18). This large difference is not highlighted to the same extent if we compare mean Cohen’s kappa coefficient based on time and epochs.

In Table 4.2 general agreement (if both annotators agree that an epoch is apneic or non-apneic-inter) is high. However, this alone does not include sufficient information since prevalence of apnea is low compared to the prevalence of non-apnea. Furthermore, it should be underlined that the first annotator agrees on average (for epochs and time) 78.39% with the second annotator about apnea segments while the average percentage of apnea agreement of the second annotator to the first annotator is lower, 35.04%.

4.2. BASED ON RESPIRATORY FEATURES

Figure 4.1 compares the ROC and PR curves of the apnea detector after applying the first, second and both annotators for CFS feature selection algorithm. This comparison includes both subject-independent and subject-specific approaches. Maximum AUC (from PR curve) and Cohen's kappa coefficient were achieved for the first annotator. Detailed results can be seen on Tables 4.3-4.5. The apnea detection system achieved a sensitivity of 0.851 and a specificity of 0.758 for subject-independent approach, while for subject-specific approach sensitivity of 0.877 and specificity of 0.844 were reached. Table 4.3 illustrates that for subject-specific approach apnea detector performs better compared to
4. RESULTS

Figure 4.1: ROC and PR curves for both Subject-Independent and Subject-Specific for CFS algorithm

The subject-independent approach (based on Kappa and AUC-PR curve). However, results from both approaches do not differ significantly as shown on Tables 4.3-4.5.

Table 4.3: Classification results using respiratory features and CFS for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.778 ± 0.020</td>
<td>0.850 ± 0.033</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.715 ± 0.049</td>
<td>0.752 ± 0.065</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.881 ± 0.034</td>
<td>0.881 ± 0.032</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.851 ± 0.078</td>
<td>0.877 ± 0.077</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.758 ± 0.025</td>
<td>0.844 ± 0.030</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.512 ± 0.044</td>
<td>0.533 ± 0.052</td>
</tr>
<tr>
<td>Feature's Subset</td>
<td>R4, R5, R6, R7, R10, R11, R18, R19, R21, R27, R30</td>
<td></td>
</tr>
</tbody>
</table>

Note that features R4, R10, R18, R19, R21 and R30 are selected for all three cases of annotations by using CFS algorithm. These feature seem robust and do not depend on the annotations. Four out of the six robust features (R18, R19, R21 and R30) belong to the new features that are proposed in this work. Feature selection for both annotators include more features compared to the ones selected for the first and second annotator individually. This happens due to the fact that feature selection tries to combine and include the physiology.
Table 4.4: Classification results using respiratory features and CFS for the second annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.800 ± 0.024</td>
<td>0.843 ± 0.027</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.501 ± 0.140</td>
<td>0.503 ± 0.168</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.881 ± 0.034</td>
<td>0.881 ± 0.032</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.809 ± 0.054</td>
<td>0.808 ± 0.107</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.799 ± 0.026</td>
<td>0.845 ± 0.031</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.338 ± 0.115</td>
<td>0.359 ± 0.113</td>
</tr>
</tbody>
</table>

**Feature's Subset** R1, R4, R6, R9, R10, R16, R18, R19, R21, R30

Table 4.5: Classification results using respiratory features and CFS for the both annotators

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.849 ± 0.026</td>
<td>0.882 ± 0.032</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.641 ± 0.168</td>
<td>0.649 ± 0.162</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.937 ± 0.015</td>
<td>0.939 ± 0.017</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.886 ± 0.054</td>
<td>0.842 ± 0.117</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.848 ± 0.030</td>
<td>0.885 ± 0.031</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.457 ± 0.148</td>
<td>0.494 ± 0.137</td>
</tr>
</tbody>
</table>

**Feature's Subset** R1, R4, R5, R9, R10, R11, R15, R16, R18, R19, R20, R21, R26, R27, R30

highlighted from both annotators.

Table 4.6: Classification results using respiratory features and mRMR for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.778 ± 0.022</td>
<td>0.847 ± 0.024</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.722 ± 0.050</td>
<td>0.727 ± 0.078</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.881 ± 0.035</td>
<td>0.882 ± 0.033</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.847 ± 0.078</td>
<td>0.877 ± 0.075</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.760 ± 0.023</td>
<td>0.840 ± 0.022</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.512 ± 0.044</td>
<td>0.532 ± 0.054</td>
</tr>
</tbody>
</table>

**Feature's Subset** R23, R31, R6, R20, R4, R5, R11, R26, R18, R30, R10

Figure 4.2 compares the ROC and PR curves of the apnea detector after applying the first, second and both annotators for mRMR feature selection algorithm. This comparison includes both subject-independent and subject-specific approaches. Maximum AUC (from PR curve) and Cohen's kappa coefficient were achieved for the first annotator. The apnea detection system achieved a sensitivity of 0.847 and a specificity of 0.76 for subject-independent approach, while for subject-specific approach sensitivity of 0.877 and specificity of 0.84 were reached. Detailed results can be seen on Tables 4.6-4.8.
4. RESULTS

(a) Subject-Independent

(b) Subject-Specific

(c) Subject-Independent

(d) Subject-Specific

Figure 4.2: ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm

Note that features R4, R10, and R30 are selected for all three cases of annotations by using mRMR algorithm. Note that R31 was selected from mRMR algorithm for both the first and second annotators. These features seem robust and do not depend on the annotations.

