Non-contact detection of cessations of breathing in pre-term infants



Dineo Serame
Hertford College

A thesis submitted for the degree of
Doctor of Philosophy

Supervised by
Prof. Lionel Tarassenko, Prof. Mauricio Villarroel

Department of Engineering Science
University of Oxford

16 January 2026

Declaration

I declare that this thesis is entirely my own work and, except where stated, describes my own research.

Dineo Serame
Hertford College

Acknowledgements

To the Rhodes Trust, thank you for the extraordinary honour of being selected as a Rhodes Scholar and for welcoming me into your remarkable global community. This experience has been deeply transformative. Your belief in me made it possible for me to come to Oxford, pursue my passions with courage, and grow both professionally and personally. The opportunities, perspectives, and lifelong connections fostered through the Rhodes Trust have shaped not only this thesis, but also my sense of purpose and responsibility in the world. I am profoundly grateful for your support and for the faith you placed in me.

I had the privilege of being supervised by two remarkable educators throughout my DPhil journey. I am deeply grateful to Professor Lionel Tarassenko for taking a chance on me and for introducing me to the biomedical space, where I have been able to pursue impactful research. Your confidence in me opened doors I had never imagined walking through, and your steady encouragement pushed me to grow beyond what I thought possible.I also owe sincere thanks to Professor Mauricio Villarroel. Your mentorship is something I will carry with me for the rest of my career. Your insight, patience, and thoughtful guidance shaped not only this thesis but the researcher I have become. Working with you both, and within the Biomedical Signal Processing Group, has been a true privilege.

My time as a graduate student would not have been the same without the discussions, collaboration, and shared perseverance of past and present members of the Biomedical Signal Processing Group and the Laboratory for Computational Medicine and Technology. Dr. Sittichok Chaichulee, thank you for your invaluable contributions to this work and for your involvement in the data collection prior to my joining the group. Dr. Eoin Finnegan, Dr. Navin Cooray and Dr. Cristian Roman, thank you for helping me find my footing in the department during my early days. Your generosity with your time, guidance, and informal support made a new and unfamiliar environment feel navigable, and I am grateful for the role you played in helping me settle into the group. Dr. Mirae Harford, I am sincerely grateful for your careful annotation of my dataset and especially for opening your home to me when home felt very far away. Alexis Dorier, Abdullah Bin-Obaid, and Marshia Shaily, thank you for your resilience, camaraderie, and good humour throughout a demanding journey. I am grateful for our conversations, your constant motivation, and the many ways you shared your cultures with me. I wish you strength, clarity, and every success as you continue navigating the challenges of a DPhil.

I am grateful to the Principals, Fellows, staff, and students of Hertford College for providing a welcoming academic home during my time in Oxford. To the members of Hertford College's Middle Common Room, particularly those who had the joy (or occasional misery) of serving on committee with me, it was a pleasure to work alongside you. To my mental health mentor, Ms. Julia Hill, the past six years would have been far more challenging without your guidance and support. You helped me navigate the complexities of my mental health struggles with patience, wisdom, and understanding. I am deeply grateful for your steady presence, for the hope you offered when I needed it most, and for giving me the courage to keep moving forward.

Finally, I want to celebrate the people who carried me with laughter, love, and unwavering support through this journey. This poem is for all of you: for constancy, for love, for laughter, and being the quiet proof that I am held. Ke a leboga.

Dr Majd Abdulghani
you cross deserts with your laughter
carry light across spires and stone
teach me that friendship is its own kind of home.

We traded poems like currency.
Our words stretched between continents,
soft as wind over the Kalahari
firm as the hands that lift the weary.

Mmakgosi Tau and Kagiso Makwatse
you've spilt tea into afternoons like sun.
Checked on me like rain checks the earth,
reminding me that love is not just a word,
but a staying.

My sister, Gorata Angela Serame
you've taught me to carry my roots
even when the ground shifts beneath my feet.
So, I carry you folded into my bones,
a map back to where the earth blushes
beneath my feet.

My life partner, Dr. Shaun Davidson
when the Oxford night folds over tired streets,
you arrive in the spaces the day leaves empty.
remind me of water, of rest, of small necessities
guide me through code that lingers long,
sit beside me in hours that feel endless,
where I forget how to breathe.

You mark the hours with care
hold my hands
while the world insists I walk alone.
You do not rescue with noise-
but with presence,
and your presence is enough.

My grandmother and mother, Botlhale le Peggy Serame
Your voices rose first in my world
like morning breaking over red dust.
You planted belief in soil too stubborn to yield,
watched my small hands and promised greatness.
You prayed me awake,
taught me to hold knowledge like a gift,
and when my roots trembled
you held the earth steady beneath me.

Abstract

Current technologies for detecting Cessation Of Breathing (COBE) episodes in pre-term infants, such as Impedance Pneumography (IP) and electrocardiography (ECG), require direct skin contact through adhesive electrodes. These contact-based methods can cause discomfort and skin irritation in pre-term infants with fragile skin and are prone to motion artefacts, resulting in false alarms and alarm fatigue that can compromise infant safety. Video-based physiological monitoring offers a non-contact alternative for continuous respiratory assessment without disrupting routine care. However, these methods have largely been developed and validated in adults, and their translation to pre-term infants, who present distinct physiological and care-related challenges, remains underexplored.

This thesis proposes and evaluates a framework for non-contact detection of COBE events in pre-term infants. I developed and validated the methodology using data from a clinical study of 30 pre-term infants hospitalised in the high dependency area of the Neonatal Intensive Care Unit (NICU) at the John Radcliffe Hospital, Oxford. The dataset comprised 90 recording sessions of RGB video recordings and reference physiological signals collected under routine clinical conditions, totalling over 430 hours.

I extracted respiratory information from video recordings during periods characterised by motion and physiological instability surrounding COBE episodes. I developed camera-based respiratory rate estimation methods with motion-based filter switching to ensure reliable performance in the NICU. I then trained deep learning models to detect COBE episodes using contact-based physiological signals, video-derived signals, and their combinations. Models were developed using five-fold cross-validation on 83.3% of the dataset, with 16.7% reserved as an independent held-out test set. Among contact-only models, ConvNeXt achieved the highest performance on combined IP and photoplethysmography (PPG) signals, with a test accuracy of 88.7%. Hybrid models integrating IP with video-derived pixel intensity achieved the highest performance (90.6%), outperforming all unimodal approaches. Camera-only models achieved a test accuracy of 76.9%, outperforming models trained on ECG or PPG alone.

These results demonstrate the feasibility of non-contact COBE detection in pre-term infants and show that integrating video-derived features with traditional physiological signals improves detection performance. The proposed hybrid framework enhances reliability without replacing existing monitoring infrastructure and supports improved respiratory monitoring in high-resource NICU settings. Adapting this approach to widely available devices, such as smartphone cameras, could further enable scalable monitoring in low-resource environments where conventional monitoring equipment may not be available.

Table of contents

Table of contents ix

List of figures xiii

List of tables xv

List of abbreviations xvii

List of symbols xix

1 Introduction 1
1.1 Pre-term birth 1
1.2 Apnoea of prematurity 2
1.3 Current methods used to detect apnoea of prematurity 3
1.4 Non-contact detection of episodes of cessation of breathing 5
1.5 Objectives 7
1.6 Contributions to knowledge 7
1.7 Outline of thesis 8
1.8 Statement of originality 8

2 Literature review 11
2.1 Introduction 11
2.2 Developmental physiology 11
2.2.1 Cardiovascular system 11
2.2.2 Respiratory system 13
2.2.3 Anatomy of the skin 17
2.3 Vital signs 18
2.3.1 Heart rate 18
2.3.2 Respiratory rate 18
2.3.3 Peripheral oxygen saturation 19
2.4 Cardio-respiratory monitoring in the NICU 20
2.4.1 Electrocardiogram (ECG) 21
2.4.2 Impedance pneumography 22
2.5 Pulse oximetry 24
2.5.1 Skin-light interactions 24
2.5.2 Photoplethysmography 27
2.5.3 Pulse oximetry 28
2.6 Non-contact vital sign monitoring 31
2.6.1 Photoplethysmography imaging (PPGi) 31
2.6.2 Vital-sign estimation 32
2.7 Detection of COBE from conventional contact-based signals 34
2.7.1 Respiratory signal extraction and pre-processing 34
2.7.2 COBE detection in adults 36
2.7.3 COBE detection in pre-term infants 37
2.8 Detection of COBE from video camera signals 39
2.8.1 Non-contact COBE detection in adults 39
2.8.2 Non-contact COBE detection in pre-term infants 39
2.9 Deep learning approaches for COBE event detection 40
2.9.1 The architecture of a residual block 41
2.9.2 The architecture of ConvNeXt 42
2.9.3 Challenges of small physiological datasets 44
2.10 Discussion 45

3 Neonatal High Dependency Unit study **49**
3.1 Introduction ... 49
3.2 Clinical study design and protocol ... 49
3.2.1 Study participants ... 50
3.2.2 Informed consent ... 51
3.2.3 Participant confidentiality ... 51
3.2.4 Data handling and record keeping ... 52
3.3 Instrumentation ... 52
3.3.1 Hardware for data acquisition ... 53
3.3.2 Software for data acquisition ... 56
3.4 Cohort description ... 57
3.4.1 Participant demographics ... 57
3.4.2 Recording statistics ... 58
3.4.3 Reference vital signs ... 59
3.5 Conclusion ... 61

4 Contact monitoring of respiratory rate **63**
4.1 Introduction ... 63
4.2 Overview of the respiratory rate estimation process ... 63
4.3 Resampling ... 64
4.4 Signal filtering ... 64
4.5 Peak and trough detection ... 66
4.6 Signal quality assessment ... 67
4.6.1 Physiological bounding ... 67
4.6.2 Spectral density distribution ... 67
4.6.3 Spectral Purity Index ... 68
4.6.4 Combining signal quality measures ... 69
4.7 Computing reference respiratory rate ... 70
4.8 Signal alignment ... 71
4.9 Results ... 72
4.10 Discussion ... 74

5 Creating a labelled dataset for COBE detection **77**
5.1 Creating the COBE training dataset ... 77
5.1.1 Selection of data for annotation ... 77
5.1.2 Manual annotation ... 79
5.1.3 Inter-reviewer agreement and consensus ... 83
5.2 Defining analysis windows ... 84
5.3 Discussion ... 85

6 Detection of COBE using physiological signals in the NICU **87**
6.1 Introduction ... 87
6.2 Overview of the methodology ... 87
6.3 Respiratory signal extraction ... 89
6.3.1 IP signal ... 90
6.3.2 ECG signal ... 90
6.3.3 PPG signal ... 91
6.4 Construction of the machine learning dataset ... 93
6.4.1 Data windowing ... 93
6.4.2 Data splitting ... 94
6.5 A CNN for COBE episode detection ... 94
6.5.1 Network architecture ... 94
6.5.2 Network training ... 95
6.6 A ResNet for COBE episode detection ... 97
6.6.1 Network architecture ... 97
6.6.2 Network training ... 98
6.7 A ConvNeXt for COBE episode detection ... 98

6.7.1 Network architecture 99
6.7.2 Network training 99
6.8 Results 100
6.8.1 Model evaluation 100
6.8.2 Baseline detection of COBE using IP-derived RR 101
6.8.3 COBE episode detection using a CNN 102
6.8.4 COBE episode detection using a ResNet 103
6.8.5 COBE episode detection using a ConvNeXt 104
6.9 Discussion 105
6.10 Conclusion 107

7 Non-contact estimation of RR during periods of cessation of breathing 109
7.1 Overview of the RR estimation process 109
7.2 Video segment extraction 110
7.3 Segment quality assessment 110
7.4 Dynamic ROI selection 110
7.4.1 Torso ROI 111
7.4.2 RR ROI 111
7.5 Computing the motion signal from a video camera 112
7.6 Computing respiratory rate from video camera data 114
7.6.1 Respiratory signal extraction 115
7.6.2 Signal filtering 116
7.6.3 Peak and trough detection 118
7.6.4 Signal quality assessment 121
7.6.5 Respiratory rate estimation 124
7.7 Results 124
7.8 Discussion 126
7.9 Conclusion 127

8 COBE episode detection using camera signals 129
8.1 Introduction 129
8.2 Overview of the methodology 130
8.3 Segment extraction 131
8.4 Segment quality assessment 132
8.5 Construction of the machine learning dataset 133
8.5.1 Time alignment 133
8.5.2 Data windowing 134
8.5.3 Data splitting 134
8.6 Baseline detection of COBE using camera-derived RR 135
8.7 ResNet for COBE detection using camera signals 136
8.8 ConvNeXt for COBE detection using camera signals 136
8.9 Hybrid COBE episode detection using camera and physiological signals 136
8.9.1 Network architecture 137
8.9.2 Network training 137
8.10 Results 138
8.10.1 COBE episode detection using ResNet 138
8.10.2 COBE episode detection using ConvNeXt 139
8.10.3 Hybrid COBE episode detection 139
8.11 Discussion 140
8.12 Conclusion 143

9 Conclusion and future work 145
9.1 Summary of key findings 145
9.1.1 Major results and contributions 145
9.2 Implications of the work 147
9.2.1 Scientific and methodological implications 147
9.2.2 Clinical and practical implications 147

9.3 Limitations 148
9.4 Future work 148
9.4.1 Heart rate estimation during COBE 148
9.4.2 Physiological indicators of impending COBE 150
9.4.3 Detection of respiratory patterns associated with health decline in LMIC infant populations 150
9.5 Concluding remarks 150

Appendix A Neonatal HDU Study Supplementary Data **153**
A.1 Participant Demographics and Recording Details 153
A.2 Recording Timelines 155

Appendix B MATLAB annotation interface **157**

Appendix C Convolutional neural network **159**
C.1 The architecture of a simple CNN 159
C.2 Network elements 159
C.3 Network training 162

Appendix D Supplementary results: COBE detection using CNNs **163**
D.1 Individual signal training 163
D.1.1 IP signal 163
D.1.2 ECG signal 164
D.1.3 PPG signal 165
D.2 Multi-signal training: late fusion 167

Appendix E Supplemetary results: COBE detection using ResNets **169**
E.1 Training using individual signals 169
E.2 Multi-modal training 170
E.2.1 Late Fusion 170
E.2.2 Early Fusion 171

Appendix F Supplementary results: COBE detection using a ConvNeXt **173**
F.1 Individual signals 173
F.1.1 Multi-modal training 174
F.1.2 Early Fusion 175

Appendix G Supplementary results: COBE detection using a combination of camera and physiological signals **177**

Bibliography **179**

List of figures

2.1 Diagram of foetal and postnatal circulatory systems. 12
2.2 Diagram of the respiratory system and the lungs. 14
2.3 Anatomical diagram of the skin showing major layers and structures. 17
2.4 Monitoring of vital signs in the Neonatal High Dependency Unit (HDU). 20
2.5 The ECG waveform 22
2.6 Measurement of respiratory rate (RR) using impedance pneumography. 23
2.7 Electromagnetic spectrum. 24
2.8 Absorption spectra of the skin at different wavelengths. 26
2.9 Components of the PPG signal. 27
2.10 Absorption spectra of oxy-haemoglobin (HbO_2) and deoxy-haemoglobin (Hb). 29
2.11 Operational modes of pulse oximetry 30
2.12 Illustration of the Photoplethysmographic Imaging (PPGi) signal. 32
2.13 Example processes for extracting respiratory signals from ECG and PPG signals. 35
2.14 The structure of a residual block. 41
2.15 Comparison of block architectures for a Swin Transformer, ResNet, and ConvNeXt. 43

3.1 Set-up for overhead video acquisition in the NICU study. 50
3.2 The JAI®AT-200CL 3-CCD video camera used for colour video acquisition in the study. 54
3.3 Patient monitoring equipment used in the NICU study 55
3.4 Distributions of baseline demographics on the first day of recording. 58
3.5 Distributions of vital signs for the entire Neonatal HDU study. 59
3.6 4-hour sample from a recording session. 60
3.7 Sample waveforms from a typical recording session. 61

4.1 Overview of the RR recomputation process. 64
4.2 Examples of IP signals extracted from a typical recording session. 65
4.3 Signal filtering of a 60-second IP waveform. 65
4.4 Peak and trough detection on a 30-second respiratory signal. 66
4.5 Spectral density distribution for a 60-second IP segment 68
4.6 Signal quality assessment of a 60-second IP waveform from a noisy period. 70
4.7 Alignment of two RR time-series 72
4.8 RR estimation using the breath counting method on a 5-minute segment of the IP signal 73
4.9 Comparison between the Philips and recomputed respiratory rate values over the entire dataset 74

5.1 Synthetic waveform showing how desaturations are defined and merged in our dataset. 78
5.2 Decision tree for manual annotation of potential COBE episodes by three annotators. 80
5.3 Reference signals extracted during a period of COBE. 81
5.4 Reference signals extracted from a period where there is a desaturation without any sustained accompanying decrease in RR. 81
5.5 Reference signals extracted during a period of non-desaturation COBE. 82
5.6 Reference signals during a period of normal breathing. 83

6.1 Workflow for COBE episode detection using conventional clinical signals. 88
6.2 Block diagram outlining respiratory rate estimation from the ECG and PPG signals. 89
6.3 Extraction of the EDR waveform from a 60-second ECG segment 91
6.4 Extraction of the PPG envelope waveform from a 30-second PPG segment 92
6.5 Example respiratory waveforms extracted from IP, ECG, and PPG signals. 93
6.6 Modified Convolutional Neural Network (CNN) for COBE episode detection. 95
6.7 Representative CNN learning curves during cross-validation. 96
6.8 Designed Residual Network (ResNet) for COBE episode detection. 98
6.9 Architecture of the proposed 1D ConvNeXt for COBE episode detection. 99

7.1 Overview of the processing pipeline for non-contact RR estimation. 110
7.2 Landmark and dynamic Regions of interest (ROI) detection 111
7.3 Example BlazePose failure mode during clinical intervention. 112
7.4 Motion signal extracted from the video camera around a COBE episode. 114
7.5 Flow diagram outlining the process of computing RR from the video camera data. 115
7.6 PPGi signals extracted from three colour channels. 116
7.7 Signal filtering of a 30-second respiratory waveform during normal activity. 117
7.8 Signal filtering of a 30-second respiratory waveform during low activity (COBE episode). . . . 118
7.9 Boxed Slope Sum Function (BSSF) peak and trough detection for a 30-second PPGi signal. 120
7.10 Peaks and troughs detection on a 30-second clean respiratory waveform. 121
7.11 Peak agreement for a 30-second respiratory signal 122
7.12 Signal quality assessment for a 30-second respiratory signal. 123
7.13 Window expansion for RR estimation. 124
7.14 Comparison between recomputed respiratory rate and the camera-derived respiratory rate. 125

8.1 Workflow for COBE episode detection using video-derived signals and hybrid models combining camera and physiological inputs. 131
8.2 Annotated 80s segment showing camera-derived respiratory signals. 132
8.3 Hybrid COBE detection network. 137

9.1 Video-based estimation of heart rate during a COBE episode. 149

A.1 Recording times for each session in the Neonatal HDU study. 155

B.1 MATLAB annotation interface used for COBE dataset review. 157

C.1 A basic CNN architecture. 159

List of tables

2.1 Comparison of core architectural components across ResNet, ConvNeXt, and Transformer blocks. 43

3.1 List of physiological parameters recorded during a typical recording session 56
3.2 Summary of population demographics. 57
3.3 Statistics for video camera and reference recordings 58
3.4 Statistics for the reference vital-sign data recorded from the Philips monitor. 59

4.1 Summary of error metrics for the proposed algorithms for recomputing RR from the IP signal. 74

6.1 Confusion matrix for baseline COBE detection using RR_{ip} compared against manual reference labels. 102
6.2 Summary of CNN model performance for single and multi-signal inputs. 103
6.3 Summary of final ResNet model performance for single- and multi-signal inputs. 103
6.4 Summary of final ConvNeXt model performance for single- and multi-signal inputs. 104
6.5 Summary of best-performing models by input signals. 105

7.1 Summary of error metrics for respiratory rate (RR) estimation from video camera data, evaluated separately for normal breathing (RR $>$ 20 breaths/min) and potential COBE events (RR $\leq$ 20 breaths/min). 125

8.1 Summary of the machine learning dataset 134
8.2 Training and test set window distribution. 135
8.3 Confusion matrix comparing camera-based method outputs against reference labels. 135
8.4 Cross-validation performance of ResNet models trained on camera signals. 138
8.5 Test set performance of the selected ResNet models. 138
8.6 Cross-validation performance of ConvNeXt models trained on individual signals. 139
8.7 Performance of multi-modal ConvNeXt models combining camera-derived signals. 139
8.8 Summary of hybrid COBE detection using camera and physiological signals. 140

A.1 Demographics of infants enrolled in the Neonatal HDU Study. 153

D.1 5-fold cross-validation performance of various CNN model configurations trained on the IP signal. 163
D.2 Final evaluation of Model B using the IP signal. 164
D.3 5-fold cross-validation results for multiple CNN model configurations trained on the ECG-derived respiration (EDR) signal 165
D.4 Final evaluation of Model B using the EDR signal. 165
D.5 5-fold cross-validation results for multiple CNN model configurations trained on the PPG signal 166
D.6 Final evaluation of Model E using the PPG signal. 167
D.7 5-fold cross-validation performance from training selected models on multiple signals using a late fusion approach. 167
D.8 Performance evaluation of late fusion multi-signal models on the unseen test set. 167

E.1 Cross-validation performance of ResNet models trained on individual signals. 169
E.2 Test set performance of the selected ResNet models. 169
E.3 Cross-validation performance of ResNet models using late fusion. 170
E.4 Test set performance of late fusion ResNet models. 170
E.5 Cross-validation performance of early fusion ResNet model. 171
E.6 Test set performance of early fusion ResNet models. 171

F.1 Cross-validation performance of ConvNeXt models (individual input signals) 173

F.2 Test performance for COBE detection using ConvNeXt models. 174
F.3 Cross-validation results for late fusion ConvNeXt models with multiple signals. 174
F.4 Test performance of late fusion ConvNeXt models trained on various combinations of physiological signals. 174
F.5 Cross-validation performance of early fusion ConvNeXt models on combined IP, ECG and PPG signals 175
F.6 Test performance of early fusion ConvNeXt models on combined IP, ECG and PPG signals. 175

G.1 Cross-validation performance of hybrid models trained by combining camera-derived signals with physiological inputs. 177
G.2 Test set performance of the hybrid models. 178

List of abbreviations

CO_2	Carbon dioxide
HbO_2	Oxy-haemoglobin
Hb	Deoxy-haemoglobin
SpO_2	Peripheral oxygen saturation, as measured by a pulse oximeter
AC	Pulsatile component of the signal
AOP	Apnoea of prematurity
BRC	Oxford Biomedical Research Centre
BSSF	Boxed Slope Sum Function
CNN	Convolutional Neural Network
COBE	Cessation of breathing
CRFs	Case Report Forms
DC	Non-pulsatile component of the plethysmograph signal
ECG	Electrocardiography
EDR	ECG-derived respiration
FFT	Fast Fourier Transform
FPR	False Positive Rate
HDU	High Dependency Unit
HR	Heart rate
IIR	Infinite Impulse Response
IP	Impedance Pneumography
IR	Infra-red
LMICs	Low and Middle Income Countries
MAC	Mean Average Curve
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error

NHS National Health Service

NICU Neonatal Intensive Care Unit

NIR near-infrared

PPG Photoplethysmography

PPGi Photoplethysmographic imaging

PSG Polysomnography

ReLU Rectified Linear Unit

ResNet Residual Network

RMSE Root Mean Square Error

ROI Region of interest

RR Respiratory rate

TPR True Positive Rate

UK United Kingdom

List of symbols

f	Frequency
t	Time
$X(f)$	Frequency-domain representation of a signal
$x(t)$	Time-domain physiological signal
RR_{ip}	Respiratory rate derived from impedance pneumography
RR_{philips}	Respiratory rate from Philips patient monitor
RR_{inst}	Instantaneous respiratory rate
RR_{estimate}	Respiratory rate estimated using the proposed contact-based method
SQI	Signal Quality Index
SQI_{bin}	Signal quality index computed per frequency bin
SQI_{breath}	Signal quality index per breath
SQI_{phys}	Physiological signal quality index
SQI_{RR}	Signal quality index for respiratory rate estimate
SPI	Spectral Purity Index
κ	Fleiss' kappa statistic for inter-rater agreement
P_0	Observed agreement
P_e	Expected agreement by chance
p_i	Observed reviewer agreement for the i-th observation
p_j	Proportion of ratings in category j
N	Number of observations
R	Number of reviewers
C	Number of categories
r_{ij}	Number of reviewers assigning category j to observation i
RR_{cam}	Camera-derived respiratory rate
HR_{cam}	Camera-derived heart rate
$RR(w)$	Respiratory rate estimated over window w
N_{breaths}	Number of breaths in a window
L_{exp}	Length of expanded window
w	Time window
r	Correlation coefficient

Chapter 1

Introduction

1.1 Pre-term birth

Pre-term birth occurs when an infant is born before 37 weeks of gestation [1, 2]. It is the leading cause of death in children under five years of age globally, with an estimated 1 million deaths across 13.4 million pre-term infants in 2020 [1, 3]. Complications associated with pre-term birth, such as respiratory distress, brain damage, and infections, are the strongest predictors of mortality within the first month of life [4, 5, 6]. Among these complications, apnoea of prematurity is a major contributor to respiratory morbidity in neonatal care settings, affecting up to approximately 85% of infants born at or before 34 weeks of gestation [7].

Despite advances in neonatal care, disparities in access and outcomes persist, even within high-income nations in the G7 group. Across these countries, pre-term birth rates generally ranged between 5% and 10% in 2022 [8]. In the United Kingdom (UK), the pre-term birth rate was approximately 7.4% of live births in 2020 [9], rising to 7.6% in 2021 and 7.9% in 2022 [10, 11]. Japan had the lowest pre-term birth rate of approximately 5.7% [8, 12], while the United States recorded the highest at 10.0% in 2020, rising to 10.4% by 2022 [3, 13]. In Low and Middle Income Countries (LMICs), approximately 50% of pre-term infants born at 32 weeks or earlier face mortality due to insufficient access to affordable healthcare. Basic care for infections and breathing difficulties, such as management of apnoea of prematurity, as well as breastfeeding support, are often unavailable [14]. Southern Asia and sub-Saharan Africa have the highest rates of pre-term birth globally, accounting for over 65% of the world's pre-term births, along with the highest mortality risk [3]. In Southern Asia, 13.2% of all live births in 2020 were pre-term, representing approximately 4.8 million pre-term infants, with an 8% mortality rate [1, 3]. Similarly, sub-Saharan Africa recorded a pre-term birth rate of 10.1%, equating to 3.9 million pre-term births, with an 11% mortality rate [3].

Given the regional disparities in pre-term birth rates and survival, the availability and quality of neonatal care are important factors in determining long-term health outcomes. Depending on the severity of their condition and the availability of resources, pre-term infants are usually nursed in the Neonatal Intensive Care Unit (NICU) or the High Dependency Unit (HDU). While technological advances in neonatal medicine have improved survival rates, pre-term infants remain at risk of long-term health problems, including motor

and cognitive impairments, as well as cardiac, renal, endocrine, and neurological issues [15, 16, 17, 18]. Due to their immature neurological systems and underdeveloped respiratory controls, pre-term infants are susceptible to respiratory disorders like apnoea of prematurity, which can persist for weeks to months and often require continuous monitoring and intervention [19, 20, 21].

1.2 Apnoea of prematurity

Apnoea of prematurity (AOP) is widely defined as either the cessation of breathing exceeding 20 seconds or a pause in breathing for at least 10 seconds accompanied by a heart rate (HR) of under 100 beats per minute (bradycardia) and a peripheral oxygen saturation (SpO_2) below 80% [22, 23]. Apnoea is typically broken into three categories based on its cause [24]. Central apnoea is characterised by the cessation of airflow and the absence of respiratory effort due to the loss of the central respiratory drive. Obstructive apnoea occurs when there is a complete absence of airflow despite ongoing respiratory efforts, typically resulting from an obstruction in the airway that hinders adequate ventilation. Mixed apnoea, the most common type observed in pre-term infants [19], encompasses a combination of central and obstructive components. In this form, obstructive apnoea may precede or follow a period of central apnoea [25].

AOP is a prevalent condition among pre-term infants, affecting approximately 7% of infants born at 34–35 weeks, 15% of those born at 32–33 weeks, and 54% of infants born at 30–31 weeks of gestation [26]. Because of the immature neurological system and the poorly developed mechanisms for respiratory control [19], pre-term infants are subject to sudden episodes of Cessation Of Breathing (COBE), defined as a complete stop of airflow into and out of the lungs. While not all episodes of COBE meet the clinical criteria for AOP, prolonged or repeated occurrences can contribute to its diagnosis [7, 27]. AOP can result in significant health complications, including an increased risk of hypoxaemia (reduced level of oxygen in the blood), hypoglycaemia (abnormally low level of glucose in the blood), neurological injury, and sepsis [28, 29]. AOP can increase the risk of death and disabilities in affected infants [30].

Accurate detection of apnoea is important, as its presence or absence is often a key factor in the clinical decision to discharge an infant from the hospital [27]. However, current monitoring guidelines for detecting apnoea of prematurity and threshold-based systems may miss clinically significant desaturation events [31, 32]. These overlooked desaturation events are critical because prolonged or frequent episodes of hypoxaemia can lead to insufficient oxygen delivery to vital organs such as the brain and the heart. The resulting hypoxic stress can lead to lasting damage, including neurological impairment, developmental delays, cardiorespiratory instability, and organ dysfunction [26, 33, 34].

Three distinct breathing patterns have been associated with hypoxaemic episodes in AOP: i) absent breathing movements, ii) continued breathing movements but absent airflow, and iii) continuous, but reduced airflow and breathing movements [27, 32, 35]. These observations suggest that desaturation episodes can occur even when some respiratory movement is present [36]. However, these respiratory movements are distinct from normal breathing, highlighting the variability and complexity of AOP presentations. This variability presents significant challenges to both the detection and management of the condition.

1.3 Current methods used to detect apnoea of prematurity

Currently, hospitals use a combination of cardiorespiratory monitors and pulse oximetry to detect episodes of COBE in pre-term infants [37]. These techniques complement each other to enable continuous monitoring of respiratory function, oxygenation status, and heart rate. Cardiorespiratory monitors typically rely on electrocardiography (ECG) to record the electrical activity of the heart, which is then used to compute heart rate and detect bradycardia [38]. The ECG electrodes are often additionally used for Impedance Pneumography (IP) to monitor respiratory patterns and estimate respiratory rate [39]. IP is a non-invasive method that measures variations in electrical impedance across the chest. These variations reflect changes in lung aeration during breathing, which alter the electrical conductivity of the thoracic tissues as current flows between electrodes. Pulse oximetry estimates SpO_2 by measuring the differential absorption of red and infrared light by oxygenated and deoxygenated haemoglobin in pulsatile arterial blood [40]. This method provides real-time insights into the infant's oxygenation status, allowing clinicians to identify desaturation events commonly associated with COBE episodes.

Although IP and pulse oximetry provide complementary information for COBE detection in the NICU, both methods are not without significant limitations. It can be difficult to detect obstructive apnoea using IP, particularly when respiratory effort persists but airflow is blocked. False alarms may also occur when cardiac artefacts affect the respiratory signal, particularly when the IP signal is contaminated by ECG interference or motion artefacts that resemble thoracic movements [39]. Pulse oximetry, while essential for identifying hypoxaemia, can introduce a 3 - 8 second delay in detecting desaturation events in pre-term infants. This delay arises from factors such as probe location, signal averaging, and approximately 3 - 4 seconds required for changes in central blood oxygen levels to manifest in peripheral circulation [40, 41, 42, 43]. Consequently, these lag times result in a gap between changes in a patient's oxygenation status and their detection by pulse oximetry, which may potentially delay timely clinical intervention during brief apnoea episodes in pre-term infants [40].

Pulse oximeters and cardiorespiratory devices are often integrated into patient monitoring systems, which enable the display of these data streams concurrently [44, 45]. This integration allows clinicians to analyse trends and interact with real-time parameters through user-friendly interfaces. Patient monitors commonly use data processing techniques, such as signal averaging, to reduce noise and stabilise displayed values [46, 47, 48]. However, this processing can delay detection of rapid physiological changes. Because alarm thresholds and signal averaging times are typically set by clinicians or nurses, brief but clinically significant desaturations may go undetected, especially when transient events fail to cross the pre-set thresholds or are smoothed out by averaging [32, 41, 46, 49]. While these settings help reduce false alarms and alarm fatigue, the extent of signal averaging plays a critical role in determining event detectability. Excessive averaging can mask brief but physiologically relevant changes in oxygen saturation, reducing the likelihood that physiological trigger thresholds, such as SpO_2 below 80%, are reached and detected [41, 49].

Bradycardia and oxygen desaturation can be easily identified from conventional methods such as ECG and pulse oximetry. However, apnoea detection requires the use of specialised methods. Currently, polysomnography (PSG) is considered the reference standard for the diagnosis and stratification of apnoea. PSG simultaneously records multiple physiological signals to assess breathing, including respiratory effort, airflow, heart rate, oxygen saturation, and sometimes the electroencephalogram (EEG), which records the brain's electrical activity to assess sleep patterns. Despite being the gold standard diagnostic tool, it is resource-intensive and expensive, both in terms of monetary costs and the need for specialised personnel and facilities [50, 51, 52]. Traditional technologies, such as transthoracic impedance plethysmography and strain gauge measurement of thoracic circumferences, can also be used to identify and diagnose apnoea in infants [37]. Transthoracic impedance plethysmography detects changes in electrical impedance across the chest during breathing by using electrodes typically placed on the infant's chest and back. These changes in impedance correspond to variations in thoracic volume during the respiratory cycle, enabling the monitoring of respiratory effort and the detection of COBE episodes [53, 54]. Conversely, strain gauge measurement tracks chest wall movement by using a belt equipped with strain gauges that wrap around the infant's chest. The belt measures changes in the chest's circumference as the infant inhales and exhales [55]. This technique is useful for identifying subtle respiratory changes but can be affected by movement and other artefacts [37]. These thoracic impedance sensors, however, often fail to detect obstructive apnoea, may falsely alarm when the infant is breathing normally and may confuse cardiac artefacts with valid respiratory impedance values [50]. This leads to frequent false alarms in the NICU, reported to be as high as 87.5%, which significantly contributes to alarm fatigue among clinicians [56]. Alarm fatigue occurs when staff become desensitised to the constant stream of non-actionable alerts,

leading them to underestimate the urgency of alarms. As a result, alarms may not be taken as seriously as they should, which can delay responses to real clinical complications and compromise patient safety [57].

Most current technologies, such as IP and PSG, require direct contact with the subject, using adhesive electrodes or sensors to track changes in the patient's vital signs. The placement and removal of these sensors can cause discomfort, stress, pain and even epidermal stripping in individuals with vulnerable skin, such as pre-term infants [58, 59]. Wires normally connect these sensors to a monitoring unit outside of the neonatal incubator, disrupting the delivery of clinical care and making physical interaction between parents and their babies difficult, especially during kangaroo care, a practice involving skin-to-skin contact between the infant and the parent's chest, which promotes physiological stability, bonding, and often facilitates breastfeeding [60, 61]. The current methods have also been reported to be inaccurate, as they often fail to detect COBE events reliably [32, 62]. This is due to the non-stationary, noisy, and unstable nature of infant respiratory signals [63]. As a result, current monitoring techniques can limit the quality and reliability of clinical care, potentially leading to missed diagnoses, false alarms, and delays in clinical interventions, highlighting the need for improved detection systems in neonatal care. These challenges are further compounded in LMICs, where high costs, limited specialised personnel, and resource-intensive infrastructure further constrain implementation of conventional monitoring systems [64, 65, 66].

1.4 Non-contact detection of episodes of cessation of breathing

The possibility of measuring back-scattered light remotely using a video camera sensitive to visible or near-infrared light (rather than a reflectance probe in contact with the body, such as a pulse oximeter) was first discussed in the scientific literature in 2005 [67]. Since then, extensive research has been done to validate the use of video cameras to monitor vital signs [68, 69, 70, 71]. These advances in the use of digital cameras in medicine have enabled research in non-contact vital sign estimation and, subsequently, contactless apnoea detection. Although different non-contact apnoea detection methods such as radar [72, 73, 74], thermal imaging [75, 76] and video camera [77, 78, 79] have been proposed, challenges remain in achieving the accuracy, reliability, and robustness required for widespread clinical use, particularly in dynamic and complex environments such as the NICU.

Thermal imaging methods detect respiration-induced temperature changes across the body's surface and allow for inferences about respiration and consequently COBE [76, 80]. Thermal cameras capture temperature variations near the nostrils or mouth, where the differences between inhaled and exhaled air can be observed. These variations cause pixel intensity changes in thermal images, which can be analysed to determine respiratory rate (RR). Thermal imaging systems, while effective, tend to be expensive due to

the high price of infrared sensors and the significant power requirements associated with their operation, making them less accessible compared to other non-contact technologies [81]. Conversely, radar-based systems operate by emitting an electromagnetic signal, and the reflected echoes from the signal allow the detection of small movements on the surface of the chest or abdomen during breathing [82, 83]. These movements, resulting from both heart and respiratory activity, can be observed whether the subject is standing, seated, or lying down. These slight displacements can be detected using the Doppler effect, providing a real-time representation of respiratory function. These measurements can then be analysed to assess the respiratory rate, evaluate respiratory volume, and identify patterns associated with COBE [82]. Radar systems offer several advantages: they are unaffected by skin pigmentation or ambient light and can operate through common materials like clothing, bedding, or incubator walls [84]. These features make them particularly appealing for use in neonatal care settings where direct skin contact is not always feasible. Despite these advantages, radar-based systems face challenges that have slowed their clinical adoption. Like thermal imaging, radar systems are costly due to specialised hardware and integration requirements, though low-cost commercial options are now increasingly available. These systems have yet to be extensively tested and validated in pre-term neonates [84]. Additionally, radar-based methods are limited due to concerns over the long-term effects of continuous electromagnetic exposure in vulnerable populations. While most devices generally operate within established non-ionising radiation guidelines set by bodies like the International Commission on Non-Ionising Radiation Protection (ICNIRP), specific safety data for pre-term neonates remains limited [85]. Radar systems reflect off the body as a whole, making it challenging to localise movements exactly (e.g., chest versus abdomen), particularly in pre-term infants whose small size and overlapping cardiac and respiratory motions complicate signal interpretation. These systems are also sensitive to motion artefacts, which can further reduce measurement accuracy [84, 86].

Unlike radar which mainly detect movement, visible-spectrum (RGB: red, green and blue) video cameras can detect both subtle chest or diaphragm movements and skin colour changes by analysing back-scattered light from the skin's surface [87]. Because they offer a visual perspective, video cameras can provide a standardised objective metric to assess respiratory patterns, allowing for the detection of neonatal disorders such as apnoea. They are cost-effective and capable of simultaneously monitoring multiple physiological signals such as RR and HR [88, 89]. The hardware and software of video cameras are well-established and accessible because of their widespread use, making it easier to integrate into existing telemedicine systems or hospital environments [90]. Video cameras are sensitive to lighting conditions, and their application in health monitoring introduces challenges regarding patient privacy and data protection, particularly in sensitive environments such as the NICU. Detection accuracy using RGB cameras may also be influenced by variations in skin tone [90].

Non-contact monitoring methods, including video, radar, and thermal imaging, provide significant advantages for use in the NICU by eliminating the risk of sensor detachment and minimising the potential for skin injuries, often associated with traditional contact-based sensors [88, 91, 92]. These non-contact monitoring technologies enable continuous, contact-free monitoring of pre-term infants. These systems reduce the clutter of wires around the incubator, improving accessibility for caregivers and parents and minimising interference with routine clinical procedures. These methods could also support other bedside assessments, such as tracking physical activity, distress, or pain [88, 93]. However, further development and validation are needed to ensure their reliability and suitability for clinical care in the NICU environment.

1.5 Objectives

The objectives of the thesis are:

1. To develop non-contact algorithms to estimate vital signs during periods of physiological distress. These algorithms aim to address the limitations of conventional contact-based monitoring systems by ensuring minimal physical interference with pre-term infants while maintaining high diagnostic accuracy.
2. To evaluate the potential and effectiveness of continuous video camera-based detection of cessation of breathing in multiple pre-term infants nursed in the high-dependency area of the NICU.
   The aim is to establish the clinical feasibility and reliability of RGB video monitoring for detecting COBE events in the studied cohort.
3. To assess the application of advanced deep learning techniques in detecting COBE.
   This involves leveraging state-of-the-art neural network architectures and training methodologies to improve the precision and reliability of apnoea detection.

1.6 Contributions to knowledge

This thesis makes the following contributions to knowledge in neonatal physiological monitoring and automated detection of COBE in pre-term infants:

1. Development of methods for non-contact respiratory-rate estimation from RGB video recordings during periods of physiological instability and infant movement associated with COBE episodes in the NICU.
2. Development and evaluation of deep learning models, including Convolutional Neural Networks (CNNs), Residual Networks (ResNets), and ConvNeXt architectures, for automated COBE detection using physiological signals routinely monitored in neonatal care.

3. Demonstration of the feasibility of camera-only COBE detection in pre-term infants using video-derived physiological signals acquired under routine clinical conditions.

4. Demonstration that hybrid multimodal models combining video-derived signals with conventional physiological signals improve COBE detection performance compared with unimodal monitoring approaches.

1.7 Outline of thesis

In Chapter 2, I review the literature on human physiology and current vital sign monitoring techniques used in clinical practice. Additionally, I examine existing methods for apnoea detection using machine learning algorithms. In Chapter 3, I provide an overview of the clinical study and the dataset used for this thesis.

Chapter 4 discusses the reference vital sign monitoring in the Neonatal HDU study. I present the signal processing methods I developed to recompute RR from the reference clinical patient monitor. In Chapter 5, I introduce a COBE dataset by annotating respiratory pauses according to both definitions of apnoea: pauses exceeding 20 seconds, and pauses of at least 10 seconds accompanied by desaturation. This dataset is validated using a clinically annotated subset of the NICU dataset. In Chapter 6, I utilise conventional signals readily available in the hospital, such as IP and ECG, to classify COBE using deep learning methods. The results from this chapter are compared to current state-of-the-art models.

Chapter 7 uses video camera data, extracting relevant signals from defined Regions of interest (ROIs) to quantify infant motion and estimate RR during periods of COBE. The chapter evaluates the feasibility of non-contact respiratory signal extraction under real NICU conditions. In Chapter 8, I employ the best-performing model from Chapter 6 to classify COBE episodes from video camera-derived signals. The results from this analysis are compared to the results obtained from COBE event classification using conventional signals only, as well as results from classification using a combination of conventional and camera signals. Finally, Chapter 9 presents this thesis's conclusions, outlining the implications of its findings for clinical practice and suggesting future research directions.

1.8 Statement of originality

This thesis contains original research conducted by the author within the Biomedical Signal Processing and Machine Learning research group at the University of Oxford. The clinical study infrastructure, including the Neonatal HDU study setup, video acquisition platform, and elements of the physiological data acquisition pipeline, was developed as part of the group's wider research programme before the

commencement of this thesis. The NICU dataset used in this thesis was collected within this established study framework. Existing signal acquisition systems and prior physiological monitoring methodologies developed within the group also informed parts of this work.

The original contributions presented in this thesis include:

- the development of methods for estimating respiratory rate from video-derived signals during periods of cessation of breathing and physiological instability in pre-term infants;
- the development of methodologies for constructing and clinically annotating a labelled COBE dataset from synchronised physiological and video recordings;
- the development and evaluation of deep learning approaches, including CNNs, ResNets, and ConvNeXt architectures, for automated COBE detection in pre-term infants;
- the investigation of hybrid monitoring approaches combining conventional physiological signals with video-derived signals for COBE detection;
- the experimental evaluation and comparative analysis of contact, non-contact, and hybrid COBE monitoring approaches in pre-term infants.

Except where explicitly stated, the analysis, methodology development, experiments, and interpretation presented in this thesis are the original work of the author.

Chapter 2

Literature review

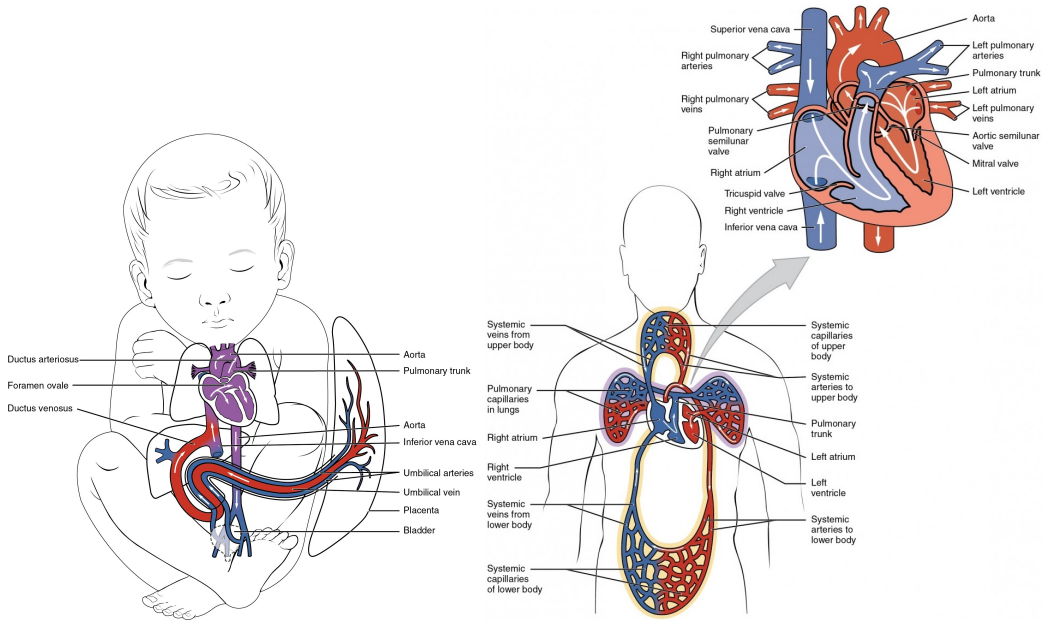
2.1 Introduction

This chapter provides a comprehensive review of the literature relevant to the detection of cessation of breathing (COBE), a defining feature of apnoea of prematurity (AOP). The review begins by discussing developmental physiology, including the cardiovascular and respiratory systems, which provide the foundation for the study of AOP. The chapter then discusses the principles of key physiological technologies such as photoplethysmography (PPG) and pulse oximetry, and describes respiratory monitoring using impedance pneumography (IP). Subsequently, the chapter provides an in-depth discussion of current state-of-the-art techniques for detecting and predicting COBE episodes, using both conventional and non-contact signals. Through this structured review, the chapter establishes a theoretical foundation for the development of improved COBE detection and monitoring methodologies.

2.2 Developmental physiology

2.2.1 Cardiovascular system

The cardiovascular system undergoes significant structural and functional changes during the transition from foetal to neonatal life, enabling the shift from placental gas exchange to independent pulmonary respiration [94, 95]. After clearing lung fluid at birth, the infant's blood pressure increases, and the pulmonary pressure significantly decreases. This reduces the need for the ductus arteriosus (see figure 2.1a), an opening between the aorta and pulmonary artery that bypasses the lungs during blood circulation in utero. As a result of these changes, the pressure in the heart's left atrium increases, which decreases the pressure in the right atrium. This shift in pressure stimulates the closure of the foramen ovale, an opening in the atrial wall that allows blood to pass from the right to the left atrium. The closure of the ductus arteriosus and foramen ovale signifies the transition from foetal to newborn circulation [95, 96], allowing blood to circulate from the lungs and enabling a new system of gas exchange that helps oxygen enter the blood from the air.



(a) (b)

Figure 2.1: Diagram of foetal and postnatal circulatory systems. a) Foetal circulatory system: In utero, blood bypasses the lungs via the ductus arteriosus and the foramen ovale, as oxygen is provided through the placenta. b) Extra-uterine (postnatal) circulatory system: After birth, the ductus arteriosus and foramen ovale close, enabling pulmonary respiration. Oxygenated blood flows from the lungs to the left atrium, then to the left ventricle, and is pumped through the aorta to systemic circulation. Deoxygenated blood returns to the right atrium, passes through the right ventricle, and is sent to the lungs via the pulmonary artery for re-oxygenation. Images reproduced from J. Gordon Betts *et al.* [97, 98].

The cardiovascular system, or extra-uterine circulatory system, transports blood throughout the human body via arteries, veins, and capillaries as illustrated in figure 2.1b. This system ensures that blood moves through the pulmonary circulation to and from the lungs, enabling oxygen absorption from tiny air sacs (alveoli) and the removal of carbon dioxide through exhalation. Once oxygenated, blood travels from the lungs through the pulmonary veins and enters the left atrium of the heart. From there, blood passes through the bicuspid valve into the left ventricle, which contracts to push oxygen-rich blood through the aortic valve into the aorta and systemic circulation [99, 100].

After systemic arterial distribution, oxygenated blood reaches the capillaries, where oxygen is released to the tissue cells in exchange for carbon dioxide (CO_2). The resulting deoxygenated blood is then collected by venules and veins, which transport it back to the heart. All systemic venous blood enters the heart through the right atrium via the superior and inferior vena cava. From the right atrium, blood flows through the tricuspid valve into the right ventricle, which contracts to pump blood through the pulmonary valve into the pulmonary artery and onward to the lungs for re-oxygenation [99, 101]. This continuous sequence of blood flow through the four chambers of the heart (right atrium, right ventricle, left atrium, and left ventricle) constitutes the cardiac cycle. It includes systole, the phase of atrial and ventricular

contraction during which blood is ejected, and diastole, the phase of atrial and ventricular relaxation when the chambers fill with blood. This cycle ensures effective transport of oxygen and nutrients to tissues while removing metabolic waste products.

While these cardiovascular and pulmonary adaptations typically occur efficiently in term infants, it is common for pre-term infants to struggle with adapting to life outside the uterus. Foetal oxygen partial pressure (PO_2) ranges from 20-30 mmHg, substantially lower than adult arterial oxygen (80–100 mmHg). This means foetal oxygenation is close to physiological hypoxia in adult tissues [102, 103], implying that foetal development occurs in a state of relative hypoxia, with foetal tissues having a lower threshold for oxygen insufficiency [104]. Though foetal development happens in a relatively low oxygen environment, at birth, the lung expands with the initiation of breathing, and oxygen availability to tissue rises almost twofold (from around 50%–60% to 90%–95%). Pre-term infants must, therefore, maintain a high oxygen supply to their circulatory system and other organs as they transition from placental gas exchange to breathing through their lungs [94, 105]. However, due to their underdeveloped systems, some pre-term infants may require immediate delivery room interventions, such as assisted ventilation or chest compressions, to support their cardiorespiratory transition [106].

Despite improvements in medical care, pre-term infants are at a greater risk of long-term adverse cardiovascular outcomes ranging from congenital heart defects to dysfunctional blood flow autoregulation [107]. Patent ductus arteriosus (PDA), which occurs when the ductus arteriosus fails to close within 72 hours after birth [108, 109], is one of the most common congenital heart defects, and is usually associated with immaturity of the cardio-respiratory system in pre-term infants [110]. This immaturity can also present as prolonged cessation of breathing episodes with associated bradycardia (low heart rate) or oxygen desaturation [27].

2.2.2 Respiratory system

The respiratory system plays a critical role in ensuring optimal oxygen levels in the body and the efficient removal of CO_2. Oxygen enters the lungs through inhalation and diffuses across the alveolar membrane into the bloodstream. Conversely, CO_2, a waste gas, diffuses from the blood into the alveoli and is expelled from the body during exhalation [111, 112, 113]. Structurally, the respiratory system is divided into the upper and lower respiratory tracts, which together form a network of organs and tissues responsible for gas exchange. Key components of this system include the airways (bronchi), the windpipe (trachea), the lungs, and blood vessels, as shown in figure 2.2. The lungs serve as the primary site of gas exchange, where de-oxygenated blood is enriched with oxygen before being returned to the heart for systemic circulation [112, 114].

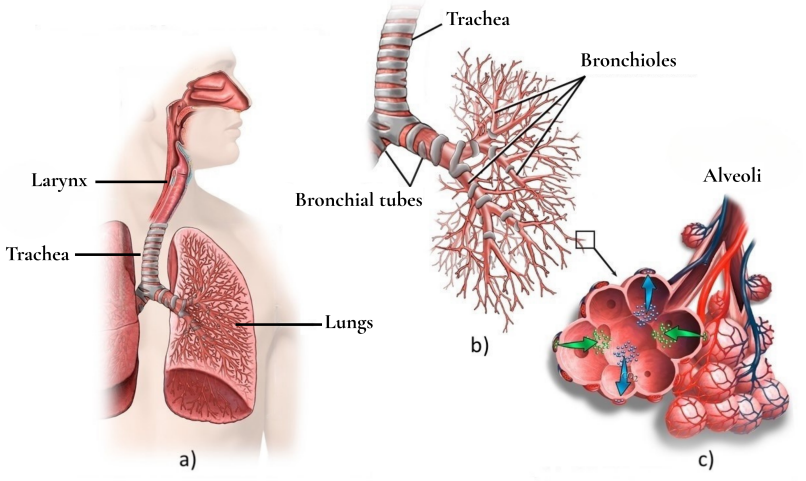


Figure 2.2: Diagram showing the anatomy of the respiratory system. a) The airways and the lungs. These structures direct inhaled air through the respiratory tract and deliver it to the bronchioles and alveoli for gas exchange. b) View of the trachea branching into bronchial tubes and smaller bronchioles, which serve as passageways for airflow throughout the lungs to the alveoli. c) Close-up view of gas exchange at the thin-walled alveoli. The blue arrows represent oxygen in inhaled air passing into the bloodstream via the pulmonary capillaries, and the green arrows show the carbon dioxide leaving the bloodstream into the lungs. (Image reproduced from The National Heart, Lung, and Blood Institute (NHLBI), courtesy of [115]).

Lung development

The lungs are the primary organ in the respiratory system, responsible for gas exchange. Their development is divided into three main periods: the early embryonic period, the foetal period, and the postnatal period [116, 117].

The embryonic period

The embryonic period, occurring from the 4th to the 7th week of gestation, marks the start of lung development and establishes the foundation for structural and functional development of the respiratory system. During this stage, the lung bud is formed from the foregut endoderm and begins branching into primary bronchial buds. Concurrently, the formation of the pulmonary vasculature begins, connecting the lungs to the heart and preparing for future circulation [116, 118]. Disruptions during this period, whether caused by genetic abnormalities or exposure to harmful substances such as tobacco, can result in congenital anomalies [117, 118]. These anomalies include pulmonary hypoplasia, a condition for which lung development is incomplete, resulting in reduced lung volume, fewer alveoli, and insufficient capacity for effective gas exchange [119]. Additionally, developmental delays during the embryonic period may hinder the normal maturation of neural and muscular components that are essential for regulating an infant's breathing patterns. These disruptions may increase the infant's risk of developing conditions such as apnoea [30, 117].

By the end of the embryonic period, the lungs exhibit a basic structure, including primary bronchial branches and early vascular connections, though they remain non-functional until the foetal period [116, 117]. While this period establishes the foundational framework of the respiratory system, the foetal period focuses on refining and expanding these structures to prepare for functional gas exchange after birth.

The foetal period

The foetal period is characterised by the formation of all the conductive airways and the appearance of the gas-exchanging (acinar) unit. This period is categorised into three distinct sub-periods: the pseudo-glandular stage (5–16 weeks of gestation), the canalicular stage (16–26 weeks of gestation), and the saccular/alveolar stage (26 weeks to early childhood) [117].

During the pseudo-glandular stage, numerous smaller bronchi, bronchioles, and terminal bronchioles are generated, creating lung tissue that cannot yet exchange gases [117, 120]. The airways are lined with ciliated cells and mucus-secreting goblet cells, which help to clear the respiratory passages. The airway epithelium produces and secretes a chloride-rich fluid that fills the lungs to support their growth [121]. This fluid, often referred to as lung fluid, serves as a medium to support the development and expansion of the lung tissue. However, it does not facilitate gas exchange, which is performed through the placenta during this period.

The canalicular stage occurs around 16 to 26 weeks of gestation. The terminal bronchioles further divide into respiratory bronchioles, and the development of the pulmonary blood vessels progresses significantly. At the same time, thin-walled capillaries surround the developing airways, creating a rudimentary air-blood interface that allows limited and inefficient gas exchange [122]. The respiratory bronchioles also give rise to primitive alveolar ducts and sacs, the foundation for future alveoli. While these gas exchange structures are forming during this stage, they remain non-functional until later in development.

During the saccular and alveolar stages, the lungs continue to grow in utero. However, they remain functionally dormant until birth because, up until then, the gas exchange is done through the placenta. The alveoli, responsible for gas exchange outside the womb, develop from around the 26^{th} week until early childhood (around 2 to 5 years of age). During this period, the alveolar ducts and sacs mature and grow, increasing the lung's surface area and optimising the efficiency of gas exchange. The capillary network surrounding the alveoli becomes denser, allowing for the exchange of oxygen and carbon dioxide to sustain life [117, 119].

The postnatal period

The first inhalation is critical in ensuring the lung fluid is cleared from the lungs and replaced with air. This inspiration effort generates a pressure gradient that moves the fluid into the interstitial tissue (the extracellular space surrounding the alveoli, airways, and blood vessels), where the pulmonary and lymphatic circulations gradually remove it [123]. Although this marks a quick adaptation to life outside the womb (extra-uterine), an infant's respiratory physiology continues to undergo significant developmental changes. Alveolar growth continues into the postnatal period to approximately 2 years of age [124].

When lung development is not interrupted by premature birth, a fully functional respiratory system capable of sustaining life is formed. However, pre-term infants are born before completing their intra-uterine lung development. Because of underdeveloped alveoli in their lungs, pre-term infants delivered before 28 to 30 weeks of gestation may need to rely on their terminal bronchioles and primitive alveoli to breathe. The alveolar structures responsible for gas exchange are not fully developed, and surfactant production is insufficient to support normal lung function [15, 124]. Consequently, their respiratory system cannot function as efficiently as that of full-term infants. Pre-term infants may have several respiratory deficiencies, such as immature respiratory control centres, a highly compliant chest wall or small respiratory passage, leading to respiratory complications such as AOP, bronchopulmonary dysplasia, and increased vulnerability to infections like pneumonia [125].

Breathing

The respiratory system is vital in moving oxygen into the bloodstream and removing carbon dioxide. This process, known as pulmonary ventilation, involves two distinct phases: inhalation and exhalation. Inhalation occurs when the diaphragm contracts and moves downwards, expanding the thoracic cavity. This action reduces the pressure inside the lungs, allowing air to enter the lungs through the respiratory passages. Exhalation occurs when the diaphragm relaxes, and the thoracic cavity volume decreases, increasing the pressure within it. The elasticity of the lungs and chest wall causes them to return to their resting shape, pushing air out of the lungs through the airways when the diaphragm relaxes.