Table 4.7: Classification results using respiratory features and mRMR for the second annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.801 ± 0.025</td>
<td>0.839 ± 0.028</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.507 ± 0.133</td>
<td>0.506 ± 0.157</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.891 ± 0.025</td>
<td>0.893 ± 0.025</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.816 ± 0.056</td>
<td>0.799 ± 0.145</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.799 ± 0.028</td>
<td>0.841 ± 0.029</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.342 ± 0.114</td>
<td>0.360 ± 0.114</td>
</tr>
</tbody>
</table>

Feature's Subset      R31, R1, R16, R21, R8, R9, R30, R19, R4, R10

Tables 4.3 - 4.5 and 4.6- 4.8 illustrate that there was no significant difference between
## 4.2. Based on Respiratory Features

Table 4.8: Classification results using respiratory features and CFS for the both annotators

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.85 ± 0.027</td>
<td>0.872 ± 0.031</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.645 ± 0.164</td>
<td>0.661 ± 0.196</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.937 ± 0.014</td>
<td>0.938 ± 0.019</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.885 ± 0.055</td>
<td>0.831 ± 0.193</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.849 ± 0.031</td>
<td>0.875 ± 0.036</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.458 ± 0.148</td>
<td>0.492 ± 0.134</td>
</tr>
</tbody>
</table>

**Feature's Subset**: R27, R17, R16, R5, R1, R9, R8, R20, R21, R26, R30, R4, R19, R18, R10

The two feature selection algorithms in terms of performance of the apnea detection system. This means that the developed system is robust. In addition, feature subset R4, R10, R18, R19, R21 and R30 was the most discriminative since it was selected in almost all cases (see Table 4.9). These features are very robust and do not depend on the annotations and the choice of feature selection algorithm. Note that four out of the six robust features (R18, R19, R21 and R30) belong to the new features that are proposed in this work.

Table 4.9: All new respiratory features selected at least once

<table>
<thead>
<tr>
<th>Features</th>
<th>CFS</th>
<th>mRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Annotator</td>
<td>Second Annotator</td>
</tr>
<tr>
<td>R17</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R18</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R19</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R20</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R21</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R23</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R26</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R27</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R30</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>R31</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
4.3. BASED ON CARDIAC FEATURES

Detection performance based on cardiac features deteriorates compared to the performance based on respiratory features. This was expected due to the fact that apnea is not always followed by cardiac changes. As a result discriminative power of cardiac features (see Figure 3.15) is lower compared to discriminative power of respiratory features (see Figure 3.14).

![Figure 4.3: ROC and PR curves for both Subject-Independent and Subject-Specific for CFS algorithm](image)

Figure 4.3 compares the ROC and PR curves of the apnea detector after applying the first, second and both annotators for CFS algorithm. This comparison was performed for both subject-independent and subject-specific approaches. Note that the apnea detection performance between the first and second annotator differs significantly, since it seems that apnea episodes annotated by the second annotator do not seem to be strongly reflected in the cardiac activity. This is also the reason why for the first annotator three cardio-respiratory features and one cardiac feature are selected, while for the first annotator nine features were selected, six cardiac and three cardio-respiratory (see Table 4.11). Detailed results can be seen in the Appendix A.2. Apneas annotated by the first annotator seem to correlate well with cardiac changes and Table 4.10 shows that the system performs well enough based on cardio-respiratory information. In order to investigate the cardio-respiratory correlation and relation, classification performed based on cardiac and respiratory features follows.
### Table 4.10: Classification results using cardiac features and CFS algorithm for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.685 ± 0.047</td>
<td>0.767 ± 0.045</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.547 ± 0.064</td>
<td>0.567 ± 0.107</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.761 ± 0.066</td>
<td>0.761 ± 0.066</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.699 ± 0.086</td>
<td>0.693 ± 0.065</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.687 ± 0.042</td>
<td>0.785 ± 0.053</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.318 ± 0.075</td>
<td>0.331 ± 0.080</td>
</tr>
</tbody>
</table>

**Feature's Subset** C17, C19, C20, C21

### Table 4.11: Classification results using cardiac features and CFS algorithm for the second annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.656 ± 0.027</td>
<td>0.598 ± 0.034</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.239 ± 0.085</td>
<td>0.303 ± 0.159</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.717 ± 0.068</td>
<td>0.709 ± 0.060</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.656 ± 0.104</td>
<td>0.572 ± 0.210</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.658 ± 0.024</td>
<td>0.600 ± 0.038</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.130 ± 0.048</td>
<td>0.131 ± 0.053</td>
</tr>
</tbody>
</table>

**Feature's Subset** C2, C8, C9, C11, C13, C18, C19, C20, C21

### 4.4. Based on respiratory and cardiac features

The system performs relatively the same for both features selection algorithms. However, the mRMR algorithm relies on the user to specify the number of features to be selected and, thus, the performance of the classifier could be degraded when the number of features used is less or more than the optimal one. Therefore, in order to avoid these problems, the CFS algorithm, which does not require manual selection of the number of features, will be used from now on.

Figure 4.4 compares the ROC and PR curves of the apnea detector after applying the first, second and both annotators for both subject-independent and subject-specific approaches. Maximum AUC (from PR curve) and Cohen’s kappa coefficient were achieved for the first annotator. Detailed results can be seen on Tables 4.12 - 4.14. The apnea detection system achieved sensitivity of 0.851 and specificity of 0.76 for subject-independent approach, while for subject-specific approach a sensitivity of 0.883 and a specificity of 0.877 were reached. Table 4.12 illustrates that for subject-specific approach apnea detector performs better compared to the subject-independent approach. However, results from both approaches do not differ significantly as shown on Tables 4.12-4.14.
The performance of the apnea detection system based on a combination of respiratory and cardiac features exceeds the performance achieved when only cardiac features are used. We notice that the system performs equally well and relatively the same when only respiratory features or a combination of respiratory and cardiac features are used.

Table 4.12: Classification results using respiratory and cardiac features and CFS for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.779 ± 0.021</td>
<td>0.851 ± 0.025</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.723 ± 0.052</td>
<td>0.741 ± 0.069</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.882 ± 0.076</td>
<td>0.883 ± 0.034</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.651 ± 0.078</td>
<td>0.677 ± 0.076</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.760 ± 0.022</td>
<td>0.645 ± 0.034</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.513 ± 0.043</td>
<td>0.537 ± 0.07</td>
</tr>
</tbody>
</table>

Feature's Subset: R4, R5, R6, R10, R11, R18, R19, R21, R26, R27, R30, C2, C21
4.5. IMPACT OF EPOCH DURATION AND PERCENTAGE OF APNEA PER EPOCH

In this sector the impact of epoch duration and percentage of apnea per epoch in order to be considered apneic epoch will be investigated. This investigation will take place for classification based on both respiratory and cardiac features. Results based on cardiac features are presented in section A.3. For this purpose annotations from the first annotator and CFS algorithm will be considered.

4.5.1. RESPIRATORY FEATURES

Figure 4.5 demonstrates AUC from ROC and PR curves, sensitivity and specificity achieved for varying epoch duration. Maximum Cohen’s Kappa coefficient of 0.569 was achieved for epoch duration of 60 second.
Small window sizes may not contain enough information to sufficiently reflect the current state of the epoch and could split an apnea activity leading to suboptimal information for an activity classification algorithm. Longer window sizes can perhaps consider more information which is helpful to recognize the current state.

Another parameter that influence the apnea detection performance is the percentage of apnea included per epoch which defines apnea and non-apnea distributions. Figure 4.6 illustrates the apnea detection performance for varying percentage of apnea consisted in an epoch of 20 seconds. Increasing the percentage of apnea for instance to 65% means that all epoch which include apnea percentage between zero and 64.9% will be considered non-apnea. As a result, this will deteriorates the detection performance and specificity since features will capture apneic behavior/physiology. This is clearly demonstrated by AUC from PR curve on Figure 4.6. Maximum Cohen's Kappa coefficient of 0.536 was achieved for a percentage of 45%.
4.6. COMPARISON

Figure 4.7: ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm

Lee et al [3] developed an algorithm for detecting central apnea by filtering the cardiac artifact from chest impedance, as mentioned in section 2.3. In this section we use chest impedance that results after filtering cardiac artifact as it was developed by Lee et al [3]. Then we will perform feature selection and classification. This analysis will be carried out for both respiratory and cardiac features. Then we will compare the performance of our system with the Lee’s approach by using annotations performed from the first annotator. The objective of this comparison is to investigate whether or not chest impedance after removal of cardiac interference includes more information than the classic approach of preprocessing that is used in this work.