The chest walls visibly move up and outward during inhalation and down and inward during exhalation, reflecting changes in the volume of the thoracic cavity. This movement is especially evident in pre-term infants, whose chest walls have greater flexibility than older children and adults, leading to more unusual breathing patterns. Differences between the chest wall of premature and full-term infants include anatomical factors such as the pre-term infant having less bone formation in their ribcage, resulting in increased flexibility [126]. Pre-term infants also have a higher chest wall compliance, making their chest

walls more pliable and prone to both collapse and paradoxical movement or recession, where the chest wall moves inward during inspiration instead of expanding outward as it should. This can hinder the full expansion of the lungs, making it difficult for the infant to generate sufficient lung volumes and maintain effective ventilation [127, 128]. The intercostal and accessory muscles of respiration of pre-term infants are less mature in contractile activity, making their diaphragmatic contraction less efficient [126, 129, 130]. As a result, this may cause muscular fatigue and apnoea [131, 132, 133].

2.2.3 Anatomy of the skin

The skin is the largest organ in the body, accounting for approximately 15% of total body weight [134]. It plays an important role in protection, thermo-regulation, and fluid balance. Structurally, the skin consists of the epidermis, dermis, and subcutaneous layer, as illustrated in figure 2.3. The epidermis contains melanocytes, which produce melanin and influence skin pigmentation and optical absorption properties. Beneath the epidermis, the dermis contains blood vessels and connective tissue that contribute to light scattering and absorption within biological tissue [135, 136].

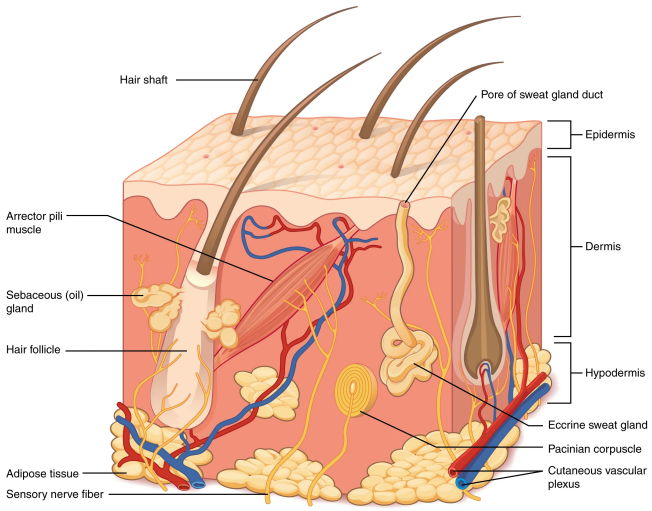


Figure 2.3: Diagram of the structure of the skin showing the epidermis, dermis, and subcutaneous layers. The epidermis comprises epithelial cells. The dermis contains connective tissue, blood vessels, and sensory structures, while the hypodermis consists mainly of loose connective and fatty tissue. (Image courtesy of [137].)

Skin maturation progresses with gestational age, with the epidermal barrier typically reaching full maturity by the 34^{th} week of gestation [142]. Consequently, pre-term infants often have thinner, more translucent skin with incomplete epidermal development and reduced melanin deposition. These characteristics influence how light is absorbed, reflected, and scattered within the tissue, affecting the performance of optical monitoring methods such as pulse oximetry and photoplethysmographic imaging (PPGi). The fragile skin of pre-term infants is also susceptible to damage from adhesive sensors, motivating the development of non-contact physiological monitoring approaches [142, 143].

2.3 Vital signs

Vital signs are a crucial aspect of assessing a patient's health status. They encompass a group of physiological measurements, including blood pressure, body temperature, heart rate (HR), respiratory rate (RR), and peripheral oxygen saturation (SpO_2). These measurements provide the first indication of abnormal physiological body changes [144] and are important in monitoring a patient's recovery progress during hospitalisation and determining a treatment plan [145]. In the NICU, continuous monitoring of vital signs can significantly improve neonatal outcomes by enabling precautionary interventions [146]. Considering their central role in clinical care, understanding the trends in vital signs is critical in supporting clinicians in making a diagnosis or treatment decision [147]. This thesis focuses on three important vital signs for detection of COBE events: HR, RR and SpO_2.

2.3.1 Heart rate

HR is one of the most frequently measured vital signs. It represents the frequency of heartbeats or cardiac contractions per minute. HR provides crucial information about the cardiovascular system's performance and the patient's overall physiological state. In pre-term infants, normal resting heart rate varies with gestational age, ranging from approximately 120 to 180 beats/min at the 28^{th} week of gestation. In infants at one month of age, normal HR decreases, varying between 90 and 160 beats/min as autonomic regulation matures. In adults, the typical resting heart rate is 60 - 100 beats/min [148]. Heart rate is used as an indicator of neonatal health, reflecting baseline cardiac function and the infant's response to physiological stressors such as activity, sleep state, pain, or illness. Deviations from normal ranges, including tachycardia or bradycardia, may signal underlying cardiac, respiratory, or metabolic issues that require prompt clinical attention [146, 149, 150]. Continuous monitoring of heart rate in the NICU therefore provides critical real-time information to guide interventions and support optimal neonatal outcomes.

2.3.2 Respiratory rate

RR is the number of breaths or respiratory cycles per minute. It is an important indicator of various pathological conditions such as pneumonia, adverse cardiac events, and apnoea. Respiratory rate is widely regarded as one of the earliest physiological indicators of patient deterioration and can provide an early indication of clinical decline before significant changes are observed in other vital signs such as heart rate or oxygen saturation [151, 152]. Because changes in RR are highly sensitive to physiological deterioration, medical professionals often monitor respiratory trends to ensure timely intervention and the prevention of adverse outcomes [53, 147]. For pre-term infants, a universally agreed-upon range for respiratory rate during regular breathing has not been established because of the dynamic and unstable

nature of their respiratory physiology. Nevertheless, several studies have reported that typical breathing rates in pre-term infants can range from 40 to 70 breaths/min depending on gestational age, with higher values observed during respiratory distress [153, 154, 155]. At one month old, infants typically breathe at 30 to 60 breaths/min, whereas adults' respiratory rate can range from 12 to 20 breaths/min [84, 156]. The variability and instability of respiratory patterns in pre-term infants therefore make continuous respiratory monitoring an important component of neonatal care, particularly during periods of physiological instability.

Despite its clinical significance as an early indicator of deterioration, RR remains one of the least reliably measured vital signs in routine clinical practice, partly because respiratory assessment commonly relies on manual breath counting [157]. Manual breath counting is time-consuming, prone to observer variability and decrease in accuracy, particularly in pre-term infants with irregular breathing patterns and spontaneous body movements [158, 159]. Consequently, RR has historically been under-recorded [158] or inaccurately documented in some healthcare settings [157, 160]. These limitations motivate the development of automated respiratory monitoring systems capable of continuously and objectively assessing respiratory activity in pre-term infants.

Automated respiratory monitoring systems may be implemented using either real-time or off-line analysis approaches. Real-time systems are designed to provide continuous, live monitoring of physiological signals and therefore require computationally efficient algorithms with minimal processing latency to ensure timely detection of respiratory deterioration [161]. Consequently, these systems must balance algorithm complexity and processing speed to support rapid clinical intervention, particularly in pre-term infants whose respiratory patterns can change rapidly during periods of physiological instability [162]. In contrast, off-line respiratory monitoring emphasises retrospective analysis on collected datasets, enabling more computationally intensive signal processing, extensive feature engineering, and the development of more complex machine learning models [88, 163, 164]. The distinction between real-time and off-line respiratory monitoring is therefore important when evaluating the clinical applicability of respiratory monitoring algorithms and influences both algorithm design and model complexity.

2.3.3 Peripheral oxygen saturation

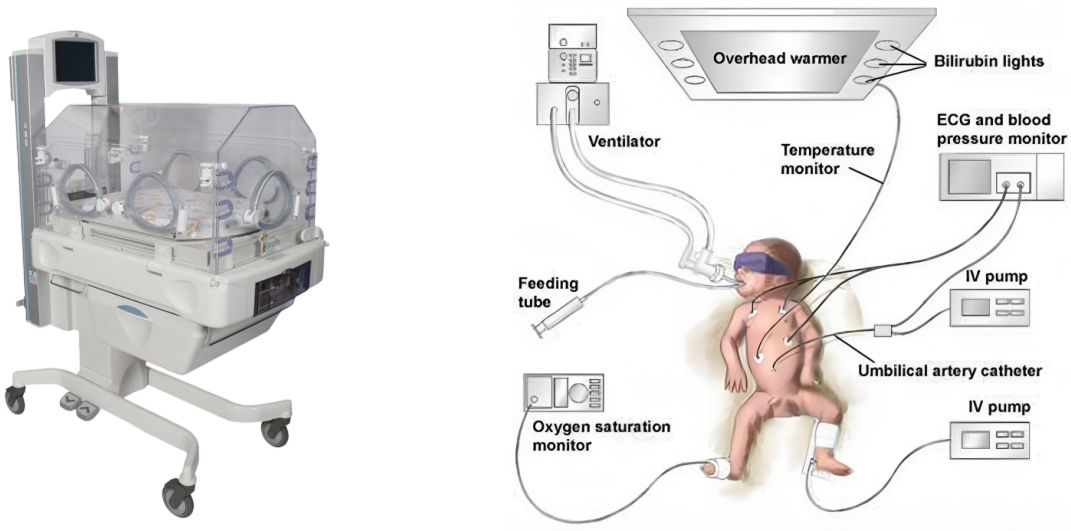
SpO_2 is a vital sign that indicates the proportion of oxygenated haemoglobin relative to total haemoglobin in the blood. It reflects the body's ability to maintain adequate oxygen levels essential for cellular metabolism and organ function, making it crucial for assessing a patient's overall health [165]. Monitoring of oxygen levels is important, especially in pre-term infants, as too high or too low levels can be detrimental [166]. A target oxygen saturation level between 90% and 94% is generally recommended for ongoing management of pre-term infants in the NICU to balance the risks of hypoxia and oxygen toxicity [166, 167, 168]. However,

specific targets and alarm limits may vary across institutions, reflecting ongoing uncertainty and differences in clinical care practice [169].

Accurately estimating oxygen saturation is important for detecting potential changes in a patient's condition. For instance, hypoxia, a condition where the oxygen level decreases below 80%, indicates a shortage of oxygen supply to the body and must be promptly identified. Monitoring and maintaining optimal SpO_2 levels is also a critical component of treating respiratory diseases such as bronchopulmonary dysplasia, pneumonia and apnoea [170].

2.4 Cardio-respiratory monitoring in the NICU

Pre-term infants often require admission to the NICU immediately after birth. Because their physiological systems are underdeveloped, these infants may experience medical complications that necessitate specialised care. The NICU is therefore equipped with incubators (see figure 2.4a) and complex medical equipment that is tailored to the specific needs of premature babies, with continuous clinical monitoring until discharge. Monitoring typically involves the use of dedicated monitors with sensors that provide comprehensive information about an infant's HR, breathing patterns, oxygen saturation levels and blood pressure. This information is essential for assessing the infant's well-being and body functions during their hospitalisation.



(a) (b)

Figure 2.4: Monitoring of vital signs in the Neonatal HDU. a) Typical incubator used in the NICU. Image courtesy of GE Healthcare. b) Schematic of NICU support provided to neonate in the incubator as published by [171]. An oxygen tube is attached to the ventilator to provide supplemental respiratory support. Electrodes are attached to the body to monitor cardiovascular physiology using an ECG. A pulse oximeter, placed around the foot, monitors the oxygen supply. An intravenous (IV) pump provides the infant with nutrients or fluids. The image was used with permission from the Mayo Foundation for Medical Education and Research; all rights are reserved.

Figure 2.4b shows the medical equipment utilised in the NICU to assist in nursing an infant in an incubator. Patient monitors are used to quantify physiological parameters by means of electrodes and sensors that are directly attached to the infant's body. These devices continuously monitor vital signs, which are then documented by clinical personnel in the patient records at specific intervals. As depicted in figure 2.4b, there are several commonly employed tools, including the placement of three electrodes on the infant's chest for recording the electrocardiogram (ECG) and impedance pneumography (IP) signals. These signals are used to evaluate cardio-respiratory physiology and estimate the HR and RR values. A pulse oximeter is typically placed on the infant's foot to monitor SpO_2 levels in the blood. Pulse oximeters help in the detection of potential apnoea events, as any major decreases in oxygen saturation levels are recorded and reported. Typically, alarm thresholds using monitored vital signs specific to neonates are set, and healthcare providers are alerted in case of a threshold breach, allowing timely intervention. Though rarely used after the first seven days of life, umbilical artery catheterisation is commonly used in extremely premature neonates to provide direct access to the arterial system, thus enabling the measurement of blood pressure [172].

The use of these standard-of-care sensors in early pre-term infants can cause skin irritation or infection due to the infants' fragile and sensitive skin. Due to the current technology's heavy reliance on cables and wires, there is a possibility of wires becoming tangled around the infant's body. Furthermore, cables and wires can act as a hindrance to parent-infant interactions and prevent skin-to-skin contact [60, 173].

2.4.1 Electrocardiogram (ECG)

The ECG is considered the standard technique for measuring heart rate in the clinic. It is a non-invasive method that records the voltage difference between two points on the body surface, generated by the action potentials in the heart muscle with each heartbeat [174]. The ECG measures this electrical activity through electrodes placed on the patient's skin [175], enabling clinicians to monitor heart rate trends and identify cardiovascular abnormalities or irregular heart rhythms. A normal cardiac cycle, represented by a typical ECG, begins with the P wave, indicating atrial depolarisation (contraction), followed by the main processes of ventricular depolarisation and repolarisation (relaxation). The QRS complex represents ventricular depolarisation and activation, whereas ventricular repolarisation is represented by the T wave [176].

QRS detection

The QRS complex is a key feature in an ECG representing the rapid depolarisation of the ventricles. It appears as a sharp, triangular-shaped wave on the ECG trace and is composed of three distinct

parts: the Q wave represents the initial negative deflection, the R wave represents the subsequent sharp positive deflection, and the S wave represents the final sharp descending deflection following the R wave, completing the complex [177, 178]. Figure 2.5 shows a typical ECG waveform of an adult with the P-wave, QRS-complex, and T-wave labelled. These features provide a detailed representation of cardiac electrical activity and allow clinicians to assess heart rate, rhythm, and conduction pathways [176].

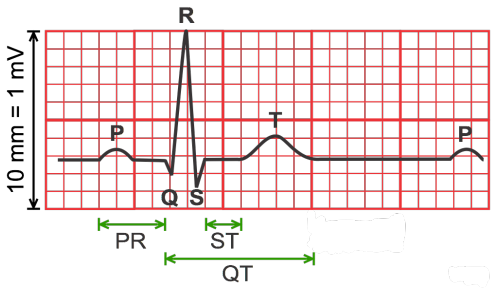


Figure 2.5: A typical ECG waveform of an adult contains a P wave, followed by a QRS complex and a T wave. Image adapted from [179].

Application in the NICU

Continuous ECG monitoring in the NICU allows clinicians to detect arrhythmias and identify early signs of cardiovascular compromise in real time. In pre-term infants, whose cardiovascular systems are still maturing, features derived from QRS complexes and the ECG amplitude can aid in apnoea identification [180, 181, 182]. ECG data can also be combined with pulse oximetry or impedance pneumography to monitor cardiorespiratory events such as apnoea-associated bradycardia.

Limitations

Despite its advantages, ECG monitoring in neonates presents several challenges. The fragile skin of pre-term infants limits electrode placement and can cause skin irritation or damage [53, 183]. Electrical power-line interference, baseline wander, and motion artefacts can also contaminate the signal, leading to inaccurate heart rate recordings and potentially affecting the diagnosis of cardiac abnormalities [184]. Furthermore, interpreting neonatal ECG requires expertise, as typical waveform morphology and HR ranges differ significantly from adults and vary with gestational age [185].

2.4.2 Impedance pneumography

Within a clinical setting, several techniques are available for monitoring respiratory rate, including using a strain-gauge transducer, collecting airflow to and from the lungs through a nasal cannula or mouthpiece, manual breath counting, and using IP [186, 187]. Among these, IP is the current widely accepted method for measuring respiratory rate [188]. It is based on the principle that the resistance to an electrical current

passing through the thorax changes due to cyclic changes in air content during breathing [189]. Utilising the same electrodes as for the ECG, IP requires the injection and transmission of high-frequency current through electrode pairs on the chest [190]. During inhalation, air fills the lungs, increasing the electrical impedance of the thoracic cavity, as the lung's effective resistivity rises due to the presence of air, which is much less conductive than the surrounding tissues. Alternatively, during exhalation, impedance decreases as air leaves the lungs. This change in electrical properties can be detected by electrodes placed on the skin (see figure 2.6). Subsequent processing of the IP signal allows extraction of respiratory rate and estimation of tidal volume, providing continuous monitoring of breathing patterns in neonates.

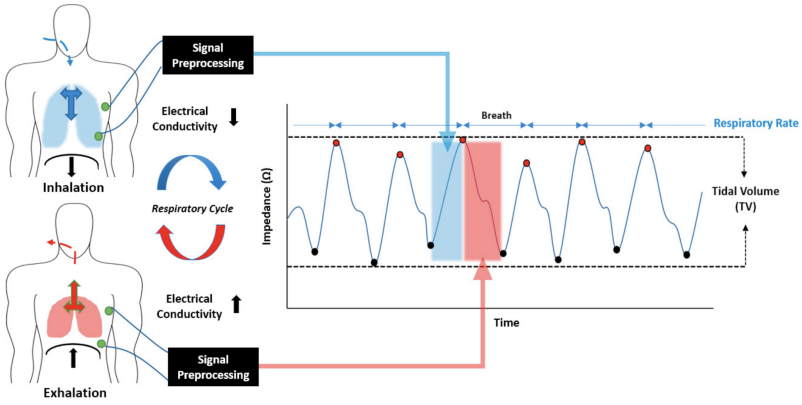


Figure 2.6: Measurement of RR using impedance pneumography. Image reproduced from [191].

Applications

In clinical settings like the NICU, continuous IP monitoring allows for the detection of changes in breathing patterns, enabling prompt intervention in cases of respiratory deterioration. This non-invasive method is particularly valuable for monitoring pre-term infants at risk of COBE events and other respiratory complications. Because IP is commonly used for continuous bedside monitoring in the NICU, respiratory rate estimation from the IP signal is typically expected to operate in real time under computational and latency constraints to support timely clinical intervention. In neonates, IP is also widely used in retrospective off-line analysis to study neonatal respiratory physiology and assess sleep-related respiratory events. For instance, studies have utilised IP to analyse tidal breathing flow volume profiles during sleep in wheezing infants, providing insights into respiratory patterns and potential airway obstructions [192]. Additionally, IP has been used to measure interbreath intervals in infants, aiding in the assessment of respiratory function and the identification of abnormal breathing patterns [163].

Limitations

Various factors can negatively impact the recording of the IP signal. These include electrode placement, infant's orientation or position, electrodes peeling off the skin due to infant movement, and very dry

skin leading to low conductance between electrodes and skin [53, 193]. Unrelated thoracic movements such as coughing or crying can also lead to false breath estimations [194]. The IP signal is, therefore, susceptible to noise and artefacts, which can significantly impact the accuracy of RR estimation [188]. In pre-term infants, these limitations may contribute to false detections or missed identification of clinically significant respiratory events such as central sleep apnoea [194, 195].

2.5 Pulse oximetry

Pulse oximetry relies on how light interacts with the skin and underlying tissues. As light travels through the epidermis, dermis, and subcutaneous layers, its transmission is affected by pigments such as melanin and by haemoglobin in the blood. These interactions, which depend on the light wavelength, form the basis for optical sensing methods such as PPG and pulse oximetry. These methods exploit pulsatile changes in haemoglobin absorption to provide non-invasive measurements of SpO_2 and related cardiovascular parameters.

2.5.1 Skin-light interactions

The electromagnetic spectrum includes the full range of electromagnetic radiation across a wide range of wavelengths, from gamma rays with wavelengths below 1×10^{-11} m to radio waves that measure in metres [196]. Within this spectrum, the human eye can only perceive visible light, which exists within the 400 to 780 nm wavelength range. Figure 2.7 shows the full electromagnetic spectrum with corresponding wavelength, frequency, and photon energy.

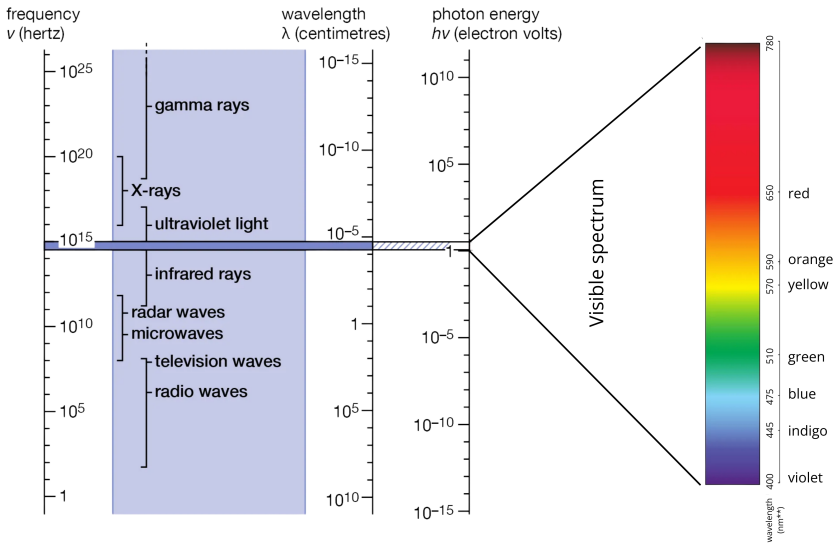


Figure 2.7: Electromagnetic spectrum with wavelength, frequency, and photon energy. The visible region is shown enlarged on the right. Image adapted from Encyclopædia Britannica, Inc.

Light can be refracted, reflected, scattered, or absorbed when it interacts with matter. Refraction is the bending of light rays after passing into a medium. Reflection entails the return of light rays to their original medium. Scattering occurs when light penetrates a medium and disperses in multiple directions. Light absorption occurs when a medium's interior absorbs the energy of photons [197]. Light intensity decreases as it passes through a substance that absorbs it, such as skin and underlying tissues. This attenuation of light can be described mathematically using the Beer-Lambert law. As light I_o, with a wavelength of λ, enters a homogeneous absorbing medium, some of the light is absorbed while the rest is transmitted through the medium [198]. The intensity of the transmitted light, I, decreases with respect to its travelling distance and the medium's absorption properties, as defined by:

$$I(\lambda) = I_o(\lambda) \cdot exp\{-\alpha(\lambda)L\} \tag{2.1}$$

where $\alpha(\lambda)$ defines the extinction coefficient of the absorbing medium at the wavelength λ, L is the light path length [199]. The extinction coefficient α quantifies the fraction of light absorbed or attenuated by the medium.

Although the Beer-Lambert law assumes a homogeneous, non-scattering medium, biological tissues are complex, with multiple layers, varying composition, and dynamic optical properties. Nonetheless, the Beer-Lambert law provides a useful conceptual basis for understanding how light interacts with skin and underlying tissues, and governs the science for optical sensing methods such as pulse oximetry [200].

In the context of skin, the physical properties of light, including its wavelength and intensity, have a significant impact on its propagation and absorption. The epidermis plays a crucial role in reflecting, transmitting, and absorbing light. About 5% - 7% of the light incident on the stratum corneum is reflected to the environment, whereas the majority permeates into the body tissues [135]. The epidermis's absorption property primarily depends on skin pigments, specifically melanin. The human skin has two distinct types of melanin: brownish-black eumelanin and reddish-yellow phaeomelanin. The absorption spectra of phaeomelanin and eumelanin are broad, with absorption higher at shorter wavelengths (400 nm) of the visible spectrum (see figure 2.8a).

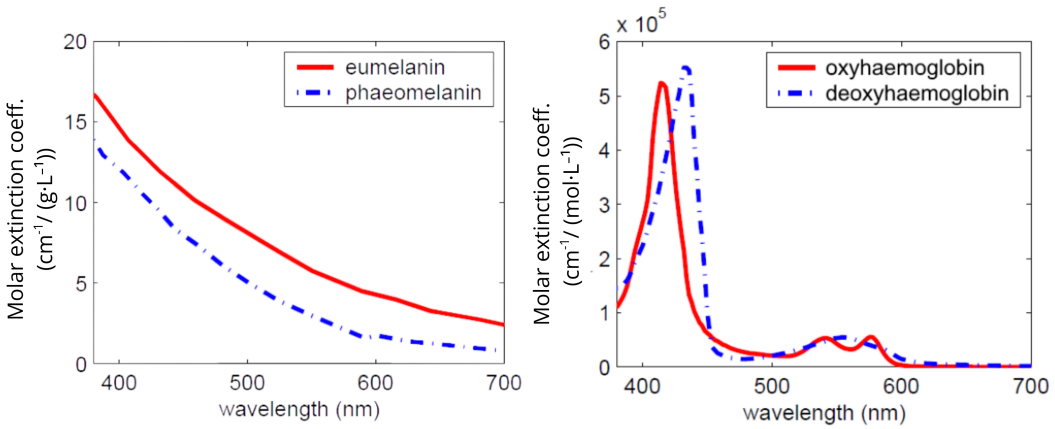


Figure 2.8: Absorption spectra of the skin at different wavelengths. a) Absorption spectrum of eumelanin and phaeomelanin. b) Absorption spectrum of oxyhaemoglobin and deoxyhaemoglobin. Images courtesy of S. Prahl and the Oregon Medical Laser Centre (OMCL).

While melanin primarily influences light absorption at the skin's surface, the deeper layers, such as the dermis, also interact with blood components, particularly haemoglobin. The concentration of melanosomes, which are membranous particles that contain melanin, affects how much light is absorbed by the epidermis [135]. This absorption property can vary significantly between individuals and among different body parts. The dermis also reflects, propagates and absorbs light. It primarily comprises connective tissues with blood vessels and nerves. Blood is heterogeneously distributed within the dermis, and the amount of blood in the tissues can vary from 0.2% to 7%. As the depth of the dermis increases, the amount of light absorbed decreases exponentially [201].

Oxygenated and deoxygenated haemoglobin exhibit distinct absorption characteristics, as figure 2.8b shows. Approximately 90%-95% of the haemoglobin in arteries is oxygenated. Veins contain less oxygenated haemoglobin than arteries, reflecting oxygen uptake by surrounding tissues [135, 202, 203]. The subcutaneous or hypodermis layer contains adipose tissues that exhibit minimal light absorption in the visible spectrum. This layer predominantly comprises white fat deposits, wherein the cells are grouped, forming clusters. Due to these deposits, most of the visible light that reaches this tissue is reflected back toward the upper layers [135]. In pre-term infants, skin reflectance in visible light increases as gestational age progresses and the skin matures [204]. This reflectance level depends on several factors, including the skin's developmental stage, structural properties such as epidermal thickness and vascularisation, gender, and skin pigmentation [142, 204]. During the first postnatal days, light penetration through the skin is greater because of incomplete epidermis development and a thin stratum corneum. As gestational age advances and the skin matures, light penetration decreases and optical absorption within the skin gradually increases [142, 205].

2.5.2 Photoplethysmography

One of the earliest light-based continuous vital-sign monitoring methods developed is PPG, first investigated in the 1930s [206]. PPG is a non-invasive technology that uses a light source and a photodetector at the skin's surface to measure the volumetric variations of blood circulation. The light source illuminates the tissue, and the photodetector records the small variations in reflected or transmitted light intensity associated with changes in perfusion. The fundamental principle of PPG relies on the differences in sensitivity of different optical wavelengths to blood and other tissue components such as skin and bone [207]. When the heart pumps blood to the body during systole, the amount of blood that reaches the capillaries in the skin surface increases, resulting in increased light absorption. Blood then travels back to the heart through the venous network, leading to decreased blood volume in the capillaries and a decrease in light absorption [208]. These fluctuations in blood volume, primarily arterial, along with changes in scattering properties, influence the measured PPG waveform. It comprises a pulsatile absorption signal that can be used to evaluate SpO_2 (using measurements of two wavelengths) and other vital signs such as HR and RR.

The PPG waveform consists of a pulsatile component (AC), synchronous with the beating heart, and a non-pulsatile component (DC), which reflects the baseline blood volume. Most of the DC signal is static and represents light that has not been modulated by arterial pulsation, including contributions from ambient light, tissues, venous blood, and non-pulsatile arterial blood [210]. Slower variations in the DC component are modulated by respiration, sympathetic nervous activity, and thermoregulation. The PPG waveform is shown schematically in figure 2.9.

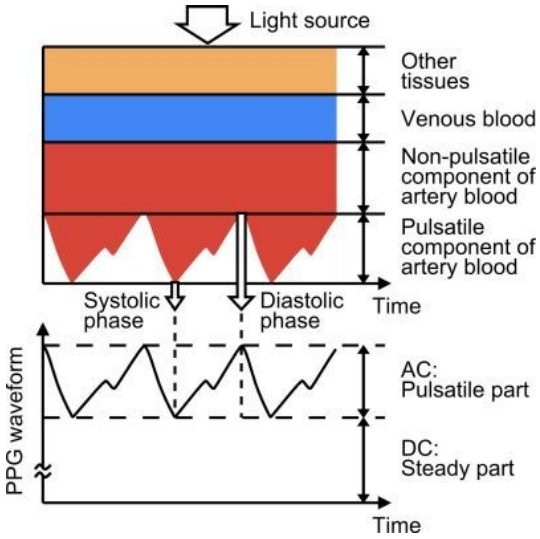


Figure 2.9: Components of the PPG signal. This signal comprises an AC pulsatile part and a steady DC part. Image reproduced from [209].

The standard Beer-Lambert law defined in equation 2.1 governs the light absorption in a homogeneous medium. In a dynamic medium, the absorption coefficient, optical path length, and scattering properties vary with blood flow, requiring a more complex model than the static form of the Beer-Lambert law. The human tissue is dynamic, and the intensity of light passing through the tissue varies as the blood volume changes beneath it [211]. When a light I_o of wavelength λ is transmitted into a non-scattering dynamic medium with an absorption coefficient of $\alpha_a(\lambda)$, the resulting intensity I detected by a photodetector can be modelled by including a time-varying effective optical path length $Lp(t)$. This effective path length reflects changes in the volume of pulsatile arterial blood beneath the sensor.

$$I(\lambda) = I_o(\lambda) \cdot exp\left\{-\alpha_a(\lambda) Lp(t)\right\} \tag{2.2}$$

The optical path length can be broken down into two components: static and dynamic, representing the optical path length of the DC and AC components, respectively (see figure 2.9):

$$I(\lambda) = I_o(\lambda) \cdot exp\left\{-\left[\alpha_{a(static)}(\lambda) L_{static} + \alpha_{a(arterialblood)}(\lambda) L_{dynamic} p(t)\right]\right\} \tag{2.3}$$

The AC component, which typically accounts for approximately 5% of total intensity [211], is useful in assessing cardiac performance, vascular disease, and arterial compliance [212, 213]. The DC components provide equally valuable information about temperature changes, sympathetic outflow, venous volume fluctuations, and other regulatory mechanisms [211].

2.5.3 Pulse oximetry

Haemoglobin is the primary chromophore responsible for light absorption in tissue within the red and near-infrared regions. However, other chromophores, such as melanin and myoglobin, can also absorb light within these regions. Venous blood, which contains less oxygenated haemoglobin, also absorbs light in the same spectral region as arterial blood. The need to isolate the contribution of arterial blood's haemoglobin to total absorption has led to the development of pulse oximetry [214]. Pulse oximetry is a non-invasive technology widely accepted as the standard for detecting hypoxaemia, an often fatal result of apnoea [215, 216]. Based on PPG, pulse oximetry measures peripheral arterial oxygen saturation by calculating the amount of light absorption by blood and its components, particularly haemoglobin.

Principles of pulse oximetry

The depth to which light can penetrate the human skin is directly related to its wavelength. The longer the wavelength, the deeper the penetration. Therefore, it is possible to measure deep-tissue blood flow by using long wavelengths of light, such as red (660-700 nm) and infrared light [217]. Pulse oximetry relies

on the principle that oxy-haemoglobin (HbO_2) and deoxy-haemoglobin (Hb) absorb the red and infra-red (IR) parts of the spectrum differently, as illustrated in figure 2.10.

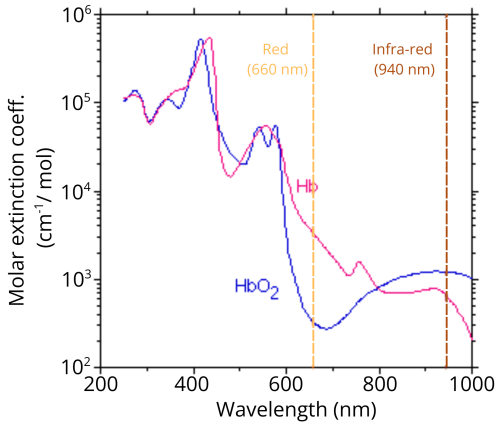


Figure 2.10: Absorption spectra of HbO_2 and Hb. Image adapted from [214], courtesy of S. Prahl and the Oregon Medical Laser Centre.

Oxygenated blood is more absorbent at a wavelength of 660 nm (red), while de-oxygenated blood is more absorbent at 940 nm (infrared). Taking advantage of this difference in light absorption properties between HbO_2 and Hb, pulse oximeters use small light-emitting diodes to emit two wavelengths of light (red and near IR). The light travels through the finger and is captured by a photodiode. By measuring the quantity of red and IR light absorbed or reflected, the pulse oximeter then determines the percentage of haemoglobin bound to oxygen [215, 218].

The ratio of AC to DC components at the red and infra-red wavelengths, known as the ratio of ratios R (see equation 2.4), is related to SpO_2 using previously derived empirical curves, allowing continuous non-invasive monitoring by pulse oximetry [215, 218].

$$R = \frac{AC_{RED}/DC_{RED}}{AC_{IRED}/DC_{IRED}} \tag{2.4}$$

This ratio governs the fundamental measurement principle behind pulse oximetry devices. By exploiting the wavelength-dependent absorption properties of oxygenated and deoxygenated haemoglobin, pulse oximeters use the PPG signal to estimate SpO_2 in real time.

Mode of operation

Generally, there are two operational modes: transmission and reflectance. These modes are determined by the positions of the light source (LED) and photodetector (PD), as illustrated in figure 2.11. In transmission mode, the photodetector and light source are on opposite sides of the tissue. The tissue is irradiated, and the photodetector on the other side of the tissue measures the light intensity transmitted (see figure 2.11a).

Transmission mode pulse oximeters are typically limited to body segments with high vascularisation and less tissue, such as fingertips, toes or earlobes. When using the transmission mode, the pulse oximeter should be secured to minimise probe-tissue movement artefact and light interference [212].

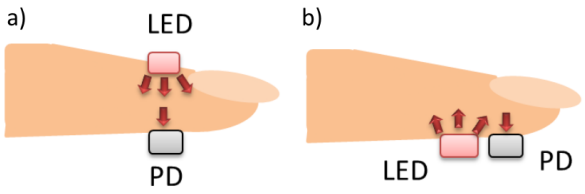


Figure 2.11: Two operational modes of pulse oximetry. a) Transmission mode; b) Reflectance mode; Images reproduced from [209].

In reflectance mode, the photodetector and LED are positioned adjacent to each other on the same side of the tissue (see figure 2.11b). The photodetector captures the intensity of the backscattered light from the skin. Reflectance mode provides greater flexibility of which body site to use. With this mode, pulse oximeters are suitable for application on skin surfaces with high superficial vascularisation, such as the forehead, temple, limbs and chest [219].

Transmission mode pulse oximetry generally provides a better signal-to-noise ratio than reflectance mode. This is because of the smaller fraction of the emitted light that reaches the photodetector in reflectance mode pulse oximetry [220, 221]. However, transmission mode pulse oximetry requires a higher operation current for light emission through the tissues. Reflectance-mode pulse oximetry is, therefore, more favourable for longer-time usage in a wearable medium [222]. Some studies have suggested using reflectance-mode pulse oximetry in pre-term infants by placing sensors on the forehead, as it offers stable perfusion, reliable pulsatile signals, and minimal interference during resuscitation [223, 224].

Applications

Pulse oximeters are commonly used in clinics to provide a continuous measurement of SpO_2 by measuring light absorbance at two wavelengths. This technology is also widely used to measure heart rate by detecting changes in the PPG signal's AC component [225, 226, 227]. Pulse oximeters can also measure respiratory rate by analysing slow variations in the PPG signal that occur with each breath. These amplitude modulations, superimposed on both the baseline (non-pulsatile component of the plethysmographic signal (DC)) and pulsatile (pulsatile component of the signal (AC)) components of the waveform, reflect changes in blood volume, thoracic pressure, and venous return associated with respiration. Modulations occur at both cardiac and respiratory frequencies, representing a pulsatile phenomenon at multiple timescales. Pulse oximetry enables healthcare professionals to detect life-threatening conditions such as apnoea, hypoxemia, cardiorespiratory decompensation, and bronchopulmonary dysplasia [146], facilitating prompt

intervention. Patent ductus arteriosus in pre-term infants can also be identified and evaluated through the measurement of PPG waveforms from multiple sites [228].

Limitations

Despite its advantages, the accuracy of pulse oximeter readings depends greatly on how well the sensors are positioned. If the sensors are poorly placed, it can lead to optical shunting or the penumbra effect. This happens when the LED light from the pulse oximeter reaches the photodetector without passing through an arterial bed. As a result, falsely low SpO_2 readings are generated, affecting the quality of the signal [229, 230].

Pulse oximetry relies on the measurement of light absorption to estimate oxygen saturation, making it susceptible to various other factors that can lead to inaccurate readings. Factors such as poor peripheral perfusion, ambient light, motion artefacts, and absorbance of light by other tissue and blood components can all introduce inaccuracies in the measurement. Skin pigmentation and body temperature also influence the accuracy of pulse oximetry [215, 231]. In pre-term infants, differences in skin reflectance and optical absorption due to immature epidermis and thin stratum corneum can further affect the accuracy of pulse oximetry readings, requiring adjustments in clinical monitoring practice [142]. Though rarely, pulse oximeters have also been reported to cause complications such as infections, burns, and pressure necrosis (skin sores caused by pressure) [232].

2.6 Non-contact vital sign monitoring

Research into non-contact vital sign monitoring methods, such as Photoplethysmographic Imaging (PPGi), has led to the development of new ways to measure and evaluate important physiological parameters in healthcare without the need for direct physical contact with the patient.

2.6.1 Photoplethysmography imaging (PPGi)

Although conventional PPG has a wide range of applications, it has several significant limitations that PPGi aims to address. Current monitoring systems available to track changes in patients' vital signs require contact with the subject by using adhesive electrodes or sensors [233]. These can, however, damage the fragile skin of young infants or cause stress and discomfort. The introduction of digital video cameras into clinical imaging systems, along with the desire to reduce physical restrictions and gain new insights that might come from perfusion imaging, inspired the evolution of conventional PPG technology to PPGi [208].

The underlying principles of PPGi closely resemble the reflectance mode PPG as described in section 2.5.3. For PPGi, ambient light, usually regular room lighting, serves as the light source, while a video camera functions as an array of photodetectors. Both tissues and blood within the skin absorb the light penetrating the surface, allowing the video camera to capture variations in the intensity of light reflected from the skin's epidermal layer, which reflects changes in blood volume during systole and diastole. However, unlike the reflectance-mode PPG that measures light intensity reflected from deeper tissue layers (such as arterioles), PPGi measures light intensity reflected from shallower vasculature, such as capillaries [68, 200]. Figure 2.12 shows a representation of the PPGi technique overlaid with the interactions of light with the skin.

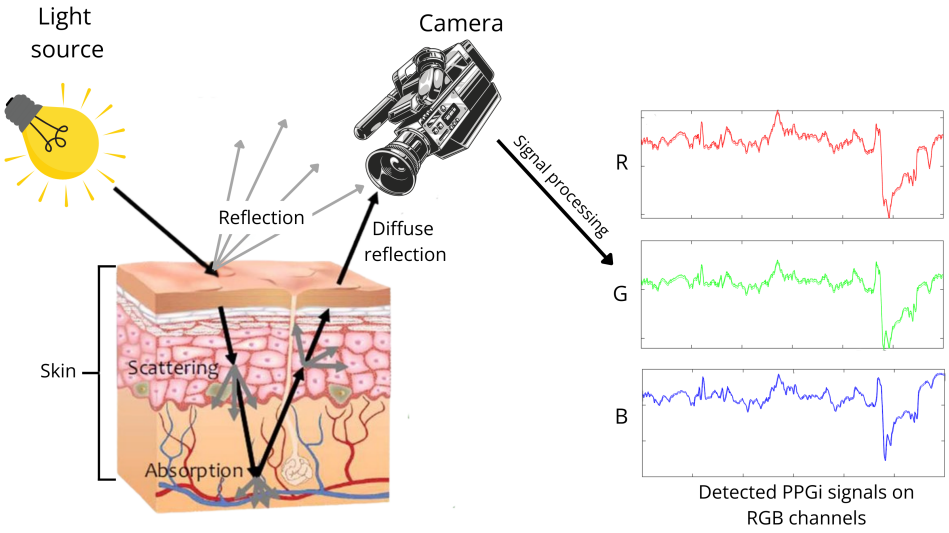


Figure 2.12: Illustration of the PPGi signal. The setup normally consists of a light source, camera, and pulsating skin area [234]. This method measures light reflected from the skin and tracks changes in red, blue, and green light caused by blood flow. By using a suitable computational approach, a physiological signal is obtained, which can be used to compute vital signs.

2.6.2 Vital-sign estimation

Video-based vital-sign monitoring extends the principles of traditional PPG, using the multiple photosites present in the imaging sensor of a video camera to record the blood volume changes associated with the cardiac cycle. This technology leverages imaging systems to derive critical physiological parameters from images of the human body. By analysing volume changes in blood flow across larger skin areas, PPGi enables the estimation of vital signs such as heart rate and oxygen saturation. These physiological changes result in a signal from which vital signs such as HR, RR, SpO_2 and others can be estimated [235, 236]. Various researchers have investigated estimating vital signs from camera-derived signals.

Blazek *et al.* [237] proposed one of the first computer-based imaging systems using near-infrared (NIR) technology for non-invasive measurement of cardiopulmonary signals. Their research demonstrated the

feasibility of measuring peripheral blood volume changes without direct contact, utilising an image sensor. However, the sensitivity of the proposed system to motion artefacts and environmental conditions limited its operational effectiveness in dynamic settings. Later, Wieringa *et al.* [67] discussed the possibility of measuring back-scattered light remotely using a video camera. This 2005 paper demonstrated that signals obtained from a video camera at different wavelengths could be used to derive arterial oxygen saturation and HR. While their findings affirmed the validity of non-contact vital sign estimation, the study faced several limitations, including motion artefacts, ambient light interference, and tissue glare. The study also involved only seven healthy Caucasian human volunteers, which limits the generalisability of the findings to broader populations [67].

Verkruysse *et al.* [68] demonstrated that PPG signals could be remotely acquired from the human face with a simple, consumer-level video camera as the detector positioned up to 1 m away from the subject. The authors presented evidence that the reflectance signals were pulsatile cardiac signals by showing that signals corresponding to movement of facial areas with no exposed skin (edge of the face and hair above the ear) were not predominantly at the heart rate frequency. The green channel provided the strongest plethysmographic signal amplitude, corresponding to an absorption peak by oxyhaemoglobin, but the red and blue channels were also shown to contain plethysmographic information. Heart rate was extracted using the frequency content of acquired images using the Fast Fourier Transform (FFT) over 10-second windows. Although the study indicated the feasibility of performing vital sign monitoring from a distance of one meter, the authors also faced challenges related to motion artefacts and the influence of non-physiological signals such as ambient lighting changes, which could affect heart rate estimation. Tan *et al.* [238] presented a vision-based respiration monitoring system. The method involved image and signal processing techniques to extract chest and abdominal movement information from a sequence of video images recorded using a video camera. While the system could produce a respiration signal, it was sensitive to subject movement, which distorted the signal, and background motion could mask respiratory motion. Similarly, Bai *et al.* [239] designed an embedded monitoring system for breathing detection using a webcam. Their design employed a temporal differencing algorithm to detect moving objects by subtracting subsequent frames. The designed system was used to identify chest movement and determine the RR. Both methods demonstrated proof-of-concept for motion-based RR estimation, however, their performance degrades significantly under subject motion, variable illumination, or occlusions, highlighting the limitations of early motion-based methods in realistic clinical environments.

Non-contact vital sign estimation has progressed from early explorations to more robust and efficient algorithms, driven in part by the work of the Biomedical Signal Processing & Machine Learning (BSP-ML) group at Oxford. The group has made significant advances in extracting reliable physiological signals

from non-contact modalities, including video cameras, through the development of sophisticated signal processing techniques and machine learning models. For example, Tarassenko *et al.* [235] employed ambient light and auto-regressive models to enhance the signal extraction process from video recordings, achieving reliable estimates of HR and RR from patients undergoing treatment in controlled environments, such as the Renal Unit for dialysis treatment. Villarroel *et al.* [88] combined video camera technology with machine learning algorithms to enable continuous non-contact physiological monitoring of pre-term infants in the NICU. Their approach demonstrated the capability of PPGi systems to address unique challenges posed in neonatal health monitoring, while highlighting the ongoing difficulties in dealing with movement, skin pigmentation variability, and low-light conditions. More recently, Jorge *et al.* [240] expanded the application of video-based monitoring to postoperative patients in intensive care. These advancements reflect the critical transition from controlled experimental setups to practical, extended use in dynamic clinical environments. Developments in deep learning for skin detection and vital-sign extraction have also enhanced reliability and signal quality, demonstrating the potential for adaptable monitoring systems in diverse populations [241, 242].

2.7 Detection of COBE from conventional contact-based signals

The detection of apnoea has been a significant area of focus in research due to its impact on patient health and outcomes. Various methodologies for detecting apnoea in both adults and pre-term infants have been extensively studied and documented in the literature. This section presents a comprehensive survey of these methods, categorising the approaches based on the population studied. The section specifically focuses on COBE detection methods that utilise physiological signals routinely monitored in the hospital, such as ECG, IP, and PPG data.

2.7.1 Respiratory signal extraction and pre-processing

Various pre-processing approaches have been proposed to extract respiratory modulations from clinical signals. Researchers have used filtering techniques to isolate the relevant frequency ranges associated with breathing, along with advanced signal processing methods such as wavelet transforms to extract respiratory components from the overall signal [243, 244, 245]. Additionally, feature-based methods like amplitude analysis, baseline wander, and frequency modulation have been utilised in these efforts [164, 246, 247]. Figure 2.13 shows the extraction of respiratory signals from ECG and PPG signals using feature-based techniques.

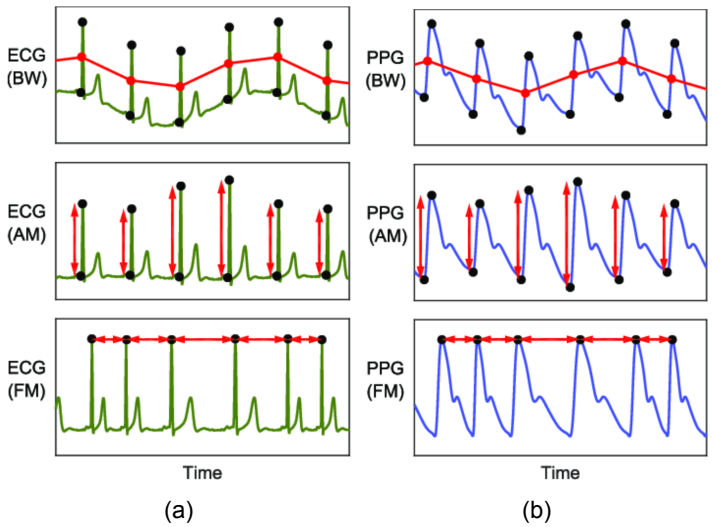


Figure 2.13: Example processes for extracting respiratory signals with measurement of the baseline wander (BW), amplitude modulation (AM), and frequency modulation (FM). Measurements are extracted for each beat from fiducial points (shown as dots). a) respiratory signals from ECG. b) respiratory signals from PPG. Figures adapted from Charlton *et al.* [244].

Respiratory signal extraction from ECG

Respiratory-induced variations in ECG signals typically manifest as changes in signal amplitude and morphology, which reflect breathing patterns [53], as illustrated in figure 2.13a. For example, the QRS complex can be analysed for respiratory variations by examining changes in the amplitude and intervals (e.g., R-R intervals) between successive QRS complexes. These variations include ECG-derived respiration (EDR) and respiratory sinus arrhythmia (RSA), which manifest as beat-to-beat changes and amplitude fluctuations of the QRS complex. Both EDR and RSA signals are derived from fiducial points (e.g., QRS peaks) in the ECG and provide indirect measures of respiration, which can be used for COBE detection [164, 177].

The EDR signal is particularly useful for COBE detection because it captures direct morphological changes in the ECG waveform caused by respiratory chest motion. Unlike RSA or R-R interval-based surrogates, EDR provides a mechanical correlate of respiratory effort, rather than relying on indirect heart-rate modulations [39, 81, 248]. During COBE, the absence of chest motion may lead to a flattening of the EDR waveform, providing a more interpretable indicator of respiratory cessation. In contrast, RSA or R–R interval-based signals may still exhibit periodicity due to cardiac modulation even when breathing stops, reducing their specificity for COBE event detection [249, 250].

Respiratory signal extraction from IP

IP signals inherently exhibit amplitude modulation characteristics corresponding to respiratory effort [251]. Frequency modulation of this signal through Respiratory Sinus Arrhythmia (RSA) can capture the RR [252,

253], while the baseline wander reflects both the respiratory cycle and factors such as body position or electrode placement. Pre-processing typically involves bandpass filtering to remove cardiac and motion artefacts and baseline correction to enhance respiratory peaks.

Respiratory signal extraction from PPG

PPG signals contain respiratory information due to the modulation of peripheral blood volume by chest movements. As shown in figure 2.13b, the signal is modulated in its amplitude, baseline, and frequency. These modulations are caused by respiratory movements affecting the hemodynamics of peripheral arterioles and capillaries, and can be analysed to detect COBE [244, 254]. Several studies have proposed filtering and demodulation techniques to extract this respiratory information embedded in the PPG signal [244, 255, 256, 257, 258]. Bandpass filtering is typically followed by envelope extraction or other demodulation techniques to isolate the respiratory component [255, 258].

2.7.2 COBE detection in adults

Apnoea in adults typically presents as a chronic condition, assessed using the Apnoea-Hypopnea Index (AHI). The AHI is defined as the average number of apnoea and hypopnoea events per hour during sleep [259, 260]. While apnoeas involve complete cessation of airflow, hypopneas are partial reductions in airflow that can result in oxygen desaturation [261, 262]. Because adult COBE events often occur over extended periods, detection models usually estimate AHI severity rather than identifying acute events. As a result, commonly used datasets, such as the PhysioNet Apnea-ECG dataset, use long observation windows which are suitable for adults but may fail to capture short, acute COBE events [263]. This contrasts with neonates, where COBE episodes can occur during wakefulness and represent acute events that must be detected rapidly to enable timely clinical intervention.

The detection of adult COBE events primarily relies on conventional physiological signals such as ECG, IP, and PPG, from which features are extracted to train machine learning models. Early event detection efforts involved the use of machine learning algorithms such as support vector machines (SVMs), decision trees, and random forests applied to handcrafted features [258, 264, 265, 266, 267]. Although these approaches demonstrated proof of concept and are interpretable, they were often constrained by short recording duration, as well as poor generalisability to diverse populations and noisy clinical environments. The handcrafted features used in these models often fail to capture subtle or complex modulations in respiratory patterns in real-world data, which reduces model reliability [268].

In contrast, more recent studies have adopted deep learning architectures, including CNNs, ResNets, Long Short-Term Memory models, and a combination of these, which automatically extract relevant

hierarchical features from raw signals [38, 269, 270, 271]. Though these models improve robustness against noise and inter-subject variability, and report high accuracy, their effectiveness mostly depends on the availability of large, well-annotated datasets to train them. Furthermore, interpretability remains limited, and models trained on adult datasets may not generalise to the neonatal populations because of differences in respiratory physiology, signal amplitude, noise characteristics and differences in diagnostic requirements. To improve model reliability, researchers have also explored data fusion by integrating multiple physiological signals at different stages of model training [272, 273, 274, 275]. Data fusion can improve model robustness by compensating for missing or noisy data [273]. However, misalignment between input signals and variable data quality can introduce artefacts.

Overall, improved signal processing, machine learning, and multimodal methods have advanced COBE episode detection. These developments provide a technical basis for developing neonatal monitoring systems. However, adaptation of the developed systems is essential to account for the short-duration, intermittent nature of COBE events in pre-term infants, as well as their distinctive physiological characteristics.

2.7.3 COBE detection in pre-term infants

The majority of research on apnoea detection has focused on adult populations. The primary dataset for apnoea detection was historically the Physionet Apnea-ECG dataset, created in 2000. While other datasets and detection methods have emerged over time, they have continued to rely on signals acquired from adult study participants. However, there is a growing interest in adapting detection methods to the unique physiological characteristics of pre-term infants.

Similar to adults, contact-based monitoring modalities, including ECG, IP, and pulse oximetry, form the basis for COBE detection systems for pre-term infants. Early detection demonstrated proof-of-concept using classification methods like equal prior quadratic classifiers, Bayesian adaptation of Gaussian mixture models and cumulative sum (CUSUM) with Shannon entropy [276, 277]. These methods highlighted the potential of cardiorespiratory and movement signals as predictors and reported good performance. However, their extremely small datasets, limited to only six patients, mean their reported performance is unlikely to generalise to broader NICU populations. Furthermore, these studies focused on predicting COBE events at birth but did not address the persistence of apnoea episodes in neonates already diagnosed with the condition.

Recently, machine learning algorithms have been incorporated into COBE detection to improve diagnostic accuracy. Various researchers have used SVMs, auto-encoders [278, 279, 280] and a combination

of machine learning techniques [281, 282]. Some studies have used larger datasets, comprising up to 367 infants, which allowed the use of more sophisticated detection algorithms. For example, Shirwaikar *et al.* [281] used autoencoder-based methods and achieved an area under the receiver operating characteristic curve (AUC) of around 0.83 [281]. However, the details about the number of episodes recorded per infant and the overall duration of recordings were not consistently reported, thereby limiting the interpretability of the results. More recently, Adjei *et al.* [163] introduced an adaptive threshold algorithm to measure inter-breath intervals. This method adjusts the threshold used to define a breath, based on the unique breathing patterns of each infant. By integrating machine learning with this thresholding method, their framework improved the ability to distinguish true COBE events from normal variability in respiratory signals. They reported an accuracy of 75% when using IP features, which increased to 87% when incorporating SpO_2 and cardiac features. While the study validated its findings by comparing the algorithm outputs with visually identified breaths, the reliance on visual confirmation can be subjective and may introduce bias. An important consideration when comparing performance across different models is that there is no standard approach for labelling COBE for automated detection. Different studies use different physiological window lengths and overlaps (e.g., 60 s non-overlapping windows in the PhysioNet dataset, or 30 s windows with 5 s overlap by Varisco *et al.* [280]). Studies with longer non-overlapping windows naturally tend to report higher accuracies, but such models have obvious trade-offs in terms of temporal sensitivity to COBE events.

In addition to algorithmic advancements, researchers have explored alternative sensing modalities to improve the detection of COBE events that may not be reliably identified using conventional physiological signals [37, 283]. Signals derived from electrical impedance tomography (EIT) and respiratory inductance plethysmography have been investigated as complementary sources of physiological information for COBE detection. These modalities have been reported to improve the detection of subtle or short-duration COBE events that may be difficult to identify using traditional impedance monitoring systems [282, 283, 284]. However, the integration of these techniques into routine clinical practice remains challenging because they often require specialised equipment and expertise for signal interpretation, limiting widespread adoption.

Although recent advances have improved neonatal COBE episode detection, accuracy still varies because of the dynamic physiology of pre-term infants [285]. Studies using traditional clinical signals in pre-term infants have relied on relatively simple models, such as SVMs and autoencoders, likely limited by small dataset sizes. Furthermore, the lack of consistent definitions of what constitutes a COBE event makes it difficult to establish standardised monitoring protocols [286].

2.8 Detection of COBE from video camera signals

In recent years, there has been growing interest in non-contact approaches for detecting COBE episodes in both adult and neonatal populations. Unlike contact-based methods that rely on ECG, IP, or pulse oximetry, these approaches analyse respiratory motion, optical physiological changes, or multimodal video features to identify respiratory pauses associated with COBE episodes.

2.8.1 Non-contact COBE detection in adults

Video-based approaches for adult COBE detection typically analyse respiratory motion patterns, thoracic displacement, or optical physiological signals extracted from image sequences recorded during sleep. Early systems such as “SleepWise” used a conventional digital video camera with infrared LEDs to automatically assess breathing patterns from body movements and position changes during sleep [287, 288]. While developed in an experimental, controlled environment with control subjects, “SleepWise” was later validated against home sleep apnoea testing in a cohort of 38 sleep clinic outpatients. The performance of the system can still be affected by lighting conditions, camera placement, and subject movement. Similarly, smartphone-based approaches have been investigated for respiratory event detection, using either repurposed smartphones as sonar systems that emit and analyse frequency-modulated sound [289], or using built-in cameras to track chest motion through optical flow or motion features [290, 291]. Studies on the use of smartphones have been limited to controlled breathing experiments with small datasets, often simulating apnoea rather than capturing true COBE episodes.

To improve event detection capabilities using video camera signals, various machine learning algorithms have been developed [78, 291, 292]. Akbarian *et al.* [292] used CNNs to analyse infrared video data. The researchers tracked thoracic and chest movements from 42 participants and distinguished between obstructive and central apnoeas. Multi-modal machine learning approaches have also been used to capture depth video and audio of patients for the detection of respiratory events that characterise apnoea [78]. Although the proposed classifiers could detect respiratory events based on the reference data, the system required large storage for video and audio recordings, and could be computationally intensive. The trained classifiers were also person-specific, limiting generalisability.

2.8.2 Non-contact COBE detection in pre-term infants

Although several video-based approaches for adult COBE detection have demonstrated feasibility, translating these methods to pre-term infants presents additional challenges because of irregular respiratory motion, spontaneous body movement, and the NICU environment.