Figure 4.7 compares the ROC and PR curves of ours and Lee’s approach for subject-independent and subject-specific system based on respiratory features. A comparison between Table 4.15 and Table 4.16 yields that the automated apnea detection system implemented in this work performs slightly better. This suggests that cardiac interference in chest impedance does not constitute the general case. It may also mean that the developed features extract efficient information and do not misinterpret apneic events. Note that for Lee’s approach 13 features are selected, while for our approach 11. For Lee’s approach eight out of 13 belong in to the group of new features proposed in this work. There are eight features (R5, R7, R10, R18, R19, R21, R27, R30) selected from both approaches. Features R10, R18, R19, R21 and R30 proved to be the most powerful and discriminative features, as it was also mentioned in section 4.3. Only feature R10 belongs to the existing features of
Table 4.15: Classification results using respiratory features and CFS for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.778 ± 0.020</td>
<td>0.850 ± 0.033</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.715 ± 0.049</td>
<td>0.752 ± 0.065</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.881 ± 0.034</td>
<td>0.881 ± 0.032</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.851 ± 0.078</td>
<td>0.877 ± 0.077</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.758 ± 0.025</td>
<td>0.844 ± 0.030</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.512 ± 0.044</td>
<td>0.533 ± 0.052</td>
</tr>
</tbody>
</table>

**Feature's Subset**: R4, R5, R6, R7, R10, R11, R18, R19, R21, R27, R30

Table 4.16: Classification results using respiratory features from chest impedance after applying Lee’s et al [3] method

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.769 ± 0.022</td>
<td>0.839 ± 0.034</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.703 ± 0.052</td>
<td>0.743 ± 0.073</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.875 ± 0.037</td>
<td>0.872 ± 0.036</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.846 ± 0.078</td>
<td>0.872 ± 0.077</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.747 ± 0.025</td>
<td>0.830 ± 0.030</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.496 ± 0.045</td>
<td>0.513 ± 0.055</td>
</tr>
</tbody>
</table>

**Feature's Subset**: R1, R5, R7, R9, R10, R17, R18, R19, R20, R21, R26, R27, R30

On the other hand, Figure 4.8 compares the ROC and PR curves of this work and Lee's approach for subject-independent and subject-specific system based on cardiac features. A comparison between Table 4.17 and Table 4.18 (Cohen’s kappa coefficient) yields that the automated apnea detection system implemented in this work performs slightly better. The main difference the two approaches is the features that are selected in each case. Since with Lee’s approach chest impedance is synchronized to heartbeats and then the cardiac interference is removed, feature C19 cannot yield any information for apnea detection, and therefore is not selected. However, in our approach C19 includes efficient information for apnea detection. This can be seen clearly on Figure 4.9. Note that for Lee’s approach three features are selected, while for our the same three features are selected (C17, C20, C21) and feature C19. Features C20, and C21 proved to be the most powerful and discriminative features, as it was also mentioned in section A.2. These features belong to the new proposed cardiac features introduced in this work. Lee et al [3] removed the cardiac artifact while we extract discriminative information from it (C19).

In conclusion, chest impedance after the removal of cardiac interference do not include more information than the classic approach of preprocessing that is used in this work, since the apnea detection system performs relatively the same in both approaches.
4.6. Comparison

(a) Subject-Independent

(b) Subject-Specific

(c) Subject-Independent

(d) Subject-Specific

Figure 4.8: ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm

Table 4.17: Classification results using cardiac features and CFS algorithm for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.685 ± 0.047</td>
<td>0.767 ± 0.045</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.547 ± 0.064</td>
<td>0.567 ± 0.107</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.761 ± 0.066</td>
<td>0.761 ± 0.066</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.699 ± 0.086</td>
<td>0.693 ± 0.065</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.687 ± 0.042</td>
<td>0.785 ± 0.053</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.318 ± 0.075</td>
<td>0.331 ± 0.080</td>
</tr>
</tbody>
</table>

Feature's Subset: C17, C19, C20, C21
Table 4.18: Classification results using cardiac features and chest impedance after applying Lee’s et al. [3] method

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.686 ± 0.050</td>
<td>0.737 ± 0.028</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.545 ± 0.066</td>
<td>0.556 ± 0.097</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.756 ± 0.068</td>
<td>0.756 ± 0.066</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.676 ± 0.080</td>
<td>0.706 ± 0.061</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.697 ± 0.046</td>
<td>0.741 ± 0.036</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.313 ± 0.081</td>
<td>0.313 ± 0.087</td>
</tr>
<tr>
<td>Feature’s Subset</td>
<td>C17, C20, C21</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.9: Histograms of apnea and non-apnea classes and mean value of apnea epoch and of nine epochs preceding and following apnea
Discussion and Conclusions

Apnea of prematurity is common in the NICU. It is known from literature that apnea is under detected in the current patient monitors. In this work, we developed a method for advanced apnea detection. The data set includes ten very preterm babies and 90 hours (nine hours per baby) of continuously monitored raw waveforms of one-lead ECG and chest impedance signals. This dataset was annotated by two neonatologists.

In this study, two annotators defined periods of apnea in the signals. It is known from literature that there is usually a large degree of inter-observer variability. The system performed better based on annotations incurred from the first annotator, than those annotations from the second annotator for all feature categories (respiratory, cardiac and combination of respiratory and cardiac). Furthermore, apnea annotated from the first annotator seemed to be strongly related to the cardiac activity, while this relation was limited for the second annotator. Annotations based on the first annotator were then considered for the remaining analysis. It should be pointed out that a disagreement between annotators is to be expected since there is lack of a clear definition of AOP. As mentioned in section 3.3, annotation was performed based on the chest impedance, as well as other vital signs that are associated with apnea such as oxygen saturation and heart rate. This caused the annotation procedure to be complex since it includes more than one signal. In addition, Vergales et al. [38] also mentioned that nursing records do not provide sufficiently reliable annotation/documentation of AOP. Lommen et al. [76] highlights a low inter-observer agreements for annotating electroencephalogram (EEG) signals.