Studies have implemented advanced video analysis techniques to monitor pre-term infants during sleep. Cattani *et al.* [77] applied RGB video recordings to detect apnoea events in a single pre-term infant. Their method involved converting the recordings to grayscale, then emphasising movements through frame differencing and binarisation. Subsequently, they amplified subtle respiratory movements using Eulerian Video Magnification, then analysed periodicity with Fourier-based techniques. Although the study is technically innovative, its focus on only one subject with recurring episodes limits the broad applicability of the results. Moreover, techniques like frame differencing and frequency analysis can be affected by non-respiratory movements, changes in lighting, and camera positioning, which could result in incorrect detection of COBE episodes.

Lorato [293] proposed a multi-modal approach by combining RGB and NIR imaging to explore the feasibility of camera-based approaches for detecting cessation of breathing in pre-term infants. Their work focused on proposing solutions that could improve detection and classification of cessation of breathing, the automatic identification of respiratory pixels, and the analysis of non-respiratory motion. However, the study was limited to five infants and single-observer annotations, introducing potential bias. Common NICU factors, such as blankets, variable lighting, and routine clinical interventions, were not systematically evaluated, leaving questions about the real-world applicability of the proposed methods.

These studies illustrate the feasibility of video-based COBE monitoring in pre-term infants but also highlight challenges of small sample sizes, susceptibility to motion artefacts, and limited robustness to environmental variability. Unlike adult monitoring applications, monitoring pre-term infants requires algorithms that are highly sensitive to subtle, irregular respiratory patterns while remaining robust to confounding motion and environmental variability. Because of small datasets, machine learning algorithms have not yet been investigated for video-based detection of COBE episodes.

2.9 Deep learning approaches for COBE event detection

Building on the classical machine learning approaches discussed for both adult and pre-term populations, recent studies have increasingly adopted deep learning architectures to improve the detection of COBE events from contact-based signals such as ECG, IP, and PPG. Unlike classical machine learning, deep learning models automatically learn hierarchical features from raw or minimally processed signals, reducing the reliance on handcrafted features and potentially improving robustness to noise and inter-subject variability. CNNs have been the most commonly applied deep learning architecture for COBE event detection, particularly in adults, though similar architectures are being explored for neonatal populations where short, intermittent events present unique challenges [38, 269, 294, 295, 296]. By applying convolutional

filters across input waveforms, CNNs can capture local temporal patterns associated with respiratory modulations during COBE events. Their hierarchical structure allows the combination of low-level features into higher-level representations, supporting the detection of subtle or short-duration events that may be missed by traditional methods. A detailed description of CNN building blocks is provided in Appendix C.

To address the limitations of CNNs, subsequent architectures such as ResNets and, more recently, ConvNeXt, have extended the foundations of CNNs. Traditional CNNs faced significant challenges as networks grew deeper due to vanishing or exploding gradients [297, 298]. To overcome this challenge, ResNets introduced residual blocks with skip connection, which allows the network to learn residuals or differences between the input and the desired output, rather than learning the output directly, as illustrated by figure 2.14. This design enables the training of deeper networks and improves the capture of complex temporal patterns, which is particularly relevant for short, intermittent COBE events.

2.9.1 The architecture of a residual block

The fundamental innovation in ResNets is the residual block, which consists of convolutional layers with shortcut (skip) connections. In this residual block, the input x goes through convolutional layers, followed by rectified linear unit (ReLU) activation. The resulting output is then combined with the original input, forming the residual function. This approach enhances the network's learning efficiency by simplifying the learning process. This design allows the smooth flow of gradients through the network, addressing the vanishing gradient problem and enabling the training of deep networks.

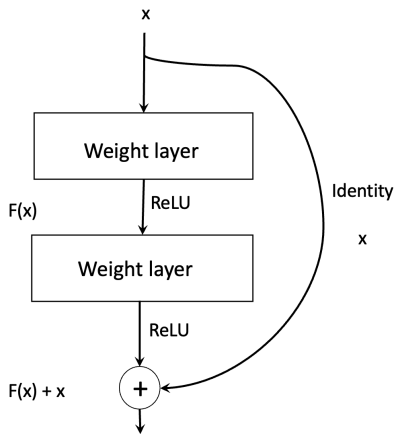


Figure 2.14: The structure of a residual block, showing the skip connection that bypasses the convolutional layers. This design helps maintain gradient flow and improves the training of deep networks.

Liu *et al.* [299] introduced ConvNeXt, an updated version of ResNets that combines elements of the ResNet design with techniques from transformer models. While CNNs and ResNets have been explored for contact-based COBE event detection, more recent architectures such as ConvNeXt have not yet been applied in the context of pre-term infants, but are considered in this thesis.

2.9.2 The architecture of ConvNeXt

ConvNeXt introduces several key architectural improvements over traditional ResNets, modernising the convolutional backbone while retaining its core principles. A significant change is the use of grouped convolution in place of standard convolutions. Grouped convolution serves a similar role to multi-headed attention in transformer models. Each group, analogous to a head, learns relationships within a subset of features, enabling broader feature diversity at a similar computational cost. The model can, therefore, learn richer internal representations by clustering related features during training. Inside each residual block, ConvNeXt combines large-kernel depthwise convolutions and grouped convolutions. The standard ReLU activation functions are replaced with GELU, which provides a smooth, continuous activation around zero, and the traditional bottleneck structure is replaced by an inverted bottleneck structure (depthwise conv → feature expansion → contraction), closely resembling the MLP block in transformers.

ConvNeXt also incorporates several other transformer-like changes beyond the block structure. These include changing the state compute ratio from [3, 4, 6, 3] to [3, 3, 9, 3], aligning with hierarchical designs in sliding window Transformer models, and replacing Batch Normalisation with Layer Normalisation, which improves training stability in deeper networks. The architecture deviates from the standard ResNet stem (a 7×7 convolution with a stride of 2 followed by max pooling). Instead, it uses a patchify stem (kernel size 4, stride 4), mimicking the non-overlapping patch input of vision transformers. These changes enable ConvNeXt to model long-range dependencies more effectively, making it a more expressive and adaptable architecture, especially in settings where attention mechanisms are unsuitable or computationally prohibitive.

In figure 2.15, the multi-head self-attention (MSA) box in the Transformer block corresponds to the d7×7 grouped conv in ConvNeXt. Both are the first operations inside their respective blocks, and layer norms and an MLP-style expansion-contraction setup follow both. Both MSA and depthwise grouped convolution aim to model complex, potentially long-range relationships between input features. MSA achieves this through attention heads, while ConvNeXt achieves it through grouped channels. Table 2.1 provides a summary comparison of the architectural features across the depicted ResNet, ConvNeXt and Transformer blocks.

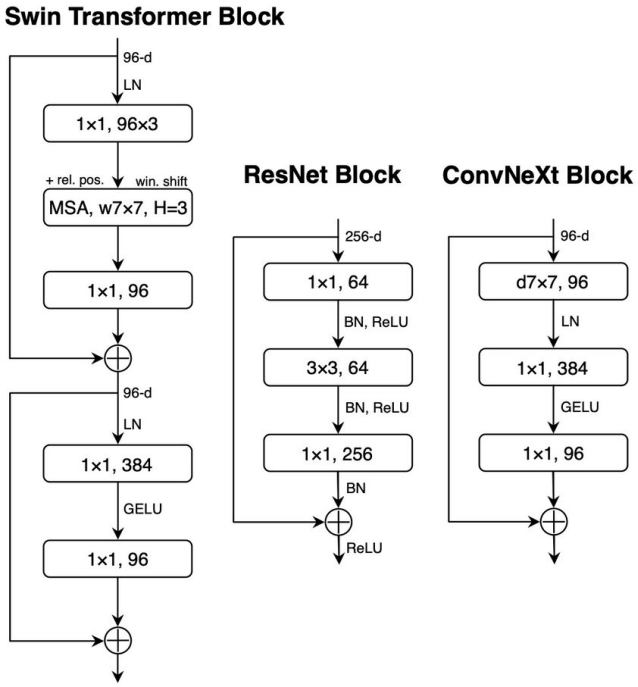


Figure 2.15: Comparison of block architectures for a Swin Transformer, ResNet, and ConvNeXt. While the Swin Transformer includes multi-head self-attention (MSA), ConvNeXt removes these self-attention mechanisms but structurally mimics the Transformer block using grouped convolution, GELU activation, and an inverted bottleneck. ResNet, in contrast, uses standard convolution and ReLU activations.

Table 2.1: Comparison of core architectural components across ResNet, ConvNeXt, and Transformer blocks.

Feature	ResNet	ConvNeXt	Transformer
Convolution	Standard	Grouped	-
Attention	-	-	Multi-head Self-Attention (MSA)
Activation	ReLU (multiple layers)	GELU (single layer)	GELU (single layer)
Norm type	BatchNorm	LayerNorm	LayerNorm
Bottleneck Type	Standard	Inverted	Inverted
Residuals	Single	Single	Double
Feature Expansion	64 → 256	96 → 384 → 96	96 → 384 → 96

Beyond structural changes, ConvNeXt integrates modern training strategies, many of which are inspired by transformer training pipelines to complement its increased parameter count and architectural complexity. These include the AdamW optimiser, Exponential Moving Average (EMA) of model weights, label smoothing, weight decay, and a learning rate schedule with linear warmup followed by cosine decay. These enhancements contribute to better regularisation and more stable convergence, especially when training on smaller or noisier datasets.

ConvNeXt has not yet been applied to COBE episode detection in pre-term infants. Its ability to capture long-range temporal dependencies, combined with adaptable training strategies, makes it well-suited for

modelling the short, intermittent respiratory events characteristic of this population. These architectural and training advantages motivate its investigation in this thesis and inform the methodological approach to applying deep learning to contact-based COBE event detection. Despite these advantages, however, ConvNeXt presents limitations that must be considered in the context of neonatal monitoring. The architecture is relatively large and computationally intensive, which may constrain deployment in real-time or in resource-limited clinical environments without additional optimisation. Like other deep learning models, its performance is dependent on the availability of sufficiently large and diverse training datasets, a resource that is currently lacking for pre-term infant populations.

2.9.3 Challenges of small physiological datasets

The development of machine learning models for neonatal physiological monitoring is often constrained by the availability of sufficiently large and annotated datasets. In the context of COBE detection, acquiring large datasets from pre-term infants presents several practical and ethical challenges, including limited access to NICU environments, patient privacy considerations that limit data sharing across institutions, and the time-intensive nature of expert annotation [300, 301, 302]. As a result, many studies investigating automated detection of respiratory events in pre-term infants rely on comparatively small datasets, which can limit model generalisability and increase the risk of overfitting [303]. Neonatal cohorts often exhibit substantial physiological and pathological variability across individuals, with cardiorespiratory signals varying considerably because of differences in gestational age, pathology, movement, and clinical condition. Consequently, models trained on small datasets may learn subject-specific or recording-specific characteristics rather than clinically meaningful physiological patterns, limiting their ability to generalise to unseen patients.

To mitigate these challenges, several validation strategies have been adopted in the literature. Cross-validation approaches, including k-fold cross-validation and leave-one-subject-out validation, are commonly used to maximise the use of limited data while providing estimates of model generalisation performance [87, 279, 304, 305]. In k-fold cross-validation, the dataset is partitioned into k number of folds, whereby the model is trained on $k-1$ folds and evaluated on the remaining fold. This process is repeated k times so that each fold serves as the validation set once. Leave-one-subject-out validation instead reserves data from a single subject for validation, while the remaining subjects are used for training [306, 307]. In physiological monitoring studies, these validation approaches are often implemented using subject-wise splitting strategies, where data from the same patient is not shared between training and validation sets. This is particularly important because it reduces data leakage, where information from either the validation or held-out test set unintentionally influences the training process, leading to

unrealistically optimistic estimates of model performance during evaluation [308]. Subject-wise splitting strategies provide more realistic assessments of model robustness across different patients [87, 279, 304].

Data augmentation techniques have also been widely investigated to improve the performance of machine learning models trained on small physiological datasets. In physiological time-series analysis, augmentation methods such as additive noise, amplitude scaling, signal masking, and time warping have been used to increase the diversity of training data and improve robustness to noise and physiological variability [309, 310]. These augmentation approaches can help address class imbalance, reduce overfitting, and improve model performance on unseen physiological data. In addition to augmentation, regularisation approaches such as dropout, early stopping, and transfer learning from larger datasets have also been explored to improve deep learning performance in small-data settings. More recently, generative approaches such as generative adversarial networks and diffusion-based models have been investigated for synthesising physiological time-series data when real data is limited [311, 312, 313]. Although these approaches may improve the availability of training data, ensuring that synthetic physiological signals remain clinically realistic and physiologically plausible remains an important challenge [313]. Nevertheless, despite these methodological advances, the availability of sufficiently large, diverse, and clinically representative neonatal datasets remains a major limitation in the development of reliable automated COBE detection systems.

2.10 Discussion

ECG, IP and pulse oximetry are contact methods currently being used in medical settings to monitor the cardiorespiratory status of patients. In pre-term infants, a pulse oximeter, usually placed around the foot, is used to provide continuous information on arterial oxygen saturation and pulse rate. ECG uses chest electrodes to record cardiac electrical activity and heart rate, while IP uses the same chest electrodes to measure the electrical impedance changes caused by breathing. When used together in the NICU, these methods provide complementary information on oxygenation, cardiac activity, and respiration of infants, aiding in the diagnosis and management of conditions such as COBE. However, the adhesive electrodes and sensors required can cause discomfort and skin irritation.

Beyond their routine use in clinical care, conventional signals such as ECG and IP have formed the reference standard against which newer non-contact methods are benchmarked. These signals have also driven much of the research on automated COBE detection. The availability of large-scale adult datasets, particularly the Physionet Apnea-ECG database, has facilitated the development of many algorithms for COBE episode detection. However, comparable publicly available datasets do not exist for neonates.

This gap has motivated researchers to explore a wider range of physiological signals, including thoracic movements and electrical impedance tomography, when detecting COBE episodes in pre-term infants. The absence of neonate-specific datasets, combined with the discomfort caused by electrodes and belts, reinforces the need for alternative monitoring strategies.

Non-contact methods for detecting COBE episodes, particularly video camera-based monitoring, have recently been introduced as unobtrusive solutions for continuous monitoring of respiratory patterns. Both RGB and depth cameras can capture subtle chest and abdominal movements, which can then be analysed to estimate respiratory patterns. Several studies have demonstrated the feasibility of this approach in controlled laboratory settings and, more recently, in clinical and home environments.

This transition from contact to non-contact monitoring does present significant challenges. Most of the current studies using non-contact signals involve healthy adult volunteers in controlled settings, whereas real hospital studies involve smaller sample sizes and require longer recording lengths. The detection of COBE episodes using the non-contact technologies is also a challenge because of factors like motion artefacts, changes in lighting, body movements and positioning. Variations in skin tone and anatomical differences can also affect the accuracy of extracted camera signals, leading to potential errors in COBE event detection. Developing reliable algorithms that perform well in a real-world hospital environment, therefore, remains a challenging task.

Recent advances in deep learning have improved the potential for both contact- and video-based COBE event detection. Deep learning models, such as CNNs and ResNets, can automatically extract hierarchical features from physiological signals, improving robustness to noise and inter-subject variability. More modern architectures, including ConvNeXt, offer additional advantages by capturing long-range temporal dependencies and enabling more stable training, which may be particularly relevant for detecting short, intermittent COBE events in pre-term infants. At the same time, these models introduce challenges. ConvNeXt is computationally demanding, relies on large and diverse datasets, and raises interpretability concerns that are especially important in clinical contexts. Incorporating modern deep learning approaches into monitoring algorithms could therefore improve detection accuracy and reduce reliance on handcrafted features, potentially enabling more reliable continuous respiratory monitoring in the NICU. Their clinical deployment, however, will require careful optimisation, validation, and attention to explainability.

Detecting COBE episodes in pre-term infants presents unique challenges, which may contribute to the lack of research in this field. Infants display irregular, spontaneous movements leading to potential inaccuracies in vital sign monitoring. The NICU environment adds further complications, with its low light

conditions, the presence of various medical equipment and wires, and kangaroo care. Nevertheless, the reviewed literature suggests that using video cameras for COBE episode detection can provide a non-invasive and effective monitoring approach. Further research focused on adapting algorithms to the unique physiology of pre-term infants and the constraints of the NICU environment will improve method accuracy, providing better respiratory monitoring and COBE event detection solutions across various age groups. Therefore, thorough validation against established reference standards and well-designed clinical studies are essential before the widespread clinical adoption of non-contact COBE detection technologies. Without robust validation and clinical evidence, integrating these methods into standard clinical practice will remain difficult.

Chapter 3

Neonatal High Dependency Unit study

3.1 Introduction

This chapter describes the Neonatal High Dependency Unit (HDU) study, which was designed to examine the efficacy of using video cameras to monitor the vital signs of pre-term infants. The clinical study was conducted in the high-dependency unit of the NICU at the John Radcliffe Hospital in Oxford, as part of a research programme in the Oxford University Hospitals, National Health Service (NHS) Foundation Trust and the Oxford Biomedical Research Centre (BRC). The research complied with the regulations of all relevant institutional and government guidelines. Ethical approval was granted by the South Central–Oxford Research Ethics Committee (reference 13/SC/0597).

3.2 Clinical study design and protocol

For this study, pre-term infants were nursed in a designated study incubator in the high-dependency area of the NICU at the John Radcliffe Hospital in Oxford. The incubator was similar to the standard incubators used in the hospital, with modifications made to accommodate a camera (see figure 3.1). The participants were continuously double-monitored with a digital video camera and standard patient monitoring devices. The video camera was securely mounted on a camera stand that was clamped to a custom-built trolley and then positioned directly over the incubator. Both the camera and camera stand were CE-marked and compliant with European health, safety and environmental protection standards. Detailed information about the study was provided to the infant's parents prior to the start of the video recording. Throughout the study session, standard patient monitoring and ongoing care were maintained. The research equipment (video recording and storage) was set up to minimise inconvenience to clinical staff. There were no supplementary sensors attached to the infants, and the positioning of the video camera and associated equipment did not restrict access to the incubator.



Figure 3.1: A video camera positioned over a specifically drilled hole in the top surface of the study incubator.

The clinical study was conducted during the daytime under normal ambient light conditions without interfering with routine patient care. As part of the standard care protocol, a six-hour designated quiet period was incorporated each day to provide the infant with an uninterrupted resting period, free from any interference by clinical staff or parents, except in the case of emergencies. During some clinical procedures, such as phototherapy and cannula insertion, or when the infants were taken out of the incubator for kangaroo care with their parents, video recording could temporarily be paused by clinical staff, or the camera could be covered as deemed necessary. Video recording was terminated if the infant needed to be transferred to another unit.

This Neonatal HDU study on non-contact monitoring in the NICU differs from other similar studies [91, 314, 315] in the following aspects: 1) it was conducted on a larger cohort of 30 pre-term infants, (2) the infants were continuously monitored for several hours a day, in parallel with regular patient care, providing a more realistic and accurate representation of the clinical environment, and (3) unlike previous studies, no restrictions were placed on gestational age, infant weight, ethnic group, or skin tone ensuring that our findings are applicable to a diverse range of patients.

3.2.1 Study participants

This study double-monitored 30 pre-term infants using both the Philips® IntelliVue MX800 patient monitor, the standard of care in the NICU, and a digital video camera during both day and night times for up to seven days. Each infant was nursed in a designated incubator in the high-dependency area of the NICU. The medical team recruited each of the infants based on the British Association of Perinatal Medicine's Categories of Care 2011 [316]. The study protocol permitted monitoring for up to seven days to capture longitudinal physiological changes, although in practice most infants were monitored for approximately

three to four consecutive days depending on their clinical stability and duration of stay in the neonatal HDU. If an infant was transferred from the neonatal unit, recording was terminated and only the data acquired up to that point was retained for analysis.

Inclusion criteria

To be eligible for the study, participants needed to satisfy all of the following criteria:

- born at less than 37 weeks of gestation.
- requires high-dependency care.
- requires continuous monitoring of heart rate, respiratory rate and oxygen saturation.
- requires to be nursed naked.

Exclusion criteria

The study excluded infants who presented with life-threatening conditions that prevented continuous monitoring in the high-dependency area of the NICU.

3.2.2 Informed consent

Parents with infants who met the inclusion criteria outlined in section 3.2.1 were given comprehensive verbal and written details of the study by the NICU clinicians. They were also given the opportunity to meet with the research team and address any concerns they may have. The parents were informed that the study would use anonymised video camera images and vital sign reference data solely for research purposes. They were also assured that their participation was voluntary and that they could withdraw their baby from the study at any time without explanation and without affecting patient care. Prior to the commencement of recording, parents were required to review and sign an informed consent form confirming their agreement to their child's participation.

3.2.3 Participant confidentiality

The study adhered to the UK's Data Protection Act. The data was promptly anonymised and kept in an encrypted cluster, accessible only to authorised members of the research team. Participant identification was limited to a unique number consisting of the letters PN followed by a three-digit code (e.g. PN001 for the first participant). Additionally, researchers were not granted access to any personally identifiable information to ensure the safety and privacy of the participants.

3.2.4 Data handling and record keeping

Patient demographics, physiological records, clinical events, and patient case notes were recorded anonymously in electronic case report forms (CRFs). OpenClinica, a clinical trial software package, was used to store the clinical research data and capture the CRFs. The following participant demographic information was recorded: ethnicity, gender, gestational age at birth (in weeks), adjusted gestational age (in weeks), birth weight (in grams) and current weight (in grams). Additionally, the clinical staff recorded information about the respiratory support the participants required (e.g. none, low-flow nasal cannula, high-flow therapy), any medications administered, the rate of enteral feeding (the delivery of nutrients directly into the gastrointestinal tract via a tube) in ml/kg/day, and the rate of continuous infusions in ml/kg/day, if applicable. All clinical procedures or interventions performed during the recording sessions were also documented in the CRFs. These procedures may include actions such as adjusting sensing probes, taking blood pressure measurements, collecting blood samples, or visits by parents. Episodes of clinical deterioration were also noted in the CRFs.

3.3 Instrumentation

The continuous recording of video images and vital sign data in a real hospital environment poses several challenges. For instance, the monitoring equipment used must be compact, easy to transport, and safe for use in hospitals. Additionally, the video recording software must be capable of reliably handling the high-throughput data produced by the camera. The Neonatal HDU study used specialised hardware and software developed by the research team to ensure accurate data acquisition. The custom acquisition system was designed to support continuous, high-resolution synchronised recording of physiological and RGB video data in the NICU environment while minimising disruption to routine clinical care. The system incorporated dedicated video and physiological data acquisition workstations, frame buffering mechanisms for high-throughput image capture, and custom acquisition software for synchronised recording and storage of uncompressed video and physiological data streams. These hardware and software components were specifically designed to support long-duration multimodal recording while preserving the temporal fidelity required for retrospective physiological signal analysis.

After conducting a comprehensive health and safety evaluation, the John Radcliffe Hospital's Department of Medical Physics and Clinical Engineering found the designed hardware suitable for use in the NICU. The developed software met the study's requirements, storing video images in their raw format without altering the data recorded by the video camera sensor. A data recording session was successful if

- there were no hardware or software malfunctions,

- the equipment did not interfere with clinical care and
- all the required data, including reference and camera data, were successfully recorded and stored.

3.3.1 Hardware for data acquisition

Throughout the study, the research team ensured that all research equipment, including video recording and data storage, was set up in a manner that did not disrupt the clinical staff. To accomplish this, the team created a data acquisition system, depicted in figure 3.1. This system consists of a trolley carrying a video camera and two recording workstations. One workstation was used to capture video data from the camera, and the other workstation was connected to the NICU's patient monitors to record the reference physiological signals.

Trolley

A technician team assembled a study-specific trolley in the mechanical workshop of the Institute of Biomedical Engineering at the University of Oxford. The trolley was designed to facilitate the research team's moving all the monitoring equipment quickly inside the hospital.

Incubator

To record a video of infants in the Giraffe OmniBed Carestation incubator (General Electric, Connecticut, USA), a small hole was drilled in the top panel of the canopy. This modification allowed for a video camera to be positioned inside the incubator's chamber without any reflection or attenuation from the perspex layer (see figure 3.1). After the modification of the study incubator, the health and safety team at the Department of Medical Physics and Clinical Engineering at the John Radcliffe Hospital conducted a series of tests on temperature and humidity for two weeks to ensure that the modified incubator posed no risk to patients. The team verified that the incubator maintained the same environmental control within the chamber as a standard incubator. This modification to the incubator was approved by the Medical Research Ethics Committee (MREC).

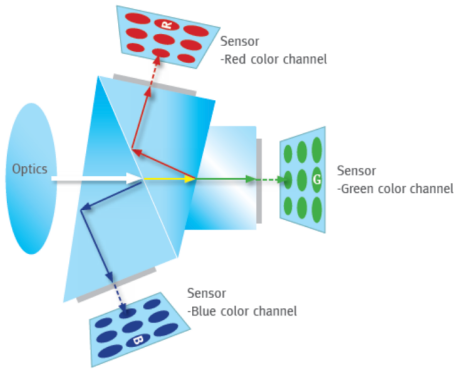
Workstations

During the data acquisition process, two workstations were used. Workstation A was responsible for video recording, while Workstation B was used to record medical data from the patient monitor. Workstation A was a custom-built Dell PowerEdge® server with a 24-core Intel® Xeon® processor (Intel, California, USA), 32 GB memory, 40 TB hard disk, and Nvidia® Quadro graphics card (Nvidia, California, USA). This workstation ran on the Fedora 20™ operating system. Workstation B was a Dell OptiPlex® equipped with

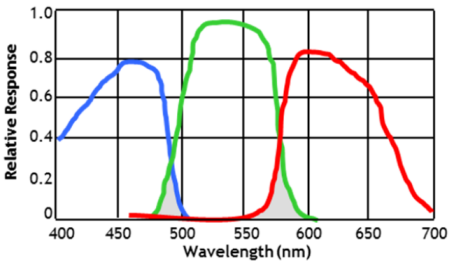
an Intel® Core™ i7 Processor, 8GB of memory, a 4TB hard disk, and the Microsoft Windows® 7 operating system. Both workstations utilised medical-grade Accumed® keyboards and mouse manufactured by Accuratus in the UK. These devices conform to the IP67 standard for dust and water protection, as well as the JIS Z 2801 test for the antimicrobial activity of plastics.

Video camera system

During the study, video footage was recorded using a JAI® AT-200CL progressive scan digital video camera that has three separate CCD sensors to measure red, green and blue light intensity separately. The camera was affixed to the trolley using a Manfrotto® 224RC camera arm (Manfrotto, Italy) and then positioned over the incubator where the infants were being nursed (refer to figure 3.1). This camera was set to acquire 24-bit true colour images, with 8 bits per pixel, at a resolution of 1628 x 1236 pixels and a frame rate of 20 frames per second. The camera incorporates three Sony® ICX274AL 31/1.8" image sensors (manufactured by Sony, Japan) to measure the light intensity of each colour channel (red, green, and blue). Figure 3.2a depicts a block diagram of a 3-CCD camera, and figure 3.2b illustrates the spectral responses of the camera sensors.



(a)



(b)

Figure 3.2: The JAI®AT-200CL 3-CCD camera used in the study. a) The prism is used for splitting the incoming light into three different colour channels: red, green and blue. b) Spectral response curve for the three imaging sensors. (Images courtesy of JAI Inc.)

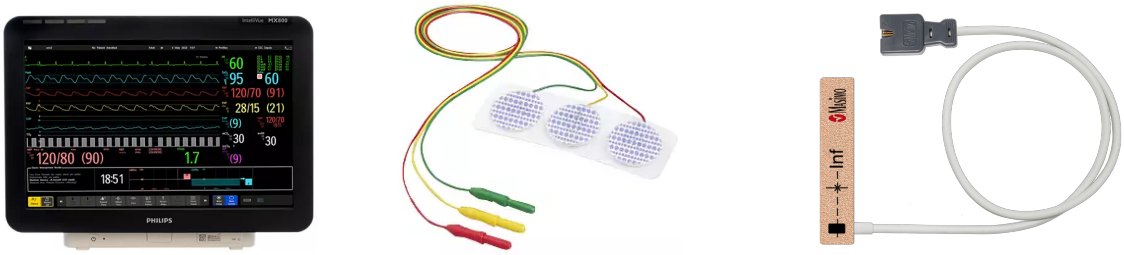
The prism within the 3-CCD camera optically separates incoming light into independent red, green, and blue channels before detection by the individual CCD sensors. This hardware-based separation enables simultaneous acquisition of full-resolution RGB information without the spatial interpolation required in conventional Bayer-filter cameras [317]. Preserving independent colour-channel information was important for this study because signals derived from PPGi are wavelength-dependent, and respiratory and cardiovascular modulations produce subtle changes in light absorption across different wavelengths [318]. Implementing channel separation in hardware therefore improves spectral fidelity and reduces interpolation artefacts compared with software-based colour reconstruction from a single image sensor [317]. This is

particularly important for camera-based estimation of physiological parameters such as oxygen saturation, where information from multiple colour channels can be used to capture differences in haemoglobin light absorption characteristics [319]

Throughout the recording session, a VS Technology SV-0614H lens kit (VS Technology, Japan) was used with the video camera. This lens kit allowed complete control over the focal length and lens aperture. The clinical staff present made necessary adjustments to these parameters as needed, both before and during the session, to ensure optimal focus on the infant's skin and appropriate brightness levels. A microEnable IV AD4-CLR frame-grabber card (Silicon Software, Germany) was used to record the video on Workstation A. To prevent image corruption or data loss during busy periods when the workstation was unable to request an additional video image from the camera, a video frame buffer was implemented using a Xilinx®Spartan Field Programmable Gate Array (FPGA) board (Xilinx, California, USA). Typically, a 1-hour recording resulted in 408 GB of data.

Patient monitor

Reference vital signs (HR, RR and SpO_2) were recorded using a Philips® IntelliVue MX800 patient monitor (Philips, Amsterdam, Netherlands), shown in figure 3.3a. The patient monitor was installed with a Philips®IntelliVue MMS measurement module and a Masimo® SET SpO_2 Vuelink IntelliVue measurement module (Masimo, California, USA). All participants in this study were equipped with a set of neonatal electrodes (in figure 3.3b) for recording ECG and IP signals. Additionally, a Masimo LNCS (Low Noise Cabled Sensors) Neo SpO_2 adhesive sensor was used to record oxygen saturation levels (see figure 3.3c).



(a) (b) (c)

Figure 3.3: Patient monitoring equipment used in the NICU study: a) Philips® IntelliVue MX800 patient monitor (Image courtesy of Philips); b) a set of neonatal electrodes (Image courtesy of Philips) and c) Masimo® LNCS Neo SpO2 adhesive sensor (Image courtesy of Masimo).

The patient monitor was connected to Workstation B via an RS232 serial interface. The ixTrend® software (Ixellence GmbH, Germany) was used to record the generated data streams. Table 3.1 summarises the physiological parameters recorded by the Philips monitor in the Neonatal HDU study. For HR, RR, and SpO_2, the reported value of 1 Hz corresponds to physiological estimates generated once per second

by the Philips monitor, rather than the sampling rates of the raw ECG, IP, and PPG signals from which these measurements were derived. Blood pressure measurements were manually obtained by neonatal nurses and recorded in the CRFs.

Table 3.1: List of physiological parameters recorded during a typical recording session

Name (units)	Description	Sampling rate(Hz)
ECG (mV)	1- lead electrocardiography signal	500
PPG (a.du)*	Photoplethysmography signal	125
IP (ohms)	Impedance pneumography signal	62.5
RR (breaths/min)	Respiratory rate derived from the IP by the Philips monitor	1
HR (beats/min)	Heart rate derived from the ECG by the Philips monitor	1
SpO_2 (%)	Oxygen saturation provided by Masimo SET pulse oximeter	1

* Arbitrary units (a.du.) represent relative changes in light absorption due to blood volume fluctuations.

3.3.2 Software for data acquisition

Apart from the ixTrend software for recording medical data using Workstation B, the BSP-ML team developed software to implement image acquisition and processing on an Xilinx Spartan Field-Programmable Gate Array (FPGA) board. The software ran on a workstation with the Fedora Linux Operating System. To ensure compliance with the MREC approval and to safeguard patient confidentiality, all images were stored in a custom binary format and saved in an encrypted storage cluster.

Video collector

The video recording software was developed to run on Workstation A. Its direct interface with the camera manufacturer's API (Application Programming Interface) allowed for full camera control during recording. Automatic gain control (AGC) was deactivated in all colour channels to ensure good video quality. The video data was saved in a 24-bit raw format in one-minute blocks (1200 video frames) without any compression or encoding.

Time synchronisation

Synchronisation was achieved using the network time protocol (NTP). A time server was set up on Workstation A. At the start of every recording, Workstation B automatically synchronised its clock with that of Workstation A. Since the recording of the video and reference data was carried out by two separate workstations, time synchronisation was necessary to ensure accurate timestamps across the two workstations. Although NTP synchronisation ensured initial alignment between the video and physiological recording workstations at the start of the recording, clock drift accumulated during long recordings because the systems operated independently after initial synchronisation. For recording sessions lasting over 5 hours, this drift could reach approximately 15 seconds. Consequently, additional

retrospective time-alignment procedures based on physiological signal matching were applied during later stages of analysis using the cross-correlation approach described in Section 4.8, which was subsequently used for synchronising camera-derived and physiological respiratory signals in Chapter 7.

3.4 Cohort description

This section presents demographic information on the research participants, as well as relevant recording statistics.

3.4.1 Participant demographics

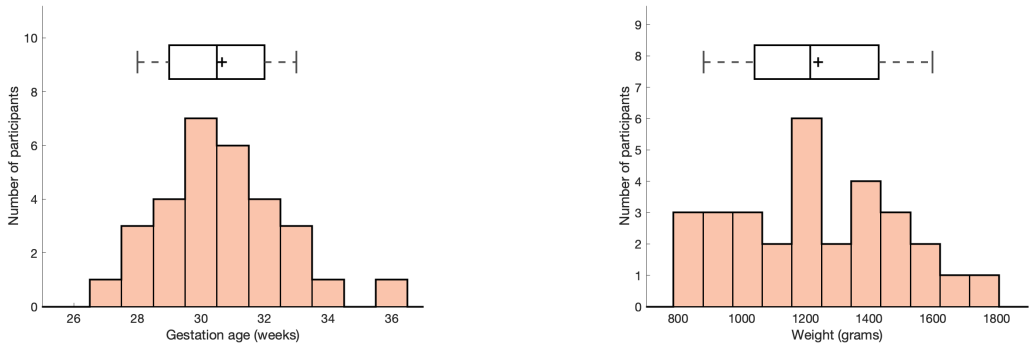
The Neonatal HDU study was conducted from February 2014 to May 2015, recruiting a total of 30 pre-term infants for monitoring. A total of 90 study sessions were conducted, resulting in over 400 hours of camera and reference data. Out of the 30 participants, 18 were male (60%). The study recruited infants from a diverse range of ethnic backgrounds, with 18 of the 30 infants being White British (60%). The mean gestational age was 31.1 weeks, with 66% of the recording done when the infants were very pre-term (28 - 32 weeks). Gestational age was overall categorised as extreme pre-term (<28 weeks), very pre-term (28 - 32 weeks), moderate pre-term (32 - 34 weeks) and late pre-term (34 - 37 weeks). The mean infant weight (taken on the first day of recording) was 1,240 grams. Table 3.2 provides a summary of participant demographics. For detailed information on each participant's recording sessions, gestational age, class, ethnicity, and gender, see Appendix A.1.

Table 3.2: Summary of population demographics (N = 30).

Characteristic	**Value**
Total participants	30
Total recording sessions	90
Gestational age (weeks)[1,2]	31.1 (±1.8)
Gestational age categories (by recording session)	
Extreme pre-term	3
Very pre-term	59
Moderate pre-term	23
Late pre-term	5
Weight (grams)[1,2]	1,240 (±252)
Gender (number of participants)	
Male	18
Female	12
Ethnicity (number of participants)	
White British	18
Asian or Asian British - Pakistani	2
Black British or Black African	1
Mixed - White and Asian	2
Mixed - White and Black Caribbean	2
Mixed - White British and Japanese	1
Mixed - Any Other Mixed	2

[1] Mean (±SD) [2] Measured on the first day of recording

Figure 3.4 shows the distribution of gestational ages and infant weights. These values were collected from each participant during the first day of the study. The range of gestational age was 27.6 − 36.4 weeks, and the range of infant weights was 830 − 1746 grams.



(a) (b)

Figure 3.4: Histograms of baseline demographics on the first day of recording for each infant: a) gestational age and b) infant weight. Box plots above the histograms show the 25% and 75% quartiles, whiskers indicate the 9% and 91% percentiles, the middle line represents the median, and the '+' symbol denotes the mean.

3.4.2 Recording statistics

Studies typically started in the morning and ended in the early afternoon. However, due to equipment malfunction during the recording of session PN014 Day 1, the reference data for this session were not recorded. In total, the participants spent 442.0 hours in the study, during which video camera or reference data were being recorded. Of these hours, 426.6 hours (96.5%) were the periods during which there was simultaneous recording (overlapping recording times) with both video camera and reference monitoring equipment. Appendix A.2 shows a timeline plot of the recording times for all sessions. Table 3.3 summarises these recording times and overlaps, presenting the statistics for video and reference vital sign acquisition. The total video recording time was slightly higher than that of the reference equipment because video recording generally started before the reference system, and one session (PN014 Day 1) had no reference data due to equipment malfunction.

Table 3.3: Statistics for video camera and reference vital-sign recordings.

Recording type/Metric	Recording time (hours)
Camera	
Total recording time	439.1
Average recording time per session[1]	4.9 (±1.6)
Average recording time per subject[1]	14.9 (±5.7)
Reference vital-signs	
Total recording time	429.5
Average recording time per session[1]	4.8 (± 1.6)
Average recording time per subject[1]	14.3 (±5.2)

[1] Value shown as Mean (±SD)

3.4.3 Reference vital signs

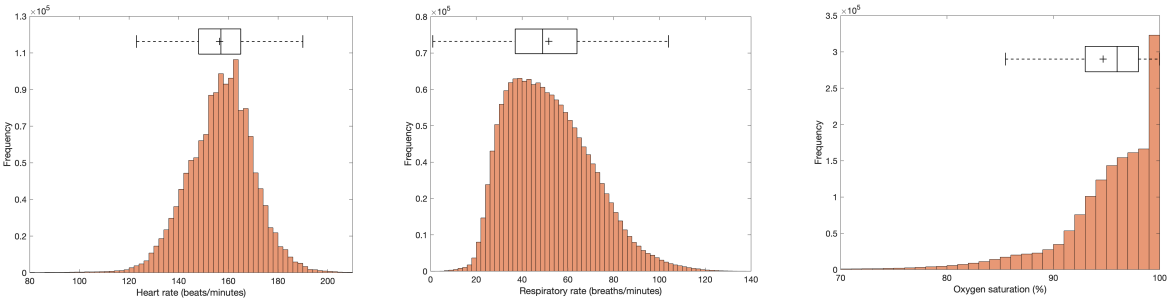
The Philips patient monitor records the vital signs of patients in the Neonatal HDU study at a frequency of 1Hz. Table 3.4 provides descriptive statistics for the reference vital-sign data recorded from the monitor. The mean heart rate provided by the electrocardiogram was 156.5 beats/min. The mean respiratory rate was 51.6 breaths/min, and the mean oxygen saturation was 94.7%.

Table 3.4: Statistics for the reference vital-sign data recorded from the Philips monitor.

Vital sign	Mean (±std)	Median (IQR)
HR derived from ECG (beats/min)	156.5 (±13.8)	157 (17)
RR (breaths/min)	51.6 (±18.5)	49 (27)
SpO_2 (%)	94.7 (±5.1)	96 (5)

std= standard deviation; IQR= interquartile range

The distribution of vital sign values (HR, RR, and SpO_2) recorded by the monitor throughout the study is presented in Figure 3.5. An analysis of these variables shows that the HR values followed a typical normal distribution. However, RR and SpO_2 distributions were skewed, resulting in a difference between their respective means and medians, as shown in table 3.4. The RR ranged from 3 bpm to a maximum of 161 bpm, with a mean rate of 51.6 bpm, a standard deviation of 18.5 bpm, and a median of 49 bpm. Since pulse oximetry became part of standard care in the NICU, pre-term infants have been subject to tightly controlled SpO_2 targets. As a result, the distribution of this vital sign is mostly confined to the 90-100% range, with an evident mode on the 99 - 100% histogram bin.



(a) (b) (c)

Figure 3.5: Histograms of reference vital signs for the entire Neonatal HDU study. a) HR derived from ECG by the patient monitor; b) RR derived from IP by the patient monitor, and c) SpO_2 from the pulse oximeter. Above the histograms, box plots bound the 25% and 75% quartiles, and whiskers bound the 9% and 91% of the data. The middle line indicates the median, while the '+' mark represents the data mean.

The Neonatal HDU dataset consists of a broad range of physiological values. An infant's vital signs can fluctuate significantly during a typical 4-hour study session, as shown by the example time series in figure 3.6. When the infant is undergoing clinical interventions such as a nappy change or blood sampling, clinicians may temporarily remove physiological sensors or pause vital sign monitoring. Signals such as HR and SpO_2 could therefore occasionally contain transient artefacts caused by temporary

sensor disconnections, infant movement, or poor sensor contact. These artefacts typically appeared as abrupt spikes, physiologically implausible values, or sudden drops to zero. During retrospective offline analysis, periods containing sensor disconnection or corrupted measurements were identified through visual inspection and excluded from further analysis. Subsequent signal quality assessment and preprocessing steps described in later chapters further reduced the impact of transient artefacts on downstream analyses. Figure 3.7 shows short segments of 30-second and 5-second time-series waveforms.

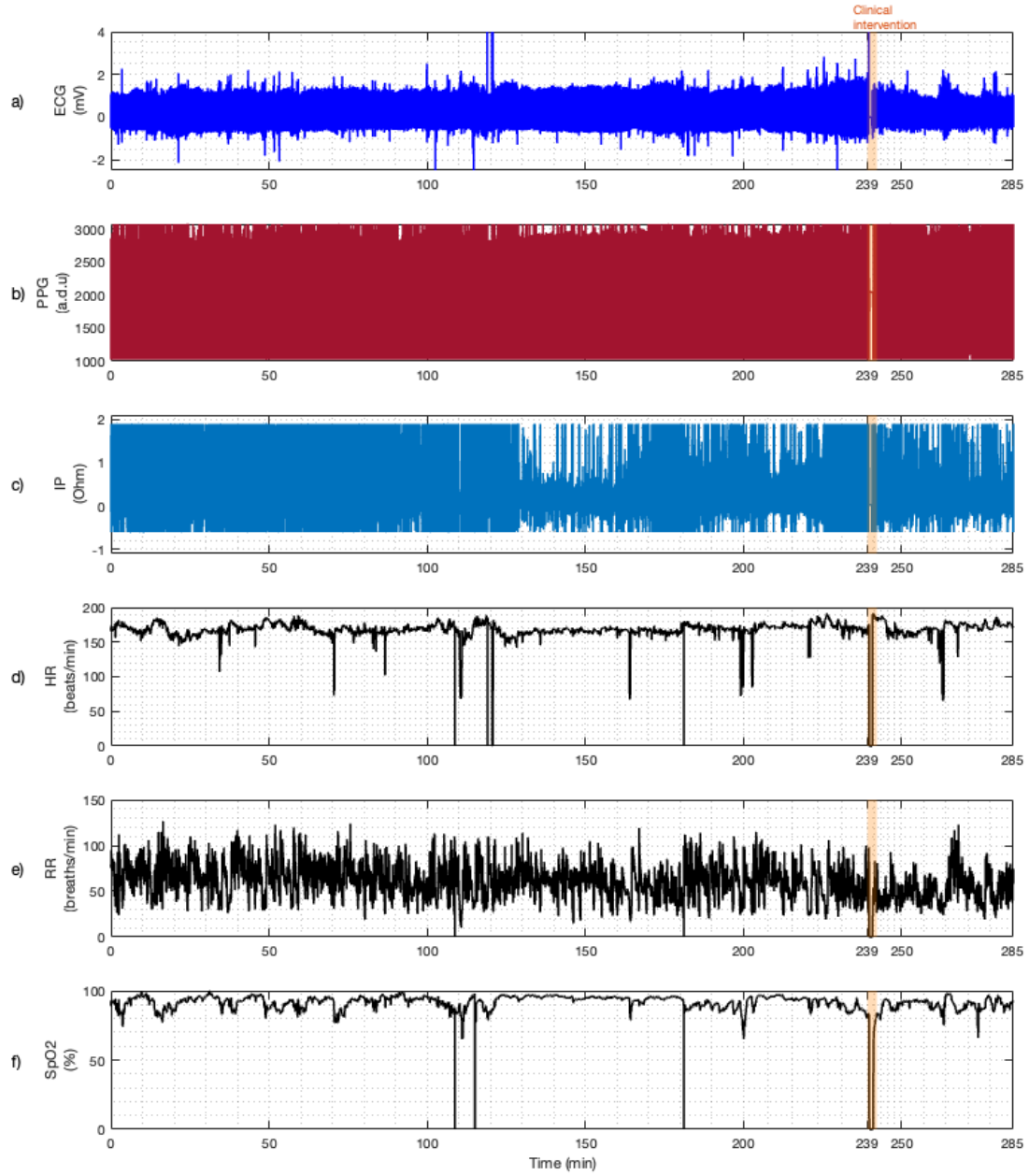


Figure 3.6: 4-hour sample from a recording session: a) ECG waveform; b) PPG waveform; c) IP waveform; d) Heart rate derived from ECG; e) respiratory rate from the Philips monitor, derived from IP signal and f) oxygen saturation from the pulse oximeter. Note that the infant was undergoing clinical interventions at minute 239, during which the monitoring was temporarily paused.

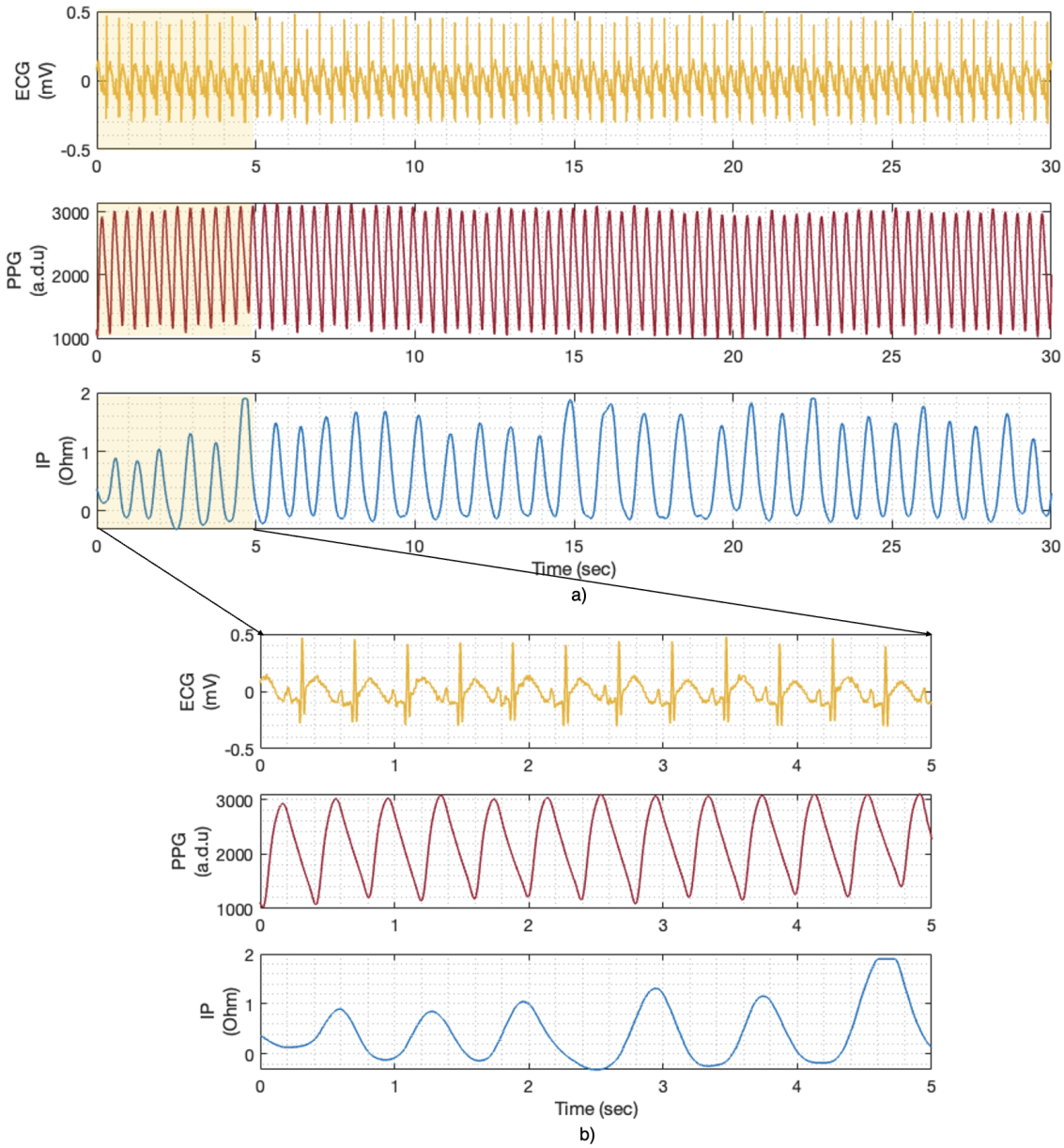


Figure 3.7: Sample waveforms from a typical recording session: a) 30-second recording and b) the first five seconds of the recording session in a).

3.5 Conclusion

The Neonatal HDU study involved double-monitoring participants using a video camera and a standard patient monitor without disrupting their routine medical care. This study generated a large dataset of video recordings and reference signals collected from participants of diverse backgrounds. The dataset was utilised to assess the efficacy of non-contact vital sign estimation algorithms. It could further be used to develop algorithms for detecting critical clinical events (such as COBE) and estimating physiological parameters.

Chapter 4

Contact monitoring of respiratory rate

4.1 Introduction

Impedance pneumography (IP) is the standard technology commonly used for monitoring respiratory rate in clinical and neonatal settings [88, 146]. It uses the same electrodes used to record the electrocardiogram to measure changes in the electrical impedance of the patient's thorax during each respiratory cycle [189]. Despite its widespread use, the IP signal is susceptible to noise and artefacts, which can significantly impact the accuracy of respiratory rate estimation [188]. Factors negatively affecting the recording of the IP signal include improper contact of electrodes with the skin, electrodes peeling off the skin due to infant movement, very dry skin leading to low conductance between electrodes and skin, shallow and irregular breathing, and variations in the infant's orientation or positioning [53, 193]. In the study, the Philips monitor provided RR estimates derived from the IP signal using proprietary algorithms without a corresponding quality measure. Using data from the same study, Jorge *et al.* [164] highlighted significant discrepancies between the derived respiratory rate and manual breath counting, indicating the unsuitability of monitor-derived values as reference values. Therefore, a new reference respiratory rate is required.

In this chapter, I use the impedance pneumogram signal recorded by the Philips IntelliVue patient monitor to compute a new reference respiratory rate. I selected the IP signal because it is the standard clinical measure of respiration and is most directly linked to thoracic motion associated with breathing [164].

4.2 Overview of the respiratory rate estimation process

The estimation process starts with resampling the IP signal. Across the entire Neonatal HDU study, the RR values from the Philips monitor range between 3 and 161 breaths/min, including during periods of cessation of breathing. Based on these values, I filtered the resampled IP signal using filter cut-off values in the frequency range between 0.033 Hz and 2.83 Hz, corresponding to respiratory rates of 2 breaths/min and 170 breaths/min, respectively. I then identified the peak and trough of each breath and assessed the signal quality for each breath cycle. After extracting, filtering, and assessing signal quality, I estimated RR using breath counting. This time-domain technique counts the number of breaths within a window. Figure 4.1 presents an overview of the process for estimating the reference RR from the IP signal.

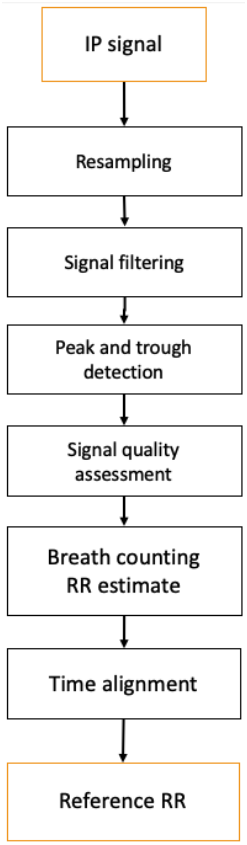


Figure 4.1: Flow diagram outlining the process of recomputing RR from the IP signal.

4.3 Resampling

I resampled the IP signal from 62.5 Hz to 24 Hz using cubic spline interpolation, reducing the number of samples and improving the speed of subsequent processing steps such as peak detection and signal quality assessment, while preserving physiologically relevant respiratory information.

4.4 Signal filtering

Pre-term infants have underdeveloped respiratory centres, often resulting in rapid, shallow, and irregular breathing [164]. Consequently, accurately estimating RR becomes more challenging within this specific demographic. In the Neonatal HDU dataset, the respiratory rates range from 3 to 161 breaths/minute depending on the gestational age and health condition of the infant. The variability in respiratory rates is large compared to term infants (30 - 60 breaths/min) and adults (12 - 20 breaths/min). Notably, previous research using this dataset has excluded periods of cessations of breathing where RR is less than 20 breaths/minute [88, 318]. Figure 4.2 shows 30-second IP segments from a recording of a 27.7-week-old infant. These examples illustrate the variety of signal characteristics observed, including regular breathing (figure 4.2a), clipped signals during deep breathing (figures 4.2b), and artefacts caused by

body movement (figures 4.2c and 4.2d). These variations highlight the need for careful filtering and signal quality assessment.

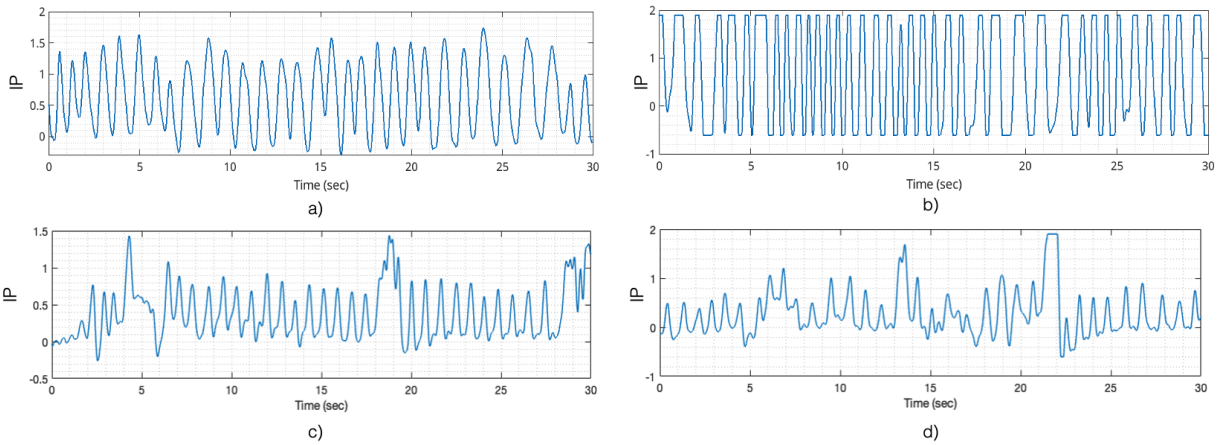


Figure 4.2: 30s segments of IP waveforms extracted from a recording session of a 27.7 week-old infant. The IP signal: a) regular breathing; b) clipped during deep breathing; c - d) containing artefacts from body movement

To account for the range of respiratory rates observed in our dataset, I first de-trended the signal to remove the DC component, then further processed the data by applying an 8^{th} order high-pass Butterworth Infinite Impulse Response (IIR) filter followed by a 6^{th} order low-pass Butterworth IIR filter. These filters were designed with cut-off frequencies of 0.033 Hz and 2.83 Hz, respectively. Figure 4.3 depicts a 30-second IP waveform before and after the filtering process. This comparative visualisation highlights the effectiveness of the IIR filters in isolating the specific frequency band associated with respiratory activity while attenuating frequencies outside the intended range.

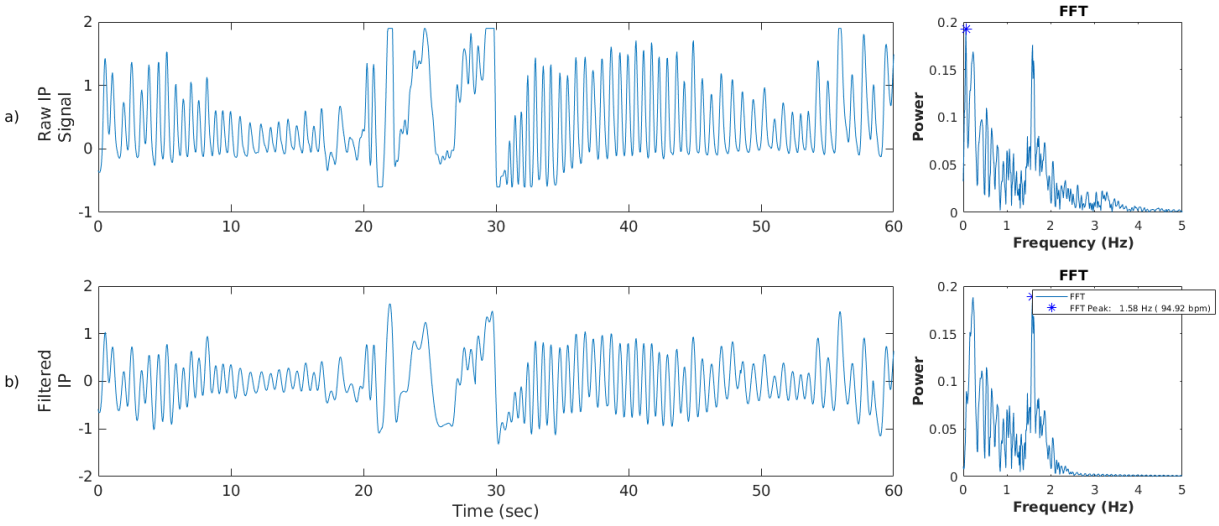


Figure 4.3: Signal filtering of a 60-second IP waveform. a) Original waveform; b) Waveform after the filtering with cut-off frequencies of 0.05 Hz and 2.83 Hz. This waveform shows a dominant frequency of 1.27 Hz, corresponding to 94.92 breaths/min.

4.5 Peak and trough detection

I detected peaks and troughs corresponding to each breathing cycle in the filtered respiratory signal. Peak-and-trough detection was selected instead of zero-crossing approaches because neonatal IP signals frequently exhibit low-amplitude fluctuations, clipped waveforms, and motion artefacts, which can produce false zero crossings and unstable respiratory estimates [320].

To perform peak detection, I used a method based on the calculation of the Mean Average Curve (MAC) as documented by Lu *et al.* [321]. The MAC defines a dynamic threshold for peak detection by averaging the local signal over a filter length L. To determine the length of the filter, L, I divided the respiratory signal into ten small learning periods of 5 minutes, distributed sparsely over the signal for each recording session. From these learning periods, I computed the respiratory frequency as the peak of the FFT power spectrum. Finally, I set the length of the filter to be the median of respiratory durations (calculated from the respiratory frequency) across all learning periods. This moving average appears as a line that intercepts each breath twice, once on inspiration and again on expiration. I use this signal as a dynamic threshold to find peaks and troughs in the respiratory signal. I defined the peak as the local maximum of values above the threshold, and the trough as the local minimum between two successive peaks that have already been detected. I rejected a peak if its amplitude was below 20% of the median amplitude calculated from the learning periods. Figure 4.4 shows the results of using the MAC method to detect peaks and troughs on a 30-second IP signal.

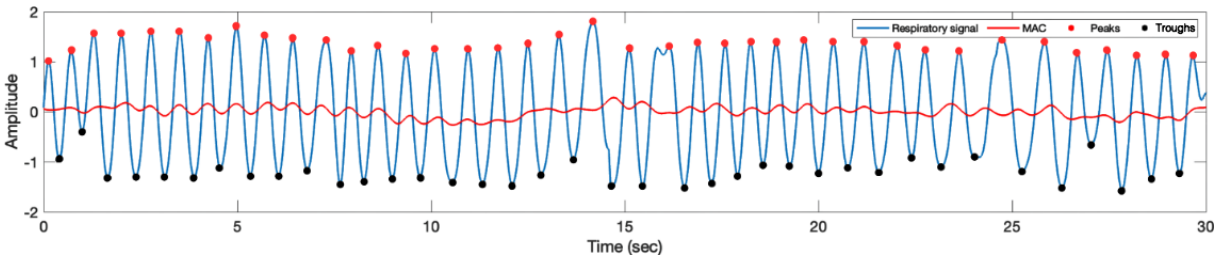


Figure 4.4: Peak and trough detection on a 30-second respiratory signal. The MAC signal, in red, is computed as a dynamic threshold. Peaks correspond to the local maxima above the threshold, and troughs correspond to the local minima found between two successive peaks.

In practice, the MAC-based peak-and-trough detection method performed reliably during periods of regular breathing and moderate signal variability, producing consistent breath-to-breath respiratory estimates across most recordings. Detection performance decreased primarily during periods of substantial body movement (e.g., around 21 seconds in the later figure 4.6), as well as during poor electrode contact or signal clipping, where missed or falsely detected respiratory cycles could occur. These limitations motivated the subsequent signal quality assessment procedures described in section 4.6, which were

introduced to identify and exclude periods of poor-quality respiratory signals before respiratory rate estimation.

4.6 Signal quality assessment

A common method for improving parameter estimation from physiological signals involves assessing the quality of the signal. The presence of artefacts caused by subject movements and clinical interventions can corrupt the data. After detecting the peaks and troughs in section 4.5, I analysed each respiratory cycle to compute measures of signal quality based on Li *et al.* [322]. The signal quality methodology involves three measures: physiological bounding (SQI_{phys}), spectral density distribution (SQI_{bin}), and signal purity index (SPI). I combined these measures to determine the overall signal quality index of the respiratory signal, which is referred to as SQI_{breath}. This process assigns a score ranging from 0 (poor quality) to 1 (good quality) to each respiratory cycle.

4.6.1 Physiological bounding

Physiological bounding measures whether the instantaneous RR falls within the physiological range of typical pre-term infants while accounting for the low RR values during episodes of COBE. For our dataset, I defined the physiological range as being from 2 to 170 breaths/min. To calculate the physiological bounding measure, $SQI_{phys}(i)$, I use the instantaneous respiratory rate of the i^{th} peak, which is denoted as $RR_{inst}(i)$:

$$SQI_{phys}(i) = \begin{cases} 1 & if\ 2\ \leq RR_{inst}(i)\ \leq 170\ breaths/min \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

4.6.2 Spectral density distribution

Spectral density distribution assesses how signal power is distributed across different frequencies within the signal's power spectrum. It helps in understanding where the energy or strength of the signal is concentrated across various frequency components. Since the energy of the respiratory signal is often concentrated around the peak of the power spectrum, the ratio of frequencies around this peak to the overall power spectrum can be used as a measure of signal quality. Using the power spectrum $P(w_k)$ of a 10-second window w_k centred on the k^{th} peak, where ν_k represents the dominant peak frequency in $P(w_k)$, the ratio of the power within a bandwidth around ν_k to the overall power spectrum is defined as

$$BIN(w_k) = \frac{\int_{\nu_k - \nu_{bw}/2}^{\nu_k + \nu_{bw}/2} P(w_k)\, d\nu}{\int_0^{f_s/2} P(w_k)\, d\nu} \tag{4.2}$$

Here, ν_{bw} is the bandwidth around the dominant peak frequency ν_k (set to 0.3 Hz following Chaiculee [318]), and f_s is the sampling frequency. The measure of the spectral density distribution, SQI_{bin}, is then defined as

$$SQI_{bin}(k) = \begin{cases} 1 & if\ BIN(w_k) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

I set SQI_{bin} to 1 (good quality) if more than half of the power spectrum is concentrated within a bandwidth of 0.3 Hz around the dominant peak, as shown in Figure 4.5.

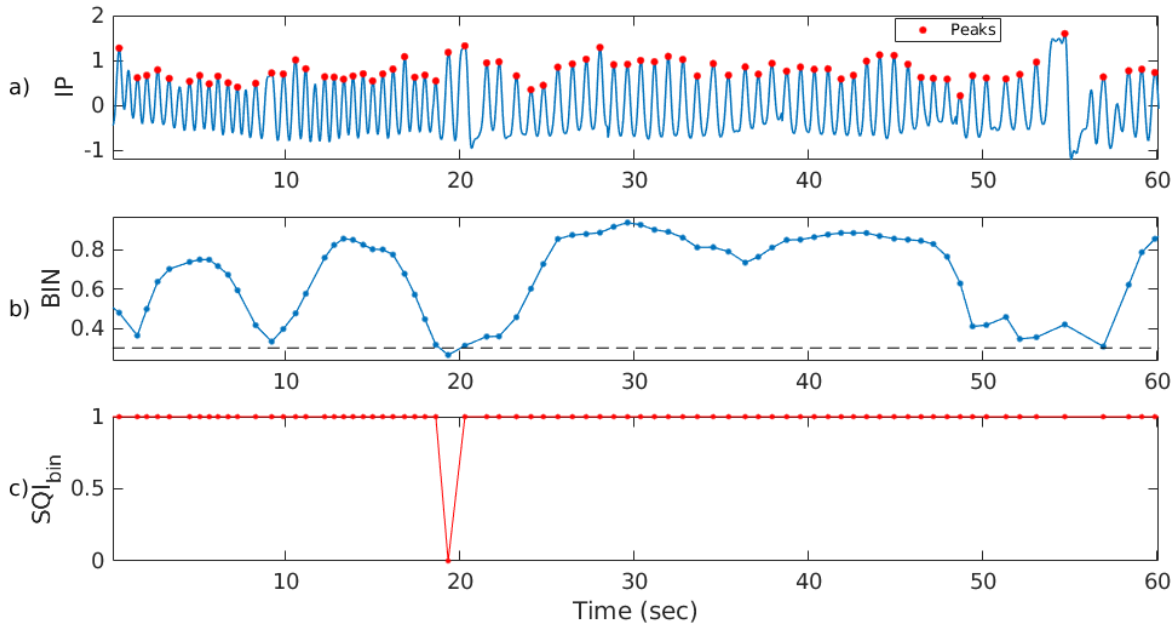


Figure 4.5: Spectral density distribution for a 60-second IP segment: a) IP signal; b) BIN values for each breath peak (dots), with the black dashed line indicating the threshold ($BIN = 0.5$) used to classify signal quality; c) SQI_{bin} indicating breaths that are considered good quality ($BIN \geq 0.5$)

4.6.3 Spectral Purity Index

Spectral Purity Index (SPI) is a metric used to describe the degree to which a single frequency can describe a signal. It measures how well-defined the dominant frequency components are within the signal's spectrum, reflecting the stability of the signal. The SPI ranges from 0 for non-sinusoidal noisy signals to 1 for pure sine wave signals. The SPI of a signal in a window, $\bar{\omega}_i$, of 10 seconds centred around the i^{th} peak is defined as:

$$SPI(i) = \frac{\bar{\omega}_2^2(\omega_i)}{\bar{\omega}_0(\omega_i)\bar{\omega}_4(\omega_i)} \tag{4.4}$$

Given the signal power spectrum $S_x(e^{j\omega})$ as a function of angular frequency ω, the nth-order spectral moment $\bar{\omega}_n$ is defined as:

$$\bar{\omega}_n = \int_{-\pi}^{\pi} \omega^n S_x(e^{j\omega}) d\omega \tag{4.5}$$

In this study, I estimated the spectral moments $\bar{\omega}_n$ in the time domain using Hjorth descriptors. The zero-order, second-order, and fourth-order spectral moments correspond to Hjorth's activity, mobility, and complexity parameters, respectively [323]. Given a discrete time series $x(n)$ of length N, spectral moments can be estimated by:

$$\bar{\omega}_i \approx \frac{2\pi}{N} \sum_{n=0}^{N-1} (x^{\frac{k}{2}}(n))^2, \quad k = 0, 2, 4 \tag{4.6}$$

In the context of respiratory signals, Hjorth activity reflects the overall variance or power of the waveform, with larger values corresponding to greater respiratory amplitude variability [323, 324]. Hjorth mobility measures the mean frequency content of the signal and therefore reflects how rapidly the respiratory waveform changes over time. Hjorth complexity quantifies variations in frequency content relative to a pure sinusoidal waveform, providing a measure of waveform irregularity and spectral spread [323, 324]. SPI, computed from the relationship between Hjorth mobility and complexity, therefore provides a measure of spectral purity by quantifying the extent to which the respiratory signal contains a single dominant frequency component with periodic waveform morphology. Respiratory signals dominated by a single respiratory frequency produce higher SPI values. Noisy or motion-corrupted signals containing multiple competing frequency components produce lower SPI values.

4.6.4 Combining signal quality measures

After computing all the necessary signal quality measures, I used $SQI_{breath}(i)$ to describe the overall signal quality of the i^{th} breath in the respiratory signal by combining all the signal quality measures:

$$SQI_{breath}(i) = SQI_{phys}(i) \times SQI_{bin}(i) \times SPI(i) \tag{4.7}$$

If an instantaneous respiratory rate was outside the expected range for pre-term infants in our dataset, the $SQI_{phys}(i)$ set $SQI_{breath}(i)$ to zero. Similarly, if there were multiple dominant frequencies in the power spectrum, $SQI_{bin}(i)$ would also set $SQI_{breath}(i)$ to zero. This meant that the SPI would only be considered reliable when the spectrum was likely to be unimodal. This formulation was intentionally conservative, as low values in any individual signal quality component substantially reduced the overall SQI_{breath} value, ensuring that respiratory cycles failing one or more quality criteria were rejected from downstream analysis. While this approach may increase data dropout during noisy periods, it reduces the likelihood of unreliable

respiratory estimates contributing to the reference signal. Alternative weighted combinations could potentially retain more data by reducing the influence of individual low-quality components. However, a weighted combination strategy would require optimisation and validation of the relative weighting coefficients for each quality component, introducing additional parameter tuning complexity and potential sensitivity to dataset-specific tuning. Therefore, the stricter multiplicative formulation was preferred in this work to prioritise reference signal reliability during periods of physiological instability. Figure 4.6 shows the assessment of the signal quality for a respiratory waveform extracted from a typical recording session. In this figure, the SQI_{breath} is high when the waveform is sinusoidal. The SQI_{bin} is 0, and the SPI decreases during the presence of body movement. Therefore, the SQI_{breath} during this period is set to 0, corresponding to a period of poor quality for which the respiratory rate cannot be reliably computed.

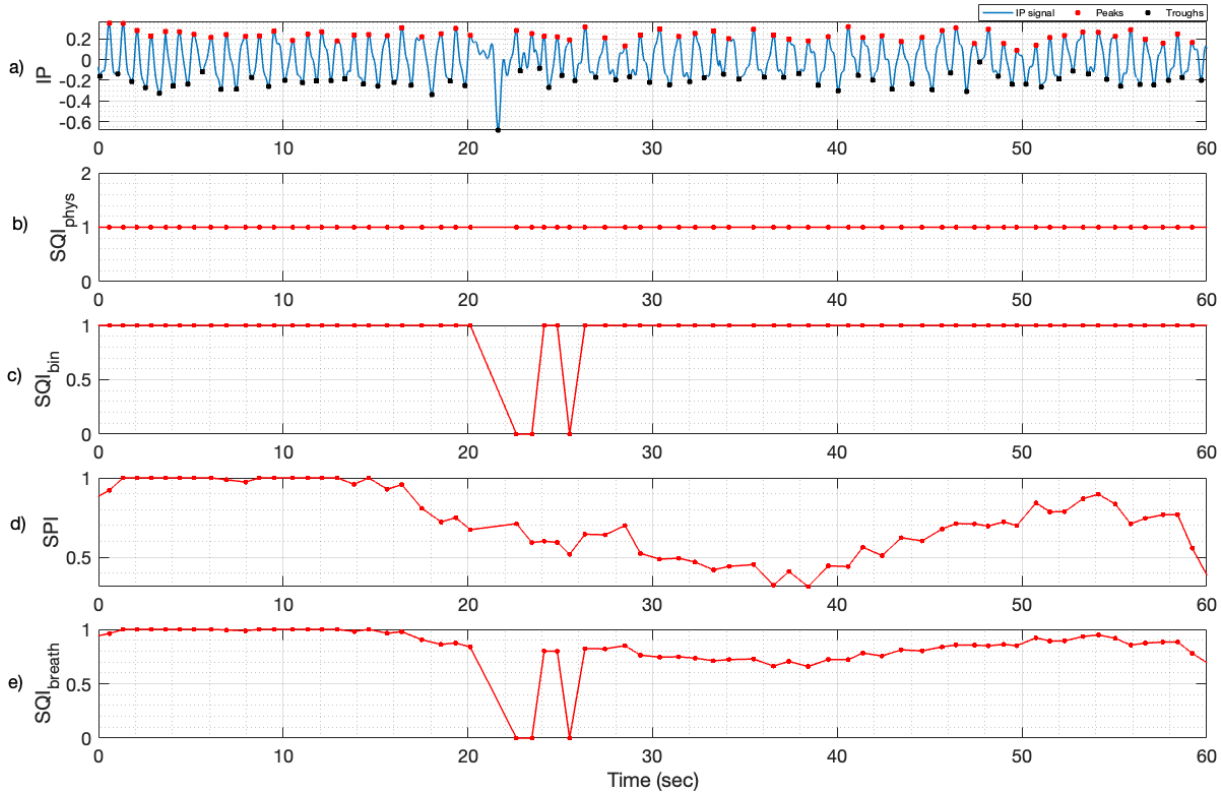


Figure 4.6: Signal quality assessment of a 60-second IP waveform from when the infant was moving. a) IP signal with the peaks and troughs detected. Movement is present at around 20 seconds; b) SQI_{phys}; c) SQI_{bin}; d) SPI; e) SQI_{breath} combining all the previous measures. The values of the SQI_{bin} and SPI decrease during the presence of motion artefacts.

4.7 Computing reference respiratory rate

After detecting the peaks and troughs of the respiratory signal and assessing the signal quality for each breath, I computed the respiratory rate using the breath counting method over a 10-second sliding window with a step size of 1 second. I counted the number of breaths in the window w and scaled to the unit of

breaths/min using equation 4.8. Thus, the RR of the window w is given by

$$RR_{ip}(w) = N_{breaths} \times \frac{60}{L} \tag{4.8}$$

where $N_{breaths}$ is the number of breaths in the window and L is the length of the window.

4.8 Signal alignment

Physiological monitoring devices used in the NICU record time measurements independently before transmitting the data to the patient monitor for display. This can cause time discrepancies between the raw signals recorded by the different systems, as well as delays in the physiological parameters derived at the device level (such as RR from IP and HR from ECG). Additionally, my research utilises a supplementary monitoring system (i.e., the video camera), which can introduce further time delays. These time discrepancies can result in inaccuracies in data analysis and interpretation; therefore, they need to be measured and compensated for before comparing signals from different sources. However, not all temporal delays observed between physiological signals arise from recording system misalignment. In particular, oxygen saturation changes measured by pulse oximetry occur physiologically after changes in respiratory activity measured by IP, because oxygen desaturation is a delayed consequence of reduced ventilation or cessation of breathing [32]. Consequently, synchronisation procedures described in this section were designed to compensate for measurement-system timing discrepancies between signals expected to represent the same underlying physiological process, such as respiratory activity derived from different modalities. Physiological delays between respiratory activity and downstream oxygenation changes were preserved and not removed during signal alignment.

To establish the accuracy of the respiratory rate readings, I took the RR_{ip} as the gold standard and aligned all the other derived signals to it. The time delay between two signals can be estimated using cross-correlation [325]. I employed this method to synchronise the RR_{philips} and RR_{ip}. The cross-correlation method determines the similarity between the two signals based on a time lag or shift applied to RR_{philips}. A systematic time lag can occur because the Phillips monitor works in real time and therefore reports RR estimates at the end of each analysis window, whereas in my work, I estimate RR at the centre of the window to improve detection of COBE episodes. To account for this, I took a 5-minute window of RR_{philips} signal and compared it to each 5-minute window of RR_{ip} within ±2.5-minutes, sliding in 1-second intervals. For each record, the time lag, corresponding to the highest correlation coefficient between RR_{ip} and RR_{philips} windows, indicated the optimal alignment between the signals as:

$$\tau = \arg\max_{\tau} \int_{-\infty}^{\infty} RR_{ip}(t) \cdot RR_{philips}(t-\tau)\, dt \tag{4.9}$$

After computing the time lag, $RR_{philips}$ was shifted by τ seconds. If $\tau > 0$, $RR_{philips}$ was shifted forward in time by τ seconds, and if $\tau < 0$, $RR_{philips}$ was shifted backwards in time by τ seconds. The aligned signal $RR'_{philips}$ was therefore determined to be:

$$RR'_{philips}(t) = RR_{philips}(t - \tau) \tag{4.10}$$

Figure 4.7 shows samples of RR_{ip} and RR_{philips} time series before and after signal alignment. The time delay between the two signals is estimated to be 8.4 seconds.

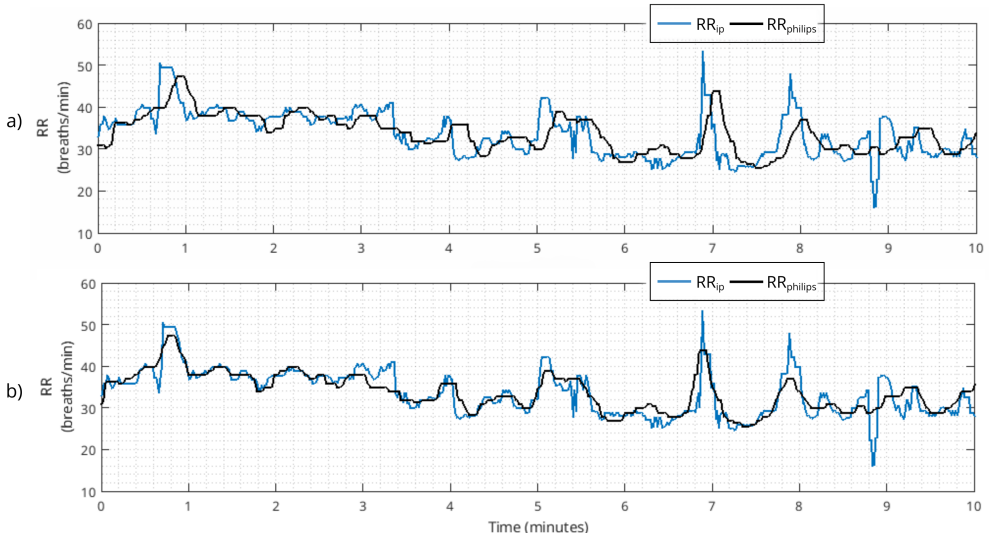


Figure 4.7: The process of aligning RR_{philips} with the reference RR_{ip}. a) before signal alignment, showing a time delay between the two signals; b) after signal alignment using cross-correlation.

4.9 Results

I performed a comparison between RR_{philips} and RR_{ip} for all recordings in the study using Bland-Altman analysis [326], Mean Absolute Error (MAE), Mean Absolute Deviation (MAD) and Pearson's correlation coefficient. I used the Bland-Altman plot to assess the agreement between two clinical measurements. I constructed this plot by plotting the mean of the IP and Philips-derived respiratory rates on the x-axis and the differences between the two measurements on the y-axis.

I used the MAE metric to measure the error between two variables, and the MAD was used to measure the absolute deviation from the mean. Given two time series RR_{ip} and RR_{philips} of length L, I defined MAE as:

$$MAE = \frac{1}{L}\sum_{i=1}^{L} |RR_{ip} - RR_{philips}| \tag{4.11}$$

Given q as the difference between the two time series, and μ as the mean of q, I defined the MAD as:

$$MAD = \frac{1}{L}\sum_{i=1}^{L}|q - \mu_q| \tag{4.12}$$

The MAE between both respiratory rate measurements was 14.3 breaths/min with an MAD of 13.7 breaths/min. I also calculated the Root Mean Square Error (RMSE) as:

$$RMSE = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(RR_{ip} - RR_{philips})^2} \tag{4.13}$$

I calculated the Pearson's correlation coefficient (**r**) using equation 4.14 to measure the linear correlation between two time series.

$$r = \frac{\mathbf{cov}(RR_{philips}, RR_{ip})}{\sigma_{RR_{philips}} \times \sigma_{RR_{ip}}} \tag{4.14}$$

where $\mathbf{cov}(RR_{philips}, RR_{ip})$ is the covariance, $\sigma_{RR_{philips}}$ and $\sigma_{RR_{ip}}$ are the standard deviation of RR_{philips} and RR_{ip} time series respectively. Figure 4.8 shows the estimation of RR for a 5-minute respiratory signal compared with RR_{philips}. The estimation is done for a segment around a period of cessation of breathing labelled by an annotator (the annotation process is described in detail in Chapter 5).

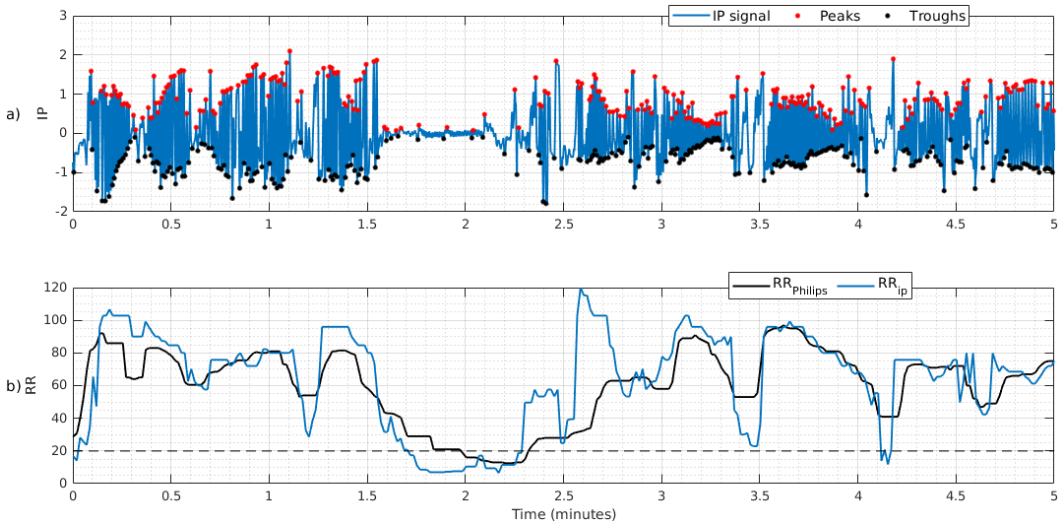


Figure 4.8: RR estimation using the breath counting method on a 5-minute segment of the IP signal: a) IP signal; b) Respiratory rate estimated using the breath counting method compared to RR_{philips}.

Figure 4.9 shows the summary Bland-Altman plots for the proposed RR re-computation algorithms. The Bland-Altman plots compare the RR_{philips} and RR_{ip} values for all 30 pre-term infants. The mean difference between the two measurements was 4.3 breaths/min. Table 4.1 then shows overall performance metrics for recomputing respiratory rate over the entire dataset of 90 sessions.

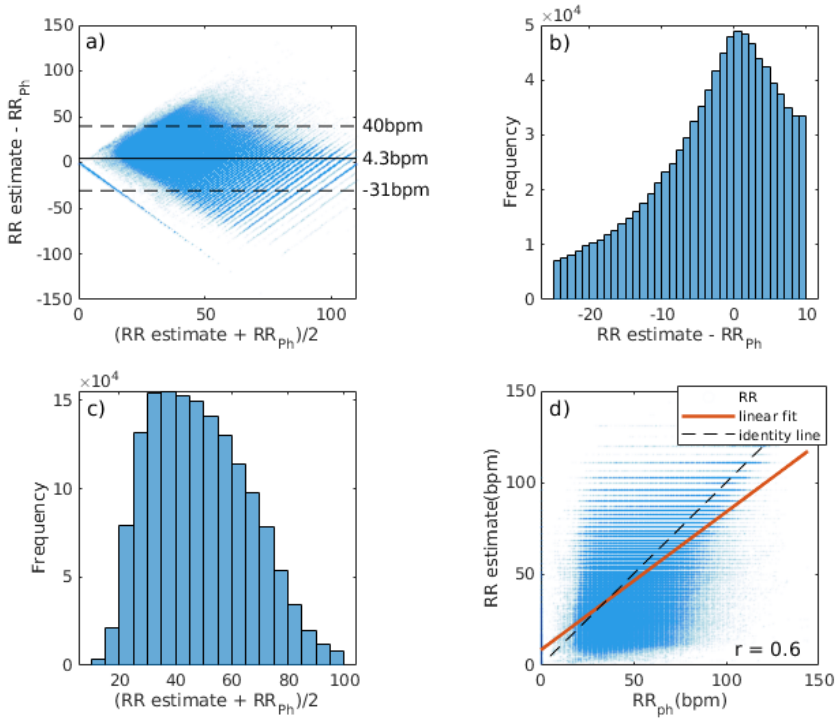


Figure 4.9: Comparison between the Philips and recomputed respiratory rate values over the entire dataset. a) Bland-Altman plot; b) Histogram of the differences between the two RR estimates; c) Histogram of the mean of the two RR estimates, and d) correlation plot showing a positive correlation between the two measurements.

Table 4.1: Summary of error metrics for the proposed algorithms for recomputing RR from the IP signal.

Metric	Value
MAE*	14.1
MAD*	13.6
RMSE*	18.8
r	0.60

* Corresponding values in breaths/min

4.10 Discussion

Patient monitoring systems in the NICU generally report the respiratory rate based on the analysis of the IP signal. This signal is prone to noise and can be disrupted by patient movement, meaning it is most reliable during periods of low patient activity, such as during sleep or quiet times. However, even during these periods, pre-term infants exhibit irregular breathing and fidgety movements [327, 328], which can interfere with the respiratory signal, making it difficult to obtain a reliable respiratory rate. In this chapter, I propose a time-domain breath counting method to compute the reference respiratory rate, even during periods of wakefulness, patient movement, and signal noise. Although many studies in the literature [21, 164, 318] tend to exclude periods of poor signal quality when computing vital signs, it is important to take these periods into account, when possible, when monitoring infants experiencing cessation of breathing. During

these times, infants may experience breathing difficulties, leading to frequent patient movement and noisy breathing signals.

I adopted a multi-step approach to recompute RR, improving the reliability of the current reference respiratory estimates. This method involved utilising advanced signal processing algorithms to filter out noise and artefacts from the IP signal, followed by employing a MAC peak detection algorithm to identify peaks corresponding to respiratory cycles. Subsequently, I employed signal quality indices to provide a quantitative assessment of the signal's quality, enabling the differentiation between respiratory-related peaks and noise-induced peaks. Compared to prior work, I relaxed SQIs like physiological bounding, particularly at low RR values, to avoid discarding cessations of breathing as artefacts. While the exact averaging window used by the Philips IntelliVue MX800 is not explicitly documented, the respiratory estimates from the monitor exhibit less short-term variability than my recomputed RR. This suggests that my choice of a 10-second window increases responsiveness to transient respiratory fluctuations. This choice produced a more variable RR signal as observed in figure 4.8, reflecting the rapid dynamics of pre-term infant breathing. Consequently, the agreement limits observed in the Bland-Altman analysis (approximately ± 10 breaths/min) were intentionally wider than those expected from bedside monitor outputs. The Philips monitor derives respiratory estimates from the IP signal using proprietary signal processing algorithms [318]. Such monitoring approaches often employ temporal averaging to provide stable bedside respiratory measurements, although this can reduce sensitivity to transient respiratory changes and short respiratory pauses [163, 329]. In contrast, the recomputed reference RR in this work was intentionally designed to respond more rapidly to short-term respiratory fluctuations and cessation of breathing events. As a result, the recomputed RR exhibited greater short-term variability but improved sensitivity to transient respiratory changes during COBE episodes.

My findings showed significant discrepancies between monitor-derived respiratory rates and my newly computed reference respiratory rate, echoing previous concerns raised by Jorge *et al.* [164] and Chaichulee [318]. The Bland-Altman plot I presented in figure 4.9a indicates a bias of 4.3 breaths per minute. The distribution of differences between the two time series shows that the values are predominantly clustered around 50 breaths per minute. Furthermore, the correlation plot in figure 4.9d shows an average positive correlation coefficient (r) of 0.60 between the respiratory rate measurements obtained from the Philips device ($RR_{philips}$) and those derived from the alternative measurement method ($RR_{estimate}$). This finding emphasises the need for a reliable gold standard reference for respiratory rate assessment in premature infants.

In conclusion, the purpose of this chapter was not to achieve a perfect correlation with RR_{philips}. Instead, my focus was on generating a reliable reference respiratory rate that could capture events, including COBE, potentially missed by the monitor. This approach addresses known limitations in patient monitor-derived respiratory rates and provides a more sensitive basis for clinical assessment and event detection.

Chapter 5

Creating a labelled dataset for COBE detection

Apnoea is widely characterised by (1) cessation of breathing for a minimum of 20 seconds or (2) a shorter pause in breathing for a minimum of 10 seconds accompanied by bradycardia (HR < 100 beats per minute), or desaturation (SpO_2 < 80 %) [22, 23]. This definition of apnoea is a subject of ongoing debate in the medical community but serves as a foundational guideline for diagnostic purposes. In this chapter, I outline the methodology used to identify periods of cessation of breathing in the Neonatal HDU dataset, and present the resulting COBE dataset created.

5.1 Creating the COBE training dataset

The process of creating the COBE dataset involved obtaining reference annotations, which served as labels for training the COBE episode detection algorithms on NICU patients.

5.1.1 Selection of data for annotation

The overall NICU dataset consists of 429.5 hours of monitoring, meaning it was not feasible to have the entire dataset manually annotated. As such, an approach was required to select relevant candidate segments of vital sign data for expert manual annotation. Because desaturation is the most clinically significant immediate consequence of apnoea, I automatically identified periods of desaturation from all of the 90 recorded sessions in the dataset. Here, I defined a desaturation event as any decrease in SpO_2 to below 80% which persists for 10 seconds or more. This definition is consistent with that used in several previous studies [330, 331].

To identify these periods of desaturation, I first identified all decreases of SpO_2 to below 80% of any duration. If two of these SpO_2 decreases occurred within 20s of one another, I considered merging them into a single event. A 20-second merging interval was introduced to reduce splitting of prolonged desaturation episodes into multiple short events caused by brief transient recoveries in SpO_2. This approach was motivated by observations in neonatal cardiorespiratory monitoring that temporally clustered desaturation or apnoea-related events may occur during periods of ongoing physiological instability rather than representing entirely independent episodes [332]. Although brief physiological recoveries in SpO_2

may occur within this interval, the objective of this stage of the analysis was to identify desaturation episodes associated with possible COBE events rather than to characterise every brief fluctuation in oxygen saturation as an independent desaturation event. The selected threshold, therefore, represents a pragmatic compromise between reducing fragmentation of prolonged desaturation episodes and preserving short recovery intervals between neighbouring events. For a possible merged SpO_2 decrease (A + C) shown using synthetic data in figure 5.1, I accepted this merge if:

$$\frac{A+C}{A+B+C} \times 100 \geq 75 \tag{5.1}$$

That is, SpO_2 was less than 80% for more than 75% of the potential merged period. This criterion was selected to ensure that the merged segment predominantly consisted of periods where SpO_2 remained below 80%, rather than isolated transient decreases in SpO_2. The 75% threshold was selected because lower thresholds would increase the likelihood of merging unrelated neighbouring desaturation events, whereas higher thresholds would prevent merging when brief transient recoveries interrupted an otherwise prolonged desaturation episode. This was considered acceptable for the present study because the merging process was intended to identify candidate desaturation episodes for subsequent expert review rather than perform precise physiological segmentation of every transient recovery in SpO_2. Figure 5.1 presents a visual illustration of this merging process. Using this approach, only 14% of initially detected desaturation episodes were merged with neighbouring events, indicating that the majority of annotated events were unaffected by this criterion.

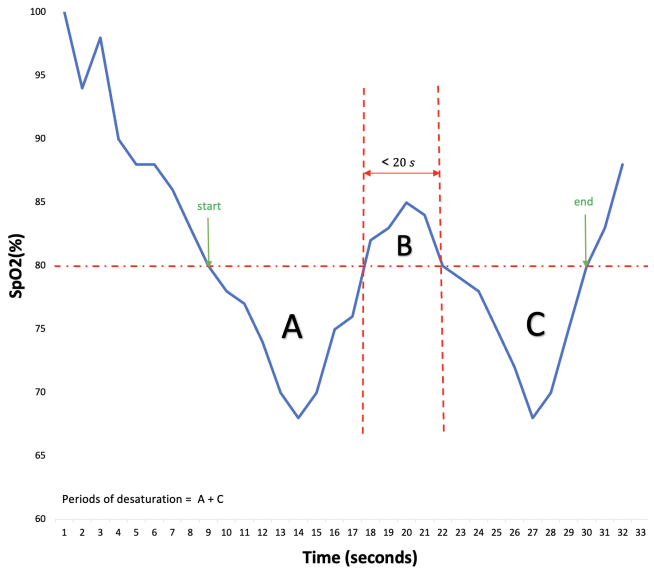


Figure 5.1: Synthetic waveform showing how desaturations are defined and merged in our dataset. The merged desaturation makes up 84.2 % of the data.

After merging, I considered a desaturation event to be valid if it had a duration of at least 10 seconds. Using this process, I identified a total of 620 desaturation events in my dataset. These 620 segments

provide manual annotators with periods of data where cessation of breathing is likely to occur, ensuring my manually annotated dataset includes a significant number of COBE. However, these segments only constitute periods where the infant experienced respiratory distress around a desaturation event, and as such they are not representative of general pre-term infant respiratory behaviour in the NICU.

To address this limitation, I constructed a second, identically sized and complementary dataset to include a full range of infant respiratory behaviours, including both pathological and non-pathological patterns, that could be feasibly annotated. Specifically, I selected 620 non-desaturation episodes from the same 67 sessions with desaturation events in the first phase of data selection. This approach ensured that the datasets were matched in terms of both size and patient representation. I excluded all overlapping segments to maintain unique annotation segments, and any episodes that intersected with previously sampled windows in the first dataset, or with another non-desaturation segment, were discarded. This second dataset, therefore, represents typical breathing episodes not associated with oxygen desaturation, reflecting the more heterogeneous distribution of respiratory behaviours seen in full-length monitoring data. Overall, this approach provided 1,240 segments of vital sign data, ready for expert manual annotation.

5.1.2 Manual annotation

To facilitate the manual annotations, I developed a user-friendly interface in MATLAB that utilised keystrokes to analyse individual segments of data. The annotation environment streamlined the review process, allowing the user to easily navigate through the recordings, observe specific events, and save the annotation results.

I presented each annotator with 5-minute segments around potential COBE episodes to provide context on the infant's physiological behaviour leading up to and following the event. The team of reviewers consisted of clinicians with expertise in neonatal medicine and engineers experienced in analysing physiological data. This interdisciplinary approach ensured that the annotations were both clinically relevant and technically accurate. To further ensure annotation accuracy, reviewers had access not only to the primary signals used for event identification (RR_{ip} and SpO_2), but also to additional raw signals, including ECG and IP recordings. The MATLAB annotation interface used to review these supplementary signals is shown in Appendix B. This allowed reviewers to check for potential artefacts or sensor detachment if there was uncertainty with the primary signals. Reviewers could request to view these supplementary signals for any segment where the quality of the primary data was in question. Although these supplementary signals improved annotation reliability, the annotation methodology did not explicitly differentiate between central, obstructive, and mixed apnoea because reliable subtype classification typically requires additional measurements of airflow and respiratory effort beyond impedance

pneumography alone, which were not consistently available in this study. Consequently, the dataset was designed to identify clinically significant episodes of cessation of breathing irrespective of underlying apnoea subtype.

Initially, all potential COBE episodes were screened using a simple oxygen desaturation test: Is SpO_2 below 80% for at least 10 seconds within the 5-minute window. This screening step divided the dataset into two major groups for subsequent review: (1) desaturation-associated events (right-hand branch of the flowchart) and (2) non-desaturation events (left-hand branch). Each group contained 620 segments, which were independently annotated using 300 seconds of data. These candidate events were then manually annotated and classified according to the decision logic summarised in Figure 5.2.

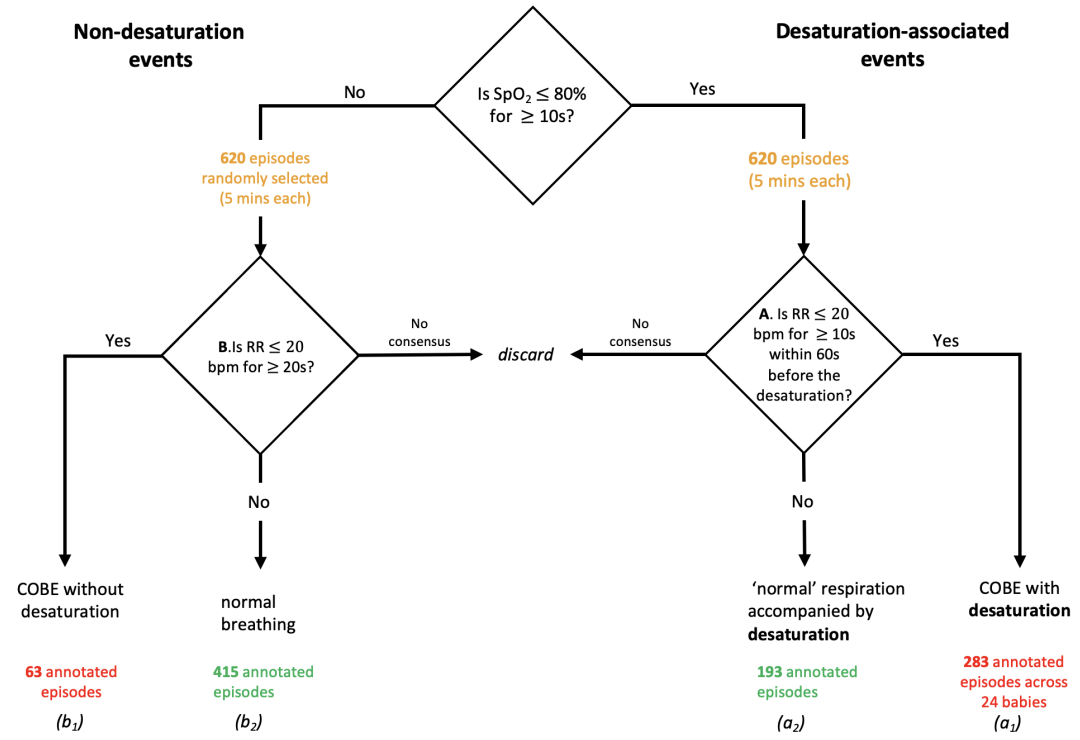


Figure 5.2: Decision tree for manual annotation of potential COBE episodes by three annotators. The reviewers examined vital signs from the clinical monitoring devices for 300 seconds around each event trigger.

Phase A – Desaturation-associated events (right-hand side of flowchart)

The three independent reviewers fully annotated 620 five-minute segments, which satisfied the SpO_2 below 80% for at least 10 seconds condition. For each segment, reviewers assessed whether RR_{ip} was ≤20 breaths/min for at least 10 seconds within 60 seconds before the desaturation episode. Segments satisfying this condition were classified as COBE with desaturation *(a1),* representing clinically significant COBE episodes characterised by both reduced respiration and a decrease in oxygen saturation. This generated 283 annotated episodes across 24 babies. Figure 5.3 illustrates a representative *(a1)*

episode, where a clear cessation of breathing is observed in the RR_{ip} signal, followed by a decrease in SpO_2.

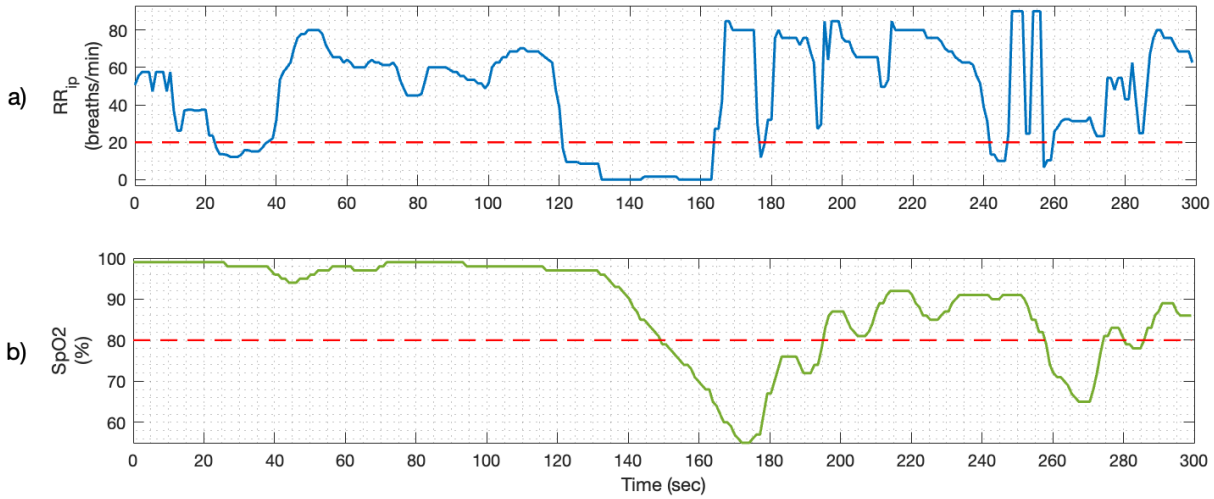


Figure 5.3: RR_{ip} and SpO_2 signals during a period of COBE *(a1)*. a) RR_{ip} signal showing cessation of breathing lasting more than 10 seconds from t = 123 s. b) SpO_2 signal reaching levels below 80% from t = 155 s. The red horizontal lines at RR_{ip} = 20 breaths/min and SpO_2 = 80% illustrate the defined thresholds for the detection of potential COBE events.

Segments that displayed desaturation without a corresponding decrease in RR_{ip} were labelled as 'normal' respiration accompanied by desaturation *(a2)*, totalling 193 annotated episodes. I discarded the remaining episodes either due to a lack of consensus among reviewers or because the data quality was insufficient, typically indicating sensor detachment or artefact. Figure 5.4 shows an example of a desaturation episode where the RR_{ip} remains above the 20 breaths/min threshold, indicating normal breathing despite desaturation.

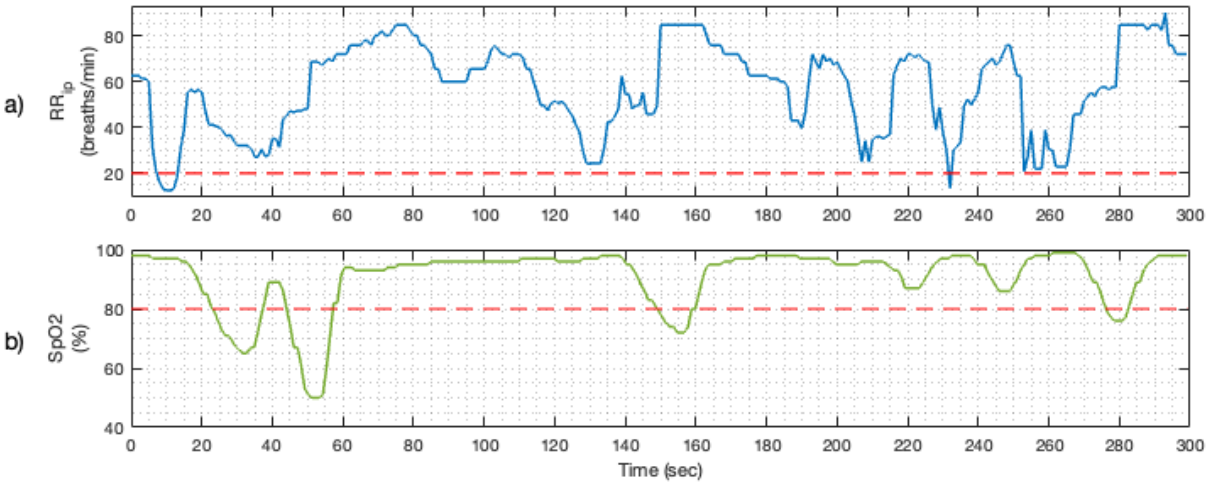


Figure 5.4: RR_{ip} and SpO_2 signals during a desaturation. a) RR_{ip} remains above the 20bpm threshold suggesting a period of 'normal' breathing (*a2)*. b) The SpO_2 signal reaches low levels of 80%, at t = 150 sec, showing a desaturation event. The red horizontal lines at RR_{ip} = 20 breaths/min and SpO_2 = 80% illustrate the defined thresholds for the detection of potential COBE events.

Phase B – Non-desaturation events (left-hand side of flowchart)

After completing Phase A annotations, the reviewers annotated another 620 segments from randomly selected 300-second windows. This phase focused on extended episodes that did not involve oxygen desaturation (SpO_2 remained above 80%). Annotators identified episodes of COBE using the criterion of RR_{ip} below 20 breaths/min for at least 20s. Segments that satisfied this condition were classified as COBE without desaturation *(b1)*, representing a clinically relevant event where the infant exhibited a prolonged cessation in breathing without a concurrent decrease in SpO_2. Figure 5.5 shows a representative *(b1)* episode, where the RR_{ip} is below 20 breaths/min for more than 20 seconds, while SpO_2 remains above the defined desaturation threshold. This category helps capture events that may still pose clinical concern but would be missed if only desaturation-based criteria were considered. Based on consensus, 63 episodes were labelled as COBE without desaturation *(b1)*.

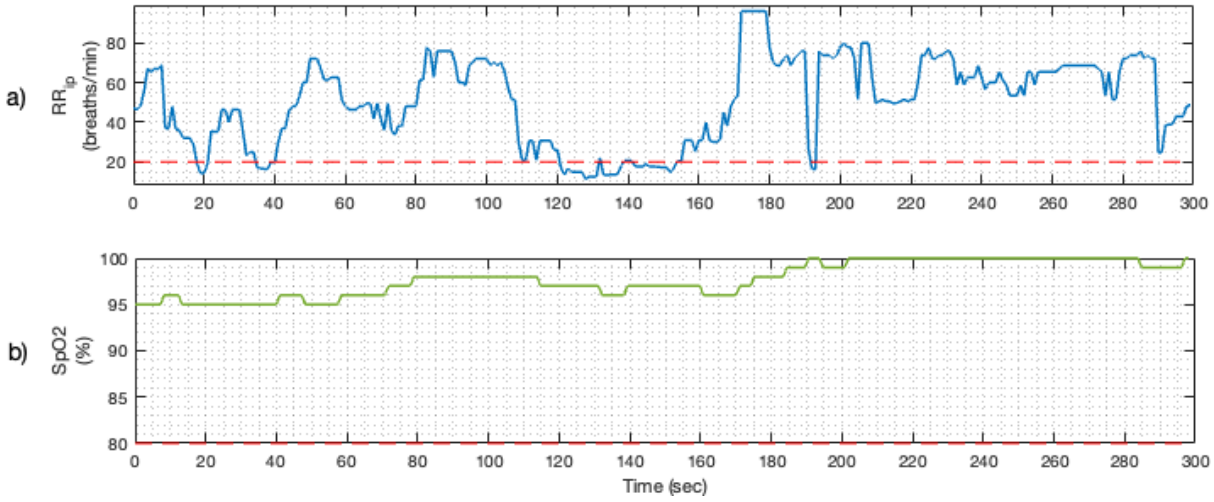


Figure 5.5: RR_{ip}, and SpO_2 during a period of COBE *(b1)*. a) RR_{ip} signal showing a cessation of breathing lasting more than 20 seconds from t = 120 s. b) SpO_2 signal remains above the 80% threshold, indicating no clinically significant desaturation. The red horizontal lines at RR_{ip} = 20 breaths/min and SpO_2 = 80% illustrate the defined thresholds for the detection of potential COBE events.

Segments that showed no decrease in RR_{ip} and no desaturation were labelled as normal breathing *(b2)*. Figure 5.6 shows a typical *(b2)* episode, where all physiological signals remain within normal limits. In total, 415 episodes were identified as normal breathing *(b2)*, while the remaining segments were discarded due to annotator disagreement or poor data quality, such as sensor detachment.

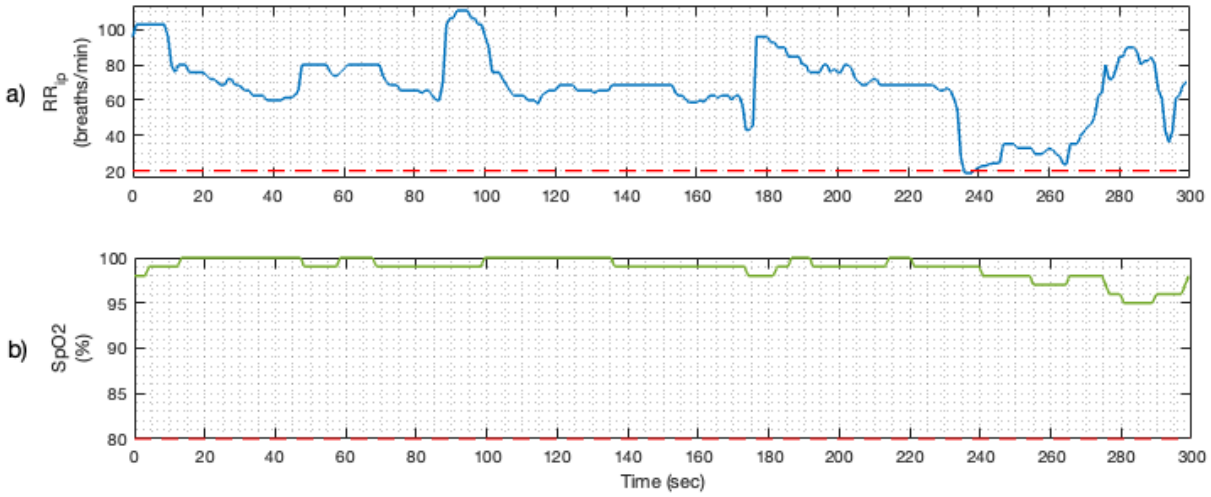


Figure 5.6: RR_{ip} and SpO_2 signals during a period of normal breathing *(b2)*. All signals remain within normal ranges, confirming the absence of any COBE event. a) RR_{ip} above 20 breaths/min and b) SpO_2 above 80%.

5.1.3 Inter-reviewer agreement and consensus

After each reviewer made their annotations, I analysed the inter-reviewer reliability of the annotations using Fleiss' kappa (κ). This coefficient is a statistical measure that assesses the reliability of agreement between two or more reviewers (or raters), removing expected agreement due to chance [333]. Fleiss' kappa takes into account both the observed agreement (P_o) and the expected agreement (P_e), providing a comprehensive measure of inter-rater reliability. The observed agreement (P_o) was computed as follows:

$$P_o = \frac{1}{N}\sum_{i=1}^{N} p_i \tag{5.2}$$

where N is the total number of observations (segments), R is the total number of reviewers, C is the number of annotation categories, and r_{ij} represents the number of reviewers assigning the j^{th} category to the i^{th} observation. Here, p_i represents the observed agreement for the i^{th} observation and is computed as:

$$p_i = \frac{1}{R(R-1)}\sum_{j=1}^{C} r_{ij}(r_{ij}-1) \tag{5.3}$$

The observed agreement represents the average proportion of agreement among reviewers across all observations. It ranges from 0 to 1, where larger values indicate greater agreement between reviewers. I calculated the expected agreement (P_e) as:

$$P_e = \sum_{j=1}^{C} p_j^2 \tag{5.4}$$

where p_j is the proportion of all observations that fall into the j-th category. The expected agreement represents the agreement that would be expected by chance alone, assuming that the reviewers' annotations were independent of each other. Finally, I computed the Fleiss' kappa (κ) using the formula:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{5.5}$$

The value of κ ranges from -1 to 1, with a value of 1 signifying complete agreement among reviewers, 0 indicating no better than chance agreement, and less than 0 denoting less agreement than expected by chance.

For Phase A, I obtained a Fleiss' Kappa value of 0.80, indicating substantial agreement beyond chance among the reviewers based on the guidelines from Altman [334]. Additionally, a P_o value of 0.892 suggests that around 89.2% of the annotations agreed, while the expected agreement (P_e) of 0.461 implies that approximately 46.1% of the annotations would agree by chance alone. These results demonstrate consistent annotation agreement among the reviewers, indicating high reliability in the dataset annotations. For Phase B, I recorded a Fleiss' Kappa value of 0.57, reflecting moderate agreement between annotators based on the guidelines from Altman [334]. The observed agreement (P_o) was 0.849, suggesting that around 84.9% of the annotations agreed, while the expected agreement (P_e) was 0.645. For both annotation phases, segments for which reviewers did not reach majority agreement were excluded from the final dataset, rather than resolved through retrospective consensus discussion, because they were considered too ambiguous for reliable labelling. Segments affected by artefact or poor signal quality that prevented reliable identification of COBE events were similarly discarded to preserve annotation reliability.

I constructed the final dataset, comprising clinical signals that captured a total of 346 COBE events (positive events) and 608 periods of normal breathing (negative events) from both Phases A and B. This carefully curated dataset provides a comprehensive representation of COBE events in the NICU dataset for further analysis and interpretation.

5.2 Defining analysis windows

After the manual annotations were completed, I defined 80-second windows around each annotated event to facilitate subsequent analyses. For desaturation-associated COBE episodes (*a1*), the window included 60 seconds preceding and 20 seconds following the beginning of oxygen desaturation, capturing both the desaturation and the associated decrease in respiration (RR_{ip} ≤20 breaths/min). Desaturation events without a respiratory decrease (*a2*) were assigned a similar 80s window spanning 60 seconds before and 20 seconds after the onset of desaturation.

For non-desaturation COBE episodes (*b1*), the 80-second window extended from 60 seconds prior to the start of the respiratory decrease to 20 seconds after. Periods of normal breathing (*b2*) were assigned an 80s window around the midpoint of the 5-minute segment.

5.3 Discussion

In this chapter, I constructed a dedicated COBE dataset from the Neonatal HDU dataset by combining automated event identification based on the recomputed RR from Chapter 4, and detailed manual annotations by three independent reviewers. The use of RR and SpO_2 enabled reviewers to accurately identify episodes of cessation of breathing.

The high inter-reviewer agreement observed in Phase A (κ = 0.80) demonstrates that the annotation methodology is reliable for events associated with desaturation. Phase B, which included non-desaturation events, exhibited moderate agreement (κ = 0.57), reflecting the inherent ambiguity in identifying subtle respiratory pauses that do not lead to immediate physiological consequences. Overall, the consensus-based approach ensured that the resulting dataset captures both positive COBE events and representative periods of normal breathing. The final dataset comprises 346 COBE episodes and 608 negative segments, providing a useful resource for the development and assessment of algorithms and methods aimed at the detection and understanding of COBE episodes in pre-term infants.

One limitation of the dataset is its relatively small size, which can be insufficient for training data-intensive algorithms. Additionally, some annotations, particularly in non-desaturation episodes, were subject to inter-reviewer variability, reflecting the challenges in consistently identifying subtle or borderline respiratory pauses. However, this dataset represents one of the biggest neonatal datasets with video monitoring. Additionally, the inclusion of both desaturation-associated and non-desaturation events ensures that the dataset represents a realistic range of neonatal respiratory behaviours observed in the NICU. In summary, this chapter presents a rigorously curated, clinically validated COBE dataset that serves as a foundational resource for the algorithmic detection of COBE episodes. The dataset bridges the gap between clinical expertise and automated analysis, enabling the development of reliable detection methods that can improve monitoring and intervention in pre-term infants.

Chapter 6

Detection of COBE using physiological signals in the NICU

6.1 Introduction

Detection of COBE episodes in NICU patients relies on traditional physiological signals such as the ECG, the PPG, and the IP. These signals are continuously recorded at the bedside by patient monitors, providing a practical foundation for automatic, data-driven detection of COBE events.

While the IP provides a direct measure of thoracic motion, respiration can also be inferred indirectly from the ECG and PPG. In the ECG, respiratory activity modulates heart rate and beat morphology, a phenomenon exploited by ECG-derived respiration (EDR) methods [164, 182, 335, 336]. Similarly, the PPG signal carries a respiratory component because breathing influences venous return and peripheral blood volume. All three signals can be affected by shallow or irregular breathing and motion artefacts. EDR is particularly sensitive to immature or variable cardiac–respiratory coupling in pre-term infants [337], whereas PPG-derived respiration may be unreliable under conditions of low peripheral perfusion [335]. Despite these limitations, combining these signals can reduce the impact of signal-specific artefacts and provide complementary information that improves the accuracy of COBE detection in NICU patients.

6.2 Overview of the methodology

In this chapter, I develop and evaluate three deep learning approaches for the automatic and continuous detection of COBE episodes using the respiratory components extracted from standard physiological signals: ECG-derived respiration (EDR), PPG envelope, and IP. The proposed methods include a baseline Convolutional Neural Network (CNN), Residual Networks (ResNets) (Section 6.6.1), and a ConvNeXt model (Section 6.7.1). I train and evaluate each method using identical pre-processing steps and consistent evaluation metrics to enable fair comparison. I conduct all training and evaluation in a Python environment using PyTorch, allowing for reproducibility and easy adaptation of the workflow for future development. Training is performed on a workstation equipped with an NVIDIA RTX A4000 GPU, providing the necessary computational power for all deep learning tasks.