A system that detects central apnea based on respiratory, cardiac and a combination of respiratory and cardiac features was developed in this study. In addition, feature selection was applied for selecting the most discriminative subset of features. For respiratory features the system obtained a sensitivity of 85.1%, a specificity of 75.8% and a Kappa of 0.512 for subject-independent approach. The subject-specific system performs better (sensitivity of 87.7%, specificity of 84.4% and Kappa of 0.533) than the subject-independent system. The false-positive rate reduced from 24.18% for the subject-independent approach to 15.64% for the subject-specific approach. Furthermore, a sensitivity of 85.1%, a specificity of 76% and a Kappa of 0.513 were achieved when applying the subject-independent approach for the combination of respiratory and cardiac features. Results for the subject-
specific approach were further improved and attained a sensitivity of 87.7%, a specificity of 84.5% and a Kappa of 0.537. For this case, the false-positive rate reduced from 23.99% for the subject-independent approach to 15.48% for the subject-specific approach. These results illustrate that for the subject-specific approach, apnea detector performs better when compared with the subject-independent approach (based on the Kappa and AUC-PR curve). However, results from both approaches do not differ significantly as shown on Tables 4.3-4.5. Since results from both approaches do not differ significantly for all annotators, it implies that the pathophysiology of apnea expressed by respiratory activity (represented by chest impedance signal) exhibits high variability between and within patients.

From this analysis the most discriminative and robust features were identified as R4, R10, R18, R19, R21, R30, C19, C20 and C21. With the exception of R4 and R10, the remaining features belong in the new respiratory and cardio-respiratory features proposed in this work. This can be attributed to the fact that the features of the existing literature have been designed for detecting sleep apnea and sleep staging in adults, whereas the new features proposed in this work are introduced for detecting AOP continuously with classification techniques. To the best of our knowledge, this is the first work in literature that deals with the automatic detection of AOP in neonates based on classification techniques. All the features that stood out investigate discriminative information based on: i) regularity of breathing (R10, R18, R30); ii) correlation of peak-toughs sequences and inhalation-exhalation cycle (R19, R21); and iii) power spectrum of chest impedance signal in HF band (0.4-1.5 Hz).

A comparison between Kappa values achieved for Lee's approach and the approach of this study based only on the preprocessing procedure of chest impedance signal (Lee implemented a different algorithm of detecting apnea that follows the preprocessing procedure of chest impedance signal), shows that the automated apnea detection system implemented in this work performs slightly better. The main difference between the two approaches is the way in which they deal with synchronization of chest impedance and heartbeats. With Lee's approach, chest impedance is synchronized to heartbeats, after which the cardiac interference is removed. In this work, synchronization of chest impedance to heartbeats and the proceeding extraction of discriminative information from this relation is performed. Although both approaches achieve almost the same performance in detecting apnea, by removing cardiac interference/cardiographic synchronization, information for apnea detection could potentially be deduced.

Several limitations of this study need to be acknowledged. Firstly, it should be noted that although 4536 (based on first annotator) apnea epoch out of 16200 epochs were used, only ten preterm infants were actually included in this study. A larger cross-section of patients is needed, perhaps divided by prematurity. Secondly, only the SVM classifier was used, exclusive focus was placed on analyzing new features for central apnea classification. Nevertheless, additional classification algorithms merit investigation in future work. Another parameter that can be optimized in the proposed approach is the number of different classes included in the analysis. As in other classification approaches, increasing the number of classes may lead to better performance. For instance, classes could have been defined to represent: 1) the regular breathing, 2) irregular breathing, 3) periodic breathing, 4) apnea-bradycardia events and 5) apnea-bradycardia-desaturation episodes. However,
increasing the number of classes leads to a higher degree of complexity.

To conclude, in this paper, respiratory, cardiac and cardio-respiratory features were analyzed, quantified and was found to differ between apnea and non-apnea segments. Furthermore, 18 novel features that characterize different aspects of chest impedance and ECG signals were extracted for automated apnea classification. Results from this work indicate that the analysis of chest impedance and ECG signals with a support vector machine can automatically detect AOP. Although this is an off-line analysis, the apnea detection system that was developed in this work can be implemented for real-time apnea detection. The system can be trained beforehand (subject-independent or subject-specific) in order be used later for real-time apnea detection. Conventional apnea monitors can therefore be considerably improved with the use of multimodal analysis and machine learning techniques.
Appendix

A.1. Feature Visualization

Feature visualization for all implemented features used in this analysis and are not presented in section 3.5.

(a) Box plot of apnea and non-apnea classes
(b) Mean apnea and of 9 epochs preceding and following apnea
(c) Histogram of apnea and non-apnea classes

Figure A.1: Spectral power of respiratory frequency (R1). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.0001.

(a) Box plot of apnea and non-apnea classes
(b) Mean apnea and of 9 epochs preceding and following apnea
(c) Histogram of apnea and non-apnea classes

Figure A.2: Spectral power in VLF (R2). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.0001.
A.1. Feature visualization

Figure A.3: Spectral power in LF (R3). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.0001.

Figure A.4: Spectral power in HF (R4). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.0001.

Figure A.5: Ratio of LF and HF bands (R5). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.0001.
Figure A.6: Respiratory frequency estimated in the frequency domain (R6). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.0001.

Figure A.7: Respiratory frequency estimated in the time domain (R7). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.0001.

Figure A.8: Breath-by-breath correlation (R8). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.0001.
(a) Box plot of apnea and non-apnea classes  (b) Mean apnea and of 9 epochs (c) Histogram of apnea and non-apnea classes

Figure A.9: Envelope Power (R9). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p<0.0001$.

(a) Box plot of apnea and non-apnea classes  (b) Mean apnea and of 9 epochs (c) Histogram of apnea and non-apnea classes

Figure A.10: Breath Length Variation (R10). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$.

(a) Box plot of apnea and non-apnea classes  (b) Mean apnea and of 9 epochs (c) Histogram of apnea and non-apnea classes

Figure A.11: Sample entropy of respiratory effort (R11). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p<0.0001$. 
Figure A.12: Standardized median of respiratory peaks (R12). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.0001.

Figure A.13: Standardized median of respiratory troughs (R13). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.0001.

Figure A.14: Respiratory peak regularity measured by sample entropy (R14). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.0001.
Figure A.15: Respiratory trough regularity measured by sample entropy (R15). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$.

Figure A.16: Median of peak-to-trough amplitude sequence (R16). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.0001$.

Figure A.17: Ratio of respiratory peak and trough regularity measured by sample entropy (R17). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. 
Figure A.18: Respiratory regularity of breath by breath duration (R18). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.19: Correlation of peak-trough breath duration (R19). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.20: Regularity of inhalation-exhalation cycle (R20). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. 

(a) Box plot of apnea and non-apnea classes
(b) Mean apnea and of 9 epochs preceding and following apnea
(c) Histogram of apnea and non-apnea classes

(a) Box plot of apnea and non-apnea classes
(b) Analysis of the apnea and non-apnea epochs
(c) Histogram of apnea and non-apnea classes

(a) Box plot of apnea and non-apnea classes
(b) Mean apnea and of 9 epochs preceding and following apnea
(c) Histogram of apnea and non-apnea classes
Figure A.21: Correlation of inhalation-exhalation cycle (R21). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.22: Median of inhalation time period (R22). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.
Figure A.23: Ratio of median value of inhalation and exhalation time periods (R23). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001.

Figure A.24: IQR of depth of breathing based on peaks distribution R(24). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001.

Figure A.25: IQR of depth of breathing based on troughs distribution (R25). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001.
Figure A.26: IQR of breathing cycle based on peaks distribution (R26). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001.

Figure A.27: IQR of breathing cycle based on troughs distribution (R27). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.28: IQR of inhalation cycle distribution (R28). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p<0.001.
Figure A.29: IQR of exhalation cycle distribution (R29). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.30: Regularity of inhalation-exhalation depth (R31). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.31: RR LF band (C1). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. 
A.1. Feature visualization

Figure A.32: RR HF band (C2). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.33: RR standard deviation (C3). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.34: RR respiratory frequency (C4). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. 
Figure A.35: RR respiratory power (C5). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.36: LF/HF Ratio (C6). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.37: Longest Shortest RR difference (C7). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.
Figure A.38: Detrended RR mean (C8). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.39: RR mean (C9). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$.

Figure A.40: Sample Entropy of RR (C10). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at $p < 0.001$. 
Figure A.41: Mean absolute deviation of RR interval (C11). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.42: SDSD measure (C12). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.
Figure A.43: RMSSD measure (C13). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.44: Inter-quartile range (C14). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.45: NN50 first (C15). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.
Figure A.46: pNN50 first (C16). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.47: NN50 second (C17). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.48: pNN50 second (C18). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.
A.1. **FEATURE VISUALIZATION**

Figure A.49: Synchronized correlation of RR interval and respiration (C19). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.50: Correlation of RR and IBI (C20). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.

Figure A.51: Correlation of RR and inhalation-exhalation time period (C21). The significance of difference (5%) was found between the two classes using an unpaired Mann–Whitney test at p < 0.001.
A.2. CLASSIFICATION RESULTS BASED ON CARDIAC FEATURES

Detection performance based on cardiac features deteriorates compared to the performance based on respiratory features. This was expected due to the fact that apnea is not always followed by bradycardia. As a result discriminative power of cardiac features (see Figure 3.15) is lower compared to discriminative power of respiratory features (see Figure 3.14).