Figure 6.1 presents a schematic overview of this workflow. I begin by pre-processing physiological signals to remove noise, reduce artefacts and enhance the respiratory patterns, ensuring that all subsequent models are trained on high-quality inputs. I then segment the signals into fixed-length windows suitable for model input. For model training and evaluation, I use a five-fold cross-validation approach with a hold-out test set to assess model generalisability. I train each architecture in both single-signal and multimodal configurations, allowing the models to learn either from individual respiratory signals (EDR, PPG envelope, or IP) or from multimodal combinations. For multimodal experiments, I investigated both early- and late-fusion strategies. Early fusion combines multiple physiological signals at the input stage before feature extraction, allowing the network to learn joint representations directly from the combined signals. In contrast, late fusion processes each signal through separate feature extraction branches before combining the resulting features later in the network. Although signals were processed separately initially, the multimodal network was still trained end-to-end using a shared classification loss, allowing gradients to jointly optimise all signal-specific branches during training.

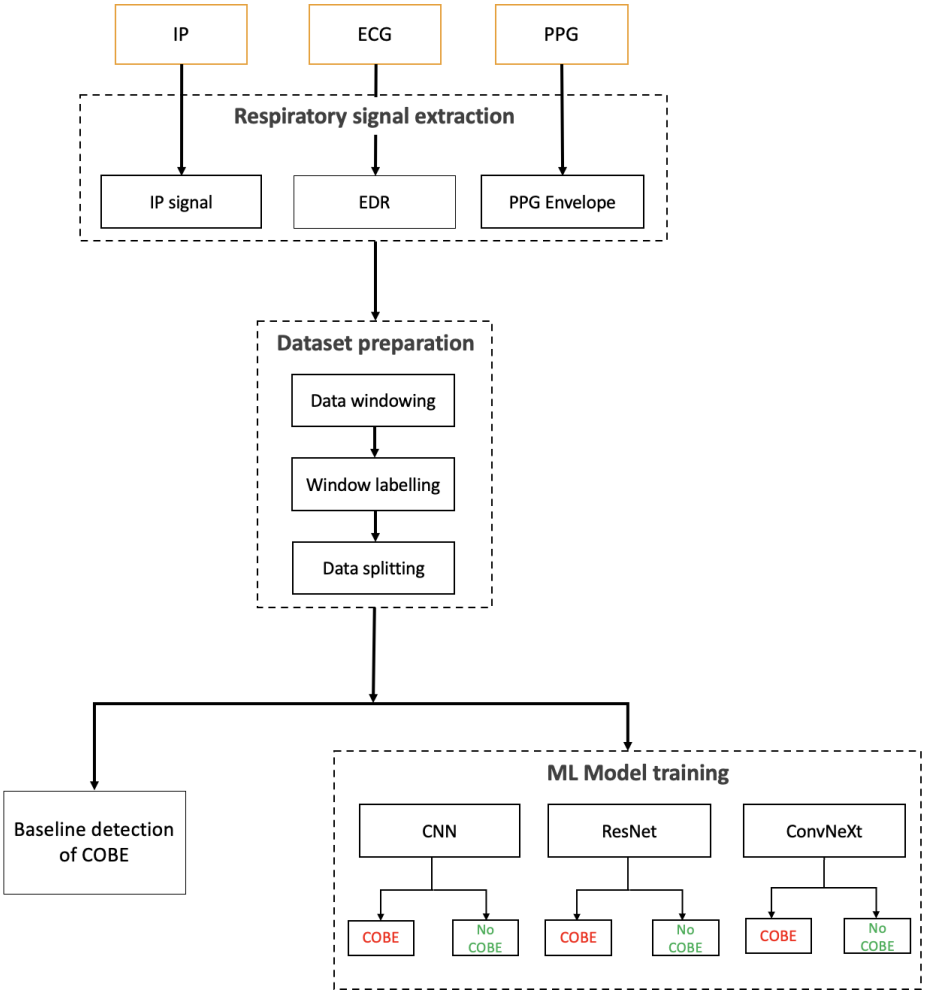


Figure 6.1: Schematic overview of the workflow for COBE episode detection using traditional physiological signals as inputs to deep learning models.

6.3 Respiratory signal extraction

Pre-processing physiological signals is a critical step for accurately classifying COBE episodes using machine learning algorithms. These steps enhance signal quality and reduce noise, ensuring the model is learning relevant physiological features from the training data [338]. I applied the pre-processing pipeline described in this section to the IP, ECG, and PPG signals in the labelled COBE dataset introduced in Chapter 5. I resampled all signals to 60 Hz, which preserves the frequency content relevant for respiratory-induced variations, ensures alignment across signals, and avoids interference from UK mains at 50 Hz. Figure 6.2 shows an overview of the steps taken to pre-process each of the signals.

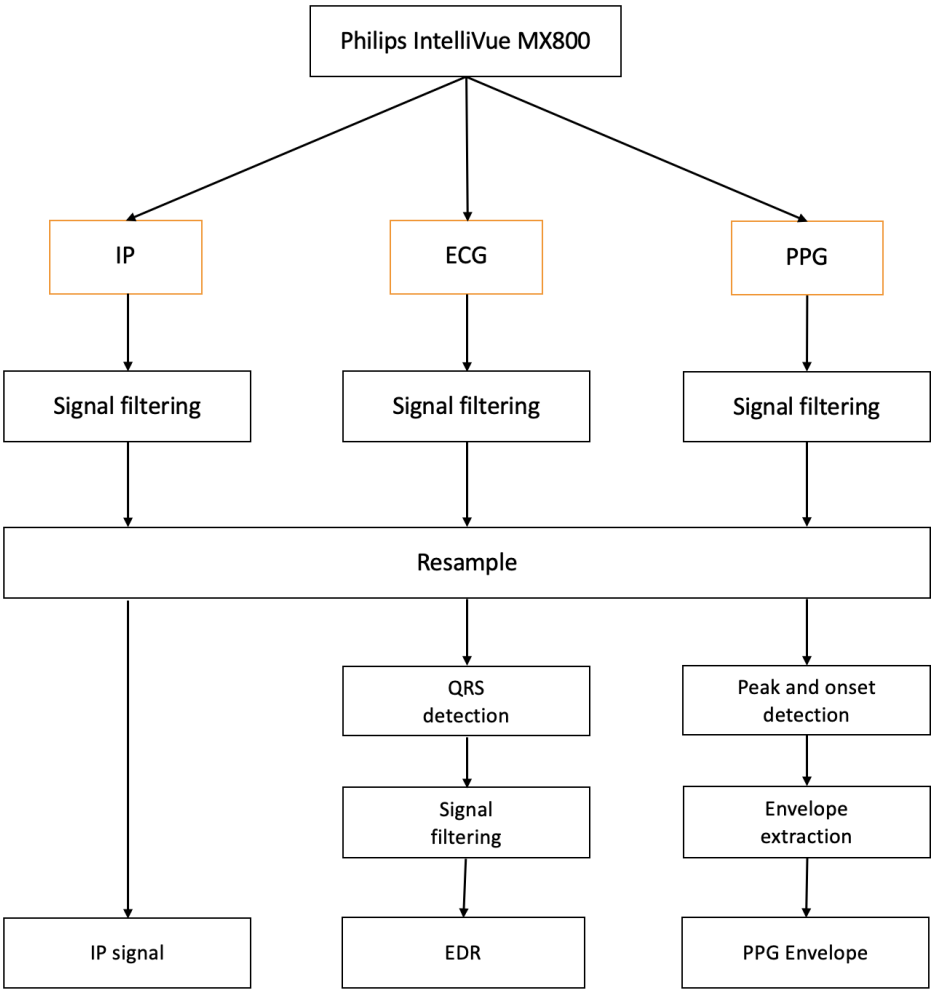


Figure 6.2: Block diagram outlining respiratory signal extraction from the IP, ECG and PPG signals. All physiological signals were acquired simultaneously from the same Philips bedside monitor.

6.3.1 IP signal

To pre-process the IP signal, I applied an 8th-order high-pass Butterworth IIR filter at 0.08 Hz (equivalent to 4.8 breaths/min) to remove slow baseline drift, followed by a 6th-order low-pass Butterworth IIR filter at 2.75 Hz (equivalent to 165 breaths/min) to suppress high-frequency components outside the respiratory range. I then resampled the filtered signal from 62.5 Hz to 60 Hz, avoiding interference from UK mains at 50Hz while preserving respiratory-related frequency content. The resulting IP waveform served directly as an input feature for machine learning models.

6.3.2 ECG signal

The raw ECG contains baseline wander, high-frequency noise, and residual mains interference, which can affect QRS detection. I pre-processed the signal to remove these artefacts and extract the EDR waveform for use as an input feature in COBE classification.

I first applied an 8th-order high-pass Butterworth IIR filter with a cut-off frequency of 0.67 Hz, followed by a 2nd-order low-pass Butterworth IIR filter with a cut-off at 4 Hz. These cut-offs correspond to minimum and maximum heart rates of 40 beats/min and 240 beats/minute, respectively, ensuring that QRS complexes are preserved for accurate detection. I then resampled the filtered signal to 60 Hz to match the sampling rate of the IP signal, ensuring alignment across input signals for subsequent machine learning analyses.

I detected QRS complexes using the Pan–Tompkins algorithm, which identifies the rapid deflections corresponding to ventricular depolarisation [177, 318]. Following QRS detection, I located the R-peaks, which represent the highest points in the QRS complex. At each R-peak, I measured the ECG amplitude to obtain a discrete set of respiration-modulated amplitudes. I interpolated these amplitudes with cubic splines to generate a continuous EDR waveform and applied a low-pass filter to retain only the low-frequency components associated with respiratory modulation, suppressing high-frequency artefacts unrelated to breathing. This amplitude-based EDR reflects slow variations in the position of ECG electrodes relative to the cardiac vector due to respiratory-induced chest movements during a breathing cycle [180, 339, 340]. Figure 6.3 shows the processing steps from the raw ECG signal to the extracted EDR waveform.

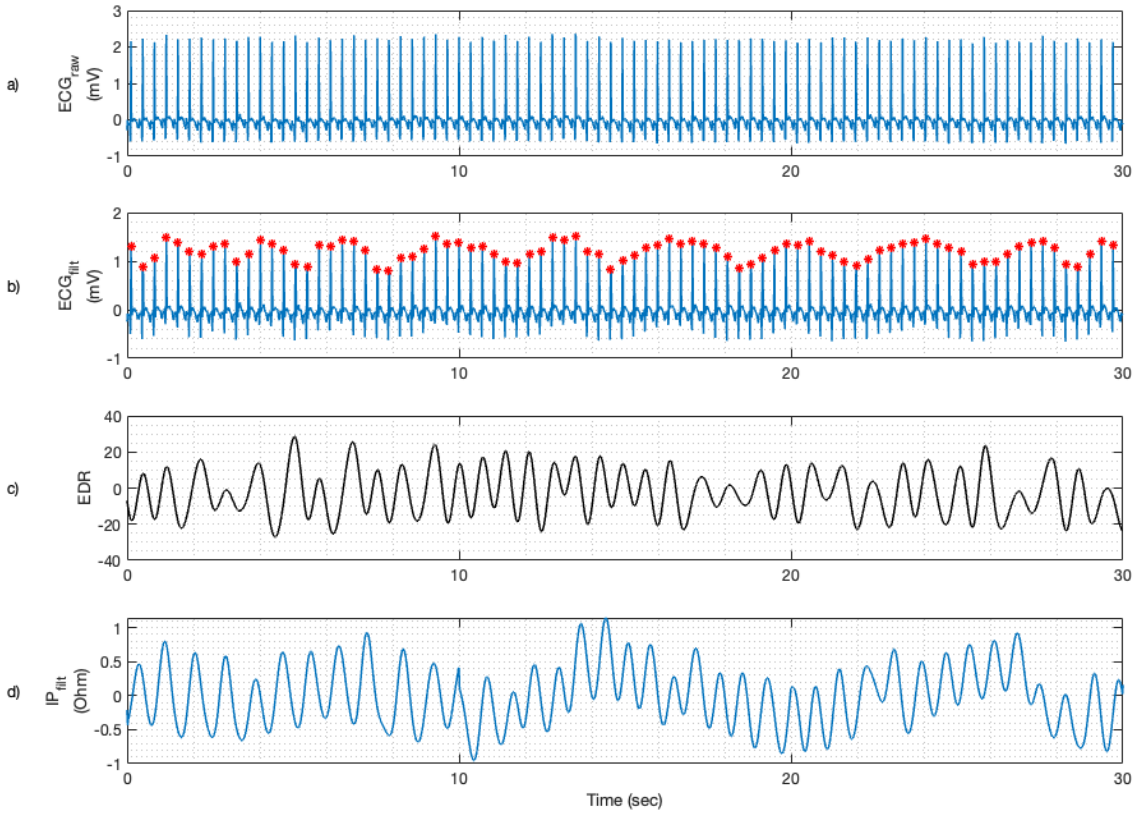


Figure 6.3: EDR waveform extracted from a 60-second ECG segment acquired from a 30.3 weeks old infant. a) Raw ECG; b) Filtered ECG waveform with detected R peaks; c) EDR signal extracted from the amplitudes of successive R peaks; d) IP signal during the same period as the ECG-derived respiratory signal.

6.3.3 PPG signal

In this study, the PPG signal was acquired from a pulse oximeter placed on the infant's foot. This raw PPG signal is often contaminated by various types of noise, such as baseline wander and motion artefacts due to the natural movement of the infant's legs, as well as high-frequency components unrelated to breathing. I applied an 8th-order high-pass Butterworth IIR and a 2nd-order Butterworth IIR low-pass filter with cut-off frequencies of 0.08 Hz (equivalent to 4.8 breaths/min) and 2.75 Hz (equivalent to 165 breaths/min), respectively, to isolate the respiratory frequency band. I then resampled the filtered signal to 60 Hz, which provides sufficient resolution to capture low-frequency respiratory-induced variations while aligning the PPG with IP and ECG signals.

After filtering and resampling, I derived the respiratory-induced variations (RIV) in the PPG by detecting peaks and troughs using the MAC peak detection algorithm (Section 4.5). I then extracted the signal envelope through cubic spline interpolation of the detected peaks, capturing amplitude modulation and baseline wander associated with respiration. The resulting RIV waveform provides a low-frequency representation of blood volume and tissue perfusion changes caused by respiration, suitable as an input

feature for machine learning models. Figure 6.4 shows the processing of a 60-second PPG segment, showing the raw signal, filtered waveform with detected peaks and troughs, and the extracted envelope representing the respiratory-induced variations.

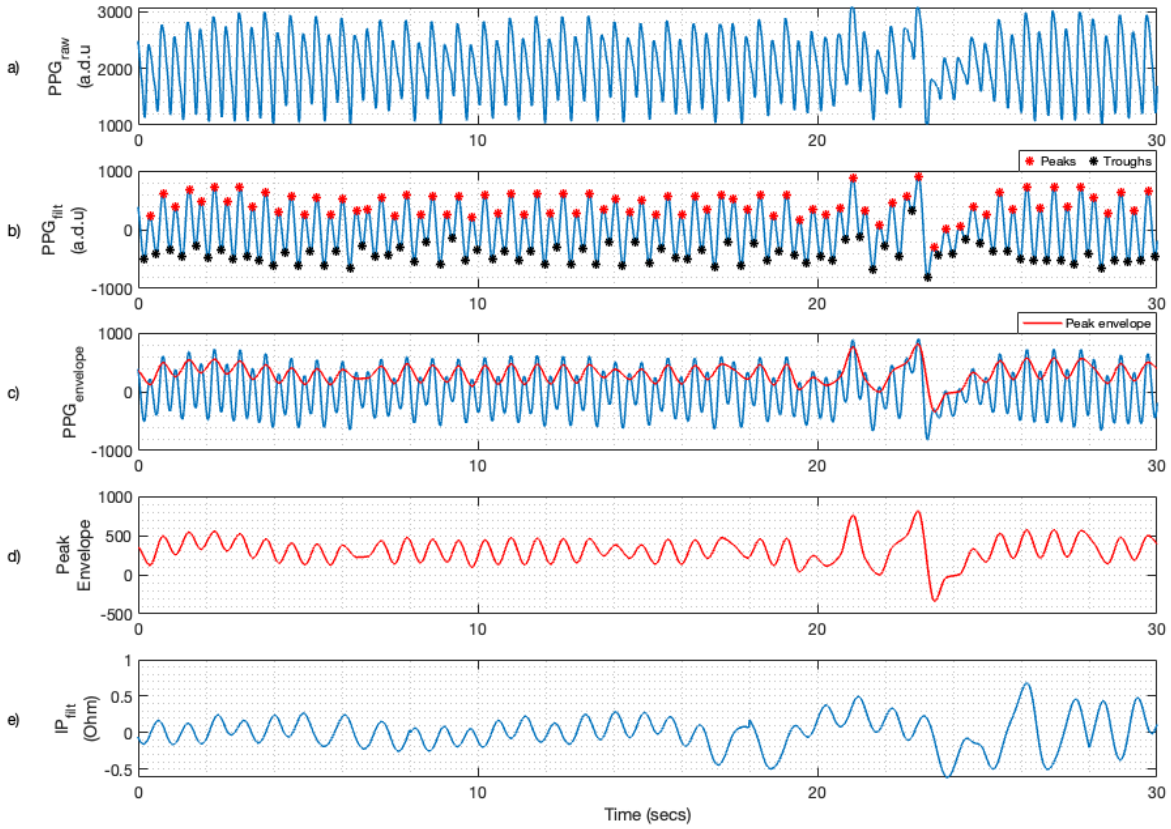


Figure 6.4: PPG envelope waveform extracted from a 30-second PPG segment acquired from a 31.9-week old infant. a) Raw PPG; b) Filtered PPG waveform with detected peaks and troughs; c) Envelope of the signal derived from the peak; d) Peak Envelope representing the extracted respiratory signal; e) IP signal during the same period as the PPG-derived respiratory signal.

Figure 6.5 shows representative respiratory waveforms derived from the IP, ECG, and PPG signals after preprocessing. Because all physiological signals were acquired simultaneously from the same Philips bedside monitoring system, no additional synchronisation between signals was required. The signals demonstrate similar respiratory periodicity and temporal behaviour, although small temporal offsets and morphology differences are visible between modalities, likely reflecting differences in the underlying physiological mechanisms and signal extraction processes.

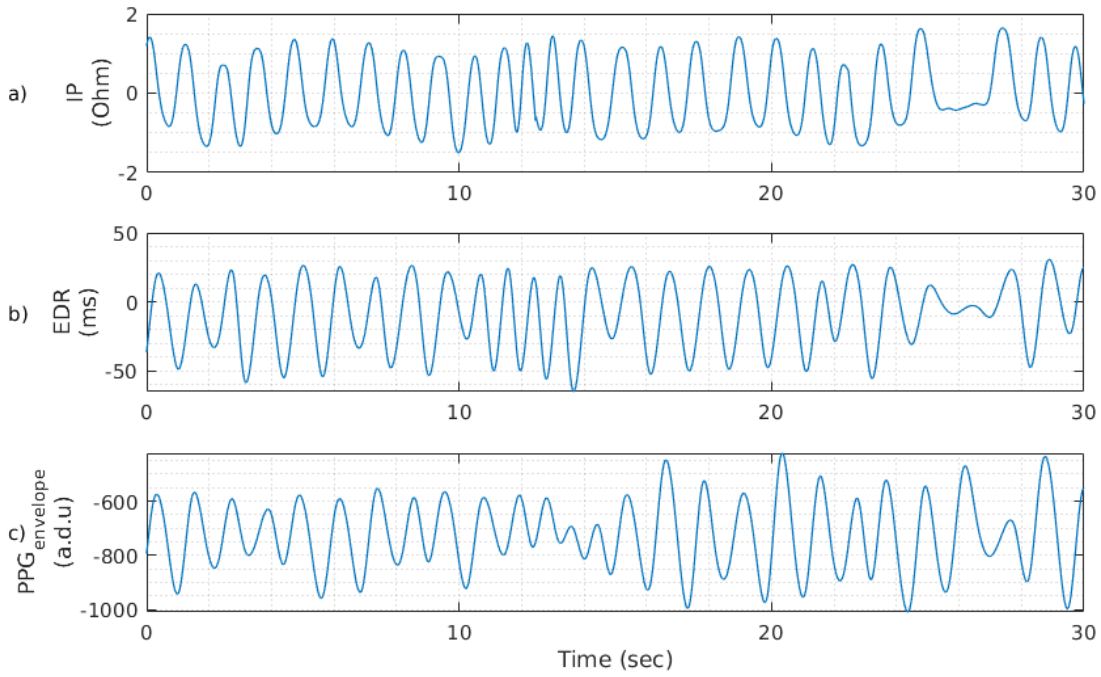


Figure 6.5: Representative respiratory waveforms extracted from physiological signals acquired from a pre-term infant. The signals demonstrate similar respiratory cycles across modalities despite minor morphology differences. a) IP waveform; b) ECG-derived respiration (EDR) waveform extracted from the ECG signal; c) PPG envelope waveform representing respiratory-induced variations extracted from the PPG signal.

6.4 Construction of the machine learning dataset

The COBE dataset described in Chapter 5 comprises 954 manually annotated 80-second segments, including 346 segments containing annotated COBE episodes and 608 segments of normal breathing. Each 80-second segment captures the entire context of an event, ensuring that the entire COBE episode is contained within the window. This section outlines the steps I took to prepare the labelled 80-second segments for input into machine learning models. Specifically, the process involved subdividing the clinically annotated periods into shorter overlapping windows to generate finer-grained labels suitable for data-driven classification of COBE episodes.

6.4.1 Data windowing

To prepare the data for machine learning, I divided each 80-second segment into overlapping 20-second windows with a 10-second step. These 20-second overlapping windows ensure that a minimum duration COBE episode (10 seconds) will be fully captured in at least one window. This produces 7 windows per 80-second segment.

For the 346 COBE segments, windows containing at least 10 seconds of RR below 20 breaths/min retained the positive label determined from the second-by-second expert annotations. Windows with less

than 10 seconds of COBE were labelled as negative, including all windows from the normal breathing segments.

The resulting training dataset comprises 954 annotated 80-second segments, corresponding to 6,678 overlapping 20-second windows across 24 pre-term infants.

6.4.2 Data splitting

I trained, validated, and tested my machine learning models using five-fold cross-validation with a held-out test set. To achieve this, I divided my labelled dataset into six subsets, each containing data from only 4 pre-term infants. When splitting the data, I ensured each of the subsets was as demographically balanced as possible, considering average infant gestational age, the number of positive and negative windows, and gender distribution.

During cross-validation, I trained each model on four subsets (80%) and validated it on the remaining subset (20%). I repeated this process five times so that each cross-validation subset served once as the validation set. After completing cross-validation, I retrained the best-performing model configuration, identified from cross-validation, using the combined data from all five cross-validation subsets and tested the resulting model on the independent test set, not used for training or validation.

6.5 A CNN for COBE episode detection

The application of CNNs to neonatal cardio-respiratory monitoring has emerged as a promising approach for the detection of COBE episodes in pre-term infants. While most prior research on sleep apnoea detection has focused on adult populations [24, 269, 338], these studies have demonstrated the potential of CNNs to automatically extract meaningful features from raw physiological signals in apnoea.

Motivated by these advances, this section presents the implementation and adaptation of a CNN architecture for the classification of COBE episodes in pre-term infants in the NICU dataset. The selected baseline model is based on the work by Urtnasan *et al.* [269], which was originally developed for apnoea detection in adult ECG recordings from the Physionet Challenge [341]. My work extends this architecture to accommodate multi-signal neonatal data, specifically, IP, ECG, and PPG. It addresses the unique physiological and signal characteristics of the pre-term infant population in the NICU setting.

6.5.1 Network architecture

The architecture proposed by Urtnasan *et al.* [269] is a six-layer CNN (plus two subsequent classification layers) designed for the detection of apnoea in adults. It comprises an input layer, followed by six

convolutional blocks (each with ReLU activation and max-pooling) and two fully connected layers for classification.

I initially replicated this 8-layer architecture to establish a baseline on my COBE dataset. Layers 1 and 2 each consist of 20 convolutional kernels (filters) of size 50 ×1, while layers 3 and 4 comprise 24 kernels of size 30 ×1. Layer 5 applies 24 kernels of size 10 ×1, and layer 6 has 12 kernels of size 10 ×1. All layers use ReLU activation functions, and dropout is applied starting from the second layer. Figure 6.6 provides a graphical representation of this baseline architecture.

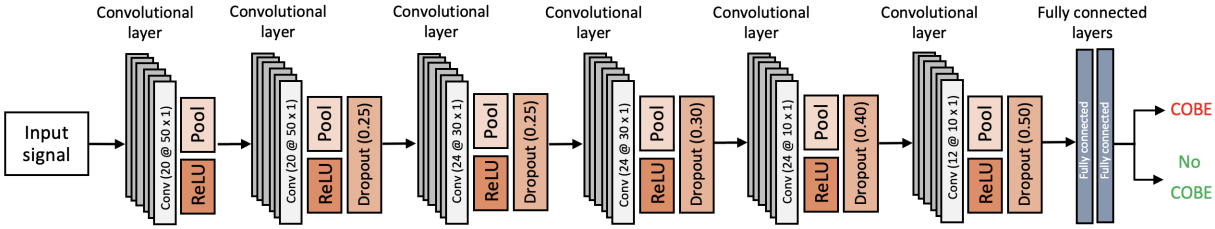


Figure 6.6: Architecture of the CNN model modified for automated COBE episode detection. The model processes 1D physiological signals (IP, ECG, or PPG) through a series of convolutional layers with ReLU activations, max pooling, and dropout, followed by fully connected layers for binary classification.

To better accommodate the distinct characteristics of NICU physiological signals, I modified the original architecture. Modifications included adjusting learning rates for each signal type, increasing dropout progressively across layers for regularisation, replacing max pooling with average pooling, selectively removing pooling operations in certain layers, and scaling kernel sizes to capture both fine-grained and broader temporal features. These changes enabled the network to learn more complex patterns from IP, ECG, and PPG signals while maintaining stable convergence and mitigating overfitting.

I subsequently trained and evaluated the modified architecture on the COBE dataset using cross-validation on single- and multi-signal inputs.

6.5.2 Network training

Using the architecture described in Section 6.5.1,I used preprocessed IP, PPG, and ECG signals as inputs. For single-signal inputs, each window is represented as a one-dimensional array of length 1,200 (corresponding to 20 s × 60 Hz). When multiple signals are stacked, the input has size $1,200 \times C$, where C is the number of concurrent signals. I used binary cross-entropy (BCE) as the loss function for binary classification, defining this loss function as:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\Big[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\Big], \tag{6.1}$$

where $y_i \in \{0,1\}$ is the ground-truth label for sample i, $\hat{y}_i \in [0,1]$ is the predicted probability, and N is the number of training samples.

To address class imbalance, I applied class weights by multiplying the loss associated with each sample by a weight inversely proportional to the class frequency. This penalises misclassification of the minority class (COBE events) more heavily and encourages the network to learn balanced representations. I optimised the model using the Adam optimiser, which provides adaptive learning rates for efficient convergence. I tuned these learning rates individually for each input signal to account for differences in temporal and spectral characteristics: 1.0×10^{-4} for IP, 1.0×10^{-5} for EDR and 5.0×10^{-4} for PPG envelope. To further regularise the network and reduce overfitting, I progressively increased dropout rates across layers (0.25, 0.25, 0.3, 0.4, and 0.5).

I trained the network on each signal using early stopping for up to 50 epochs, which was sufficient for convergence without overfitting. I used a batch size of 16. This training procedure used separate training, validation, and held-out test sets. During cross-validation, overfitting was assessed by monitoring validation performance across epochs. If validation loss did not improve for five consecutive epochs while training performance continued to improve, training was terminated using early stopping. Figure 6.7 shows representative training and validation loss curves for a multimodal CNN model during cross-validation. Both training and validation losses decrease and stabilise over successive epochs, demonstrating stable convergence behaviour and illustrating how validation performance was monitored to detect overfitting and guide early stopping. Additional measures to mitigate overfitting included dropout regularisation, class-weighted training, and final evaluation on an independent held-out test set not used during model optimisation.

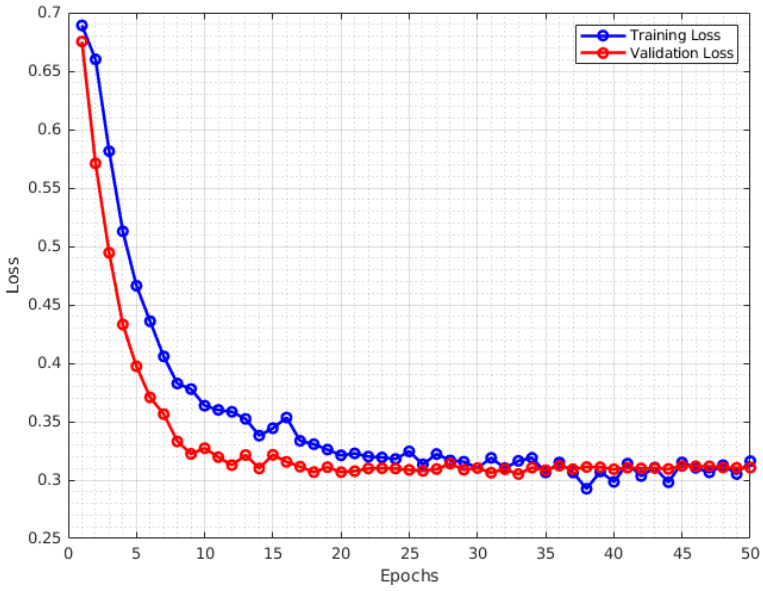


Figure 6.7: Representative training and validation loss curves from one fold of a multimodal CNN during cross-validation. Both losses decrease and stabilise over successive epochs, indicating stable convergence. The minimum validation loss occurred at epoch 33.

For the final model training on the full cross-validation dataset, I fixed the number of epochs to the average number at which early stopping occurred in the preliminary cross-validation runs. This ensured that the model was trained for a consistent number of epochs while preserving the regularisation effect of early stopping. Throughout, I systematically applied additional design choices, including kernel sizes, network depth, and pooling strategies (average versus max pooling, and selective removal of pooling layers), across all input signals. I carried out hyperparameter optimisation using random search, which efficiently explores a large parameter space [342].

6.6 A ResNet for COBE episode detection

I adapted the ResNet architecture for 1D time-series data by replacing all 2D convolutional, pooling, and batch normalisation layers with their 1D equivalents. I also implemented modifications to the network, as detailed in subsection 6.6.1, to improve the model's effectiveness in handling the unique characteristics of 1D signal data.

6.6.1 Network architecture

I reduced the number of feature channels in each ResNet block from the conventional [64, 128, 256, 512] to [32, 32, 64, 64] following work by Carter *et.al* [343]. This adjustment reflects the lower information density found in 1D signals compared to 2D images, where the spatial context is significantly richer. Instead of using standard average pooling before the fully connected layer, I applied adaptive average pooling. In conventional 2D ResNets, average pooling is used to reduce the feature map by averaging over its height and width, and this helps reduce the data size before classification. When applied to 1D data such as the IP signal, this operation compresses the temporal dimensions, potentially discarding useful timing information. Adaptive average pooling mitigates this by preserving a fixed number of regions within the respiratory feature map before classification. I implemented adaptive average pooling by dividing the feature map into several output regions and averaging values within each region, rather than compressing the entire waveform into a single pooled representation. This allowed the classifier to retain coarse information from different parts of the respiratory waveform while still producing a consistent feature representation for classification.

To further improve classification capacity, I extended the model's classifier by substituting the single fully connected layer with a multi-layer perceptron (MLP). This change helps the model better understand and capture non-linear patterns between the pooled features and the output labels. It also allows the model to learn more from the different time segments kept by the adaptive pooling, making its decisions more informed and accurate. The overall structure (shown in figure 6.8) maintains the key principle of residual

learning, skip connections are consistently preserved, but these adjustments ensure the architecture fits the specific needs of 1D time-series analysis.

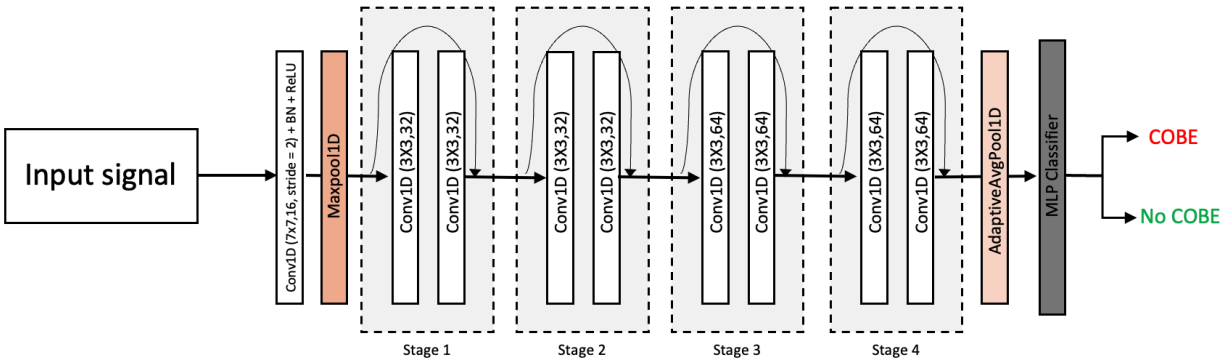


Figure 6.8: The architecture of the ResNet model I designed for automated COBE episode detection. In the simplified schematic, each of the four stages shows two convolutional layers. For ResNet-18 and ResNet-34, the actual architectures follow (2, 2, 2, 2) and (3, 4, 6, 3) basic blocks per stage, respectively. Since each basic block contains two convolutional layers, this results in 4 to 12 convolutional layers per stage. ResNet-50 also follows a (3, 4, 6, 3) configuration, but uses bottleneck blocks instead, each comprising three convolutional layers (1×1, 3×3, 1×1). The same network is applied separately to each input signal (IP, EDR, and PPG envelope).

6.6.2 Network training

I trained several standard variants of the adapted ResNet (ResNet-18, ResNet-34, and ResNet-50) for up to 100 epochs with early stopping using the Adam optimiser. To account for potential sensitivity to initial learning rates when training CNNs, I sampled values from a logarithmic scale between 9×10^{-5} and 1×10^{-4}. I selected this range based on preliminary experiments that showed larger learning rates could cause unstable training or prevent convergence. In contrast, smaller values slowed convergence without significantly improving generalisation. By sampling from a logarithmic scale, I explored greater detail in this region, where small changes in the learning rate can disproportionately affect model stability and overall performance. I applied a fixed exponential decay factor of 0.9 at every epoch to encourage stable convergence across folds.

6.7 A ConvNeXt for COBE episode detection

Although ConvNeXt was primarily developed for and evaluated on 2D image classification tasks, recent research has begun to explore its application to 1D domains such as physiological signal classification [344, 345]. My work in this section presents a modified 1D ConvNeXt pipeline for the detection of COBE episodes using unimodal and multimodal physiological signals. My initial experiments involved unimodal training on IP,ECG, and PPG signals, using the default ConvNeXt architecture without any structural changes. Results from these experiments suggest that ConvNeXt can achieve better performance than conventional CNN-based models for COBE episode classification, even without the use of

multimodal fusion or extensive architectural tuning. This is the first application of a 1D ConvNeXt model for the classification of COBE episodes in pre-term infants, highlighting its potential as an alternative to existing CNN-based approaches in the context of sleep apnoea detection.

6.7.1 Network architecture

The original ConvNeXt architecture is tailored for 2D image classification tasks. When adapting it for 1D temporal signals, such as IP, ECG, and PPG, certain components require modification or re-evaluation. In particular, the 7×7 depthwise convolution kernel, effective for spatial feature extraction in images, may be suboptimal for capturing temporal dependencies in 1D signals. To address this, I experimented with increasing kernel sizes and reducing the width of the feature set to better match the lower dimensionality of 1D signals. I retained all other key structural elements, such as the use of Gaussian Error Linear Unit (GELU) non-linearities and inverted bottlenecks from the default ConvNeXt design, and applied them uniformly across unimodal and multimodal input configurations. The resulting 1D ConvNeXt architecture, adapted for temporal physiological signals, is shown in figure 6.9.

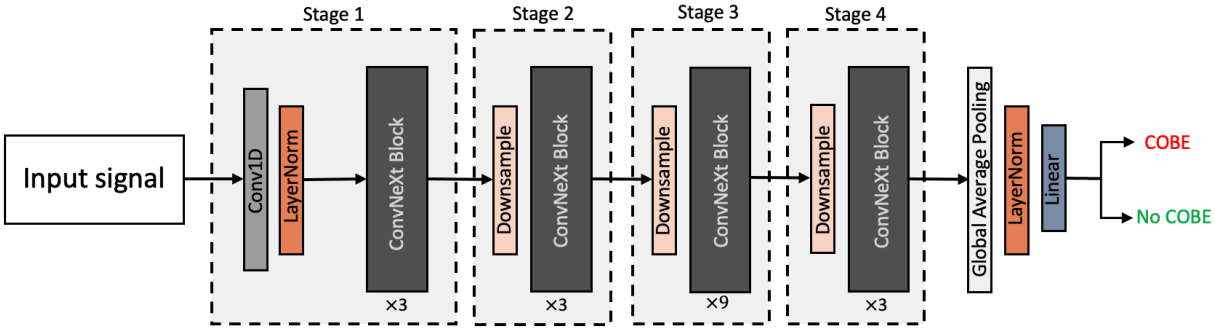


Figure 6.9: Architecture of the minimally modified 1D ConvNeXt model for automated COBE episode detection. The network processes 1D physiological signals (IP, EDR, and PPG envelope) through an initial 1D convolution and LayerNorm (commonly referred to as the “stem”), followed by four stages of ConvNeXt blocks with intermediate downsampling. The final stage output is aggregated via global average pooling, normalised with LayerNorm, and classified with a linear layer into COBE or No COBE.

6.7.2 Network training

I trained ConvNeXt using the AdamW optimiser, with learning rates sampled logarithmically between 9×10^{-5} and 1×10^{-4}. I also implemented a cosine learning rate schedule with a 5-epoch warmup period, allowing the model to gradually adapt before decaying the learning rate for stable convergence. To improve generalisation and avoid overfitting, I applied several regularisation strategies. These include exponential moving averages (EMA) applied to the model weights to stabilise parameter updates across batches and a weight decay value of 0.05 to penalise large weights. I also used label smoothing with a factor of 0.1 to prevent overconfident predictions.

The overall training methodology I used was consistent with that used for the CNN and ResNet models. This included 5-fold cross-validation, evaluation on a separate held-out test set, and the use of class-weighted cross-entropy loss to address class imbalance. Model performance was assessed primarily using (balanced) Accuracy to ensure fair evaluation across both COBE and non-COBE event classes.

6.8 Results

In the following sections, I present the results of a simple threshold-based baseline method for COBE detection using respiratory rate derived from the IP signal. I then present the results for the deep learning architectures studied in this work: the conventional CNN, the adapted 1D ResNet, and the 1D ConvNeXt. For each model, I report performance across both single-signal and multi-signal input configurations. Summary tables highlight the best-performing models for each input modality and fusion strategy. Detailed results, including per-fold cross-validation metrics, network depth comparisons, and unimodal versus multimodal performance, are provided in Appendices D, E, and F.

6.8.1 Model evaluation

I assessed the performance of all architectures using several metrics, including True Positive Rate (TPR), False Positive Rate (FPR), Precision, F1 score, Cohen's Kappa, per-class accuracies (normal and apnoea), and Balanced Accuracy. For cross-validation experiments, I report results as the mean ± standard deviation over five folds. During cross-validation, I primarily selected models with the highest mean Balanced Accuracy, while also considering models with lower variance, which typically generalise better. In some cases, I considered a trade-off between slightly higher accuracy and lower variance.

I classified outcomes as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), which form the basis for all subsequent performance metrics. Using these, I defined the key evaluation metrics as follows:

The True Positive Rate (TPR), also referred to as Sensitivity, measures the proportion of actual positive windows correctly identified:

$$TPR = \frac{TP}{TP + FN} \tag{6.2}$$

Precision quantifies the proportion of predicted positives that are actually positive:

$$Precision = \frac{TP}{TP + FP} \tag{6.3}$$

The False Positive Rate (FPR) captures the proportion of negative windows incorrectly classified as positive:

$$FPR = \frac{FP}{FP + TN} \tag{6.4}$$

The $F1$ score, the harmonic mean of Precision and TPR, offers a single measure that balances false positives and false negatives:

$$F1 = 2 \times \frac{TPR \times Precision}{Precision + TPR} \tag{6.5}$$

I also calculated the Balanced Accuracy, which averages the TPR for each class (COBE or normal), ensuring equal weight is given to both classes, regardless of class imbalance. I defined Balanced Accuracy as:

$$Balanced\ Accuracy = \frac{TPR_{COBE} + TPR_{no\ COBE}}{2} \times 100\% \tag{6.6}$$

Balanced Accuracy was selected as the primary evaluation metric because the dataset contains substantially more non-COBE than COBE windows, and standard accuracy could therefore overestimate performance by favouring the majority class. However, in the NICU setting, missed COBE episodes (false negatives) may carry greater clinical risk than false alarms because delayed or missed detection could postpone necessary clinical intervention. Balanced Accuracy was therefore interpreted alongside sensitivity (TPR) and F1 score to ensure that model performance on COBE detection itself remained clinically meaningful.

6.8.2 Baseline detection of COBE using IP-derived RR

To provide a baseline reference for evaluating my machine learning models, I implemented the simplest approach possible for COBE detection using the RR recomputed from the IP signal. In this approach, I defined a COBE event as a sustained decrease in RR_{ip} to below 20 breaths/min for at least 20 seconds. I chose 20 seconds so that the COBE detection could be done on a per-20 second window basis, enabling a like-for-like comparison between this baseline and subsequent machine learning experiments. I evaluated this baseline method by comparing its predictions against the manual annotations, which incorporate both RR and SpO_2 information. The resulting 2 × 2 confusion matrix is presented in Table 6.1.

Table 6.1: Confusion matrix for baseline COBE detection using RR_{ip} compared against manual reference labels.

		RR_{ip}		
		COBE	**No COBE**	**Total**
Reference Labels	**COBE**	712 (0.693)	316 (0.307)	1028
	No COBE	0 (0.000)	5650 (1.000)	5650
	Total	712	5966	6678

The dataset used for training and evaluation contains approximately five non-COBE windows for every COBE window. This class imbalance biases the baseline detector towards the majority class (normal breathing), resulting in high non-COBE classification accuracy but lower sensitivity to COBE events. The confusion matrix in Table 6.1 reflects this behaviour. For this baseline method, the balanced accuracy, calculated using Equation 6.6, is 84.6%.

6.8.3 COBE episode detection using a CNN

Table 6.2 summarises the performance of the final CNN models for both single and multi-signal inputs on the unseen test set (16.7% of the total dataset). For single-signal training, the IP-based model achieved the highest overall performance, with an overall Accuracy of 86.8% and a TPR of 0.91. The ECG and PPG models showed lower performance. For multi-signal training using late fusion, combinations that included the IP signal consistently performed well. The IP+ PPG combination achieved the highest F1 score (0.73) and overall Accuracy (86.6%), while adding all three signals did not improve performance beyond the best two-signal combinations. Detailed results for all CNN configurations are provided in Appendix D.

Table 6.2: Summary of CNN model performance for signal and multi-signal configurations. Reported metrics are on the unseen test set (16.7% of data). Best performing models for single and multi-input configurations in bold.

Signal(s)	Performance Metrics							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
Individual signal training								
IP	**0.91**	**0.18**	**0.60**	**0.72**	**0.62**	**82.5**	**91.1**	**86.8**
ECG_{edr}	0.64	0.38	0.32	0.43	0.19	61.6	63.6	62.6
PPG_{env}	0.61	0.30	0.36	0.46	0.24	69.6	60.7	65.1
Multi-signal training: Late fusion								
IP + ECG	0.89	0.17	0.60	0.72	0.62	83.2	89.5	86.3
IP + PPG	**0.88**	**0.15**	**0.63**	**0.73**	**0.64**	**85.2**	**87.9**	**86.6**
ECG + PPG	0.64	0.26	0.41	0.50	0.32	73.7	64.2	69.0
IP + ECG + PPG	0.88	0.18	0.58	0.70	0.59	81.7	88.5	85.1

6.8.4 COBE episode detection using a ResNet

Table 6.3 summarises the performance of the best ResNet models for both single and multi-signal inputs on the unseen test set. Among individual signals, the IP-based ResNet-34 model achieved the highest overall accuracy (87.7%) and TPR (0.93), reflecting strong sensitivity to COBE episodes. Models trained on ECG and PPG signals performed less consistently, highlighting differences in information content and noise across modalities. For multi-signal configurations, late and early fusion models showed variable performance depending on the signal combination. The IP + PPG late fusion model achieved the highest F1 score (0.73) and overall Accuracy (87.1%), while other fusion models, including early fusion, performed comparably or worse than the best single-signal models. Extended results, including cross-validation metrics and performance across all ResNet depths, are provided in Appendix E.

Table 6.3: Summary of final ResNet model performance on the unseen test set (16.7% of data). For each signal or combination, the best-performing ResNet architecture is shown.

Signal(s)	Resnet Depth	Performance Metrics							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
Individual signal training									
IP	**34**	**0.93**	**0.17**	**0.61**	**0.73**	**0.63**	**82.8**	**92.7**	**87.7**
ECG_{edr}	50	0.70	0.34	0.37	0.49	0.28	66.1	70.3	68.2
PPG_{env}	34	0.63	0.31	0.37	0.47	0.26	69.5	62.6	66.1
Multi-signal training: Late fusion									
IP + ECG	34	0.92	0.20	0.56	0.70	0.58	79.6	92.0	85.8
IP + PPG	**34**	**0.91**	**0.17**	**0.60**	**0.73**	**0.63**	**82.8**	**91.4**	**87.1**
ECG + PPG	34	0.50	0.18	0.44	0.47	0.30	81.9	49.8	65.9
IP + ECG + PPG	34	0.91	0.18	0.59	0.71	0.61	81.6	91.1	86.3
Multi-signal training: Early fusion									
IP + ECG + PPG	34	0.91	0.18	0.59	0.72	0.62	82.0	91.4	86.7

6.8.5 COBE episode detection using a ConvNeXt

Table 6.4 summarises the performance of the best ConvNeXt models for both single and multi-signal inputs on the unseen test set. Among individual signals, the IP-based ConvNeXt model with larger kernels achieved the highest overall Accuracy (88.0%) and TPR (0.91), demonstrating strong sensitivity to COBE episodes. Models trained on ECG and PPG signals showed more variable performance, reflecting differences in signal quality and the amount of discriminative information available in each modality. For multi-signal configurations, performance varied depending on the signal combination and fusion strategy. The IP + PPG late fusion model achieved the highest overall Accuracy (88.7%) and F1 score (0.75). Other fusion models, including early fusion, showed slightly reduced performance, with accuracies ranging from 86.1% to 87.0%. Complete ConvNeXt results are presented in Appendix F.

Table 6.4: Summary of final ConvNeXt model performance on the unseen test set (16.7% of data). For each signal or combination, the best-performing ConvNeXt architecture is shown.

Signal(s)	ConvNeXt variant	Performance Metrics							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
Individual signal training									
IP	**Larger Kernels**	**0.91**	**0.15**	**0.64**	**0.75**	**0.66**	**85.3**	**90.7**	**88.0**
ECG_{edr}	Default	0.77	0.37	0.37	0.50	0.29	62.7	76.7	69.7
PPG_{env}	Default	0.73	0.40	0.34	0.47	0.23	60.0	72.8	66.4
Multi-signal training: Late fusion									
IP + ECG	-	0.91	0.17	0.61	0.73	0.63	83.0	90.7	86.9
IP + PPG	-	**0.94**	**0.16**	**0.62**	**0.75**	**0.66**	**83.9**	**93.6**	**88.7**
ECG + PPG	-	0.80	0.37	0.38	0.51	0.31	62.9	79.6	71.2
IP + ECG + PPG	-	0.92	0.16	0.62	0.74	0.65	84.1	92.3	88.2
Multi-signal training: Early fusion									
IP + ECG + PPG	Default	0.88	0.14	0.64	0.74	0.65	85.6	87.9	86.7
IP + ECG + PPG	Larger Kernels	0.93	0.19	0.59	0.72	0.61	81.3	92.7	87.0
IP + ECG + PPG	Reduced Features	0.90	0.18	0.59	0.71	0.61	81.8	90.4	86.1

Table 6.5 summarises the best-performing models across all input modalities and architectures. While ConvNeXt models with larger kernels achieved the strongest single-signal performance with IP, multi-modal combinations provided only marginal improvements in test accuracy. In particular, the IP + PPG combination achieved slightly higher accuracy than the best single-signal model, whereas other multi-modal combinations offered minimal gains or similar performance. These results suggest that most of the discriminative information for COBE episode detection is already captured by IP, and adding additional signals provides limited benefit under the current modelling framework.

Table 6.5: Summary of best-performing models by input modality and signal combinations, including architecture variations and validation and test accuracies.

Input modality	Input Signal(s)	Architecture	Modification	Accuracy (%)	
				Validation	Test
Individual	IP	ConvNeXt	Larger Kernels	86.7 $\pm$ 1.9	88.0
	ECG_{edr}	ConvNeXt	default	71.9 $\pm$ 5.5	69.7
	ECG_{raw}	ResNet	ResNet-18	58.0 $\pm$ 7.0	65.0
	PPG_{env}	ConvNeXt	default	69.6 $\pm$ 3.6	66.4
	PPG_{raw}	ResNet	ResNet-34	66.0 $\pm$ 3.3	65.8
Multimodal	IP + ECG	ConvNeXt	Larger kernels (IP), default model (ECG)	85.1 $\pm$ 3.2	86.9
	IP + PPG	ConvNeXt	Larger kernels (IP), default model (PPG)	85.6 $\pm$ 3.3	88.7
	ECG + PPG	ConvNeXt	default	72.2 $\pm$ 4.8	71.2
	IP + ECG + PPG	ConvNeXt	Larger kernels (IP), default model (PPG and ECG)	85.4 $\pm$ 3.0	88.2

The comparisons across signal modalities, fusion strategies, and architectural modifications provide a brief ablation-style assessment of which components contributed most to performance. Detailed comparisons of pooling strategies, kernel sizes, network depth, architectural variants, and fusion strategies are provided in Appendices D-F.

6.9 Discussion

This chapter explored the detection of COBE episodes in pre-term infants using clinically acquired IP, PPG, and ECG signals. These signals are routinely used to monitor respiratory and cardiac function and are well established in adult sleep and apnoea research. However, their use in machine-learning approaches for neonatal and paediatric COBE episode detection has remained relatively under-explored. This is partly due to the practical and ethical challenges of collecting and annotating high-quality data from vulnerable infant populations. Neonatal signals are often noisy due to motion artefacts, poor sensor contact, and intrinsic physiological variability. Additionally, labelling COBE episodes accurately requires close clinical supervision, which limits the availability of large, annotated datasets. These constraints contribute to the scarcity of studies that apply deep learning methods in this setting.

Despite these challenges, my work demonstrates the feasibility of applying machine-learning architectures such as CNNs, ResNets, and ConvNeXt to clinical signals for COBE episode detection. The comparison between single-signal and multi-signal experiments across the different architectures shows how different input configurations influence performance. For the specific combination of signals tested, multimodal inputs provided only marginal gains over single-signal IP performance.

CNN-based classifiers provided a machine learning-based baseline for evaluating how signal-specific characteristics and architectural choices influence performance. These classifiers showed that the IP signal contains the most reliable features for detecting COBE episodes, while the EDR and PPG en-

velope signals contributed less consistent indicators of respiratory pauses. This reflects that the direct measurement of respiration captured by IP provides a clearer basis for identifying pauses in breathing, whereas the other signals offer indirect indicators of respiratory activity, which may limit their predictive value for detecting COBE episodes. In multimodal configurations, combining signals generally improved detection compared to using EDR or PPG envelope signals alone, but the improvements were small and not uniform across architectures or metrics. These results suggest that multimodal inputs may provide minor benefits in cases of poor IP signal quality but do not substantially change overall detection performance. These findings align with current NICU monitoring practices, which already use IP, PPG, and ECG signals for clinical observation. They suggest that automated detection systems could augment existing monitoring by providing more consistent and timely identification of COBE episodes, relying primarily on IP signals while optionally incorporating complementary signals to improve reliability when thoracic measurements are compromised. This approach supports early intervention without requiring additional sensors, enhancing the utility of currently available clinical data.

The adapted ResNet and ConvNeXt models provided deeper layer architectures with residual connections, allowing for more complex feature extraction. The IP signal consistently yielded the most informative features, reflecting its direct measurement of respiratory effort. By contrast, ECG and PPG-derived respiratory surrogates captured less pronounced COBE features but sometimes added complementary information when fused with IP. Across the held-out test set, the models achieved the best classification Accuracies of 86.6% for the CNN (IP + PPG, late fusion), 87.7% for ResNet on the IP signal alone, and 88.7% for ConvNeXt using IP and PPG-envelope signals in late fusion. Corresponding F1 scores were 0.73, 0.73, and 0.75, respectively. A limitation of this architectural comparison is that the ConvNeXt models were trained using a different optimisation strategy compared to the CNN and ResNet models, including AdamW optimisation, cosine learning-rate scheduling with warmup, exponential moving averages, and label smoothing. Although preliminary experiments suggested that Adam optimisation provided more stable convergence for the CNN and ResNet architectures on this dataset, differences in optimisation methodology may still partly contribute to the observed performance differences between architectures. Consequently, the superior performance of ConvNeXt cannot be attributed solely to architectural differences. Despite this optimisation-related limitation, the models still demonstrated that deep learning architectures can effectively differentiate between COBE and non-COBE episodes using widely available physiological signals. Both unimodal and fusion-based configurations indicate that model architecture and signal modality choice can influence classification performance. The results highlight the informative value of the IP signal, both alone and when combined with complementary signals. Unimodal PPG-envelope or EDR models yielded significantly lower accuracies, consistent with known limitations of these signals for

reliably capturing respiratory events.

The best-performing method presented in this chapter, a ConvNeXt fusion of IP and PPG signals (88.7%), outperformed the simple baseline approach (84.6%), highlighting the significant improvement achieved through learning-based fusion over straightforward thresholding of RR_{ip}. The baseline method, which is conceptually similar to the approach currently used in clinical monitors such as Philips, relies on a threshold-based detection of RR_{ip} and a minimum 20-second RR decrease criterion [23]. My baseline approach effectively identifies COBE events and produces very few false positives relative to the annotator labels.

Compared to the existing approaches for COBE detection in pre-term infants, which often rely on single-signal inputs, shallow classifiers, or multi-step pipelines, the proposed ConvNeXt-based approach achieves comparable performance by learning feature representations directly from data. This work moves beyond traditional convolutional networks and classical machine learning methods that rely on hybrid models that combine deep feature extractors with separate classifiers [278, 282]. By training and validating on a clinically relevant dataset of 24 pre-term infants with 346 annotated COBE episodes, I demonstrate potential applicability in neonatal care.

A key limitation of this study is its relatively small sample size, which may limit the generalisability of the findings to broader clinical populations. While signal quality assessment was used to reduce artefacts, neonatal physiological signals remain inherently noisy and can be difficult to annotate. Future work should consider incorporating more advanced signal preprocessing methods and exploring techniques such as data augmentation or semi-supervised learning to handle variability and limited labelled data better.

6.10 Conclusion

This chapter demonstrates that clinically acquired neonatal signals can support effective COBE detection using modern deep learning architectures. The IP signal consistently yielded the highest performance across all models, most likely because clinical COBE annotations were primarily based on observed respiratory activity captured by IP. While multimodal approaches incorporating PPG and ECG were explored, they provided only marginal improvements over unimodal IP-based models, suggesting that most of the discriminative information is contained within the IP signal itself. Among the models tested, ConvNeXt achieved the lowest false positive rates, while ResNet architectures showed stable generalisation in both unimodal and fused signal settings. These findings suggest a promising path for real-world deployment of COBE detection tools in neonatal intensive care. Further work is needed to validate these systems prospectively, explore real-time inference potential, and address potential bias introduced by label-source

dependence. The novelty of this work lies in the application of machine learning models to NICU-acquired time-series data and the systematic comparison of signal fusion strategies across architectures.

Chapter 7

Non-contact estimation of RR during periods of cessation of breathing

In the previous chapters, this thesis primarily focused on examining patient-monitor signals such as IP and ECG to characterise respiratory patterns during potential COBE episodes. While these signals provide useful information, prior work has shown that conventional clinical estimates of RR can be imprecise due to signal averaging by patient monitors. Motivated by this, I recomputed RR from the IP signal to provide a more reliable reference. I then used this improved reference to create an annotated COBE dataset and to evaluate machine learning models for COBE detection using only the clinical signals, including ECG, IP, and PPG.

In the next two chapters, I extend this work to video-based, non-contact monitoring. Specifically, I investigate whether camera-derived signals can be used to (1) estimate RR accurately and (2) detect COBE episodes. In this chapter, I focus on the first objective and present a framework for estimating RR around annotated COBE episodes using video camera-derived signals. I use the recomputed RR from the IP signal as the reference to evaluate these non-contact estimates. I apply the RR estimation process to the COBE dataset introduced in Chapter 5, which contains 346 annotated episodes from 24 preterm infants.

7.1 Overview of the RR estimation process

Figure 7.1 provides an overview of the processing pipeline developed in this chapter to obtain reliable estimates of RR. Starting from the MP4 video recordings, I extract 5-minute segments centred on each annotated event and apply a visual quality check to ensure that the video segments are suitable for non-contact respiratory analysis. From each valid episode, I identify upper-body and respiratory regions of interest. I then quantify patient motion by computing frame-difference and average pixel-intensity signals within these regions, capturing both gross movement and subtle breathing-related motion. I subsequently filter these signals and assess the quality before computing final estimates of RR.

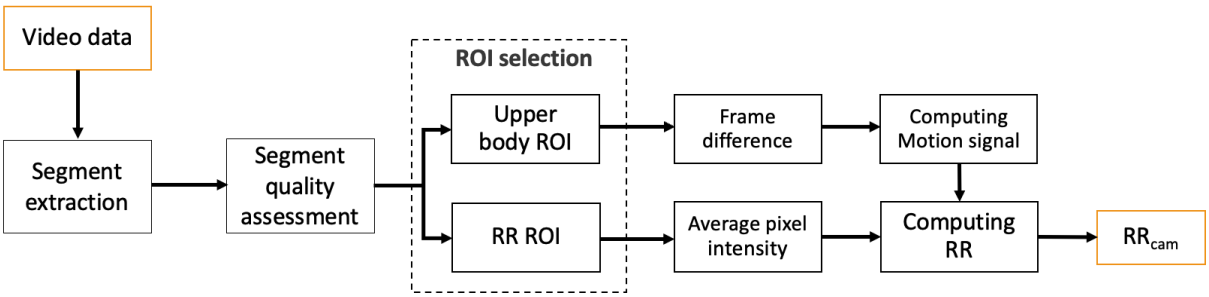


Figure 7.1: Overview of the processing pipeline for non-contact respiratory rate estimation from neonatal video recordings.

7.2 Video segment extraction

The video recordings in the NICU dataset spanned 4.9 ± 1.6 hours per infant. For the purposes of non-contact respiratory analysis, I extracted short video segments centred on each annotated COBE event. Each segment was 5 minutes long (2.5 minutes before and after the event) and served as input for subsequent ROI selection and signal extraction. These segments allow focused analysis around the time of the event while avoiding unnecessary processing of the full-length recordings.

7.3 Segment quality assessment

To ensure that only video segments suitable for respiratory signal extraction were analysed, I visually reviewed each of the extracted 5-minute segments. I excluded segments if the infant's torso was fully obstructed from the camera view, such as during clinical interventions, kangaroo care, or when the infant was moved out of frame. I also applied additional exclusion criteria, including inadequate lighting and lack of overall video clarity, following the quality assessment approach described by Chaichulee [318].

Of the 346 initial COBE annotations, I excluded 49 segments because of obstruction during clinical interventions, 50 because the video camera had been turned off or the infant was out of frame, and 1 due to the video being out of focus. The final dataset comprised 246 segments with annotated COBE episodes for camera-based respiratory analysis.

7.4 Dynamic ROI selection

I processed the remaining 246 segments using an automated dynamic ROI detection pipeline based on pose estimation. I applied the MediaPipe Pose Python toolbox, which implements the BlazePose CNN model [346]. BlazePose, primarily trained on adult humans, identifies 33 anatomical landmarks covering major joints and body segments such as the head, shoulders, elbows, hips, and knees. By detecting these key anatomical points, the model enabled localisation of infant body regions across different lighting,

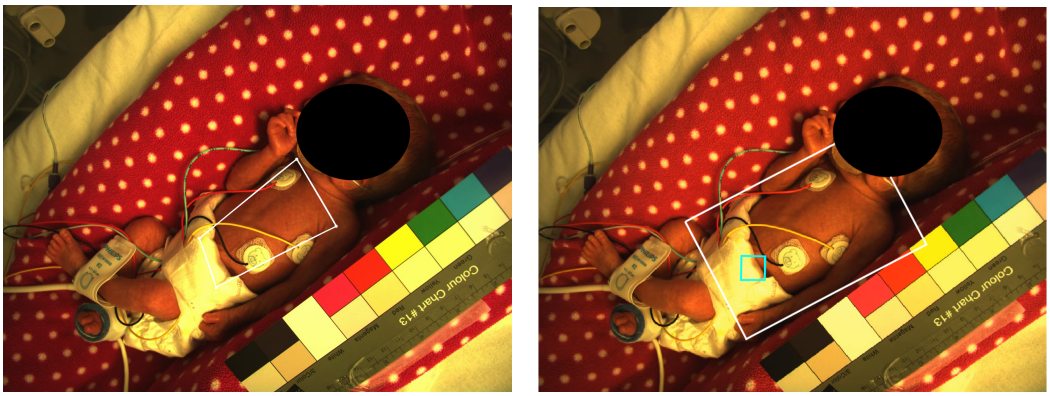
camera angles and postural conditions. From the detected landmarks, I automatically identified the torso region and RR ROIs.

7.4.1 Torso ROI

Although BlazePose estimates 33 landmarks, I focused on a subset comprising the left and right shoulders and hips. From these four landmarks, I computed the torso centre as the midpoint between the shoulder midpoint and the hip midpoint. I also used these four points to define a quadrilateral enclosing the infant's torso. I then fitted a minimum-area bounding rectangle around this quadrilateral to approximate the torso's orientation, aligning the rectangle with the line connecting the midpoint of the shoulders to the midpoint of the hips. I expanded the rectangle by 10% to ensure complete coverage of the torso to accommodate variations in posture. Finally, I smoothed the position and orientation of the bounding rectangle over time using a simple recursive filter that preserved gradual motion while suppressing abrupt frame-to-frame shifts. This dynamically updated rectangle serves as the torso ROI, capturing gross body movements and subtle respiratory motion.

7.4.2 RR ROI

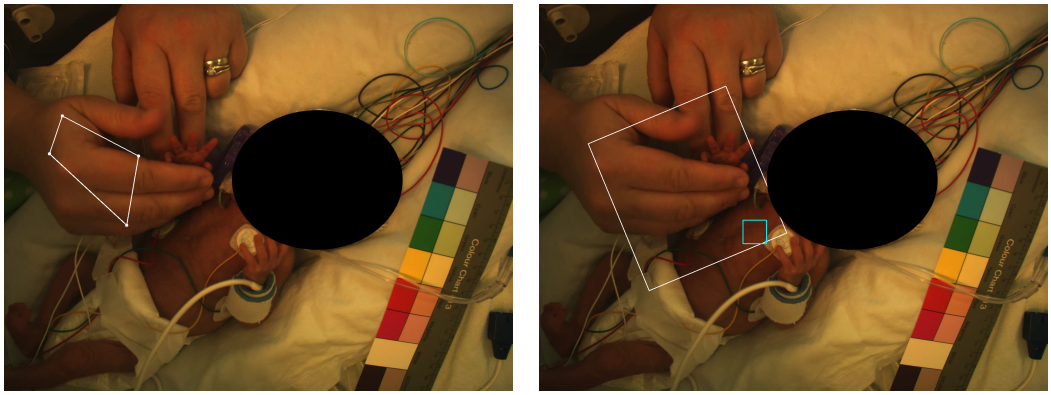
I selected a 75 ×75 RR ROI because small ROIs are reported as more likely to contain motion caused by breathing [318, 347]. This is because respiratory movements are spatially localised, and smaller regions reduce interference from unrelated body or background motion. I centred this ROI relative to the smoothed torso centre, with a fixed offset positioning it near the abdominal or nappy region where respiratory-induced motion is most pronounced. I similarly smoothed this ROI over time using the same recursive filter to ensure continuity and reduce frame-to-frame noise. Figure 7.2 shows the results of the landmark and ROI detection in my dataset.



(a) (b)

Figure 7.2: Landmark and dynamic ROI detection. a) Shoulder and hip keypoints used to guide upper body ROI detection. b) White bounding box represents the upper body ROI and the cyan box represents the RR ROI.

Several failure modes were observed when applying BlazePose to pre-term infants in incubators. Sudden infant movement occasionally caused frame-to-frame jumps in the detected landmarks, leading to transient shifts in the torso and RR ROIs. In addition, the presence of clinicians or caregivers within the video frame sometimes resulted in incorrect landmark assignment or partial tracking of non-infant body regions. Similar reductions in pose-estimation performance under challenging infant recording conditions involving occlusion, variable lighting, and non-standard body positions have been reported previously [348]. An example of landmark misassignment during clinical handling is shown in Figure 7.3. In this example, BlazePose partially tracked the caregiver's hand instead of the infant torso, resulting in displacement of the detected torso and RR ROI. To reduce these temporal artefacts, I applied recursive smoothing to the detected landmark positions and ROI coordinates, suppressing abrupt frame-to-frame fluctuations while preserving gradual changes in ROI position and orientation over time.



(a) (b)

Figure 7.3: Example BlazePose failure during clinical handling of a pre-term infant. (a) Incorrect landmark assignment to the clinician's hand and surrounding objects during caregiver interaction. (b) Resulting displacement of the detected torso ROI (white box) and RR ROI (cyan box), which may affect subsequent respiratory signal extraction.

7.5 Computing the motion signal from a video camera

Cessation of breathing can be characterised by a lack of detectable motion in a set of video frames. Following the work of Cattani *et al.* [77], I defined a motion signal by computing the pixel-to-pixel difference between successive frames over the upper body ROI and counting the pixels where this difference exceeded a given threshold. I selected the ROI to capture the body movement of the pre-term infant while excluding surrounding movements that were not related to the infant, for example, clinical staff moving around the incubator.

Let F be an image within the ROI, representing a single frame from the green colour channel of a video sequence. I computed the pixel-to-pixel difference between consecutive frames within the ROI, Z, at a

discrete time point i as:

$$Z_i[x, y] = |F_{i+1}[x, y] - F_i[x, y]| \tag{7.1}$$

where $[x, y]$ represents the pixel coordinates inside the polygonal ROI. I then computed the binary scale image sequence B as:

$$B_i[x, y] = \begin{cases} 1\ ; & \text{if } Z_i[x, y] > T_B \\ 0\ ; & \text{otherwise} \end{cases} \tag{7.2}$$

where T_B is a discrete low-value threshold chosen heuristically to select the average number of pixels that represent moving objects within the selected ROI. In our dataset, T_B values of more than 5-pixel intensity difference between successive frames resulted in almost no pixels being classified as moving. Therefore, I selected the appropriate T_B as a value below 5. Finally, I computed the non-binary motion signal M, at a discrete time point i, as:

$$M_i = 100 \times \frac{1}{|\Omega_i|} \sum_{(x,y)\in\Omega_i} B_i[x, y], \tag{7.3}$$

where Ω_i is the set of pixels within the polygonal ROI at frame i. This normalisation ensures that the motion signal is independent of the ROI shape or size. Since only one infant was present in the incubator at a time, the motion signal M represents the percentage of pixels in which infant movement occurred, as shown in figure 7.4. The respiratory rate estimate, which I recompute from the IP signal, is shown in Figure 7.4d below the motion signal. This respiratory rate falls below 20 breaths per minute for more than 10 seconds between A and B (the criterion for COBE detection), which corresponds to the motion signal (Figure 7.4c) decreasing to less than 20% at the same time.

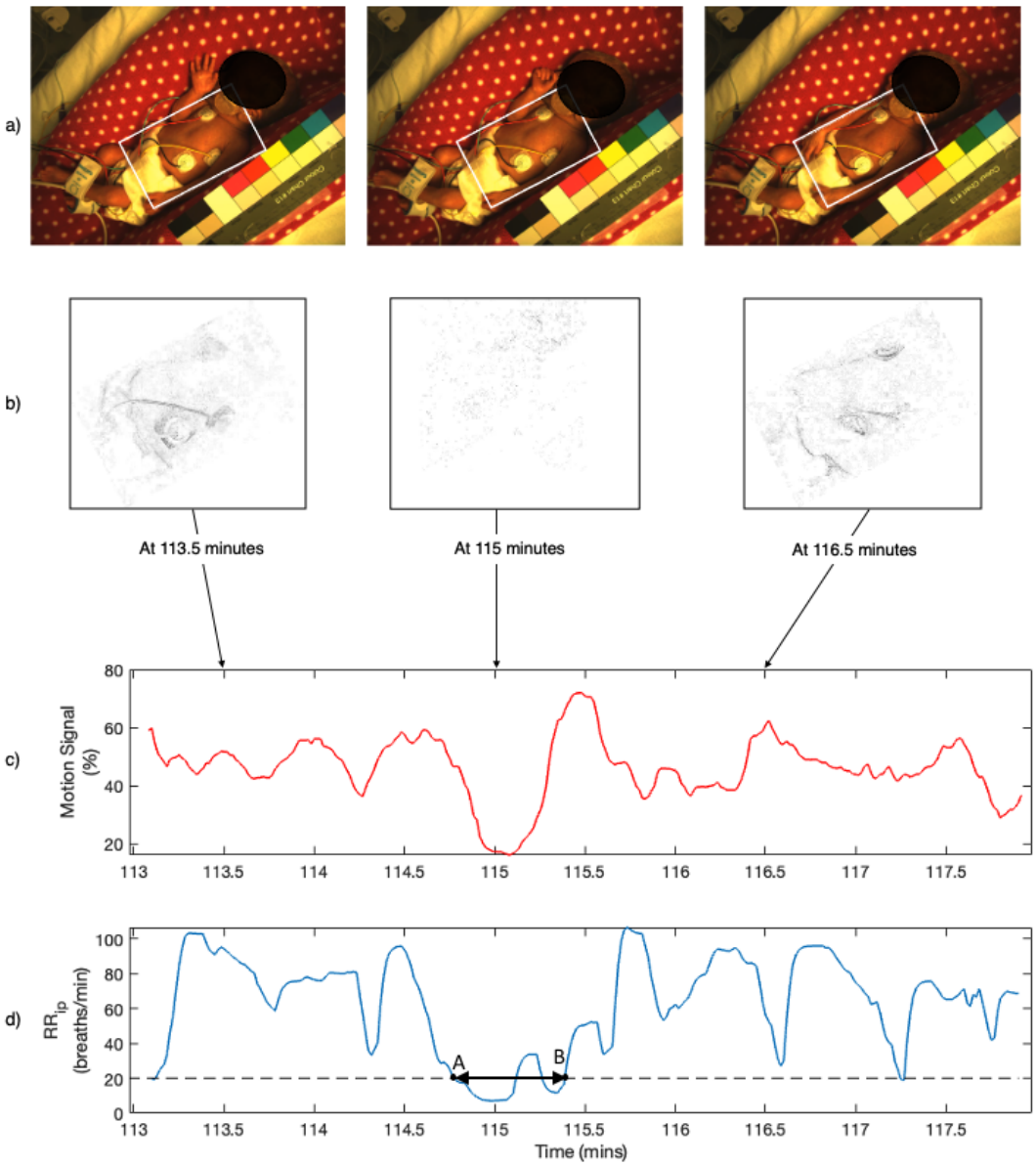


Figure 7.4: Motion signal extracted from the video camera around a COBE episode from a 28.0 week pre-term infant. a) Sample images during periods of activity and no activity, the white box represents a region of interest from which I computed the motion signal. b) Grayscale video frames; the lighter the frame, the less motion captured; c) Motion signal; d) Reference respiratory rate re-computed from the IP signal.

7.6 Computing respiratory rate from video camera data

Breathing causes volume changes in the infant's lungs, as well as body movement. This movement can be captured by a video camera from areas containing exposed skin or covered by tight-fitting clothing, such as a nappy. Figure 7.5 presents an overview of the method I used to estimate RR from video camera data. In this section, I extracted a respiratory signal by computing the average pixel intensity for every frame from a RR ROI over the infants' abdomen. Subsequently, I removed frequency components outside the respiratory rate by designing a sequence of low and high-pass filters. I then used two algorithms

to detect peaks and troughs in the filtered signal. Following peak detection, I assessed the quality of the respiratory signal to determine good-quality breaths. I then applied a Kalman filter to the initial RR estimates to reduce the effects of transient artefacts and improve respiratory rate estimation accuracy. I compared the resulting RR estimates against the reference values recomputed from the IP signal.

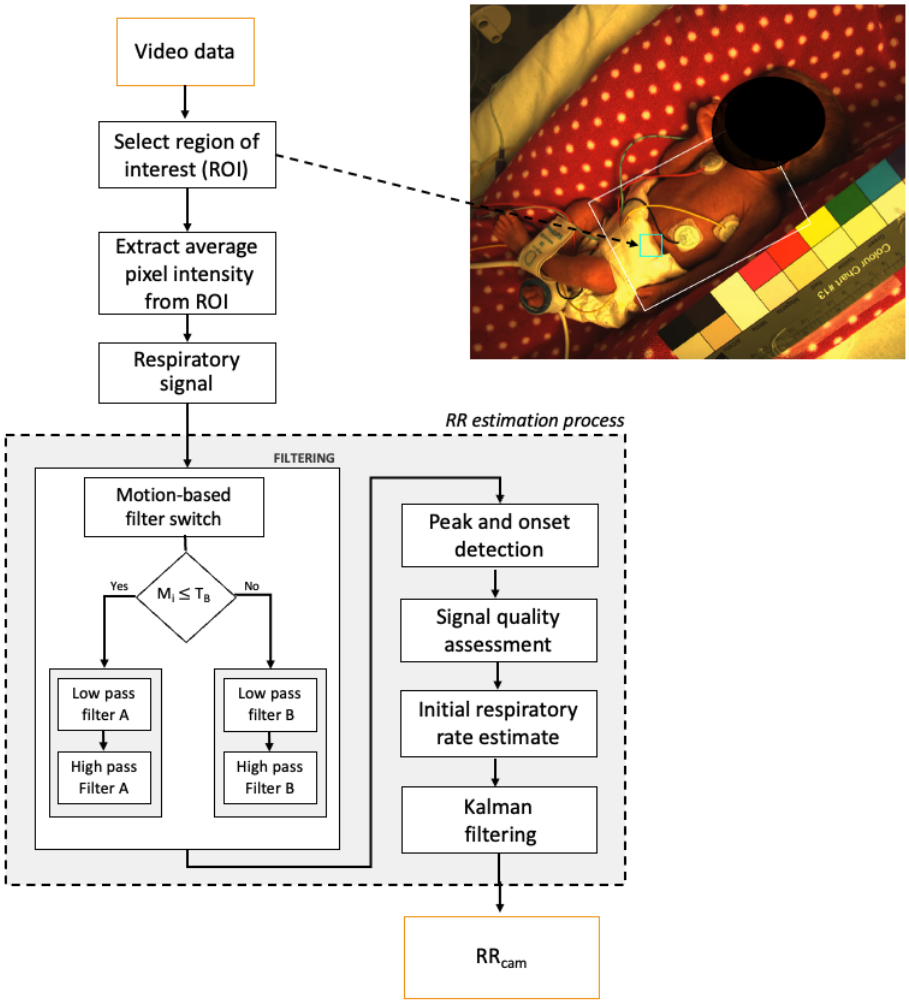


Figure 7.5: Flow diagram outlining the process of computing RR from the video camera data.

7.6.1 Respiratory signal extraction

I extracted three PPGi signals from the selected RR ROI: $PPGi_{red}$ from the red channel, $PPGi_{green}$ from the green channel, and $PPGi_{blue}$ from the blue channel. A typical segment from a video recording is shown in figure 7.6, displaying the PPGi signals extracted from each colour channel. The pre-term infants in this study were nursed with their torsos exposed, although their lower bodies were sometimes covered by nappies or blankets, resulting in more noticeable respiratory pulsations. Following previous studies, I derived the respiratory signal from the green channel, which has been shown to capture respiratory-related motion effectively in skin regions [318, 349]. Although the respiratory signal can also be obtained from

the red or blue channels, the green channel is commonly used in the literature for its reliable performance in video-based respiratory estimation.

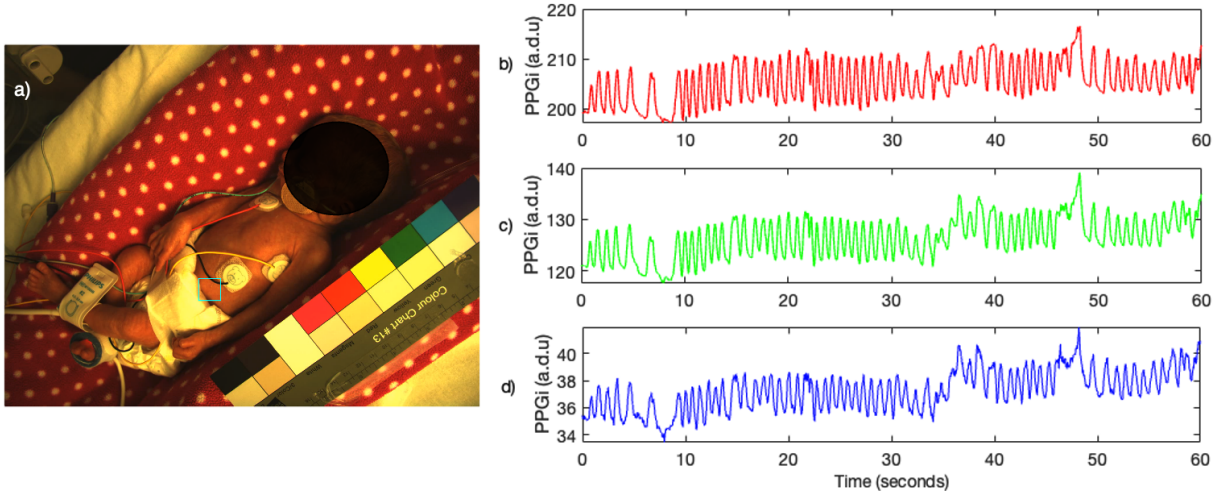


Figure 7.6: Signals extracted from three colour channels within the selected RR ROI.(a) Example video frame with an automatically selected ROI (cyan box) for respiratory signal extraction. (b-d) PPGi signals extracted from the red, green and blue channels, respectively. These signals capture respiratory-related torso motion, which I later process and compare to the reference RR derived from the IP signal.

7.6.2 Signal filtering

During regular activity periods, the respiratory rate in our dataset varies between 21 breaths/min and 138 breaths/min. However, during a COBE episode, the respiratory rate decreases significantly. Several motion artefacts are typically present in the respiratory signal extracted during a COBE episode, leading to incorrect RR estimates as reported by Jorge *et al.* [350]. Using the motion signal M computed in section 7.5, I defined the activity signal A_m to distinguish between periods of high and low activity as follows:

$$A_{m,i} = \begin{cases} 1\ ; & \text{if } M_i > T_{act} \\ 0\ ; & \text{otherwise} \end{cases} \tag{7.4}$$

where T_{act} is a dynamic activity threshold computed using sliding 60 s windows over the 5-minute motion signal. I computed this threshold as:

$$T_{act} = \mu_M + 1.5 \cdot \sigma_M \tag{7.5}$$

where μ_M and σ_M are the mean and standard deviations of the motion signal within each 60 s window, respectively. I de-trended each respiratory signal before the filtering process, removing any linear trend in the signal. I then designed two filter sets; one set was applied to the respiratory signal extracted during periods with 'normal' activity, as defined by A_m with a value of 1. I applied the second set of filters during periods of low/no activity as defined by A_m with a value of 0.

For periods during which there was normal activity, I applied an 8^{th} order high-pass and a 6^{th} order low-pass IIR filters with cut-off frequencies of 0.42 Hz and 2.75 Hz, corresponding to a respiratory rate of 25 breaths/min and 140 breaths/min, respectively. For periods with low/no activity, I applied a 6^{th} order high-pass and 2^{nd} order low-pass IIR filters with cut-off frequencies of 0.20 Hz and 1.42 Hz, corresponding to 12 breaths/min and 85 breaths/min, respectively. Figures 7.7 and 7.8 show the output of de-trending and applying the filters to the respiratory signal during periods of normal activity and low activity, respectively.

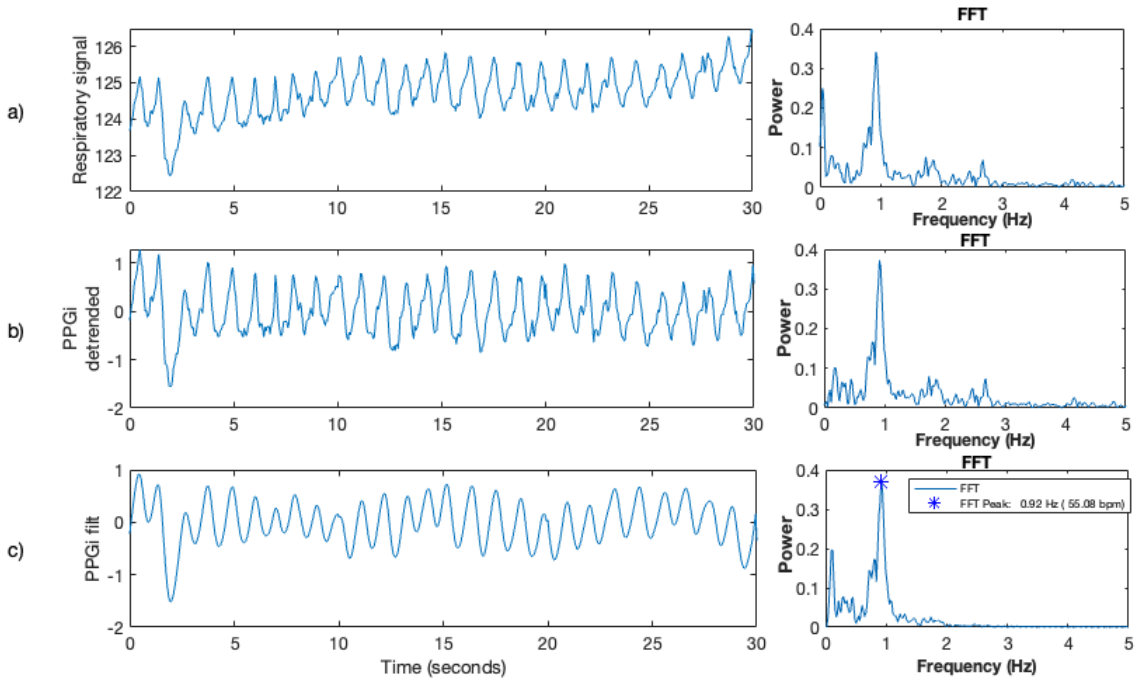


Figure 7.7: Signal filtering of a 30-second respiratory waveform during normal activity. a) Original PPGi waveform with the corresponding frequency components computed by the FFT; b) Waveform after de-trending; c) Waveform after the band-pass filter with cut-off frequencies of 0.42 Hz and 2.75 Hz. This waveform shows a respiratory rate of 55.1 breaths/min, corresponding to a frequency of 0.92 Hz.

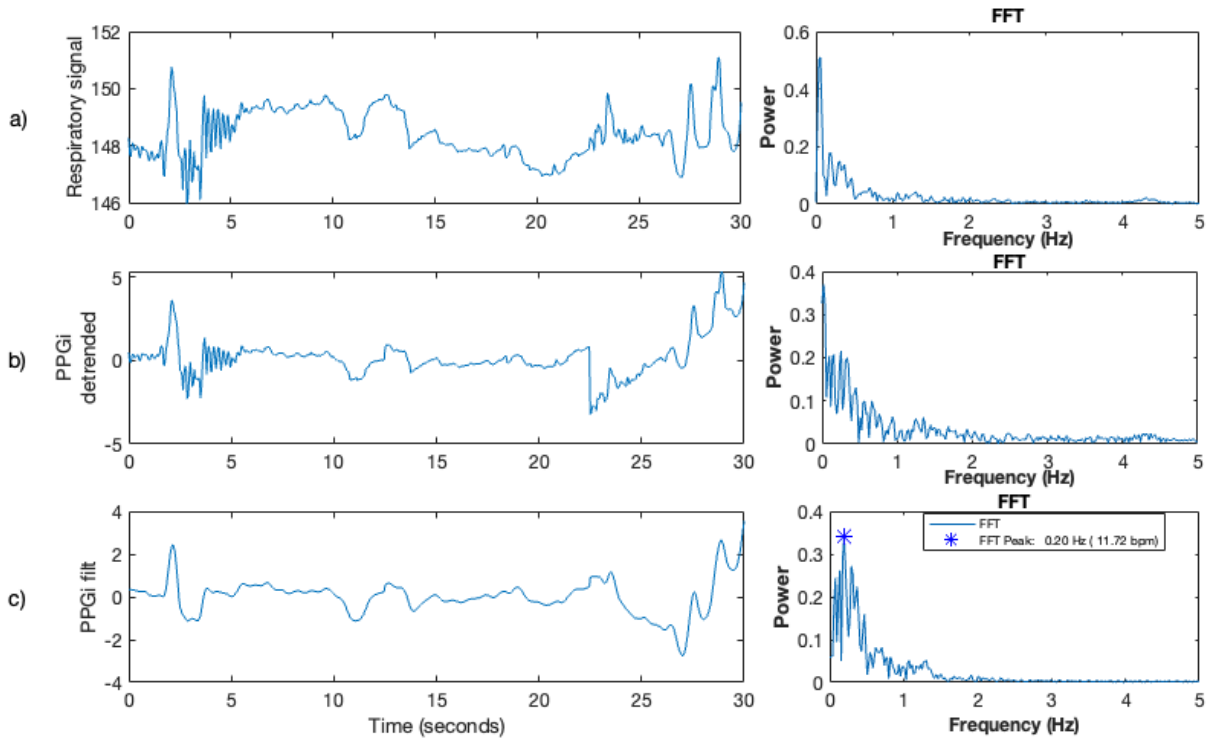


Figure 7.8: Signal filtering of a 30-second respiratory waveform during low activity (COBE episode). a) Original PPGi waveform with the corresponding frequency components computed by the FFT; b) Waveform after de-trending; c) Waveform after the band-pass filter with cut-off frequencies of 0.2 Hz and 1.42 Hz. This waveform has an RR of 11.7 breaths/min, corresponding to a frequency of 0.20 Hz.

7.6.3 Peak and trough detection

I used two algorithms, the Boxed Slope Sum Function (BSSF) and the Mean Average Curve (MAC) peak detectors, to detect peaks and troughs corresponding to breathing cycles in the respiratory signal extracted from the video camera data.

BSSF peak and trough detection

The Boxed Slope Sum Function (BSSF) algorithm was originally developed to detect the peaks and troughs of cardiac pulsatile signals, such as the photoplethysmogram (PPG) signal recorded by a pulse oximeter [233]. I modified the algorithm to work on the respiratory signal recorded from the video camera used in our study. Given r_t as the filtered respiratory signal at time t, I computed a new time series called the slope sum function (SSF) containing only the positive upslopes of each breath over a running window as:

$$SSF_i = \sum_{t=i-w}^{i} u_t \tag{7.6}$$

where w is the length of the analysis window and

$$u_t = \begin{cases} \nabla r_t \ = r_t - r_{t-1}; & \nabla r_t \ > \ 0 \\ 0; & \nabla r_t \leq 0 \end{cases} \tag{7.7}$$

Because equations 7.6 and 7.7 rely on local slope changes in the respiratory waveform, high-frequency noise and transient motion artefacts can produce rapid fluctuations in ∇r_t. Small noisy oscillations may therefore be interpreted as positive upslopes, resulting in false increases in the SSF signal. These additional upslopes can lead to incorrect identification of breath events, causing multiple detections within a single breath or missed breaths. To reduce this effect, I de-trended and band-pass filtered the respiratory signal to suppress low-frequency drift and high-frequency noise before computing the SSF signal. Subsequently, I computed the BSSF signal by finding the zero-crossing points at the upslope's beginning and the downslope's end of each breath cycle in the SSF signal. Although initially designed for different types of signals (MAC for respiratory and BSSF for cardiac), I modified the BSSF algorithm to work with respiratory signals. Instead of using the typical duration of the upslope of the cardiac pulse (w_{up} = 170 ms or 3 samples for PPGi signal at 20 Hz) for computing the BSSF signal, I modified the upslope duration of the respiratory pulse to w_{up} = 300 ms. The upslope points in the BSSF signal (uBSSF) correspond to potential troughs of a breath cycle. Figure 7.9a-d shows the detrended PPGi signal, the resulting SSF, BSSF signals and the uBSSF data points, respectively, carefully selected to show a 30-second segment of a very stable PPGi signal.

To identify the peak and trough of each breath cycle, I applied adaptive amplitude thresholding to the uBSSF data points. Initially, I computed this threshold as the median of the first 10s from the uBSSF data points. I adaptively adjusted this threshold value as the uBSSF values changed. For each subsequent uBSSF data point, if its value was greater than the computed threshold, I considered the data point as a breath onset, and I increased the threshold by 10% of the difference between the previous threshold and the current uBSSF data point. The peak of each breath cycle was then the location of the maximum value between two successive troughs. I applied a 500 ms refractory period (a recovery time after the peak) to prevent multiple trough detections.

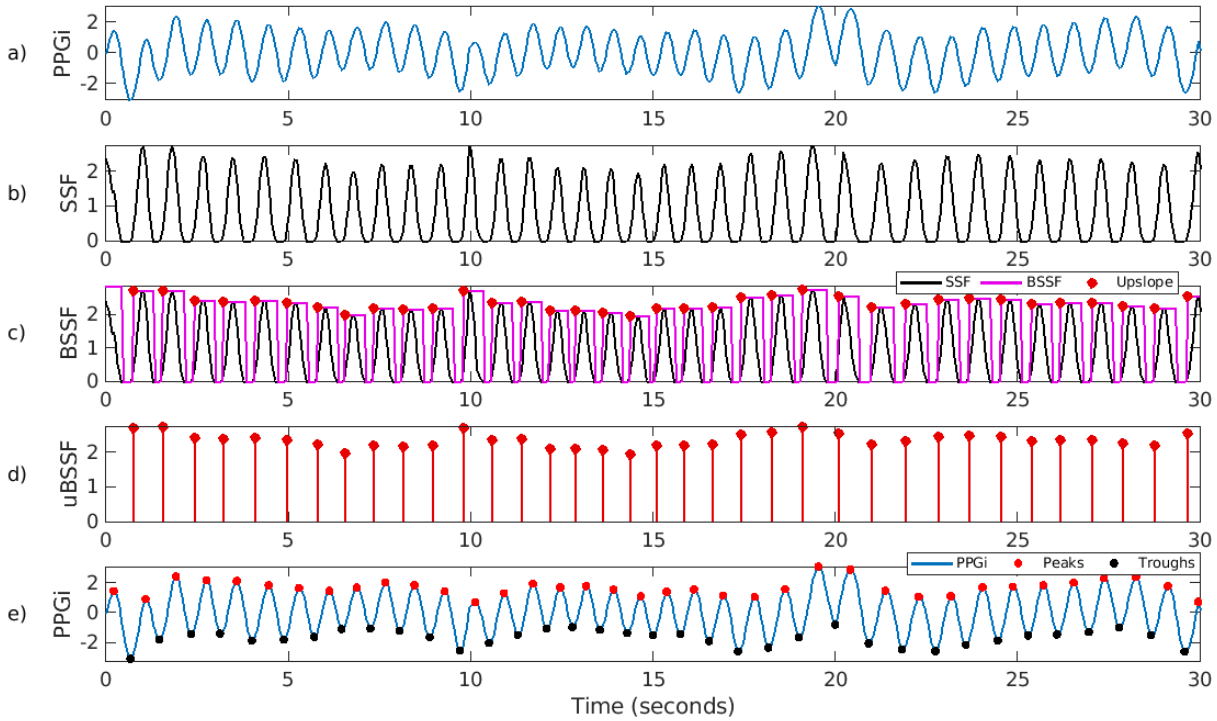


Figure 7.9: Peak and trough detection for a 30-second PPGi signal. a) The filtered PPGi signal; b) I computed the SSF signal based on the sum of the slope of the pulses; c) The BSSF signal is matched with each breath; d) The uBSSF data points correspond to the upslope points in the BSSF signal and e) I detected the breath troughs and peaks based on adaptive filtering and decision rules.

MAC peak and trough detection

Both the MAC and BSSF algorithms have been reported to have high accuracy [321, 351]. However, only the MAC-based algorithm is designed and validated specifically for respiratory waveforms. Therefore, I applied the MAC peak detector introduced in section 4.5 to the respiratory signal without any modifications. I chose this peak detector to segment each breath in the respiratory signal. I later used the BSSF algorithm as part of signal quality assessment and respiratory rate estimation. Figure 7.10 shows the comparison of peaks and troughs detected by the two different algorithms on a clean respiratory waveform.

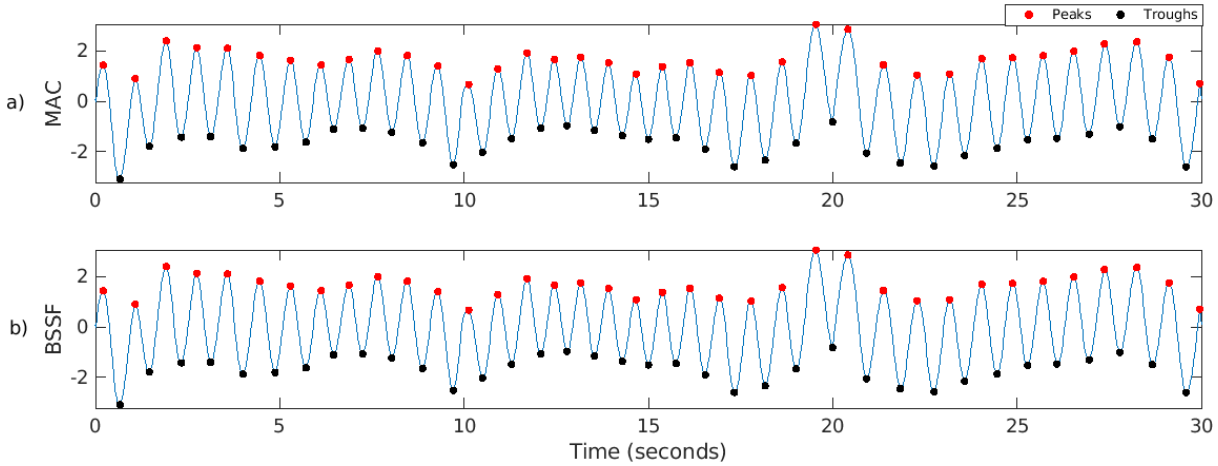


Figure 7.10: The detection of peaks and troughs for a 30-second clean respiratory waveform by a) the MAC algorithm and b) the BSSF algorithm. There are no differences between the outputs of the two algorithms for this clean respiratory signal.

7.6.4 Signal quality assessment

Video analysis for vital sign estimation faces significant challenges due to data corruption from subject movement. It is, therefore, crucial to assess the quality of the PPGi signal to estimate respiratory rate from video data accurately. In this work, I extended the assessment of the PPGi signal from that described in Chaichulee *et al.* [318]. I evaluated the quality of the respiratory signal based on two metrics: a measure of the level of activity in the video and the agreement of two peak and trough detectors.

Activity index

Changes in lighting or movement caused significant changes in light intensity recorded by the video camera. I computed the pixel-to-pixel difference between successive frames Z, at a discrete time i. I then computed the frame difference signal as follows:

$$D_i = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} Z_i[x, y] \tag{7.8}$$

Subsequently, I computed a raw activity signal quality index for each frame, SQI_{raw}, as:

$$SQI_{raw}(i) = \begin{cases} 1; & D_i \leq T_{sqi} \\ 0; & \text{otherwise} \end{cases} \tag{7.9}$$

where I defined the threshold T_{sqi} as:

$$T_{sqi} = \mu_D + 2.0 \cdot \sigma_D \tag{7.10}$$

where μ_D and σ_D are the mean and standard deviations of the frame difference signal D_i, respectively. For a window starting at sample s and containing k samples, where s denotes the starting sample index of the analysis window, I computed the final activity index SQI_{act} as:

$$SQI_{act}(i) = \frac{1}{k}\sum_{j=0}^{k-1} SQI_{raw}(s+j) \tag{7.11}$$

Peak agreement

I used the agreement between the two peak and trough detection methods (MAC and BSSF) to compute a signal quality metric, SQI_{peak}. For the i^{th} breath, I calculated SQI_{peak} as the proportion of the number of peaks in agreement with the number of all peaks detected over a 10-second window(w), centred around the i^{th} breath:

$$SQI_{peak}(i) = \frac{N_{Agreed\ peaks}(w)}{N_{All\ peaks}(w)} \tag{7.12}$$

The SQI_{peak} ranged from 0 (poor quality) to 1 (good quality). If the peaks identified by the two detectors were not located more than 5 samples (or 0.25 seconds – half of the duration of the highest frequency rate that can be estimated) apart from each other, I considered an agreement to occur, and I set SQI_{peak} to 1. Figure 7.11 shows the peaks and troughs detected by both techniques and the SQI_{peak} computed. As expected, the SQI_{peak} decreases in the presence of peak detector disagreement, for instance (by a small amount) between $t = 25$ and $t = 30$ seconds.

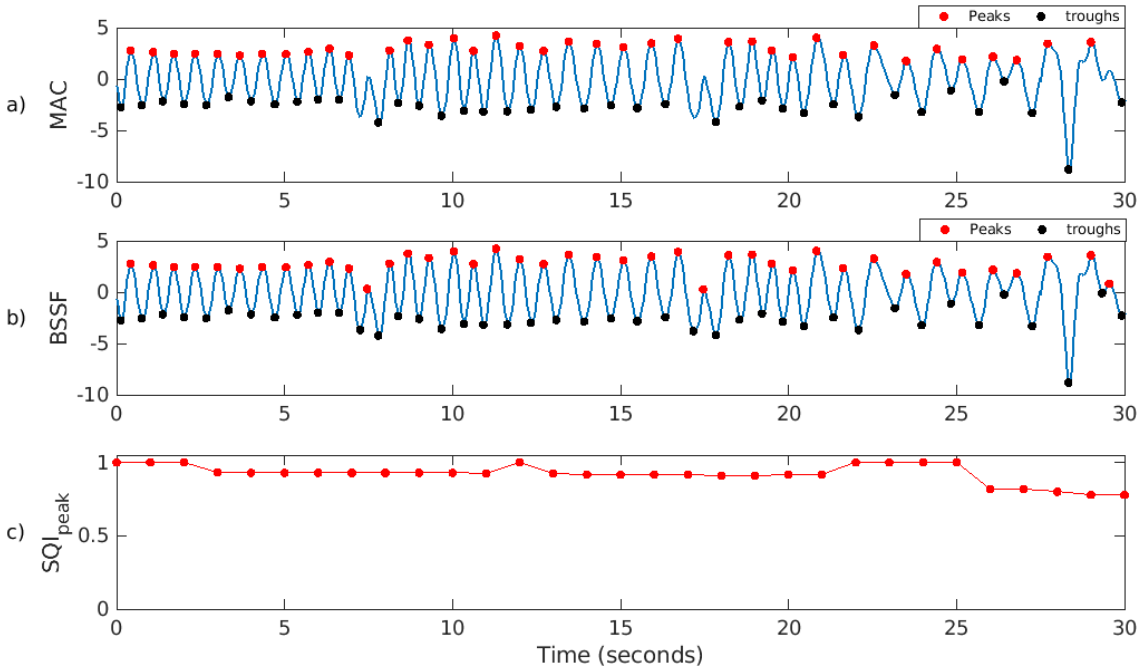


Figure 7.11: The calculation of SQI_{peak}. a) Peaks detected using the MAC algorithm; b) Peaks detected using the BSSF algorithm; c) SQI_{peak}.

Combining signal quality measures

Once I calculated the two signal quality metrics, I computed a combined signal quality index (SQI_{breath}) for the i^{th} breath, using the equation:

$$SQI_{breath}(i) = SQI_{act}(i) \times SQI_{peak}(i) \tag{7.13}$$

The SQI_{breath} metric ranged between 0 (corresponding to a poor quality signal for which respiratory rate could not be reliably computed) and 1 (corresponding to low motion periods of a good quality respiratory signal when the two peak detectors agreed). Figure 7.12d shows the output of the process of deriving SQI_{breath} from all signal quality indices for a 30-second window.

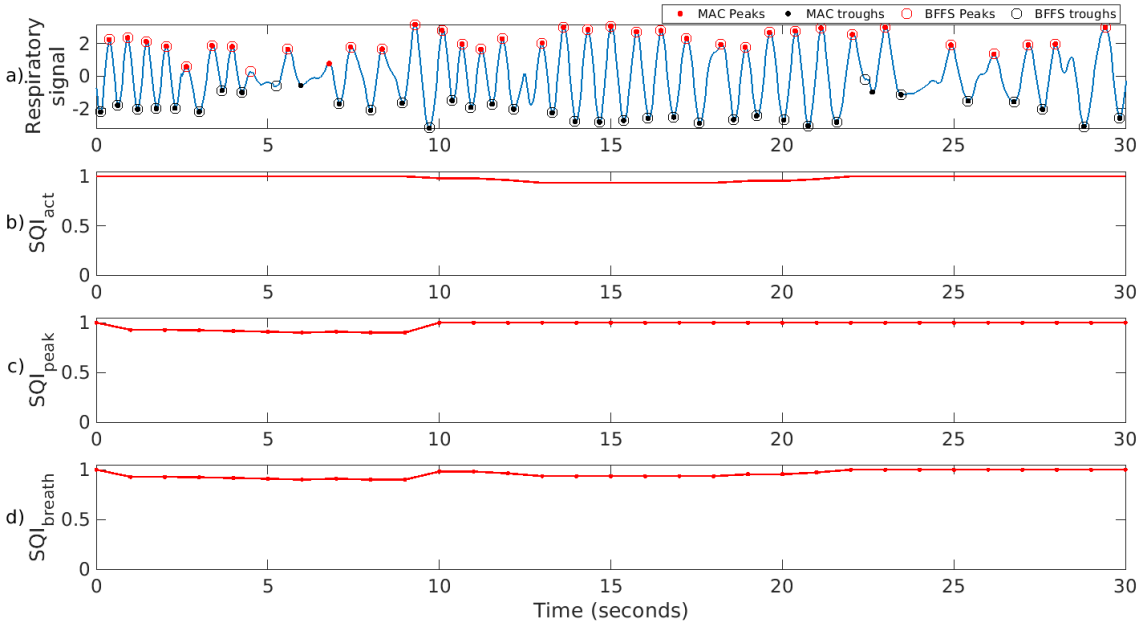


Figure 7.12: The calculation of SQI_{breath} for a 30-second respiratory signal. a) Peaks and troughs detected by the MAC and the BSSF techniques; b) SQI_{act} values; c) SQI_{peak} values; d) SQI_{breath} calculated by multiplying all the previous metrics together, as in equation 7.13.

7.6.5 Respiratory rate estimation

I computed an initial respiratory rate estimate over a 10-second sliding window with a step size of 1 second. I first expanded each 10-second time window to include the complete first and last breath pulse before computing the respiratory rate estimate RR as:

$$RR(w) = N_{breaths} \times \frac{60}{L_{exp}} \tag{7.14}$$

where $N_{breaths}$ is the number of breaths in the expanded window and L_{exp} is the length of the expanded window w. Figure 7.13 shows how I calculated respiratory rate using the breath counting technique, which relies on the accuracy of the peak and trough detectors.

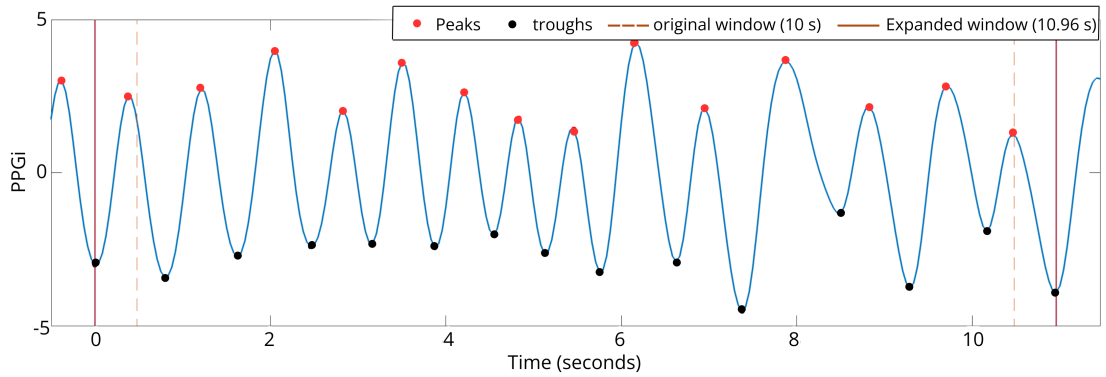


Figure 7.13: Window expansion for RR estimation. The original window w of 10 seconds was expanded to include the whole duration of the first and last breaths. The expanded window of 10.96 seconds contains 14 breaths, corresponding to RR of 76.64 breaths/mins.

Once computed, I time-synced RR_{cam} to RR_{ip} using the cross-correlation method introduced in section 4.8. I calculated the signal quality index of the respiratory rate estimate (SQI_{RR}) as the average of the SQI_{breath} values over a given window.

7.7 Results

This chapter presents the evaluation of the proposed video-based respiratory rate monitoring algorithms. I use summary plots (Bland–Altman, histograms, and correlation) to report results and quantify them with error metrics (MAE, MAD, RMSE, and correlation) for different RR ranges. I also calculated metrics across the entire dataset, to provide an overall measure of camera-based RR estimation performance around COBE episodes.

To characterise RR estimation performance across clinically relevant RR ranges, I computed error metrics separately for periods of normal breathing (RR $>$ 20 breaths/min) and periods of low respiratory rates associated with potential COBE events (RR $\leq$ 20 breaths/min). This stratification allows me

to quantify the differences in estimation accuracy between higher and lower RR ranges. Table 7.1 summarises the errors calculated when comparing RR_{ip} and RR_{cam} across the 5-minute periods around each COBE.

Table 7.1: Summary of error metrics for respiratory rate (RR) estimation from video camera data, evaluated separately for normal breathing (RR $>$ 20 breaths/min) and potential COBE events (RR $\leq$ 20 breaths/min).

Metric	Potential COBE (RR $\leq$ 20 bpm)	Normal breathing (RR $>$ 20 bpm)	All segments
MAE*	11.2	16.9	15.3
MAD*	6.6	11.2	12.0
RMSE*	13.1	20.5	19.3
r	0.20	0.57	0.62

*Values reported in breaths/min

Figure 7.14 shows the summary plots for the proposed RR estimation algorithms across all 246 annotated COBE episodes, with estimation computed over a 10-second sliding window (step size of 1 second).

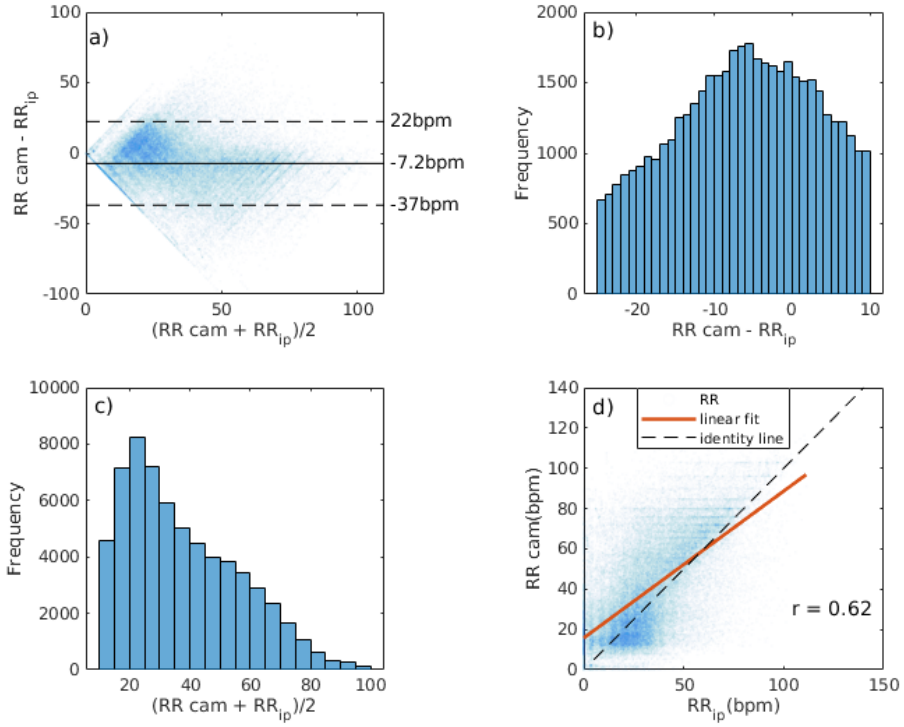


Figure 7.14: Comparison between RR_{ip} and RR_{cam} over the entire COBE episodes. a) Bland-Altman plot; b) Histogram of the differences between the two RR estimates; c) Histogram of the mean of the two RR estimates; and d) correlation plot showing a positive correlation between the two measurements.

7.8 Discussion

Although research on non-contact estimation of vital signs using video cameras has significantly increased over the past decade, most of the studies have focused on estimating vital signs during periods with good-quality camera signals and reduced motion artefacts, excluding analysis during periods such as cessation of breathing. In this chapter, I propose algorithms to estimate respiratory rate from video camera data, specifically for periods of COBE.

The level of respiration is a key factor in identifying cessations of breathing. Respiratory rate serves as an indicator of the regularity of breathing patterns, making it essential for detecting any deviations that may indicate a cessation of breathing. Therefore, any COBE detection pipeline, whether contact-based or non-contact-based, should account for this vital sign. Several studies have presented signal processing methodologies for estimating respiratory rate from camera-derived PPGi signals [21, 164, 318, 347], while others have employed computer vision techniques to analyse the respiratory-synchronous motion of the subject [88, 352, 353, 354, 355]. These methods were applied during periods of 'normal' breathing and mostly to adult subjects. The range of 'normal' respiratory rate in pre-term infants (20 -120 breaths/min) is significantly greater than that of adults. The sudden episodes of cessation of breathing experienced by pre-term infants may further increase this range. Because of these reported differences in the range of respiratory rates between adults and pre-term infants, specialised algorithms need to be developed and validated for the neonatal population.

This chapter presented a 'filter-switching' technique for estimating respiratory rate during COBE episodes. The technique involves extracting a motion signal from the camera, indicating thoracic movement in each video frame. I then used this signal to determine appropriate filters for low and high periods of motion, corresponding to periods of cessation of breathing and normal breathing, respectively. This study further draws comparisons between the recomputed respiratory rate (RR_{ip}) and the camera-derived RR.

Compared with the reference respiratory rate, the respiratory rate estimates derived from the camera gave an MAD of 12.0 breaths/min and MAE of 15.3 breaths/min for all infants in the COBE annotated dataset. These results indicate that respiratory rate can be estimated for pre-term infants using signal processing methodologies even during periods of cessation of breathing. The correlation coefficient, however, reduces during low RR periods despite lower absolute error magnitudes than those observed during normal breathing. This occurs because even modest absolute errors represent a much larger proportion of the narrow low RR range (2 - 20 breaths/min), resulting in increased variability relative to the range of values. As a result, although the MAE, MAD, and RMSE are lower for low-RR periods,

these errors represent a greater proportion of the physiological range and therefore have an effect on the reduction of the correlation coefficient.

7.9 Conclusion

I have demonstrated the feasibility of estimating respiratory rate from digital video cameras in ambient light conditions. Since pre-term infants are physiologically unstable, their vital sign values can vary substantially within a short period. Using non-contact technologies to monitor pre-term infants can provide advantages over conventional vital sign monitoring techniques. They can be integrated into existing patient wards or telemedicine systems, providing continuous and unobtrusive monitoring without the need for physical sensors or electrodes attached to the infant's body. In addition, they could be expanded to provide other bedside assessments, such as the infant's physical activity, distress or pain.

Chapter 8

COBE episode detection using camera signals

8.1 Introduction

Conventional COBE detection pipelines rely on physiological signals such as the IP, ECG and the PPG signals, all of which require physical sensors placed on the infant's body. Although these signals remain the clinical standards for identifying and quantifying breathing cessation in the NICU, they often suffer from motion artefacts and can injure the fragile skin of pre-term infants [32, 58]. These challenges are further compounded in low-resource settings, where conventional sensors may be limited by availability, recurring consumable costs, or the need for specialist staff to maintain and apply them [64, 65]. In contrast, camera-based approaches offer a promising, non-contact alternative that can be deployed using inexpensive, widely available hardware. Recent advances in computer vision and video-based physiological sensing have demonstrated the feasibility of extracting respiratory and cardiac-related information from standard video recordings without physical contact [68, 88, 89]. Such systems are particularly relevant for low- and middle-income countries (LMICs), where the cost and long-term sustainability of conventional monitoring equipment remain major barriers to delivering continuous neonatal care [65, 66]. Nevertheless, despite their potential advantages, camera-based methods need to be rigorously evaluated to determine whether they can reliably detect clinically meaningful events such as COBE episodes.

In Chapter 7, I demonstrated that respiratory-related signals can be reliably extracted from routine neonatal video recordings using dynamic ROI selection, filtering, and robust signal-quality assessment, even during periods of COBE. Building on these findings, this chapter evaluates whether the camera-derived signals themselves contain sufficient information for detecting COBE episodes without relying solely on contact-based physiological measurements. Specifically, the chapter investigates whether video-derived signals alone can support accurate COBE detection, and whether combining camera-based and traditional physiological signals improves classification performance compared with either modality alone.

By addressing these questions, the chapter assesses the potential role of low-cost camera-based monitoring in LMICs, where improving detection of COBE episodes could substantially enhance clinical

safety while reducing reliance on resource-intensive equipment. At the same time, hybrid approaches that fuse video and contact-based signals may also offer benefits in high-resource settings, providing an additional visual modality not captured by traditional sensors and potentially improving reliability during periods of poor signal quality or clinical intervention.

8.2 Overview of the methodology

Figure 8.1 presents the workflow to classify COBE episodes using video-derived signals, and, in the case of hybrid models, a combination of video and conventional physiological signals introduced in Chapter 6. The pipeline begins with extracting signals from short segments of camera recordings and assessing their quality, followed by alignment of these camera signals to the reference clinical data. I then segment the signals into fixed-length windows suitable for model input. For model training and evaluation, I use a five-fold cross-validation approach with a hold-out test set to assess model generalisability. I train ResNet and ConvNeXt architectures in both single-signal and multimodal configurations, to distinguish between COBE episodes and normal breathing. Finally, I evaluate hybrid models that integrate camera-derived features with traditional physiological signals. I construct all hybrid and multimodal models using the best-performing single-signal configurations identified during individual-modality training, rather than re-optimising network architectures for combined inputs. I summarise all resulting performance comparisons in a table.

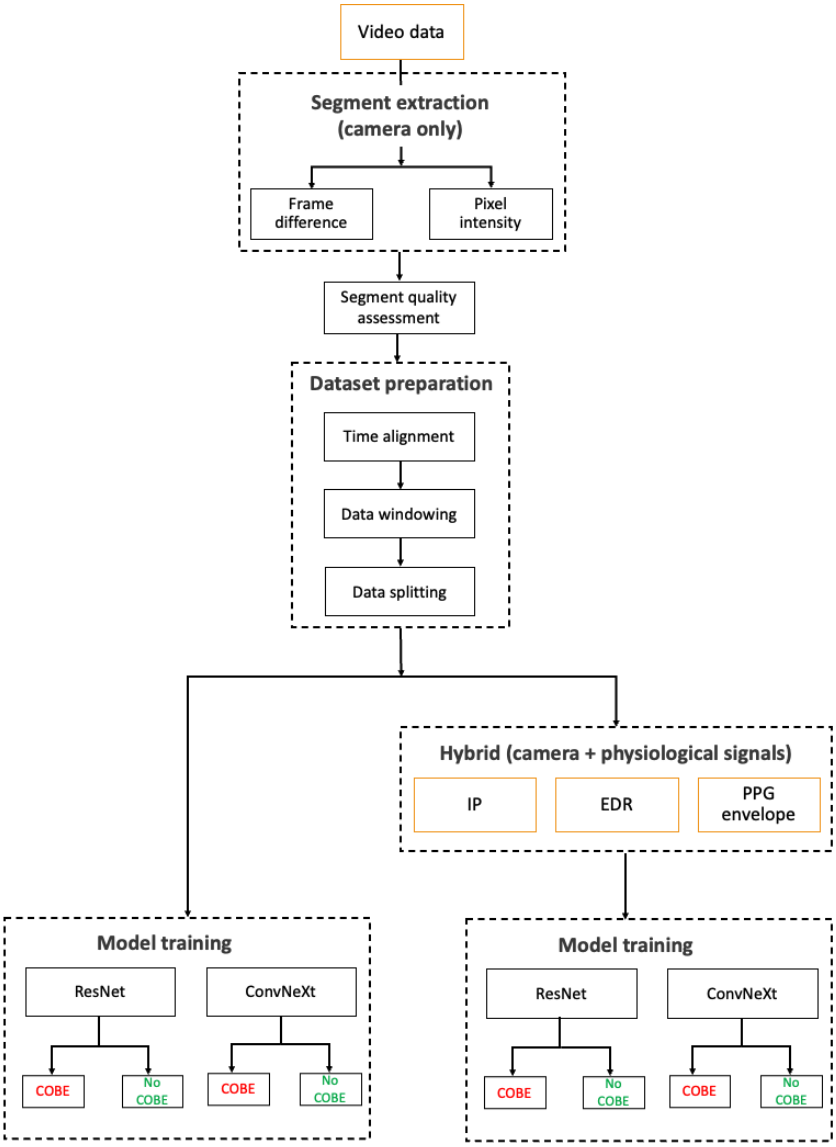


Figure 8.1: Workflow for COBE episode detection using deep learning models. I train models on individual camera-derived signals and on multimodal inputs, combining these signals. Subsequently, I develop hybrid models that integrate camera-derived features with conventional physiological signals using the best-performing single-signal configurations identified during prior training.