![ROC curve](image1)

(a) Subject-Independent

![ROC curve](image2)

(b) Subject-Specific

![PR curve](image3)

(c) Subject-Independent

![PR curve](image4)

(d) Subject-Specific

Figure A.52: ROC and PR curves for both Subject-Independent and Subject-Specific for CFS algorithm

Figure A.52 compares the ROC and PR curves of the apnea detector after applying the first, second and both annotators for both subject-independent and subject-specific approaches for CFS algorithm. Note that the apnea detection performance between the first and second annotator differ significantly, since it seems that apnea episodes annotated by the second annotator do not seem to be strongly reflected in the cardiac activity. Maximum AUC (from PR curve) and Cohen’s kappa coefficient were achieved for the first annotator, even though only four features are selected, while for the second annotator nine features were selected. Detailed results can be seen on Tables A.1 - A.3. The automated apnea detection system achieved maximum sensitivity of 0.699 and specificity of 0.687 for subject-independent approach, while for subject-specific approach sensitivity of 0.693 and specificity of 0.785 were reached. Table A.1 illustrates that for subject-specific approach apnea detector performs better compared to the subject-independent approach. However, results from both approaches do not differ significantly as shown on Tables A.1 - A.3. This means that the pathophysiology of apnea expressed by cardiac activity exhibits large differences between and within patients.
Table A.1: Classification results using cardiac features and CFS algorithm for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.685 ± 0.047</td>
<td>0.767 ± 0.045</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.547 ± 0.064</td>
<td>0.567 ± 0.107</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.761 ± 0.066</td>
<td>0.761 ± 0.066</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.699 ± 0.086</td>
<td>0.693 ± 0.065</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.687 ± 0.042</td>
<td>0.785 ± 0.053</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.318 ± 0.075</td>
<td>0.331 ± 0.080</td>
</tr>
</tbody>
</table>

**Feature's Subset**: C17, C19, C20, C21

Table A.2: Classification results using cardiac features and CFS algorithm for the second annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.656 ± 0.027</td>
<td>0.598 ± 0.034</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.239 ± 0.085</td>
<td>0.303 ± 0.159</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.717 ± 0.068</td>
<td>0.709 ± 0.060</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.656 ± 0.104</td>
<td>0.572 ± 0.210</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.658 ± 0.024</td>
<td>0.600 ± 0.038</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.130 ± 0.048</td>
<td>0.131 ± 0.053</td>
</tr>
</tbody>
</table>

**Feature's Subset**: C2, C8, C9, C11, C13, C18, C19, C20, C21

Note that features C20 and C21 are selected for all three cases of annotations by using CFS algorithm. These feature seem robust and do not depend on the annotations. C20 and C21 features belong to the new features that are proposed in this work.

Figure A.53 compares the ROC and PR curves of the apnea detector after applying the first, second and both annotators for both subject-independent and subject-specific approaches for mRMR algorithm. Maximum AUC (from PR curve) and Cohen's kappa coefficient were achieved for the first annotator. The automated apnea detection system achieved sensitivity of 0.702 and specificity of 0.686 for subject-independent approach, while for subject-specific approach sensitivity of 0.689 and specificity of 0.783 were reached. Detailed results can be seen on Tables 4.6- 4.8.

Note that features C20 and C21 are selected for all three cases of annotations by using mRMR algorithm. These feature seem robust and do not depend on the annotations.

Tables A.1- A.3 and A.4- A.6 illustrate that the automated apnea detection system performs relatively the same for both feature selection algorithms. This means that the developed system is robust. In addition, feature subset C20 and C21 were the most discriminative since it was selected in all cases (see Table A.7). These features are very robust and do not depend on the annotations and the choice of feature selection algorithm.
### Table A.3: Classification results using cardiac features and CFS algorithm for both annotators

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.672 ± 0.029</td>
<td>0.658 ± 0.050</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.299 ± 0.099</td>
<td>0.323 ± 0.107</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.760 ± 0.082</td>
<td>0.762 ± 0.077</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.704 ± 0.029</td>
<td>0.660 ± 0.198</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.673 ± 0.021</td>
<td>0.658 ± 0.050</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.161 ± 0.062</td>
<td>0.173 ± 0.061</td>
</tr>
</tbody>
</table>

**Feature's Subset** C5, C15, C17, C20, C21

### Table A.4: Classification results using cardiac features and mRMR algorithm for the first annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.685 ± 0.042</td>
<td>0.765 ± 0.034</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.550 ± 0.067</td>
<td>0.565 ± 0.118</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.764 ± 0.062</td>
<td>0.767 ± 0.059</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.702 ± 0.042</td>
<td>0.689 ± 0.058</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.686 ± 0.040</td>
<td>0.783 ± 0.032</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.320 ± 0.068</td>
<td>0.340 ± 0.069</td>
</tr>
</tbody>
</table>