8.3 Segment extraction

For training and evaluation of deep learning models, I use two camera-derived signals extracted from neonatal video recordings following the extraction pipeline described in Chapter 7. I extracted a frame difference signal, which quantifies per-frame motion of the infant's torso by computing the absolute difference in pixel intensities between consecutive frames. I also extracted a pixel-intensity based signal from a 75×75 ROI near the abdominal region, where respiratory-induced motion is most pronounced. I restricted the signal extraction to time intervals corresponding to the manually annotated segments, ensuring temporal alignment between the camera-derived signals and the reference labels used for COBE

detection. Figure 8.2 shows representative examples of the two camera-derived signals extracted from an annotated recording segment, showing periods of normal breathing and COBE episodes.

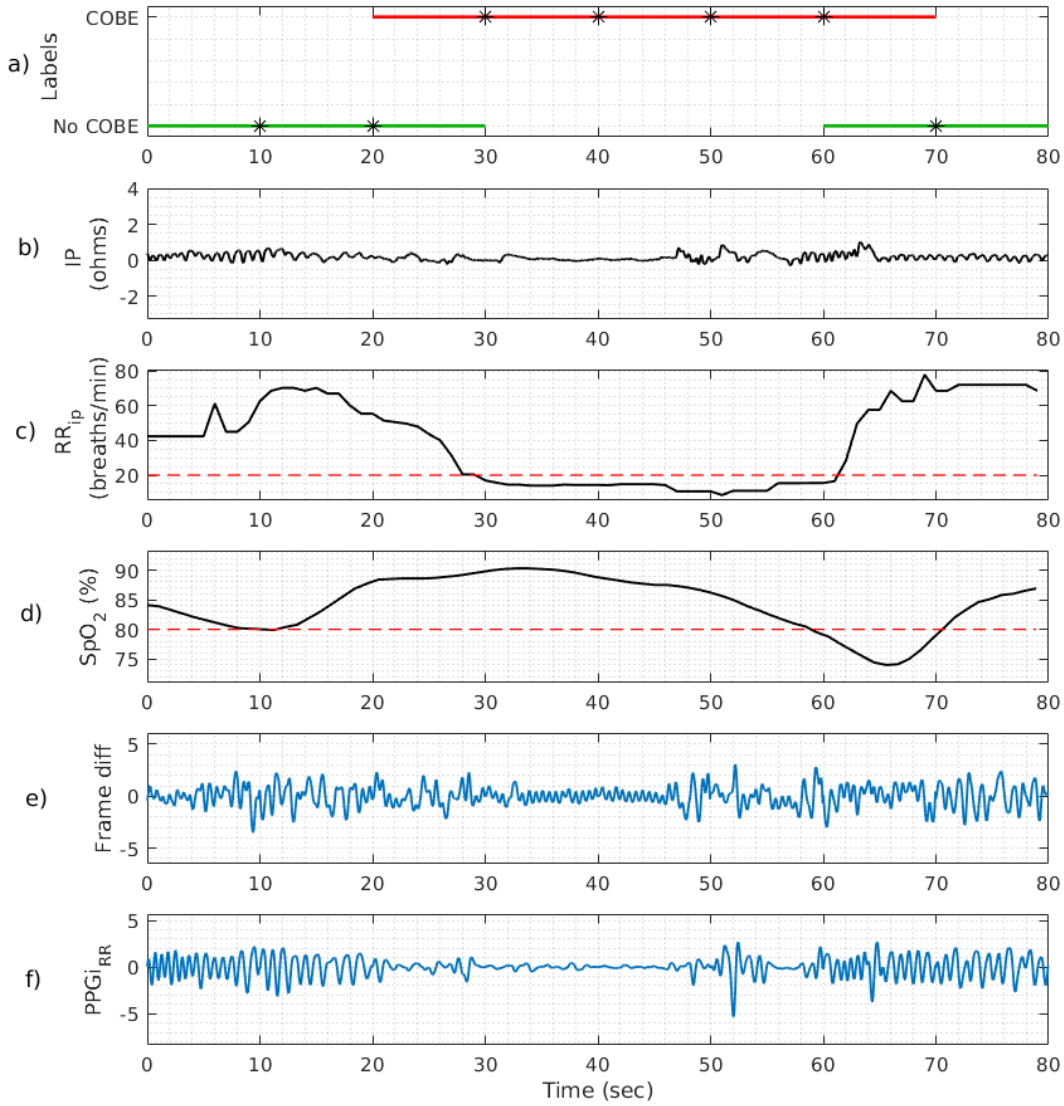


Figure 8.2: Camera-derived respiratory signals extracted from a neonatal video recording, showing an example annotated segment with normal breathing and a COBE episode. a) Label lines indicating periods of COBE (red) and no COBE (green). Asterisks mark the center of each 20 s window; b) IP signal; c) Respiratory pause (RR_{ip} below 20 breaths/min) for longer than 10 s; d) Desaturation where SpO_2 decreases to below 80% for at least 10 s; e) Frame difference; f) Pixel-intensity from abdominal motion.

8.4 Segment quality assessment

The dataset used to estimate RR in Chapter 7 comprises 246 manually annotated 80-second segments containing COBE episodes. To complete the training dataset with negative examples (normal breathing), I visually reviewed the remaining segments from the COBE dataset and excluded any segments where reliable signal extraction was compromised. Specifically, I excluded segments if the infant's torso was

fully obstructed, such as during clinical interventions, kangaroo care, or when the infant was moved out of frame. I also excluded segments if the video camera had been turned off or if the video was out of focus.

Out of the initial 608 normal breathing episodes, I excluded 59 segments in which the infant was obstructed, 91 segments in which the infant was out of frame or the camera was turned off, and 2 segments because the video was out of focus. The final training dataset, therefore, contains 246 COBE episodes and 443 normal breathing episodes from 23 infants, for a total of 689 80-second segments used for machine learning. Aside from this reduction in dataset size, I applied all preprocessing, windowing, labelling, and training procedures exactly as described in Chapter 6.

8.5 Construction of the machine learning dataset

This section describes the processes used to prepare the manually labelled 80-second segments for input into machine learning models. Following the approach in Chapter 6, I first time synchronised the camera-derived signals with the contact-derived reference signals, then subdivided each segment into shorter overlapping windows, ensuring that even brief COBE episodes are fully captured. I assigned each time-aligned window a positive or negative label according to the expert annotations, which were originally based on contact-derived physiological signals. By applying these labels to the corresponding camera-derived signals, I enabled evaluation of whether video features can support COBE detection. Finally, I split the dataset into training, validation, and held-out test sets for model development and evaluation.

8.5.1 Time alignment

Accurate alignment of the camera-derived signals with the clinical reference signals is essential for reliable evaluation. In Chapter 7, I performed time alignment only around annotated COBE episodes to evaluate respiratory rate estimation. For the current analysis, which also includes normal breathing segments, I extend this procedure to all episodes in the filtered camera dataset. For each COBE and non-COBE segment, I quantified the event-specific lag by computing the cross-correlation between the reference recomputed RR signal and the corresponding camera-derived respiratory signal. This synchronisation accounts for potential delays introduced by differences in devices, signal processing pipelines, or recording offsets. I then perform all subsequent analyses on the time-aligned signals, enabling a fair comparison between camera-derived and reference-derived metrics across all clinically annotated events.

8.5.2 Data windowing

Following the methodology in Chapter 6, I divided each 80-second segment into overlapping 20-second windows with a 10-second step. The 20-second window is long enough to fully contain the shortest clinically relevant COBE episodes (10 seconds), while remaining short enough to provide sufficient temporal resolution for the machine learning models. The 10-second step ensures 50% overlap between consecutive windows, allowing brief episodes near window edges to be captured in at least one window. Each 80-second segment, therefore, produces seven overlapping windows. All windows inherit the labels already assigned in section 5.1.2. That is, each window retains the positive or negative COBE label derived from the expert annotations, ensuring consistency with the original episode labelling. This process produces a total of 4,823 overlapping 20-second windows, including 755 positive and 4,068 negative windows across 23 pre-term infants aligned with the previously generated labels.

Table 8.1 summarises the results of the windowing procedure, showing the number of segments, overlapping windows, and the distribution of positive and negative labels across all infants. Because COBE episodes can be brief relative to the 80-second segment duration, only a subset of the overlapping 20-second windows within a COBE-labelled segment contains a cessation of breathing. Windows that fall outside the annotated COBE interval therefore capture periods of normal breathing and are labelled as negative. As a result, COBE-labelled segments may contain more negative than positive windows.

Table 8.1: Composition of the labelled machine learning dataset used for COBE classification, prior to data splitting.

Metric	Number of 80-second segments	Number of 20-second windows	Number of positive windows within 80-second segments	Number of negative windows within 80-second segments
Segments with COBE	246	1,722	755	967
Segments with normal breathing	443	3,101	0	3,101
Total	689	4,823	755	4,068

8.5.3 Data splitting

I divided the labelled dataset into a training set and an independent held-out test set. I then divided the training set into five subsets for cross-validation, with four subsets containing data from 4 pre-term infants each and one subset containing data from 3 pre-term infants due to the exclusion of low-quality segments. When splitting the data, I ensured that each subset was as demographically balanced as possible, considering average infant gestational age, the number of positive and negative windows, and gender distribution.

For five-fold cross-validation, I trained each model on four of the five training subsets, representing approximately 80% of the training data, and validated it on the remaining subset, representing approximately 20%. I repeated this process five times so that each subset served once as the validation set. After identifying the best-performing model configuration from cross-validation, I retrained the model using all five training subsets combined. I then evaluated the final model on the independent held-out test set, which was not used during training or validation, ensuring an unbiased assessment of performance. Table 8.2 shows the distribution of positive and negative windows in the training and held-out test sets after dataset splitting, providing an overview of the data used for model development and evaluation.

Table 8.2: Distribution of positive and negative windows in the training and held-out test sets.

Dataset	Positive windows	Negative windows
Training	639	3,456
Test	116	612
Total	755	4,068

8.6 Baseline detection of COBE using camera-derived RR

To provide a baseline reference for evaluating the camera-based machine learning models presented in this chapter, I implemented a simple approach for COBE detection using RR estimated from the camera. In this approach, I defined a COBE event as a sustained decrease in RR to below 20 breaths/min for at least 20 seconds. As in section 6.8.2, I evaluated predictions from this baseline method by comparison against manual reference annotations, which incorporate both contact-based RR and SpO_2 information. The resulting 2$\times$2 confusion matrix is presented in Table 8.3. When evaluated using class-balanced accuracy, the baseline camera-based method achieved a balanced accuracy of 67.9%. At the class level, the method identifies 39.6% of COBE episodes (299/755) and 96.1% of non-COBE episodes (3909/4068).

Table 8.3: Confusion matrix comparing camera-based method outputs against reference labels.

		Camera-derived RR		
		COBE	**No COBE**	**Total**
Reference Labels	**COBE**	299 (0.396)	456 (0.604)	755
	No COBE	159 (0.039)	3909 (0.961)	4068
	Total	458	4365	4823

8.7 ResNet for COBE detection using camera signals

I trained ResNet models (ResNet-18, ResNet-34, and ResNet-50) to classify COBE episodes using the camera-derived frame difference and PPGi signals. The network architectures and training procedures follow the 1D ResNet designs described in section 6.6.1, with the input modality changed to camera-derived signals, allowing a direct comparison of camera-based and contact-based sensor performance.

To assess the individual contributions of each camera-derived signal, I first trained the ResNet models using only the frame difference or only the PPGi signal. I then combined the two signals in a multimodal configuration using late fusion, evaluating whether integrating both camera-derived features improved COBE detection relative to single-signal models.

8.8 ConvNeXt for COBE detection using camera signals

I applied ConvNeXt to the frame difference and PPGi signals, largely following the architecture variants used in section 6.7.1. Although the camera-derived inputs in this chapter have lower temporal resolution (20 Hz) compared with the physiological signals (60 Hz), I retained the default ConvNeXt parameters. However, I adjusted the kernel size when experimenting with larger kernels (using 9 instead of 21), to better match the lower-resolution of camera-derived signals. I trained ConvNeXt models on each camera-derived signal individually and then on the combined frame difference and PPGi inputs.

8.9 Hybrid COBE episode detection using camera and physiological signals

I investigated whether integrating camera-based signals with physiological signals from the patient monitor improves COBE episode detection in the NICU. To achieve this, I developed a hybrid deep learning framework that combines at least one camera-derived signal with one or more physiological signals, enabling assessment of complementary information across these modalities. I trained each branch using the optimised configurations identified during single-modality experiments in Chapter 6 and section 8.7, ensuring that each input type contributed its best-performing representation. I designed this system to improve detection accuracy in real-world clinical conditions, where individual signals may be affected by noise, occlusion, or poor signal quality.

Physiological signals from the patient monitor, including the ECG, IP, and PPG, provide high temporal resolution but are susceptible to motion artefacts and electrode displacement. In contrast, camera-based measurements offer a non-contact alternative but can be affected by poor lighting, occlusions, and infant movement. To address these challenges, I fused both contact and non-contact modalities to exploit their complementary advantages, mitigate the risk of single-sensor failure and improve COBE detection.

8.9.1 Network architecture

I implemented a dual-branch architecture for hybrid COBE detection. One branch uses a ConvNeXt network optimised for physiological signals (IP, PPG, ECG), and the other uses a ResNet network, optimised for camera-derived signals (frame difference and PPGi). Each branch independently extracts modality-specific features, which I then concatenate using a late-fusion approach before the final classification layer.

I designed the fusion head with two fully connected layers and GELU activations, allowing the network to learn complex interactions between modalities rather than solely relying on fixed-weight averaging. This fusion process allows the model to emphasise more reliable signals based on quality and confidence, while preserving modality-specific information. Ultimately, a classification layer predicts the probability of a COBE episode within each analysis window. Figure 8.3 shows the dual-branch architecture and the fusion head used to integrate camera and physiological modalities.

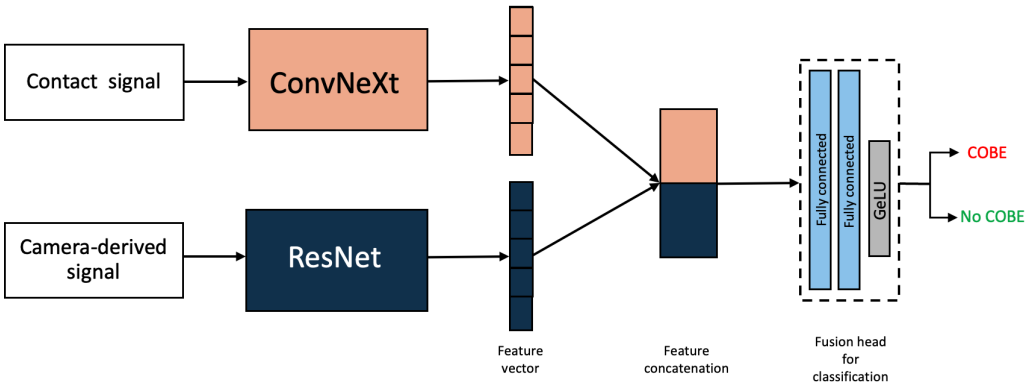


Figure 8.3: Hybrid COBE detection network. The ConvNeXt branch extracts features from physiological signals, while the ResNet branch extracts features from camera-derived signals. The resulting feature vectors are concatenated and processed by a two-layer fusion head (fully connected layers with GELU activations) before the final classification output.

8.9.2 Network training

I trained the ConvNeXt branch using the AdamW optimiser with weight decay regularisation, label smoothing, exponential moving average (EMA) of model parameters, and cosine learning rate scheduling, as described in section 6.7. I did not apply additional hyperparameter tuning, since I had already determined the optimal settings in Chapter 6. Similarly, I trained the ResNet branch using standard cross-entropy loss, without additional hyperparameter tuning, because I had already optimised these architectures for the camera-derived signals.

8.10 Results

In this section, I evaluate the performance of the ResNet and ConvNeXt models trained on the camera-derived frame difference and $PPGI_{rr}$ signals. I report results from five-fold cross-validation and from final evaluation on the held-out test set using the same performance metrics as in Chapter 6: F1 score, TPR, FPR, and accuracy. These metrics enable direct comparison between camera-derived models and their contact-based counterparts. I also present results from a hybrid model that combines the most effective camera-derived signals with patient monitor signals, to assess whether multimodal input improves COBE classification performance.

8.10.1 COBE episode detection using ResNet

Table 8.4 summarises the five-fold cross-validation performance of the ResNet models trained on the individual camera-derived signals and their multimodal combination. I evaluated multiple ResNet depths to determine whether deeper architectures improved COBE classification performance for camera-derived inputs.

Table 8.4: Cross-validation performance of ResNet models trained on individual camera-derived signals and multimodal combinations of camera-derived signals. Results are reported as mean $\pm$ standard deviation across five folds. For each signal configuration, the best-performing ResNet depth (ResNet-18, ResNet-34, or ResNet-50) is highlighted in bold.

Signal (s)	ResNet architecture	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
FD	18	0.57 $\pm$ 0.17	0.32 $\pm$ 0.10	0.27 $\pm$ 0.08	0.36 $\pm$ 0.09	0.18 $\pm$ 0.07	68.6 $\pm$ 10.4	57.1 $\pm$ 17.4	62.8 $\pm$ 4.9
	34	0.67 $\pm$ 0.10	0.41 $\pm$ 0.07	0.25 $\pm$ 0.05	0.36 $\pm$ 0.06	0.15 $\pm$ 0.03	59.0 $\pm$ 6.8	66.7 $\pm$ 9.9	62.8 $\pm$ 2.5
	50	**0.60 $\pm$ 0.12**	**0.33 $\pm$ 0.07**	**0.27 $\pm$ 0.08**	**0.36 $\pm$ 0.08**	**0.18 $\pm$ 0.07**	**66.5 $\pm$ 7.2**	**60.1 $\pm$ 12.3**	**63.3 $\pm$ 4.3**
$PPGi$	18	0.77 $\pm$ 0.09	0.38 $\pm$ 0.06	0.29 $\pm$ 0.07	0.41 $\pm$ 0.08	0.23 $\pm$ 0.07	61.7 $\pm$ 6.2	76.7 $\pm$ 9.3	69.2 $\pm$ 5.2
	34	**0.71 $\pm$ 0.14**	**0.34 $\pm$ 0.07**	**0.30 $\pm$ 0.07**	**0.42 $\pm$ 0.08**	**0.24 $\pm$ 0.07**	**66.3 $\pm$ 6.9**	**71.0 $\pm$ 14.4**	**68.7 $\pm$ 5.8**
	50	0.70 $\pm$ 0.04	0.33 $\pm$ 0.07	0.30 $\pm$ 0.09	0.41 $\pm$ 0.08	0.24 $\pm$ 0.07	67.5 $\pm$ 6.8	69.6 $\pm$ 4.3	68.5 $\pm$ 2.8
FD + $PPGi$	-	0.76 $\pm$ 0.07	0.39 $\pm$ 0.05	0.28 $\pm$ 0.07	0.41 $\pm$ 0.07	0.22 $\pm$ 0.06	61.4 $\pm$ 4.8	75.8 $\pm$ 6.8	68.6 $\pm$ 3.6

FD represent the frame difference signal

Following cross-validation, I selected the best-performing model configuration for each signal and signal combination and evaluated it on the independent held-out test set. Table 8.5 presents the resulting test-set performance metrics.

Table 8.5: Test set performance of the best-performing models identified during five-fold cross-validation (bolded in Table 8.4).

Signal (s)	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
FD	0.52	0.24	0.31	0.39	0.22	76.5	51.7	64.1
$PPGi$	0.80	0.26	0.38	0.52	0.37	73.6	80.2	76.9
FD + $PPGi$	0.70	0.24	0.37	0.49	0.34	75.9	69.8	72.9

FD represent the frame difference signal

8.10.2 COBE episode detection using ConvNeXt

Table 8.6 summarises the cross-validation performance of ConvNeXt models trained using different architectural configurations for the frame difference and $PPGi_{rr}$ signals. I evaluated modifications to kernel size and feature dimensionality to assess whether architectural adaptations improved performance for camera-derived inputs.

Table 8.6: Cross-validation performance of ConvNeXt models trained on individual signals. Results are reported as mean ± standard deviation across five folds.

Signal	ConvNeXt architecture	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
FD	**Default**	**0.71 ± 0.10**	**0.43 ± 0.06**	**0.25 ± 0.06**	**0.36 ± 0.08**	**0.16 ± 0.04**	**57.1 ± 5.5**	**70.4 ± 10.0**	**63.7 ± 2.6**
	Smaller Kernels	0.69 ± 0.09	0.43 ± 0.05	0.24 ± 0.06	0.36 ± 0.07	0.15 ± 0.03	57.0 ± 4.9	68.8 ± 8.7	62.9 ± 2.1
	Larger Kernels	0.71 ± 0.10	0.49 ± 0.04	0.23 ± 0.06	0.35 ± 0.08	0.12 ± 0.05	51.6 ± 4.4	71.2 ± 9.5	61.4 ± 2.7
	Reduced Features	0.72 ± 0.07	0.52 ± 0.09	0.22 ± 0.07	0.33 ± 0.08	0.11 ± 0.06	48.2 ± 8.9	71.3 ± 6.8	59.7 ± 3.6
$PPGi$	**Default**	**0.79 ± 0.06**	**0.42 ± 0.09**	**0.28 ± 0.06**	**0.41 ± 0.06**	**0.22 ± 0.06**	**58.4 ± 9.3**	**78.8 ± 5.8**	**68.6 ± 3.8**
	Smaller Kernels	0.79 ± 0.07	0.45 ± 0.08	0.26 ± 0.05	0.39 ± 0.06	0.19 ± 0.05	55.5 ± 8.2	79.1 ± 7.1	67.3 ± 4.1
	Larger Kernels	0.79 ± 0.07	0.44 ± 0.08	0.27 ± 0.06	0.39 ± 0.07	0.19 ± 0.06	55.8 ± 7.6	79.2 ± 6.9	67.5 ± 3.9
	Reduced Features	0.80 ± 0.07	0.45 ± 0.07	0.26 ± 0.06	0.39 ± 0.07	0.19 ± 0.05	55.1 ± 7.1	79.4 ± 6.9	67.2 ± 3.6

FD represent the frame difference signal

Based on the cross-validation experiments, I selected the best-performing ConvNeXt configurations and evaluated them on the held-out test set. Table 8.7 summarises the resulting performance metrics.

Table 8.7: Performance metrics for multi-modal models trained on camera-derived frame difference (FD) and PPGi signals. Results are shown for both cross-validation and held-out test sets.

Signal	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
FD	0.72	0.45	0.25	0.37	0.16	55.18	72.41	63.80
$PPGi$	0.76	0.34	0.31	0.44	0.27	66.08	75.86	70.97

FD represent the frame difference signal

8.10.3 Hybrid COBE episode detection

Table 8.8 summarises the performance of the hybrid models I tested on the unseen test set under different input combinations. Each configuration includes at least one camera-derived signal combined with one or more physiological signals. I report performance metrics for models containing one, two, or three physiological signals. Extended hybrid results, including additional signal combinations and cross-validation outcomes, are provided in Appendix G.

Table 8.8: Summary of final hybrid model performance on the unseen test set. Each row represents a combination of at least one camera-derived signal and at least one physiological signal.

Signal combination	Performance Metrics							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
One camera + one physiological signal								
FD + IP	0.81	0.08	0.68	0.74	0.68	92.1	81.0	86.6
$PPGi_{rr}$ + IP	0.92	0.11	0.63	0.75	0.68	88.9	92.2	90.6
Two cameras + one physiological signal								
FD + $PPGi_{rr}$ + IP	0.91	0.12	0.60	0.72	0.65	87.7	90.5	89.1
FD + $PPGi_{rr}$ + PPG	0.80	0.29	0.36	0.50	0.34	71.0	80.2	75.6
FD + $PPGi_{rr}$ + ECG	0.72	0.19	0.44	0.55	0.42	81.0	72.4	76.7
Two cameras + three physiological signals								
FD + $PPGi_{rr}$ + IP + PPG + EDR	0.76	0.08	0.66	0.71	0.64	92.1	75.9	84.0

FD represent the frame difference signal

8.11 Discussion

This chapter investigated the feasibility of detecting COBE episodes in pre-term infants using camera-derived signals, both independently and in combination with clinical physiological measurements. For this analysis, I used only video segments in which the infant was visible and the video recording was in focus. This resulted in a reduced dataset (compared to Chapter 6) consisting of 23 infants and 246 annotated COBE episodes. Despite this reduction in the size of the dataset, my experiments show that camera-based respiratory and motion features contain useful information for distinguishing COBE episodes from non-events.

Camera-only models show that non-contact monitoring captures respiratory motion patterns that can be used to detect COBE episodes, highlighting their potential application in low-resource settings where access to conventional physiological sensors may be limited. Among the camera-derived signals, frame-difference (FD) provides overall torso movement, including both respiratory and non-respiratory motion, while $PPGi_{rr}$, extracted from a small ROI in the lower abdomen, captures localised breathing motion and reflects respiratory patterns similar to thoracic impedance measurements. Across both ResNet and ConvNeXt architectures, $PPGi_{rr}$ consistently outperformed FD, achieving a test-set balanced accuracy of 76.9% compared with 64.1% for FD. This also represents a notable improvement over the simple baseline camera-based method, which relied on thresholding RR estimated from the $PPGi_{rr}$ signal using signal processing methods and achieved a balanced accuracy of 67.9%. For the camera-only models, the best-performing approach using $PPGi_{rr}$ achieved a COBE-class accuracy of 80.2%, representing a substantial improvement over the simple baseline camera-based method, which identified only 39.6% of COBE episodes (299/755). For non-COBE episodes, $PPGi_{rr}$ achieved 73.6% accuracy, lower than

the baseline method's 96.1%, reflecting a trade-off between improved detection of COBE events and a modest reduction in specificity. These results suggest that visually extracted breathing patterns provide a more direct indicator of respiratory instability than general movement alone, and that learning-based feature extraction provides more reliable detection of COBE episodes than simple signal processing and threshold-based approaches.

Combining FD and $PPGi_{rr}$ resulted in a lower test accuracy of 72.9%, suggesting that FD introduces additional variability that can obscure the respiration-specific signal captured by $PPGi_{rr}$. This likely reflects the small dataset size, which limits the model's ability to learn to differentiate between respiratory and non-respiratory motion captured in the frame difference signal. With more training data, the network may learn to better distinguish respiration-specific patterns from other sources of movement. For all signal configurations, performance on the independent test set is slightly higher than the mean cross-validation performance observed during training. However, the difference between test and mean cross-validation performance falls within the variability observed across cross-validation folds. This behaviour is consistent with training on small clinical datasets, where individual validation folds may contain a higher proportion of challenging segments or infants with noisier respiratory signals. Such fold-specific difficulty can reduce average cross-validation performance, even when the model's overall ability to generalise to unseen data remains comparable. Overall, these results demonstrate the feasibility of using camera-derived signals to detect COBE episodes in pre-term infants, and highlight the advantages of learning-based approaches over a simple threshold-based method for COBE episode detection. Such models have several potentially important applications in settings where contact sensors are impractical or unavailable due to recurring consumable costs and maintenance requirements, including in low- and middle-income countries. For example, my models could be deployed for episodic COBE detection using smartphones, which integrate cameras and on-device processing capabilities within a single, widely available platform.

My work shows that camera-derived signals can also complement existing physiological monitoring by further improving detection accuracy or reducing sensor burden by partially replacing contact-based measurements. Hybrid models combining camera signals with the IP achieved the strongest held-out test performance. In particular, pairing camera-based $PPGi_{rr}$ with contact IP provided the highest overall test accuracy of 90.6%, better than any of the test results using contact signals in Chapter 6, where IP alone achieved 88.0 % and multimodal combinations provided only marginal improvements (88.7 % test accuracy for combination of IP and PPG). In contrast, combining $PPGi_{rr}$ with PPG resulted in only a marginal increase in performance, from 76.9% ($PPGi_{rr}$ only) to 77.1% ($PPGi_{rr}$ + PPG), highlighting similarities in the physiological information captured by both signals and the importance of carefully selecting which contact and camera-based signals to combine.

Although the absolute improvement in balanced accuracy between the best contact-only model and the hybrid $PPGi_{rr}$ + IP model was modest (+2.6 percentage points), this corresponds to an approximate 21.7% relative reduction in classification error (from 12.0% to 9.4% error). At already high balanced accuracy, further performance gains become progressively more difficult to achieve, particularly when working with noisy neonatal physiological signals and relatively small clinical datasets. Clinically, even modest improvements in balanced accuracy may reduce missed COBE episodes or false alarms, potentially improving monitoring reliability and reducing unnecessary clinical interventions in the NICU. To assess whether the observed improvement was robust despite the limited dataset size, I compared fold-wise balanced accuracies between the best contact-only model and the hybrid $PPGi_{rr}$ + IP model using a Wilcoxon signed-rank test [356]. The resulting comparison did not reach the conventional 0.05 significance threshold (two-sided $p = 0.0625$). However, because the study used a relatively small cohort and only five subject-wise cross-validation folds primarily for model selection and evaluation, this statistical comparison should be interpreted cautiously. Future work using larger datasets or leave-one-subject-out evaluation strategies may enable more robust statistical comparison of hybrid and contact-only approaches.

A comparison of per-class performance further emphasises the advantages of hybrid models combining both contact and camera-based physiological signals. Combining camera-based $PPGi_{rr}$ with contact IP substantially improved performance for both the COBE and non-COBE classes compared to the best performing model using a combination of only contact signals in Chapter 6. Including camera-based signals increased normal class accuracy notably from 83.9% (IP + PPG) to 88.9% (IP + $PPGi_{rr}$) while COBE class accuracy remained comparable (93.6% with IP + PPG and 92.2% with IP + $PPGi_{rr}$).

The FD signal, while not particularly effective for COBE detection on its own, can improve performance when combined with contact signals. For example, combining FD with ECG significantly improved accuracy from 64.1% (FD alone) and 69.7% (ECG alone) to 78.2%, suggesting that gross movement context from the camera complements contact-based respiratory information from the ECG. This emphasises that camera-derived movement information provides valuable context, indicating periods of infant motion or stillness, which helps models interpret physiological signals more accurately. Such contextual information is a key advantage of non-contact monitoring and could inform future multi-modal approaches. However, combined camera- ECG setups may be less practical than camera- PPG configurations due to sensor placement and cost. Notably, ECG and IP sensors are chest-worn, while PPG is typically foot-worn in pre-term infants, non-adhesive, and affordable, making it more compatible with low-cost, wearable setups. The benefits of hybrid models also do not extend to indiscriminate fusion of all available signals because combining every camera and physiological input resulted in a lower accuracy of 84.0%, suggesting that adding signals without careful selection may introduce redundant information, leading to reduced

performance, rather than additional, informative features.

8.12 Conclusion

This chapter established the feasibility and potential use of camera-based respiratory monitoring in pre-term infants. The results show that non-contact, video-derived signals contain clinically relevant information and can be used for COBE detection in pre-term infants either independently or in combination with conventional physiological measurements. In particular, hybrid models that combine camera-derived and physiological signals outperformed models trained on clinically acquired data alone. This improvement reflects the additional context provided by visual monitoring. For example, movement captured by the frame-difference signal, including involuntary movements during attempts to breathe, provides complementary information alongside physiological signals and can improve detection of COBE episodes.

A key limitation of this work is the reduced dataset size, especially due to the exclusion of low-quality recordings, reflecting the practical challenges of deploying camera-based monitoring in neonatal environments. Despite this, the findings establish a foundation for non-contact COBE episode detection as a valuable tool in both clinical and low-resource settings, including the potential for smartphone-based monitoring. Future work should therefore focus on larger, more diverse datasets to improve model reliability and generalisability. In addition, explicitly modelling infant motion, for example as a signal quality indicator or contextual feature, may help differentiate involuntary movements during respiratory effort associated with COBE episodes. More advanced approaches, including models trained directly on raw video data, may further exploit spatial and temporal context to enhance COBE detection performance.

Chapter 9

Conclusion and future work

This thesis addressed the clinical need for accurate and scalable detection of cessation of breathing events (COBE) in pre-term infants by investigating both contact-based and non-contact physiological monitoring approaches. Pre-term neonates are particularly vulnerable to COBE and its associated complications, including hypoxaemia and adverse neurodevelopmental outcomes. Despite widespread use of impedance pneumography and cardiorespiratory monitors in neonatal intensive care units, existing contact-based systems can fail to provide reliable, continuous coverage due to motion artefacts and poor electrode adherence. These challenges are further exacerbated in low-resource healthcare settings, where access to consumables and trained personnel is limited, motivating the need for scalable and non-contact alternatives.

To address these limitations, I developed a series of signal processing and deep learning pipelines to detect COBE from both conventional monitoring data and video camera recordings. The clinical and scientific contributions of this thesis are summarised below.

9.1 Summary of key findings

The primary aim of this thesis was to investigate whether non-contact, video-derived physiological signals can be used to accurately detect COBE in neonates, and whether combining these signals with conventional patient monitor data improves detection performance. Specifically, the research focused on:

- Developing non-contact algorithms to estimate vital signs during periods of physiological distress.
- Evaluating the potential and effectiveness of continuous video camera-based detection of cessation of breathing in multiple pre-term infants nursed in the high-dependency area of the NICU.
- Assessing the application of advanced deep learning techniques in detecting COBE in pre-term infants.

9.1.1 Major results and contributions

The first contribution of this thesis, presented in Chapter 7, is the development of methods for non-contact vital sign estimation during periods of significant movement surrounding a COBE episode. These methods include the extraction of motion and respiratory signals from video, and the adaptation of signal processing

techniques to estimate respiratory rate under real-world clinical constraints. Using RGB video in this study, respiratory estimates achieved a mean absolute deviation of 12.0 breaths/min, demonstrating feasibility in the challenging context of pre-term infants experiencing movement and physiological instability.

The second and primary contribution of this thesis is the development of both contact (Chapter 6) and non-contact (Chapter 8) algorithms for detecting COBE in pre-term infants nursed in the NICU. This work included the extraction of respiratory features from both conventional sensors and video camera data, as well as the construction of an annotated dataset suitable for training deep learning models. Independent clinicians and engineers annotated each COBE event to ensure accurate ground truth. Using the dataset, I processed contact-based signals (IP, ECG, and PPG) using classical signal analysis and benchmarked against deep learning models. I also extracted camera-derived signals, including frame-difference and PPGi, using anatomically informed, dynamic ROIs in Chapter 8, enabling consistent signal extraction across varying infant poses.

Machine learning approaches for neonatal COBE detection have commonly relied on SVMs, autoencoders, and simpler CNN or LSTM models applied to small datasets, which limited their potential for generalisation to broader NICU cohorts. Using IP, ECG, and PPG, I trained CNN, ResNet, and ConvNeXt models for COBE detection, representing one of the first reported uses of ConvNeXt architecture for COBE episode detection in pre-term infants. ConvNeXt achieved the highest performance, with accuracies up to 88.7% and an F1 score of 0.75 in a late-fusion configuration of IP and PPG, demonstrating that advanced deep learning architectures can effectively integrate multiple physiological signals for COBE detection in this population.

Non-contact video-based COBE detection in pre-term infants has also been limited to small cohorts or single-subject studies, typically relying on motion analysis and frame differencing to extract respiratory signals. While demonstrating feasibility, these studies did not explore machine learning approaches for accurate episode detection across multiple infants. Extending these approaches, this thesis shows that video-derived signals can be effectively incorporated into deep learning models, with $PPGi_{rr}$ achieving up to 76.9% accuracy.

Hybrid models that combined video and conventional physiological signals achieved the best overall performance across all models, demonstrating that multimodal fusion captures complementary information unavailable to unimodal approaches. Specifically, the combination of $PPGi_{rr}$ and IP showed the highest overall performance, with 90.6% accuracy and an F1 score of 0.75. These two metrics together establish that non-contact monitoring can match or augment conventional contact-based systems in neonatal care.

A further practical contribution of this thesis is the development of a MATLAB-based annotation tool and user-friendly interface that enables rapid labelling of neonatal clinical data through keystroke input. This tool supported the generation of high-quality ground truth for COBE episode detection by allowing multiple annotators to label large volumes of data efficiently. While not a scientific output, this tool was crucial in generating the dataset used to train and evaluate the deep learning models proposed in this thesis.

9.2 Implications of the work

9.2.1 Scientific and methodological implications

This work demonstrates that deep learning can achieve state-of-the-art performance in neonatal COBE episode detection when applied to clinically acquired physiological signals. The results highlight the potential of modern architectures, particularly ConvNeXt, to handle noisy neonatal data and to benefit from multimodal fusion strategies. I establish the feasibility of non-contact monitoring, showing that video cameras can provide physiologically meaningful signals suitable for machine learning-based COBE episode detection. Together, these findings open avenues for the design of monitoring systems that are less intrusive, more flexible, and better suited to integration with existing clinical infrastructure.

9.2.2 Clinical and practical implications

Clinically, this work highlights opportunities to improve neonatal care through the development of robust and minimally invasive techniques. Non-contact monitoring could reduce reliance on adhesive electrodes and wired sensors, minimising the risks of skin injury and facilitating closer parent-infant interaction during kangaroo care. Camera-based monitoring using low-cost and widely available hardware is especially relevant in LMICs where access to specialised monitors may be limited. In the NICU, automated multimodal COBE episode detection could reduce false alarms, provide earlier alerts to clinicians, and support more timely interventions. Importantly, hybrid approaches can complement, rather than replace existing systems, offering reliability in critical care environments.

Translating this work into clinical practice will require the integration of the proposed multimodal pipeline into existing NICU monitoring systems and conducting prospective clinical trials to evaluate real-time performance. Such trials could assess not only the accuracy and reliability of COBE detection, but also usability and workflow impact for clinical staff, as well as parental acceptance and trust in non-contact monitoring. In low-resource settings, simplified video-based solutions using smartphones, which combine cameras and processing capabilities in a single device, could provide affordable and scalable alternatives to conventional monitoring, supporting early warning, triage, and prioritisation of care.

Additionally, implementation studies could explore strategies for integrating non-contact monitoring into existing electronic health record systems, automated alerting platforms, and clinical decision support tools, ensuring that these technologies enhance existing clinical workflows without adding to staff workload. Long-term evaluations could further assess whether non-contact alone in low-resource environments, or hybrid monitoring approaches in well-resourced settings, contribute to reductions in adverse outcomes, such as hypoxaemia, delayed interventions, or prolonged episodes of COBE, providing evidence for their safe and effective adoption across different NICU environments.

9.3 Limitations

Despite promising results, several limitations should be acknowledged. Camera-derived signals were sensitive to occlusion, skin tone, lighting variation, and infant movement. While dynamic ROI placement improved robustness, performance could still degrade under adverse conditions. In addition, physiological signals in neonates are inherently noisy and prone to artefacts, which complicates annotation and model training. ConvNeXt may be data-hungry and less interpretable than simpler models. Finally, the dataset, though clinically relevant, was not large or demographically diverse enough to support generalisability claims across all NICU populations or LMICs settings. Future validation across larger and more heterogeneous populations is therefore required.

9.4 Future work

While this thesis focused on detecting COBE events as they occur, future work should focus on complementary work, such as identifying early physiological markers that precede COBE events, and adapting these methods for broader paediatric populations and low-resource settings, where continuous contact-based monitoring may not be feasible. Collectively, these approaches aim to develop scalable, non-invasive monitoring tools that support early warning and triage in both high- and low-resource healthcare environments.

9.4.1 Heart rate estimation during COBE

While respiratory rate is the primary focus in monitoring COBE, heart rate provides complementary information, particularly for detecting bradycardic events associated with COBE episodes. Pre-term infants exhibit higher baseline heart rates (120 - 180 beats/min) compared to adults, which can decrease significantly during COBE episodes, sometimes resulting in bradycardia, indicating physiological stress or compromised oxygenation in the infant [146, 150]. Existing non-contact heart rate estimation methods

perform well during stable physiological periods, but may fail to capture critical changes during these events [88, 352, 240].

In preliminary work, I estimated heart rate from video using a subset of the COBE dataset with 17 clinically annotated episodes [240]. Following previous observations that heartbeats are more prominent in larger regions of interest while smaller ROIs (25 × 25 and 50 × 50 pixels) are more affected by noise and motion related to breathing [318, 347], I placed a 100 × 100 pixel region of interest near the infant's shoulder to capture exposed skin. I then computed the average pixel intensity within this ROI over time to extract cardiac signals. I adapted the filter-switching technique for heart rate estimation during COBE episodes and applied classical signal processing methods to account for sudden changes in lighting conditions and infant movement. Figure 9.1 shows the potential of this approach, showing motion, respiratory rate, heart rate, and oxygen saturation during a typical COBE episode.

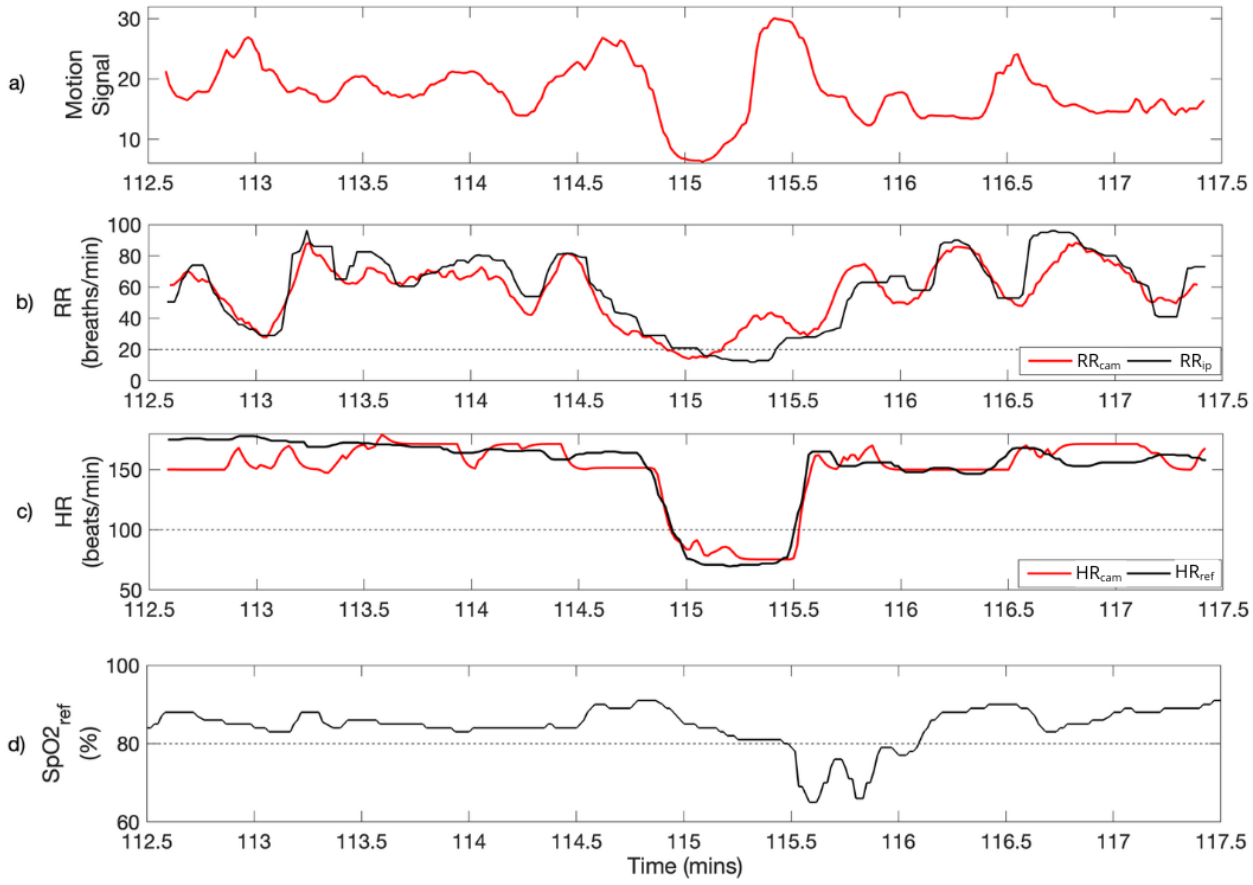


Figure 9.1: Video-based estimation of heart rate during a COBE episode of a 28 week old female pre-term infant. a) Motion signal; b) Camera-derived respiratory rate; c) HR_{cam} compared to reference HR (HR_{ref}); d) Oxygen saturation (SpO_2) during a period of desaturation.

These initial experiments demonstrate that camera-derived heart rate estimates can track changes during COBE. This work, though limited to a small subset of patients, highlights the potential for integrating

heart rate estimation with non-contact respiratory monitoring. Future work will extend these methods to the full COBE dataset, refine signal extraction techniques, and explore multimodal monitoring to detect both COBE events and associated bradycardia.

9.4.2 Physiological indicators of impending COBE

There is significant potential in identifying early physiological markers that precede COBE episodes. Subtle patterns, including decreased respiratory variability and alterations in heart rate dynamics, may indicate an impending apnoeic episode [357]. By incorporating both heart rate and respiratory features, future models could capture dynamic features over several minutes leading up to COBE events. By training models on longer temporal windows and explicitly labelling pre-COBE segments, researchers may identify latent precursors not visible in snapshot classification. While the above approaches focus on NICU infants, similar non-contact monitoring principles could be adapted for broader paediatric populations in low-resource settings.

9.4.3 Detection of respiratory patterns associated with health decline in LMIC infant populations

Developing lightweight, embedded models that can run on mobile devices may allow the deployment of smartphone tools in rural clinics with limited infrastructure. While this thesis focused on COBE episode detection in neonates, similar camera-based methods using smartphones may also help with the detection of other respiratory conditions in paediatric populations in LMICs. Collaboration with public health systems can ensure that the resulting early warning indicators are usable, interpretable, and ethically appropriate for the specific healthcare and cultural contexts in which they are deployed.

I also contributed to the setup of a clinical study in Bungoma and Mama Lucy Counties in Kenya. The study involves infants between 2 and 59 months who present with symptoms of pneumonia, including chest indrawing and respiratory distress. It investigated the feasibility of capturing video recordings of paediatric patients in a low-resource setting with the goal of developing algorithms for classifying respiratory patterns. Although analysis of this data falls outside the scope of the current thesis, it represents a valuable opportunity to extend the camera-based methods developed here to older paediatric populations and to acute respiratory conditions relevant to LMICs.

9.5 Concluding remarks

This thesis has demonstrated the feasibility of combining deep learning with non-contact video monitoring to detect cessations of breathing in pre-term infants. By establishing a clinically annotated dataset,

advancing machine learning methods for COBE episode detection, and introducing a novel pipeline for camera-based vital sign monitoring, I have been able to contribute both scientific and practical advances. The findings suggest a path toward less invasive and more scalable monitoring solutions that can benefit vulnerable infants worldwide. Ultimately, this work lays the foundation for the development of multi-modal, non-contact neonatal monitoring systems that could transform clinical practice in NICUs and extend life-saving monitoring technologies to low-resource settings.

Appendix A
Neonatal HDU Study Supplementary Data

A.1 Participant Demographics and Recording Details

Table A.1: Demographics of infants enrolled in the Neonatal HDU Study

Participant	Recording session	Gestational age (weeks)	Class	Ethnicity	Gender
PN001	Day 1	31.3	Very pre-term	White	Male
	Day 2	31.4	Very pre-term		
	Day 3	31.6	Very pre-term		
PN002	Day 1	27.6	Extreme pre-term	White	Female
	Day 2	27.7	Extreme pre-term		
	Day 3	27.9	Extreme pre-term		
	Day 4	28.0	Very pre-term		
PN003	Day 1	28.7	Very pre-term	White	Female
	Day 2	28.9	Very pre-term		
	Day 3	29.0	Very pre-term		
PN004	Day 1	33.4	Moderate pre-term	White	Male
	Day 2	33.6	Moderate pre-term		
	Day 3	33.7	Moderate pre-term		
	Day 4	33.9	Moderate pre-term		
PN005	Day 1	36.4	Late pre-term	Asian	Male
PN006	Day 1	33.4	Moderate pre-term	White	Male
	Day 2	33.6	Moderate pre-term		
	Day 3	33.7	Moderate pre-term		
PN007	Day 1	32.0	Moderate pre-term	White	Female
	Day 2	32.1	Moderate pre-term		
	Day 3	32.3	Moderate pre-term		
PN008	Day 1	30.1	Very pre-term	White	Female
	Day 2	30.3	Very pre-term		
	Day 3	30.4	Very pre-term		
PN009	Day 1	30.9	Very pre-term	Black	Male
	Day 2	31.0	Very pre-term		
	Day 3	31.1	Very pre-term		
PN010	Day 1	30.1	Very pre-term	White	Male
	Day 2	30.3	Very pre-term		
	Day 3	30.4	Very pre-term		
PN011	Day 1	28.7	Very pre-term	White	Female
	Day 2	28.9	Very pre-term		
	Day 3	29.0	Very pre-term		
PN012	Day 1	31.0	Very pre-term	Asian	Female
PN013	Day 1	28.0	Very pre-term	Mixed [a]	Female
	Day 2	28.1	Very pre-term		
	Day 3	28.3	Very pre-term		
PN014	Day 1	29.4	Very pre-term	White	Female
	Day 2	29.6	Very pre-term		
	Day 3	29.7	Very pre-term		
	Day 4	29.9	Very pre-term		
PN015	Day 1	31.7	Very pre-term	Mixed [a]	Male
	Day 2	31.9	Very pre-term		
	Day 3	32.0	Moderate pre-term		
	Day 4	32.1	Moderate pre-term		

[a] Mixed White & Asian

Participant	Recording session	Gestational age (weeks)	Class	Ethnicity	Gender
PN016	Day 1	32.7	Moderate pre-term	White	Female
	Day 2	32.9	Moderate pre-term		
PN017	Day 1	29.6	Very pre-term	White	Male
	Day 2	29.7	Very pre-term		
	Day 3	29.9	Very pre-term		
PN018	Day 1	30.4	Very pre-term	White	Male
	Day 2	30.6	Very pre-term		
	Day 3	30.7	Very pre-term		
PN019	Day 1	29.6	Very pre-term	Mixed Other	Female
	Day 2	29.7	Very pre-term		
	Day 3	29.9	Very pre-term		
PN020	Day 1	30.4	Very pre-term	Mixed Other	Male
	Day 2	30.6	Very pre-term		
	Day 3	30.7	Very pre-term		
	Day 4	30.9	Very pre-term		
PN021	Day 1	33.3	Moderate pre-term	White	Male
	Day 2	33.4	Moderate pre-term		
	Day 3	33.6	Moderate pre-term		
PN022	Day 1	29.6	Very pre-term	White	Male
	Day 2	29.7	Very pre-term		
	Day 3	29.9	Very pre-term		
PN023	Day 1	30.0	Very pre-term	White	Female
	Day 2	30.1	Very pre-term		
	Day 3	30.3	Very pre-term		
PN024	Day 1	31.7	Very pre-term	Mixed [b]	Male
	Day 2	31.9	Very pre-term		
	Day 3	32.0	Moderate pre-term		
PN025	Day 1	32.1	Moderate pre-term	Mixed [b]	Male
	Day 2	32.3	Moderate pre-term		
PN026	Day 1	34.3	Late pre-term	White	Male
	Day 2	34.4	Late pre-term		
	Day 3	34.6	Late pre-term		
	Day 4	34.7	Late pre-term		
PN027	Day 1	31.3	Very pre-term	Mixed [a]	Male
	Day 2	31.4	Very pre-term		
PN028	Day 1	31.0	Very pre-term	White	Female
	Day 2	31.1	Very pre-term		
	Day 3	32.3	Very pre-term		
PN029	Day 1	30.6	Very pre-term	White	Male
	Day 2	30.7	Very pre-term		
	Day 3	30.9	Very pre-term		
	Day 4	31.0	Very pre-term		
PN030	Day 1	32.1	Moderate pre-term	White	Male
	Day 2	32.3	Moderate pre-term		
	Day 3	32.4	Moderate pre-term		

[b]Mixed White & Black [a]Mixed White & Asian

A.2 Recording Timelines

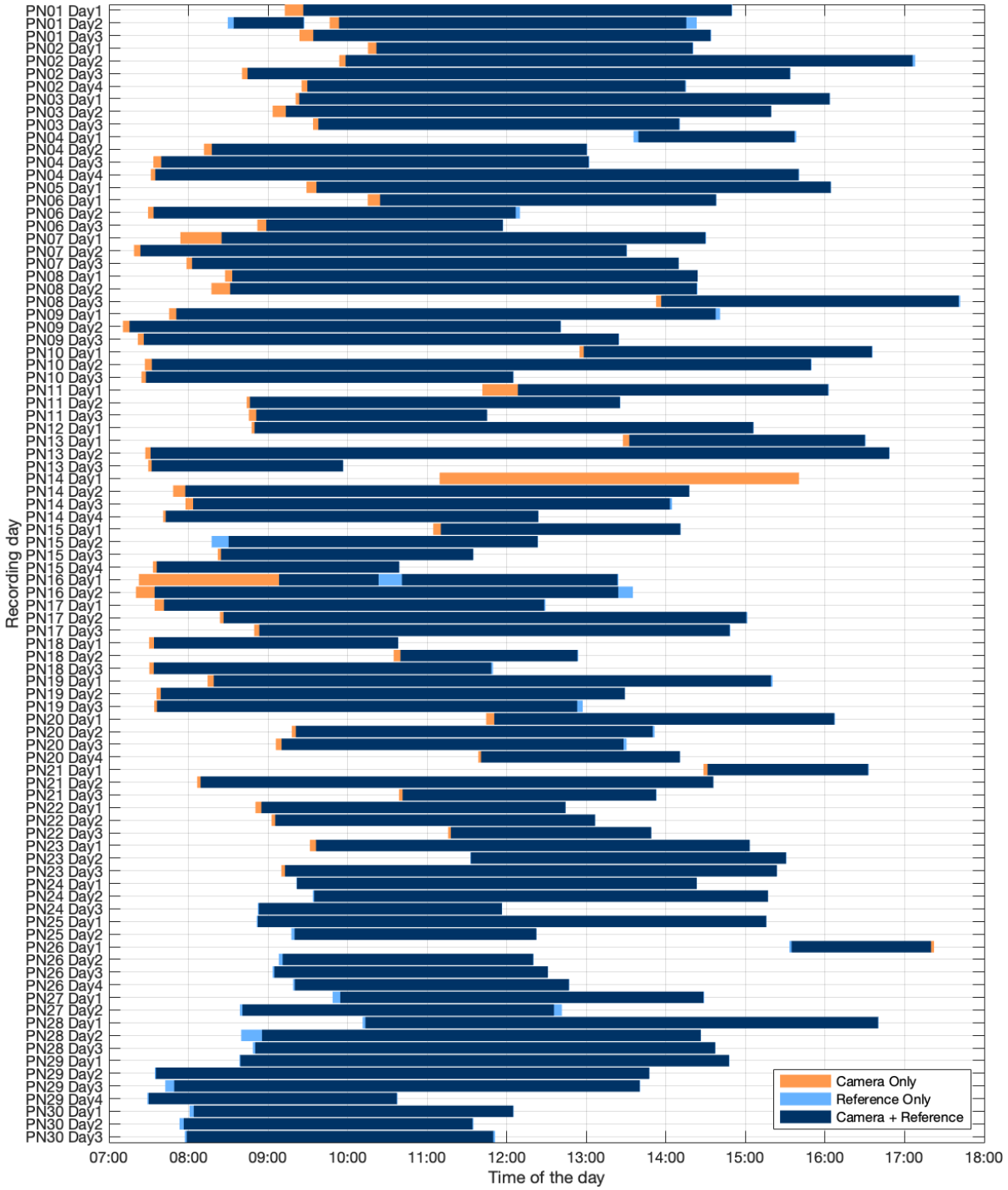


Figure A.1: Recording times for each session in the Neonatal HDU study. The recording of reference data usually started a few minutes after the video camera.

Appendix B

MATLAB annotation interface

This appendix presents the MATLAB-based graphical user interface (GUI) developed for expert annotation of candidate COBE episodes. The interface was designed to support manual review of synchronised physiological signals extracted from the Neonatal HDU dataset, including ECG, PPG, IP, recomputed respiratory rate (RR_{ip}), monitor-derived heart rate, and SpO_2. Annotators used the interface to determine whether candidate segments satisfied the study definition of COBE, characterised by either (1) periods where RR_{ip} remained below 20 breaths/min for at least 20 s, or (2) shorter respiratory pauses lasting at least 10 s accompanied by oxygen desaturation ($SpO_2 \leq 80\%$) [22, 23].

Primary annotation decisions were initially based on the RR_{ip} and SpO_2 signals. For difficult annotation cases or suspected signal artefacts, candidate segments could then be reloaded in an extended review mode containing supplementary ECG, PPG, and raw IP waveforms to support annotation decisions. Figure B.1 shows the MATLAB-based annotation interface used for this extended multi-signal review process. The interface supported zooming, signal rescaling, and detailed inspection of ambiguous segments containing possible motion artefacts or sensor disconnections.