**Feature's Subset** C2, C19, C21, C20

### Table A.5: Classification results using cardiac features and mRMR algorithm for the second annotator

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.625 ± 0.024</td>
<td>0.590 ± 0.036</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.204 ± 0.065</td>
<td>0.233 ± 0.162</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.697 ± 0.070</td>
<td>0.693 ± 0.062</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.666 ± 0.108</td>
<td>0.575 ± 0.142</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.614 ± 0.025</td>
<td>0.591 ± 0.036</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.110 ± 0.040</td>
<td>0.121 ± 0.057</td>
</tr>
</tbody>
</table>

**Feature's Subset** C2, C15, C10, C4, C5, C8, C21, C11, C20
Figure A.53: ROC and PR curves for both Subject-Independent and Subject-Specific for mRMR algorithm

Table A.6: Classification results using cardiac features and mRMR algorithm for both annotators

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent (±)</th>
<th>Subject-Specific (±)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.682 ± 0.028</td>
<td>0.663 ± 0.045</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.307 ± 0.116</td>
<td>0.337 ± 0.110</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.765 ± 0.082</td>
<td>0.768 ± 0.081</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.698 ± 0.028</td>
<td>0.650 ± 0.167</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.685 ± 0.020</td>
<td>0.664 ± 0.048</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.169 ± 0.067</td>
<td>0.186 ± 0.077</td>
</tr>
</tbody>
</table>

Feature's Subset: C4, C15, C21, C11, C20

Table A.7: All new cardiac features selected at least once

<table>
<thead>
<tr>
<th>Features</th>
<th>CFS</th>
<th>mRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flrst Annotator</td>
<td>Second Annotator</td>
<td>Both Annotators</td>
</tr>
<tr>
<td>C19</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>C20</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C21</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
A.3. IMPACT OF EPOCH DURATION AND PERCENTAGE OF APNEA PER EPOCH BASED ON CARDIAC FEATURES

Figure A.54 demonstrates AUC from ROC and PR curves, sensitivity and specificity achieved for varying epoch duration. Maximum Cohen’s Kappa coefficient of 0.33 was achieved for epoch duration of 30 seconds. Table A.8 shows the results of the maximum performance for both subject-independent and subject-specific approaches.

![Graph](image1)

(a) AUC from PR and ROC curves

![Graph](image2)

(b) Sensitivity and specificity

Figure A.54: Apnea detection performance of subject-independent approach for varying epoch duration

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.676 ± 0.044</td>
<td>0.768 ± 0.048</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.624 ± 0.068</td>
<td>0.654 ± 0.060</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.748 ± 0.056</td>
<td>0.750 ± 0.054</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.679 ± 0.064</td>
<td>0.641 ± 0.061</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.682 ± 0.050</td>
<td>0.810 ± 0.049</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.330 ± 0.071</td>
<td>0.350 ± 0.085</td>
</tr>
<tr>
<td><strong>Feature’s Subset</strong></td>
<td>C17, C19, C20, C21</td>
<td></td>
</tr>
</tbody>
</table>

As it was mentioned previously, another parameter that influence the apnea detection performance is the percentage of apnea included per epoch which defines apnea and non-apnea distributions. Figure A.55 illustrates the apnea detection performance for varying percentage of apnea consisted in an epoch of 20 seconds. Maximum Cohen’s Kappa coefficient of 0.343 was achieved for a percentage of 45%. Table A.9 shows the results of the maximum performance for both subject-independent and subject-specific approaches.

In conclusion, for respiratory features maximum performance (based on Kappa and AUC-PR curve) was achieved for apnea duration of 60 seconds, while for cardiac features
A.3. Impact of epoch duration and percentage of apnea per epoch based on cardiac features

(a) AUC from PR and ROC curves  
(b) Sensitivity and specificity

Figure A.55: Apnea detection performance of subject-independent approach for varying apnea percentage per epoch

Table A.9: Classification results using respiratory features for apnea percentage of 45% per epoch

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subject-Independent</th>
<th>Subject-Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.692 ± 0.043</td>
<td>0.777 ± 0.034</td>
</tr>
<tr>
<td>AUC-PR curve</td>
<td>0.617 ± 0.076</td>
<td>0.634 ± 0.113</td>
</tr>
<tr>
<td>AUC-ROC curve</td>
<td>0.756 ± 0.054</td>
<td>0.757 ± 0.057</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.667 ± 0.072</td>
<td>0.690 ± 0.097</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.715 ± 0.040</td>
<td>0.802 ± 0.043</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.343 ± 0.062</td>
<td>0.355 ± 0.071</td>
</tr>
</tbody>
</table>

Feature’s Subset: C17, C19, C20, C21

for 30 second. Maximum performance concerning percentage of apnea per epoch was achieved for 45% for both respiratory and cardiac features.