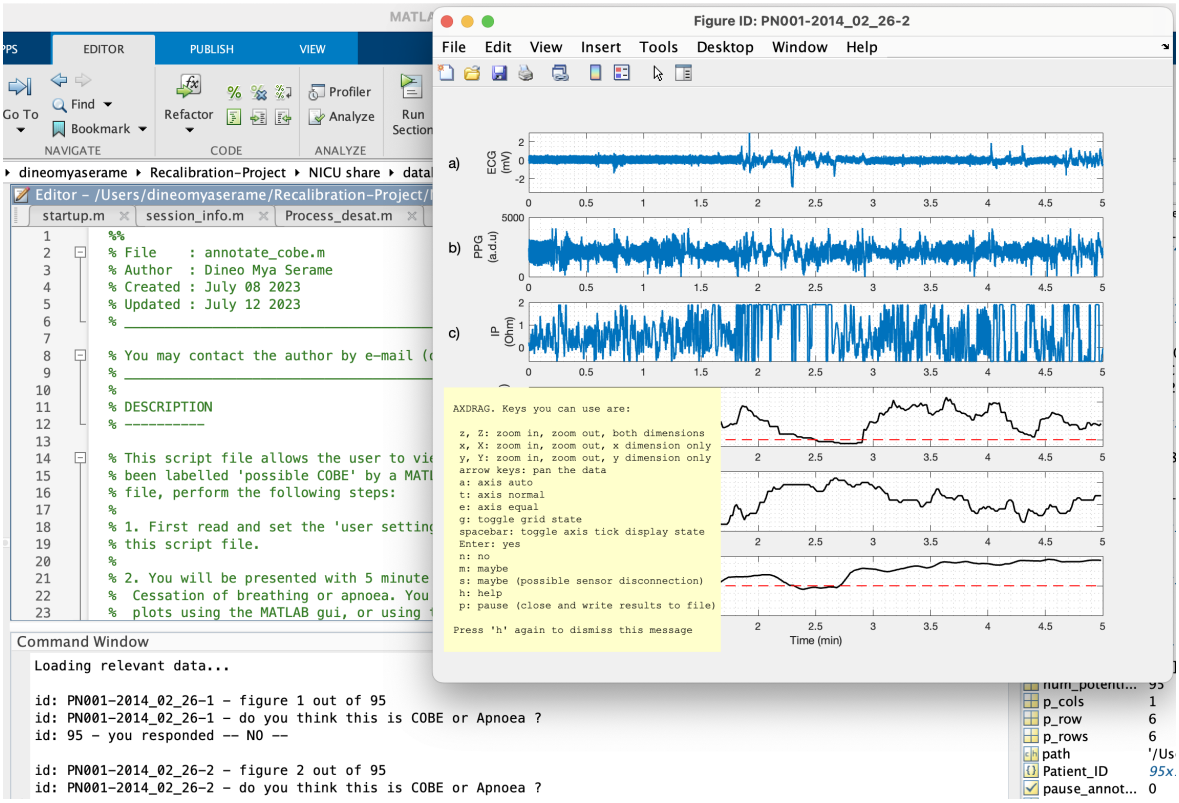


Figure B.1: MATLAB annotation interface used for expert review of candidate COBE episodes. The GUI displays multiple synchronised physiological signals, including ECG, PPG, IP, RR, HR, and SpO_2, alongside annotation controls and keyboard shortcuts. The command window shows an example segment already annotated as “NO”.

Appendix C
Convolutional neural network

Convolutional Neural Networks(CNNs) have transformed the field of computer vision and image recognition. Inspired by the biological visual cortex,CNNs were initially designed to automatically and adaptively learn spatial hierarchies of features from input images. Over time, their versatility has been expanded to process data of varying dimensionalities, including 1D data such as ECG. Unlike traditional fully connected neural networks, CNNs utilise specialised layers, including convolutional layers, pooling layers, and non-linear activation functions, to extract spatial patterns in data efficiently.

In this appendix, I provide an overview of the essential components of CNNs, showing mathematical formulations of how these layers contribute to the overall functionality of CNNs.

C.1 The architecture of a simple CNN

A typical CNN model consists of a series of layers stacked together. The main layers include convolution layers, pooling layers, and fully connected layers. Several important components, such as activation functions, flattening layers, normalisation layers, and dropout layers, complement these layers. Each layer plays a vital role in transforming input data into meaningful representations, making CNN highly effective for tasks such as image recognition and classification. Figure C.1 shows the basic CNN architecture, classifying obstructive sleep apnoea from a single lead ECG signal. A comprehensive layout of and mathematical approach to CNNs is presented in Appendix C.

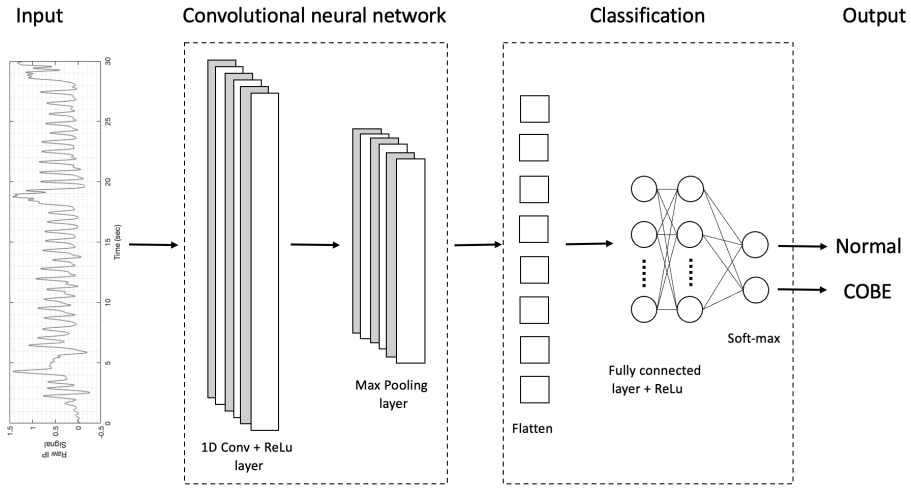


Figure C.1: A basic CNN architecture. A CNN typically consists of two main components: a feature extraction stage and a classification stage. The feature extraction stage uses convolution layers with ReLU activation to learn temporal patterns from the input signal, followed by max pooling layers for downsampling. The resulting feature maps are then flattened and passed to fully connected layers, which perform classification using a softmax activation.

C.2 Network elements

Convolutional layer

A convolutional layer is an essential part of CNNs used primarily for feature extraction from input data. In a convolutional layer, a set of trainable filters (also known as kernels) are slid over the input data, computing dot products between the filters and local regions of the input to produce feature maps. This process continues until the filter has covered the entire input sequence. The process is called the convolutional operation [269] and can be described mathematically in the l_{th} layer as follows:

$$y_k^l = b_k^l + \sum_{i=1}^{N} w_k^{l-1} \cdot X_i^{l-1} \tag{C.1}$$

where y_k represents the k_{th} feature map in layer l, b denotes the bias of the k_{th} feature map in layer l, X_i represents the i_{th} feature map in layer $l-1$, and w_k^{l-1} is the kernel from all features in layer $l-1$ to the k_{th} feature map in layer l. N denotes the total number of features in layer $l-1$.

Batch normalisation

During training, a batch normalisation layer is added between the convolutional and ReLU layers. This layer addresses issues related to internal covariate shifts within feature maps. Internal covariate shift refers to changes in the distribution of hidden units' values due to updates in model parameters, which can impede convergence and require a lower learning rate and careful parameter initialisation [358, 359]. Batch normalisation normalises each batch's activations, stabilising the learning process and allowing the use of higher learning rates. This normalisation introduces noise during training, which acts as a regulariser similar to the effect introduced by dropout. As a result, batch normalisation can reduce the need for dropout, although it does not eliminate it. Equation C.2 shows the batch normalisation process:

$$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \rightarrow y = \gamma\hat{x} + \beta \tag{C.2}$$

Where σ_B^2 and μ_B and are the batch variance and batch mean, respectively, ϵ is a random noise(for numerical stability) and γ and β are trainable parameters that scale and shift the normalised values.

Non-linear activation function

The ReLU activation function is placed after the convolutional layer to introduce non-linearity to the model [360, 361]. This function assigns zero to all negative values, and the value itself is assigned to all values above zero. It is described by

$$f(x) = max(0, x) \tag{C.3}$$

Pooling layer

The pooling layer follows the convolutional layer and reduces the spatial dimensionality of the feature maps [362]. This decreases the computational complexity and the number of parameters, which helps mitigate overfitting. Overfitting occurs when a network learns the training data too well, capturing noise and details that do not generalise to new, unseen data in the test dataset. The two most common types of pooling are max pooling and average pooling.

Max pooling

Max pooling involves sliding a window over the input feature map and selecting the maximum value within each window. This operation can be mathematically represented as:

$$y_{i,j} = \max_{(m,n)\in \text{window}} X_{i+m,j+n} \tag{C.4}$$

Where X denotes the input feature map, y represents the output feature map, and (i, j) denotes the coordinates in the feature map. The pooling window slides over the input with a certain stride, typically 2, meaning the window moves 2 pixels at a time.

Average pooling

Contrarily, average pooling calculates the average of all the values in the window. This can be expressed mathematically as:

$$y_{i,j} = \frac{1}{H \cdot W} \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} X_{i+m,j+n} \tag{C.5}$$

Where H and W represent the height and width of the filter, respectively. The pooling window is determined by the filter size $H \times W$, which defines the spatial dimensions of the pooling window. X is the input feature map, y represents the output feature map, and (i, j) denotes the coordinates in the feature map. In the equation, the indices m and n range from 0 to $H - 1$ and $W - 1$, respectively. This ensures that the summation covers all elements within the $H \times W$ pooling window.

Dropout

Dropout is a technique that introduces regularisation within the network by randomly skipping some units or connections with a certain probability. This ultimately improves generalisation by preventing the learned weight values from becoming overly tailored to the training data, thus enhancing their ability to be generalised well to test data [363].

Fully connected layer

The fully connected layer is responsible for high-level reasoning and decision-making in the network. Mathematically, the output of this layer can be described as:

$$y = wX + b \tag{C.6}$$

Where y is the output vector,w is the weight matrix, X is the input vector from the previous layer and b is the bias vector. In this layer, the input features are transformed through a linear combination of weights and biases, often followed by a non-linear activation function. This dense connectivity allows the network to learn complex patterns and representations, making fully connected layers essential for tasks such as classification and regression.

Softmax layer

The softmax layer is a crucial component in the neural network architecture, enabling multi-class classification and the generation of probabilities across multiple classes. It processes inputs using the softmax

function, which takes a vector of arbitrary real-valued scores and produces a probability distribution over these scores. Mathematically, the output probabilities are computed using the following equation:

$$P(y = i|x) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \quad (i = 1, 2, ..., N) \tag{C.7}$$

Where x_1,x_2,..., x_N are the softmax layer's input values, and $P(y = i|x)$ is the output value representing the probability that the input belongs to class i. The softmax function exponentiates each score and then normalises them by dividing them by the sum of all exponentiated scores, ensuring that the output probabilities sum up to 1. This enables the model to interpret the scores as probabilities and make predictions based on them.

C.3 Network training

Training a CNN involves optimising its parameters, such as weights and biases, to minimise a predefined loss function. This optimisation process uses a gradient descent algorithm and backpropagation.

Gradient descent

Gradient descent is an optimisation algorithm that iteratively updates each of the network's parameters in the direction of the steepest descent of the loss function [364]. During training, the network is augmented by a loss layer, where the output of the function L represents the loss with respect to the parameter w. The update is commonly performed in batches to reduce parameter variance during the update and improve learning stability. Mathematically, the update rule for the parameters can be expressed as:

$$w_{t+1} = w_t - \alpha \cdot \frac{\partial L}{\partial w_t} \tag{C.8}$$

where α is the learning rate at iteration t, and $\frac{\partial L}{\partial w_t}$ is the gradient of the loss function L with respect to the parameter w at iteration t.

Backpropagation

Backpropagation is an algorithm that computes the gradient of the loss function with respect to each parameter by utilising the chain rule of calculus [365]. The multi-layered nature of the network allows for the utilisation of the chain rule to derive the partial derivative of each layer, thereby propagating the loss backwards through the network. This approach is called backpropagation and can be defined mathematically as:

$$\frac{\partial L}{\partial w^l} = \frac{\partial L}{\partial y^l} \cdot \frac{\partial y^l}{\partial w^l} \tag{C.9}$$

Where $\frac{\partial L}{\partial w^l}$ represents the gradient of the loss function with respect to the weights of layer l, $\frac{\partial L}{\partial y}$ is the gradient of the loss function with respect to the output of layer l, and $\frac{\partial y}{\partial w^l}$ is the gradient of the output of layer l with respect to its weights. This process is repeated for each layer in the network, allowing the gradients to be efficiently computed and the parameters to be updated accordingly.

Appendix D
Supplementary results: COBE detection using CNNs

All evaluations are performed on datasets where transition regions have been excluded, and a class-balanced loss function is applied. Performance is assessed using several metrics: TPR,FPR, Precision, F1 score, Cohen's Kappa, per-class accuracies (normal and COBE), and overall Accuracy. For cross-validation experiments, results are reported as the mean and standard deviation over five folds. Model selection during cross-validation primarily prioritised the highest mean overall Accuracy. However, models with lower standard deviation were also favoured, as they typically generalise better. In some cases, a trade-off was considered between slightly higher Accuracy and lower variance. The rows highlighted in green indicate the best-performing configurations.

D.1 Individual signal training

D.1.1 IP signal

Table D.1: 5-fold cross-validation performance of various model configurations trained on the IP signal. The table compares different design choices, including pooling strategy, kernel size, and network depth. Highlighted rows indicate configurations selected at various stages of model development, with the final selected model emphasised in darker shading.

Model	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
Average Pooling Strategy								
A	0.88 ± 0.04	0.16 ± 0.03	0.48 ± 0.07	0.62 ± 0.05	0.52 ± 0.06	83.6 ± 3.6	88.3 ± 4.5	86.0 ± 2.4
B	0.90 ± 0.06	0.16 ± 0.04	0.48 ± 0.07	0.62 ± 0.06	0.54 ± 0.07	83.6 ± 4.4	90.0 ± 5.7	86.8 ± 3.7
C	0.90 ± 0.07	0.19 ± 0.04	0.44 ± 0.10	0.58 ± 0.08	0.48 ± 0.08	80.8 ± 3.7	90.2 ± 7.1	85.5 ± 2.5
D	0.89 ± 0.05	0.17 ± 0.05	0.47 ± 0.07	0.61 ± 0.06	0.52 ± 0.06	82.6 ± 5.2	88.7 ± 5.0	85.6 ± 2.8
Max Pooling Strategy								
A_{max}	0.89 ± 0.05	0.17 ± 0.02	0.46 ± 0.06	0.60 ± 0.06	0.51 ± 0.05	82.5 ± 2.2	89.3 ± 4.3	85.9 ± 2.4
B_{max}	0.91 ± 0.04	0.20 ± 0.06	0.44 ± 0.09	0.59 ± 0.08	0.49 ± 0.09	80.2 ± 5.9	91.6 ± 3.9	85.9 ± 1.9
C_{max}	0.87 ± 0.06	0.17 ± 0.05	0.47 ± 0.06	0.61 ± 0.05	0.51 ± 0.06	83.1 ± 4.8	87.4 ± 5.6	85.2 ± 3.0
D_{max}	0.92 ± 0.03	0.21 ± 0.06	0.44 ± 0.06	0.59 ± 0.06	0.49 ± 0.07	79.6 ± 6.2	92.3 ± 3.2	85.9 ± 3.2
Kernel Size change								
E	0.93 ± 0.02	0.22 ± 0.06	0.41 ± 0.05	0.57 ± 0.05	0.47 ± 0.06	77.7 ± 6.2	93.2 ± 1.9	85.5 ± 3.4
F	-	-	-	-	-	-	-	
Deeper 8-layer model								
G	0.91 ± 0.05	0.24 ± 0.04	0.39 ± 0.07	0.54 ± 0.06	0.43 ± 0.06	76.2 ± 3.5	90.7 ± 4.9	83.4 ± 3.0

Based on the proposed trial strategy in 6.5.2, several architectural decisions were made during model development and training.

Decision 1: Same vs. increasing dropout rates: I compared Model A, which uses a uniform dropout rate across all layers, with Model B, which applies progressively increasing dropout rates through the network depth. Model B achieved higher TPR, F1 score, and COBE class accuracy without increasing the FPR. Based on this improvement across key performance metrics, I selected Model B for subsequent comparisons.

Decision 2: Pooling removed in the first or last layers: To assess the impact of removing pooling operations, I compared Model B to Model C (pooling removed in the final layers) and Model D (pooling removed in the first two layers). Model B consistently outperformed both alternatives, particularly in F1 score and FPR, while also maintaining lower variation between folds. I, therefore, retained Model B.

Decision 3: Average vs. Max pooling: I evaluated Model B, which employed average pooling, against all max pooling counterparts (A_{max} through D_{max}). Although D_{max} and B_{max} showed marginally higher TPR and COBE class accuracy, they also led to elevated FPRs and reduced Accuracy in detecting normal breathing. Since minimising false alarms is a key objective in COBE detection, I continued with average pooling as implemented in Model B.

Decision 4: Different kernel sizes: To investigate the effects of kernel sizes, I compared the baseline model (Model B) with Model E, which used reduced kernel sizes. I did not explore a larger kernel size model, as it required the removal of pooling layers to preserve spatial dimensionality, an adjustment that, based on earlier results (Models C and D), did not yield consistent performance benefits. While Model E achieved a slightly higher TPR, it also resulted in the highest FPR and the lowest Normal class accuracy. Given the primary objective of reducing false alarms, I retained Model B as the more suitable architecture.

Decision 5: Model Depth: To assess the impact of network depth, I compared a deeper 8-layer model (Model G) against the 6-layer baseline (Model B). Although Model G had higher COBE class accuracy, it performed worse across several other metrics, including FPR, Precision and Normal class accuracy. I, therefore, retained the shallower 6-layer Model B as the final IP model.

Table D.2: Test performance of the selected model (Model B) trained on 83.3% of the dataset and evaluated on the remaining 16.7%, using the IP signal. This evaluation provides an estimate of the model's generalisation performance on unseen data.

Model	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Overall Accuracy (%)
B	0.91	0.18	0.60	0.72	0.62	82.5	91.1	86.8

D.1.2 ECG signal

Table D.3 shows 5-fold cross-validation results for model variants trained on the ECG-derived respiration (EDR) signal. Performance is reported as mean□±□standard deviation across metrics including TPR,FPR, precision, F1 score, Cohen's Kappa, per-class accuracy (normal and COBE), and Overall Accuracy. Variants differ by pooling method (average vs. max), kernel size, and model depth. Highlighted rows indicate selected or best-performing models.

Following the proposed training strategy in Section 6.5.2, several architectural decisions were made during model development.

Decision 1: Same vs. increasing dropout rates: I compared Model A, which used a uniform dropout rate across all layers, with Model B, which applied progressively increasing dropout rates. Model B achieved a higher F1 score, precision, and Accuracy while slightly reducing the FPR. Based on these, I selected Model B for further comparison.

Decision 2: Pooling removed in the first or last layers: I compared Model B with Models C and D, where pooling was removed either in the last layers (C) or the first layers (D). Both variants yielded lower performance in terms of Precision and Normal class accuracy. Since Model B maintained higher overall Accuracy and lower FPR, I retained it for the next evaluation stage.

Table D.3: 5-fold cross-validation results for multiple CNN model configurations trained on the EDR signal, comparing different strategies for dropout, pooling (average vs. max), kernel size, and model depth. Performance is reported as mean $\pm$ standard deviation. Highlighted models in green represent chosen or best-performing variants.

Model	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
Average Pooling Strategy								
A	0.68 ± 0.12	0.32 ± 0.09	0.26 ± 0.05	0.37 ± 0.06	0.21 ± 0.06	67.6 ± 8.8	68.1 ± 11.7	67.8 ± 5.3
B	0.68 ± 0.10	0.30 ± 0.04	0.28 ± 0.05	0.39 ± 0.06	0.24 ± 0.04	70.4 ± 3.9	68.5 ± 9.9	69.5 ± 3.7
C	0.66 ± 0.12	0.32 ± 0.07	0.26 ± 0.06	0.37 ± 0.06	0.21 ± 0.04	68.5 ± 7.5	66.6 ± 12.2	67.6 ± 3.5
D	0.71 ± 0.05	0.33 ± 0.09	0.27 ± 0.07	0.39 ± 0.08	0.22 ± 0.08	66.4 ± 8.5	71.2 ± 4.9	68.8 ± 4.6
Max Pooling Strategy								
A_{max}	0.66 ± 0.10	0.29 ± 0.08	0.28 ± 0.07	0.39 ± 0.07	0.24 ± 0.08	71.3 ± 8.3	65.5 ± 10.5	68.4 ± 6.1
B_{max}	0.69 ± 0.12	0.30 ± 0.08	0.28 ± 0.04	0.40 ± 0.06	0.25 ± 0.05	70.4 ± 8.0	69.2 ± 11.8	69.8 ± 4.4
C_{max}	0.63 ± 0.05	0.28 ± 0.08	0.28 ± 0.08	0.38 ± 0.08	0.23 ± 0.08	71.5 ± 8.3	63.7 ± 5.0	67.6 ± 4.3
D_{max}	0.66 ± 0.07	0.30 ± 0.14	0.29 ± 0.08	0.39 ± 0.08	0.24 ± 0.09	70.4 ± 14.1	66.4 ± 6.8	68.4 ± 4.8
Kernel Size change								
E	0.70 ± 0.10	0.31 ± 0.04	0.27 ± 0.07	0.39 ± 0.08	0.24 ± 0.08	68.7 ± 4.2	70.4 ± 9.6	69.5 ± 5.6
F	-	-	-	-	-	-	-	
Deeper 8-layer model								
G	0.67 ± 0.07	0.30 ± 0.03	0.27 ± 0.07	0.38 ± 0.07	0.23 ± 0.06	70.1 ± 2.9	67.3 ± 7.1	68.7 ± 4.1

Decision 3: Average vs. Max pooling: I evaluated the best average pooling model (Model B) against all max pooling counterparts (A_{max} to D_{max}). Although B_{max} matched Model B in FPR and slightly outperformed it in COBE class accuracy, it showed a marginally higher variance for all class accuracies and the overall Accuracy. I, therefore, proceeded with Model B using average pooling.

Decision 4: Different kernel sizes: To assess the effect of kernel size, I compared Model B with Model E, which used halved kernel sizes. While Model E showed a slight increase in TPR and COBE class accuracy, it also introduced a higher FPR and reduced Normal class performance. Given the goal of balancing sensitivity with specificity, I continued with Model B.

Decision 5: Model Depth: Finally, I tested a deeper 8-layer network (Model G) against the 6-layer Model B. Model G underperformed in F1 score and overall Accuracy, with no substantial improvement in false alarm rates. I, therefore, retained Model B as the final model for the ECG_{edr} signal.

Table D.4: Final evaluation of Model B using the EDR signal. Reported results reflect model performance on an unseen test set, highlighting generalisation capability.

Model	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
B	0.64	0.38	0.32	0.43	0.19	61.6	63.6	62.6

D.1.3 PPG signal

Table D.5 reports the 5-fold cross-validation results on different model configurations trained on the PPG signal. Each model's performance is measured across several metrics, including TPR, FPR, F1 Score, Cohen's Kappa, per-class accuracies (normal and COBE), and overall Accuracy. The reported values represent the mean and standard deviation across the five folds.

Based on the proposed trial strategy in Section 6.5.2, several architectural decisions were made based on the results in Table D.5:

Decision 1: Same vs. increasing dropout rate: I compared a model with a constant dropout rate across all layers (Model A) to one with increasing dropout rates through the network (Model B). While

Table D.5: 5-fold cross-validation results for multiple CNN model configurations trained on the PPG signal, comparing different strategies for dropout, pooling (average vs. max), kernel size, and model depth. Performance is reported as mean ± standard deviation. Highlighted models in green represent chosen or best-performing variants.

Model	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
Average Pooling Strategy								
A	0.69 ± 0.07	0.37 ± 0.06	0.24 ± 0.07	0.35 ± 0.07	0.18 ± 0.06	62.8 ± 6.3	68.9 ± 7.0	65.8 ± 3.5
B	0.72 ± 0.12	0.39 ± 0.04	0.24 ± 0.06	0.35 ± 0.07	0.18 ± 0.05	61.4 ± 4.1	72.0 ± 12.3	66.7 ± 4.3
C	0.72 ± 0.09	0.39 ± 0.07	0.24 ± 0.07	0.35 ± 0.07	0.18 ± 0.06	61.1 ± 6.8	72.0 ± 9.4	66.6 ± 4.2
D	0.76 ± 0.08	0.39 ± 0.01	0.24 ± 0.06	0.37 ± 0.07	0.20 ± 0.05	61.1 ± 1.5	75.8 ± 7.8	68.5 ± 4.2
Max Pooling Strategy								
A_{max}	0.79 ± 0.04	0.45 ± 0.05	0.23 ± 0.07	0.35 ± 0.08	0.17 ± 0.06	54.7 ± 5.5	79.2 ± 3.8	67.0 ± 3.5
B_{max}	0.79 ± 0.11	0.46 ± 0.06	0.22 ± 0.06	0.34 ± 0.07	0.16 ± 0.05	53.6 ± 6.0	79.4 ± 11.1	66.5 ± 3.3
C_{max}	0.74 ± 0.05	0.40 ± 0.05	0.24 ± 0.06	0.36 ± 0.06	0.18 ± 0.04	60.2 ± 4.9	74.3 ± 5.2	67.3 ± 2.7
D_{max}	0.68 ± 0.06	0.34 ± 0.07	0.26 ± 0.07	0.37 ± 0.07	0.20 ± 0.06	66.2 ± 7.3	68.0 ± 6.4	67.1 ± 2.6
Kernel Size change								
E	0.69 ± 0.05	0.34 ± 0.07	0.26 ± 0.06	0.37 ± 0.07	0.21 ± 0.06	66.4 ± 6.5	68.8 ± 4.8	67.6 ± 3.3
F	0.69 ± 0.14	0.35 ± 0.07	0.25 ± 0.07	0.36 ± 0.06	0.19 ± 0.05	65.1 ± 7.0	69.6 ± 13.4	67.3 ± 4.4
Deeper 8-layer model								
G	0.76 ± 0.07	0.37 ± 0.07	0.26 ± 0.07	0.38 ± 0.08	0.22 ± 0.07	63.4 ± 7.5	75.6 ± 7.5	69.5 ± 4.4

both models achieved similar performance in terms of overall Accuracy, Cohen's Kappa and F1 scores, I selected Model A as it yielded a lower FPR and exhibited less variance in the COBE class predictions across folds. This consistency in reliably detecting the COBE class was prioritised, given the clinical importance of reducing false alarms.

Decision 2: Pooling removed in the first or last layers: Building on Model A, I assessed the impact of removing pooling in different parts of the network by removing pooling in either the last layers (C) or the first layers (D). Although it has a higher FPR and slightly lower Normal class accuracy than Model A, Model D outperformed others in all the other metrics. I selected D for its balanced improvement across metrics.

Decision 3: Average vs. Max pooling: Using Model D as a baseline, I compared it against all max pooling variants (A_{max} to D_{max}). Although D_{max} achieved a slightly lower FPR and improved Normal class accuracy, Model D outperformed it in terms of TPR, COBE class accuracy, and overall Accuracy. Given the clinical importance of reliably detecting COBE events, I retained Model D for its stronger and more consistent performance on the target class.

Decision 4: Different kernel sizes: Using Model D as a baseline, I compared models with smaller kernels (E) and larger kernels (F). Model E had marginally better FPR, Precision, and the highest Kappa. Model E also showed a more balanced performance between the detection of the different classes, so I chose it.

Decision 5: Model Depth: Finally, I compared the best 6-layer model (Model E) to a deeper 8-layer model (Model G). Although Model G had a slightly higher F1 score and sensitivity, Model E performed more consistently and had lower variance across folds. I chose Model E to maintain efficiency and avoid overfitting while still achieving strong performance.

Table D.6: Final evaluation of Model E using the PPG signal when trained on 83.3% of the dataset (20 infants) and evaluated or tested on the remaining 16.7% (4 infants). Reported results reflect model performance on an unseen test set, highlighting generalisability.

Model	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
E	0.61	0.30	0.36	0.46	0.24	69.6	60.7	65.1

D.2 Multi-signal training: late fusion

Building on the single-signal experiments, multi-signal models were constructed using a late fusion strategy. In this approach, the best-performing model for each signal (identified in earlier trials) was retained as a dedicated sub-network. The output features from these sub-networks were concatenated and passed to a shared fully connected layer for final classification. Table D.7 presents 5-fold cross-validation results for various multi-signal configurations. Each row presents average metrics across the folds, including TPR, FPR, Precision, F1 score, Cohen's Kappa, per-class accuracy (for normal and COBE), and overall Accuracy. These results allow a comparative analysis of how signal combinations affect classifier performance.

Table D.7: 5-fold cross-validation results for multi-signal models. For each signal, the best-performing single-signal model (from previous experiments) was reused, and their outputs were concatenated and fed into a shared fully connected layer for final classification.

Signals	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG	0.90 ± 0.04	0.18 ± 0.04	0.45 ± 0.08	0.60 ± 0.06	0.51 ± 0.07	81.8 ± 4.0	90.3 ± 4.2	86.0 ± 2.5
IP + PPG	0.90 ± 0.07	0.19 ± 0.02	0.44 ± 0.06	0.59 ± 0.06	0.49 ± 0.05	81.0 ± 2.4	89.8 ± 7.3	85.4 ± 3.7
ECG + PPG	0.72 ± 0.08	0.32 ± 0.08	0.28 ± 0.07	0.40 ± 0.08	0.24 ± 0.07	67.9 ± 8.0	71.8 ± 7.5	69.9 ± 3.3
IP + ECG + PPG	0.90 ± 0.06	0.19 ± 0.02	0.44 ± 0.08	0.59 ± 0.07	0.49 ± 0.07	81.3 ± 2.6	89.5 ± 6.1	85.4 ± 3.5

Table D.8: Test set performance of late fusion multi-signal models, evaluated after training on 83.3% of the dataset. Each model combines signal-specific subnetworks through feature concatenation, followed by a shared classification layer. Metrics reported include TPR, FPR, Precision, F1 score, Cohen's Kappa, per-class accuracy (normal and COBE), and Accuracy.

Signals	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG	0.89	0.17	0.60	0.72	0.62	83.2	89.5	86.3
IP + PPG	0.88	0.15	0.63	0.73	0.64	85.2	87.9	86.6
ECG + PPG	0.64	0.26	0.41	0.50	0.32	73.7	64.2	69.0
IP + ECG + PPG	0.88	0.18	0.58	0.70	0.59	81.7	88.5	85.1

Appendix E

Supplemetary results: COBE detection using ResNets

E.1 Training using individual signals

To evaluate the capacity of different ResNet architectures for one-dimensional signal classification, I trained and validated models on individual input signals using three variants: ResNet-18, ResNet-34, and ResNet-50. While ResNet-50 uses bottleneck blocks to reduce computational cost and preserve representational power, it did not consistently outperform the shallower variants. These results (presented in tables E.1 to E.2) highlight a trade-off between model complexity and the risk of overfitting, especially in low-data, high-noise settings common in physiological motion processing.

Table E.1: Cross-validation performance of ResNet models trained on individual signals. Results are reported as mean ±standard deviation across five folds. For each signal type, the best-performing ResNet depth (ResNet-18, ResNet-34, or ResNet-50) is highlighted in bold.

Signal	ResNet architecture	Performance Measure							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP	18	0.87 ± 0.04	0.20 ± 0.04	0.43 ± 0.07	0.57 ± 0.07	0.46 ± 0.06	80.3 ± 4.1	86.4 ± 3.4	83.4 ± 1.7
	34*	**0.88 ± 0.04**	**0.19 ± 0.04**	**0.45 ± 0.05**	**0.59 ± 0.05**	**0.49 ± 0.04**	**81.6 ± 3.9**	**87.8 ± 4.0**	**84.7 ± 1.5**
	50	0.89 ± 0.07	0.19 ± 0.05	0.44 ± 0.09	0.58 ± 0.07	0.49 ± 0.08	80.9 ± 4.9	89.3 ± 7.2	85.1 ± 2.8
ECG_{edr}	18	0.70 ± 0.11	0.33 ± 0.01	0.26 ± 0.08	0.38 ± 0.09	0.22 ± 0.08	67.4 ± 1.4	69.8 ± 11.4	68.6 ± 5.8
	34	0.73 ± 0.11	0.32 ± 0.03	0.28 ± 0.08	0.40 ± 0.09	0.25 ± 0.09	68.6 ± 3.3	73.2 ± 11.4	70.9 ± 6.3
	50*	**0.72 ± 0.11**	**0.28 ± 0.02**	**0.30 ± 0.08**	**0.42 ± 0.09**	**0.28 ± 0.08**	**72.1 ± 2.3**	**71.7 ± 11.2**	**71.9 ± 5.5**
ECG_{raw}	**18***	**0.60 ± 0.36**	**0.44 ± 0.29**	**0.16 ± 0.11**	**0.25 ± 0.16**	**0.08 ± 0.09**	**56.4 ± 28.7**	**59.6 ± 35.8**	**58.0 ± 7.0**
	34	0.35 ± 0.31	0.23 ± 0.25	0.20 ± 0.13	0.21 ± 0.14	0.09 ± 0.08	76.8 ± 25.4	34.8 ± 31.2	55.8 ± 4.2
	50	0.41 ± 0.41	0.39 ± 0.40	0.12 ± 0.10	0.17 ± 0.16	0.01 ± 0.05	61.4 ± 39.6	41.0 ± 40.5	51.2 ± 3.5
PPG_{env}	18	0.76 ± 0.10	0.41 ± 0.07	0.24 ± 0.07	0.36 ± 0.08	0.18 ± 0.07	58.8 ± 6.8	75.9 ± 9.6	67.3 ± 4.3
	34*	**0.66 ± 0.09**	**0.33 ± 0.07**	**0.25 ± 0.04**	**0.36 ± 0.06**	**0.19 ± 0.06**	**66.4 ± 6.9**	**66.1 ± 8.8**	**66.3 ± 4.7**
	50	0.72 ± 0.14	0.37 ± 0.06	0.24 ± 0.06	0.36 ± 0.08	0.19 ± 0.06	62.8 ± 6.1	72.0 ± 14.2	67.4 ± 5.5
PPG_{raw}	18	0.70 ± 0.13	0.39 ± 0.10	0.24 ± 0.10	0.35 ± 0.11	0.17 ± 0.10	60.6 ± 10.7	70.4 ± 12.7	65.5 ± 4.9
	34*	**0.61 ± 0.09**	**0.30 ± 0.07**	**0.26 ± 0.09**	**0.36 ± 0.08**	**0.21 ± 0.08**	**70.5 ± 7.5**	**61.3 ± 8.9**	**66.0 ± 3.3**
	50	0.67 ± 0.08	0.37 ± 0.10	0.24 ± 0.09	0.35 ± 0.10	0.17 ± 0.09	62.9 ± 9.6	66.7 ± 8.0	64.8 ± 4.1

Table E.2: Test set performance of the best-perfoming ResNet models identified during five-fold cross-validation (bolded in Table E.1). Each model was trained on 83.3% of the dataset (20 infants) and evaluated on the remaining 16.7% (4 infants). Results reflect the final generalisation ability on held-out data.

Signal	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP	**0.93**	**0.17**	**0.61**	**0.73**	**0.63**	**82.8**	**92.7**	**87.7**
ECG_{edr}	**0.70**	**0.34**	**0.37**	**0.49**	**0.28**	**66.1**	**70.3**	**68.2**
ECG_{raw}	0.50	0.20	0.42	0.46	0.28	80.2	49.8	65.0
PPG_{env}	0.63	0.31	0.37	0.47	0.26	69.5	62.6	66.1
PPG_{raw}	**0.58**	**0.26**	**0.39**	**0.46**	**0.27**	**74.2**	**57.5**	**65.8**

E.2 Multi-modal training

To assess whether performance could be improved by incorporating complementary signal information, I extended the training setup to include multiple inputs. Multi-modal learning is particularly relevant in COBE detection, where different signals capture distinct physiological responses to the cessation of breathing. I evaluated two common fusion strategies: late fusion and early fusion.

E.2.1 Late Fusion

In the late fusion configuration, each input signal is processed independently by its own ResNet encoder. The extracted feature representations are then concatenated and passed to a multi-layer perceptron for classification. The results of the late fusion experiments, shown in Tables E.3 and E.4, compare performance across different combinations of input signals and ResNet depths.

Table E.3: Cross-validation performance of ResNet models using late fusion across different signal combinations. Results are reported as mean ± standard deviation across five folds.

Signals	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG	0.89 $\pm$ 0.07	0.18 $\pm$ 0.02	0.44 $\pm$ 0.07	0.58 $\pm$ 0.05	0.49 $\pm$ 0.04	81.4 $\pm$ 2.0	89.3 $\pm$ 7.2	85.3 $\pm$ 3.3
IP + PPG	0.88 $\pm$ 0.06	0.18 $\pm$ 0.04	0.45 $\pm$ 0.06	0.59 $\pm$ 0.05	0.50 $\pm$ 0.04	82.0 $\pm$ 4.0	88.3 $\pm$ 6.0	85.2 $\pm$ 2.6
ECG + PPG	0.62 $\pm$ 0.09	0.27 $\pm$ 0.05	0.28 $\pm$ 0.07	0.39 $\pm$ 0.07	0.23 $\pm$ 0.07	73.4 $\pm$ 5.1	61.9 $\pm$ 8.8	67.7 $\pm$ 5.1
IP + ECG + PPG	0.89 $\pm$ 0.03	0.21 $\pm$ 0.06	0.43 $\pm$ 0.07	0.57 $\pm$ 0.06	0.47 $\pm$ 0.07	79.3 $\pm$ 6.4	89.5 $\pm$ 2.5	84.4 $\pm$ 2.6

Table E.4: Test set performance of the best late fusion ResNet models trained on different signal combinations. Each configuration uses separate ResNet encoders per signal, followed by feature concatenation and classification. These results reflect model generalisation to unseen data.

Signals	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Overall Accuracy (%)
IP + ECG	0.92	0.20	0.56	0.70	0.58	79.6	92.0	85.8
IP + PPG	0.91	0.17	0.60	0.73	0.63	82.8	91.4	87.1
ECG + PPG	0.50	0.18	0.44	0.47	0.30	81.9	49.8	65.9
IP + ECG + PPG	0.91	0.18	0.59	0.71	0.61	81.6	91.1	86.3

E.2.2 Early Fusion

In the early fusion setup, all input signals are stacked channel-wise and passed through a shared ResNet encoder. This forces the network to learn joint representations from the beginning, capturing low-level interactions across modalities during the early stages of feature extraction.Tables E.5 and E.6 present the results for early fusion models trained on the full signal combination (IP, ECG, and PPG).

Table E.5: Cross-validation performance of ResNet models using early fusion on combined input signals. Results are reported as mean ± standard deviation across five folds.

Signals	ResNet architecture	Performance Measure							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG + PPG	18	0.90 ± 0.03	0.19 ± 0.03	0.45 ± 0.06	0.60 ± 0.06	0.50 ± 0.06	81.4 ± 3.1	89.9 ± 2.9	85.7 ± 1.9
	34	0.89 ± 0.05	0.19 ± 0.04	0.44 ± 0.11	0.58 ± 0.09	0.49 ± 0.09	81.1 ± 4.5	89.2 ± 5.0	85.1 ± 2.7
	50	0.87 ± 0.03	0.18 ± 0.04	0.45 ± 0.08	0.59 ± 0.07	0.49 ± 0.07	81.8 ± 3.5	87.2 ± 2.7	84.5 ± 2.4

Table E.6: Test set performance of early fusion ResNet models trained on a combination of IP, ECG, and PPG signals. Results reflect the model's ability to generalise to unseen data across varying ResNet depths.

Signals	ResNet architecture	Performance Measure							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG + PPG	18	0.85	0.12	0.67	0.75	0.66	87.7	85.3	86.5
	34	0.91	0.18	0.59	0.72	0.62	82.0	91.4	86.7
	50	0.90	0.19	0.58	0.71	0.60	81.4	89.8	85.6

Appendix F

Supplementary results: COBE detection using a ConvNeXt

F.1 Individual signals

Tables F.1 and F.2 show the performances of the ConvNeXt model for individual input signals across different experimental settings. Specifically, I tested the ConvNeXt model using default settings, larger convolutional kernels, and reduced feature dimensionality. This comparison enabled me to investigate how these design choices affect classification performance across different signal types. The results presented in Table F.1 show the model's behaviour on the cross-validation sets.

Table F.1: Cross-validation performance of ConvNeXt models for individual signals using different architectural configurations: default, larger kernels, and reduced feature dimensionality. Bolded entries indicate the best-performing setup per signal.

Signal	ConvNext architecture	Performance Measure							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP	Default	0.90 ± 0.02	0.18 ± 0.05	0.45 ± 0.07	0.60 ± 0.06	0.51 ± 0.07	81.8 ± 5.0	89.7 ± 1.8	85.7 ± 2.2
	Larger Kernels	**0.89 ± 0.02**	**0.16 ± 0.05**	**0.49 ± 0.09**	**0.63 ± 0.07**	**0.54 ± 0.08**	**84.2 ± 4.5**	**89.1 ± 1.9**	**86.7 ± 1.9**
	Reduced Features	0.89 ± 0.03	0.17 ± 0.04	0.48 ± 0.09	0.62 ± 0.07	0.53 ± 0.08	83.5 ± 4.1	88.8 ± 2.7	86.2 ± 2.1
ECG_{edr}	**Default**	**0.82 ± 0.07**	**0.39 ± 0.06**	**0.26 ± 0.06**	**0.40 ± 0.07**	**0.23 ± 0.07**	**61.5 ± 5.6**	**82.3 ± 7.3**	**71.9 ± 5.5**
	Larger Kernels	0.81 ± 0.05	0.40 ± 0.08	0.26 ± 0.06	0.39 ± 0.07	0.21 ± 0.07	60.1 ± 8.2	81.1 ± 5.1	70.6 ± 5.4
	Reduced Features	0.82 ± 0.09	0.38 ± 0.05	0.26 ± 0.06	0.40 ± 0.07	0.23 ± 0.07	62.1 ± 5.2	81.7 ± 9.0	71.9 ± 6.0
ECG_{raw}	Default	0.77 ± 0.07	0.57 ± 0.12	0.19 ± 0.05	0.30 ± 0.06	0.09 ± 0.04	42.7 ± 11.7	77.0 ± 7.5	59.9 ± 4.3
	Larger Kernels	0.79 ± 0.07	0.62 ± 0.11	0.18 ± 0.06	0.29 ± 0.07	0.07 ± 0.04	37.7 ± 11.4	78.7 ± 6.6	58.2 ± 3.8
	Reduced Features	**0.77 ± 0.11**	**0.54 ± 0.20**	**0.20 ± 0.07**	**0.31 ± 0.08**	**0.11 ± 0.07**	**46.0 ± 19.8**	**76.7 ± 11.0**	**61.4 ± 5.6**
PPG_{env}	**Default**	**0.89 ± 0.06**	**0.50 ± 0.03**	**0.23 ± 0.06**	**0.36 ± 0.08**	**0.18 ± 0.05**	**50.3 ± 3.0**	**88.7 ± 6.3**	**69.5 ± 3.6**
	Larger Kernels	0.82 ± 0.05	0.48 ± 0.03	0.22 ± 0.06	0.35 ± 0.08	0.16 ± 0.05	51.8 ± 3.2	81.9 ± 4.9	66.9 ± 3.5
	Reduced Features	0.84 ± 0.07	0.48 ± 0.04	0.23 ± 0.06	0.35 ± 0.08	0.17 ± 0.06	51.5 ± 3.7	84.3 ± 7.0	67.9 ± 4.5
PPG_{raw}	Default	0.70 ± 0.12	0.42 ± 0.07	0.22 ± 0.08	0.33 ± 0.09	0.15 ± 0.08	57.9 ± 7.0	70.3 ± 11.6	64.1 ± 6.1
	Larger Kernels	**0.71 ± 0.07**	**0.39 ± 0.05**	**0.24 ± 0.07**	**0.35 ± 0.08**	**0.17 ± 0.07**	**60.9 ± 5.3**	**70.7 ± 6.9**	**65.8 ± 5.4**
	Reduced Features	0.71 ± 0.12	0.44 ± 0.08	0.22 ± 0.08	0.33 ± 0.09	0.14 ± 0.07	56.6 ± 8.3	70.8 ± 11.8	63.7 ± 4.6

Table F.2: Test performance for COBE detection using ConvNeXt models trained on various physiological signals. Each row presents performance metrics for a specific signal-modality and model configuration. Results reflect each model's ability to generalise to unseen data.

Signal	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP	**0.91**	**0.15**	**0.64**	**0.75**	**0.66**	**85.3**	**90.7**	**88.0**
ECG_{edr}	**0.77**	**0.37**	**0.37**	**0.50**	**0.29**	**62.7**	**76.7**	**69.7**
ECG_{raw}	0.58	0.34	0.33	0.42	0.19	66.2	57.8	62.0
PPG_{env}	**0.73**	**0.40**	**0.34**	**0.47**	**0.23**	**60.0**	**72.8**	**66.4**
PPG_{raw}	0.67	0.44	0.31	0.42	0.17	56.4	67.4	61.9

F.1.1 Multi-modal training

To assess whether performance could be improved by incorporating complementary signal information, I extended the ConvNeXt training setup to support multiple input modalities. Multimodal learning is particularly relevant in COBE detection, where different signals capture distinct physiological responses to cessation of breathing. I evaluated two common fusion strategies: late fusion, which combines the outputs of independently trained single-signal models and early fusion, which combines multiple signals at the input level.

Late Fusion

In the late fusion approach, features were extracted from each ConvNeXt branch before their classification layers, concatenated, and then passed through a separate multi-layer perceptron (MLP) for final classification. Tables F.3 and F.4 summarise the results from late fusion during training (with cross-validation) and on held-out test data, respectively.

Table F.3: Cross-validation performance of late fusion ConvNeXt models trained on different combinations of physiological signals. Results are presented as mean and standard deviation across the five folds.

Signals	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG	0.89 $\pm$ 0.01	0.19 $\pm$ 0.06	0.45 $\pm$ 0.08	0.59 $\pm$ 0.07	0.49 $\pm$ 0.09	80.7 $\pm$ 5.9	89.5 $\pm$ 1.2	85.1 $\pm$ 3.2
IP + PPG	0.89 $\pm$ 0.04	0.18 $\pm$ 0.05	0.46 $\pm$ 0.09	0.60 $\pm$ 0.07	0.51 $\pm$ 0.08	82.0 $\pm$ 5.0	89.3 $\pm$ 4.0	85.6 $\pm$ 3.3
ECG + PPG	0.82 $\pm$ 0.05	0.38 $\pm$ 0.07	0.27 $\pm$ 0.05	0.40 $\pm$ 0.07	0.24 $\pm$ 0.06	62.3 $\pm$ 7.1	82.0 $\pm$ 4.5	72.2 $\pm$ 4.8
IP + ECG + PPG	0.90 $\pm$ 0.04	0.20 $\pm$ 0.05	0.44 $\pm$ 0.08	0.59 $\pm$ 0.07	0.49 $\pm$ 0.07	80.5 $\pm$ 5.2	90.3 $\pm$ 3.5	85.4 $\pm$ 3.0

Table F.4: Test set performance of late fusion ConvNeXt models trained on various combinations of physiological signals. Metrics reflect each model's ability to generalise to unseen data.

Signals	Performance Measure							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG	0.91	0.17	0.61	0.73	0.63	83.0	90.7	86.9
IP + PPG	0.94	0.16	0.62	0.75	0.66	83.9	93.6	88.7
ECG + PPG	0.80	0.37	0.38	0.51	0.31	62.9	79.6	71.2
IP + ECG + PPG	0.92	0.16	0.62	0.74	0.65	84.1	92.3	88.2

F.1.2 Early Fusion

The performance of early fusion ConvNeXt models was evaluated using different architectural variations on combined IP, ECG and PPG signals. Table F.5 presents the cross-validation results, showing the average metrics and their standard deviations across five training folds. Table F.6 reports the corresponding test set performance, demonstrating the models' ability to generalise to unseen data. Both tables highlight the impact of architectural choices such as default settings, larger convolutional kernels, and reduced feature dimensions on classification performance.

Table F.5: Cross-validation performance of early fusion ConvNeXt models trained on combined IP, ECG and PPG signals, showing mean and standard deviation for each metric across five folds.

Signals	ConvNeXt architecture	Performance Measure							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG + PPG	Default	0.92 ± 0.02	0.20 ± 0.06	0.44 ± 0.10	0.59 ± 0.09	0.49 ± 0.10	79.6 ± 6.4	92.1 ± 1.9	85.9 ± 3.2
	Larger Kernels	0.89 ± 0.03	0.19 ± 0.06	0.45 ± 0.09	0.59 ± 0.09	0.50 ± 0.10	81.4 ± 5.7	89.5 ± 3.0	85.4 ± 3.0
	Reduced Features	0.91 ± 0.02	0.19 ± 0.05	0.44 ± 0.09	0.60 ± 0.08	0.50 ± 0.10	80.6 ± 5.3	90.8 ± 1.7	85.7 ± 3.3

Table F.6: Test set performance of early fusion ConvNeXt models trained on combined IP, ECG and PPG signals using different architectural variations.

Signals	ConvNeXt architecture	Performance Measure							
		TPR	FPR	Precision	F1 Score	Cohen Kappa	**Normal class** accuracy (%)	**COBE class** accuracy (%)	Accuracy (%)
IP + ECG + PPG	Default	0.88	0.14	0.64	0.74	0.65	85.6	87.9	86.7
	Larger Kernels	0.93	0.19	0.59	0.72	0.61	81.3	92.7	87.0
	Reduced Features	0.90	0.18	0.59	0.71	0.61	81.8	90.4	86.1

Appendix G

Supplementary results: COBE detection using a combination of camera and physiological signals

Tables G.1 and G.2 show the performances of hybrid models under different input combinations. Each hybrid model combines camera-derived signals with physiological inputs (IP, ECG, and PPG envelope) to detect COBE episodes.

Table G.1: Cross-validation performance of hybrid models trained by combining camera-derived signals with physiological inputs. Results are reported as mean ±standard deviation across five folds. The best-performing combination is highlighted in bold.

Signal combination	Performance Metrics							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
One camera + one physiological signals								
FD + IP	**0.86 ± 0.04**	**0.20 ± 0.02**	**0.46 ± 0.10**	**0.59 ± 0.09**	**0.48 ± 0.09**	**79.9 ± 2.4**	**85.4 ± 4.3**	**82.6 ± 2.6**
FD + PPG	0.78 ± 0.10	0.43 ± 0.08	0.26 ± 0.04	0.39 ± 0.04	0.19 ± 0.06	56.6 ± 7.5	78.1 ± 9.7	67.4 ± 6.1
FD + ECG	0.78 ± 0.08	0.39 ± 0.08	0.29 ± 0.06	0.42 ± 0.06	0.23 ± 0.07	61.2 ± 8.1	78.1 ± 8.8	69.6 ± 5.0
$PPGi_{rr}$ + IP	**0.81 ± 0.08**	**0.23 ± 0.05**	**0.42 ± 0.12**	**0.54 ± 0.11**	**0.42 ± 0.11**	**77.4 ± 4.5**	**81.0 ± 8.0**	**79.2 ± 3.9**
$PPGi_{rr}$ + PPG	0.67 ± 0.05	0.37 ± 0.07	0.27 ± 0.05	0.37 ± 0.05	0.18 ± 0.07	62.6 ± 6.8	67.1 ± 5.4	64.8 ± 5.4
$PPGi_{rr}$ + ECG	0.68 ± 0.05	0.35 ± 0.06	0.28 ± 0.06	0.39 ± 0.06	0.21 ± 0.06	65.0 ± 6.2	68.3 ± 5.1	66.7 ± 3.5
One camera + two physiological signals								
FD + PPG + IP	0.73 ± 0.06	0.21 ± 0.03	0.40 ± 0.09	0.51 ± 0.08	0.39 ± 0.07	78.6 ± 2.7	73.0 ± 5.9	75.8 ± 2.4
FD + ECG + IP	**0.88 ± 0.04**	**0.20 ± 0.03**	**0.47 ± 0.09**	**0.61 ± 0.08**	**0.50 ± 0.08**	**80.2 ± 2.8**	**87.8 ± 4.3**	**84.0 ± 2.0**
FD + ECG + PPG	0.79 ± 0.08	0.39 ± 0.06	0.28 ± 0.05	0.42 ± 0.05	0.23 ± 0.05	60.8 ± 5.7	79.0 ± 7.6	69.9 ± 4.5
$PPGi_{rr}$ + PPG + IP	0.80 ± 0.07	0.21 ± 0.03	0.43 ± 0.08	0.55 ± 0.08	0.43 ± 0.07	78.9 ± 3.1	79.7 ± 6.8	79.3 ± 2.4
$PPGi_{rr}$ +ECG + IP	**0.78 ± 0.08**	**0.21 ± 0.03**	**0.43 ± 0.10**	**0.55 ± 0.10**	**0.43 ± 0.10**	**79.3 ± 3.1**	**77.6 ± 7.5**	**78.5 ± 3.8**
$PPGi_{rr}$ +ECG + PPG	0.71 ± 0.07	0.36 ± 0.05	0.28 ± 0.05	0.40 ± 0.05	0.21 ± 0.05	63.5 ± 5.2	71.0 ± 6.9	67.2 ± 3.8
Two camera + one physiological signal								
FD + $PPGi_{rr}$ + IP	0.80 ± 0.08	0.23 ± 0.03	0.41 ± 0.11	0.54 ± 0.11	0.41 ± 0.12	77.0 ± 3.5	79.8 ± 7.5	78.4 ± 4.5
FD + $PPGi_{rr}$ + PPG	0.84 ± 0.08	0.42 ± 0.09	0.29 ± 0.06	0.43 ± 0.06	0.24 ± 0.08	58.6 ± 8.6	83.6 ± 7.8	71.1 ± 5.7
FD + $PPGi_{rr}$ + ECG	0.74 ± 0.05	0.38 ± 0.07	0.28 ± 0.06	0.40 ± 0.06	0.21 ± 0.06	62.1 ± 6.9	73.8 ± 4.9	68.0 ± 3.8
Two camera + three physiological signals								
FD + $PPGi_{rr}$ + IP + PPG + ECG	0.90 ± 0.03	0.19 ± 0.02	0.48 ± 0.07	0.62 ± 0.07	0.52 ± 0.06	80.6 ± 2.1	90.2 ± 2.5	85.4 ± 1.7

FD represent the frame difference signal

Table G.2: Test set performance of the hybrid models, with the best models highlighted per camera signal. Each model was trained on a dataset of 19 infants and evaluated on the remaining 4 infants. Results reflect the final generalisation ability on held-out data.

Signal combination	Performance Metrics							
	TPR	FPR	Precision	F1 Score	Cohen Kappa	Normal class Accuracy (%)	COBE class Accuracy (%)	Accuracy (%)
One camera + one physiological signals								
FD + IP	**0.81**	**0.08**	**0.68**	**0.74**	**0.68**	**92.1**	**81.0**	**86.6**
FD + PPG	0.78	0.41	0.28	0.41	0.22	59.1	78.5	68.8
FD + ECG	0.89	0.32	0.36	0.51	0.36	67.7	88.8	78.2
$PPGi_{rr}$ + IP	**0.92**	**0.11**	**0.63**	**0.75**	**0.7**	**88.9**	**92.2**	**90.6**
$PPGi_{rr}$ + PPG	0.76	0.22	0.42	0.54	0.41	78.4	75.9	77.1
$PPGi_{rr}$ + ECG	0.85	0.37	0.32	0.47	0.29	63.3	85.3	74.3
One camera + two physiological signals								
FD + PPG + IP	0.85	0.07	0.71	0.78	0.73	93.0	85.3	89.2
FD + ECG + IP	**0.91**	**0.12**	**0.61**	**0.73**	**0.66**	**88.1**	**91.4**	**89.7**
FD + ECG + PPG	0.85	0.32	0.35	0.50	0.34	68.2	85.3	76.8
$PPGi_{rr}$ + PPG + IP	0.91	0.13	0.58	0.71	0.63	86.8	90.5	88.7
$PPGi_{rr}$ +ECG + IP	**0.91**	**0.13**	**0.59**	**0.71**	**0.64**	**86.8**	**91.4**	**89.1**
$PPGi_{rr}$ +ECG + PPG	0.88	0.25	0.41	0.56	0.43	74.5	87.9	81.2
Two camera + one physiological signal								
FD + $PPGi_{rr}$ + IP	0.91	0.12	0.60	0.72	0.65	87.7	90.5	89.1
FD + $PPGi_{rr}$ + PPG	0.80	0.29	0.36	0.50	0.34	71.0	80.2	75.6
FD + $PPGi_{rr}$ + ECG	0.72	0.19	0.44	0.55	0.42	81.0	72.4	76.7
Two camera + three physiological signals								
FD + $PPGi_{rr}$ + IP + PPG + ECG	0.76	0.08	0.66	0.71	0.64	92.1	75.9	84.0

FD represent the frame difference signal

Bibliography

[1] World Health Organization (WHO), *Born too soon: decade of action on preterm birth*. Geneva: World Health Organization, 2023. Licence: CC BY-NC-SA 3.0 IGO.

[2] M. Platt, "Outcomes in preterm infants," *Public health*, vol. 128, no. 5, pp. 399–403, 2014.

[3] E. O. Ohuma, A.-B. Moller, E. Bradley, S. Chakwera, L. Hussain-Alkhateeb, A. Lewin, Y. B. Okwaraji, W. R. Mahanani, E. W. Johansson, T. Lavin, *et al.*, "National, regional, and global estimates of preterm birth in 2020, with trends from 2010: a systematic analysis," *The Lancet*, vol. 402, no. 10409, pp. 1261–1271, 2023.

[4] A. G. Demisse, F. Alemu, M. A. Gizaw, and Z. Tigabu, "Patterns of admission and factors associated with neonatal mortality among neonates admitted to the neonatal intensive care unit of university of gondar hospital, northwest ethiopia," *Pediatric health, medicine and therapeutics*, vol. 8, p. 57, 2017.

[5] A. Desalew, Y. Sintayehu, N. Teferi, F. Amare, B. Geda, T. Worku, K. Abera, and A. Asefaw, "Cause and predictors of neonatal mortality among neonates admitted to neonatal intensive care units of public hospitals in eastern ethiopia: a facility-based prospective follow-up study," *BMC pediatrics*, vol. 20, no. 1, pp. 1–11, 2020.

[6] H. C. Glass, A. T. Costarino, S. A. Stayer, C. Brett, F. Cladis, and P. J. Davis, "Outcomes for extremely premature infants," *Anesthesia and analgesia*, vol. 120, no. 6, p. 1337, 2015.

[7] L. Thompson, J. W. Werthammer, and D. Gozal, "Apnea of prematurity and oxidative stress: Potential implications," *Antioxidants*, vol. 13, no. 11, p. 1304, 2024.

[8] World Health Organization, "Preterm birth rate (per 100 live births)." `https://www.who.int/data/gho/data/indicators/indicator-details/GHO/preterm-birth-rate-per-100-live-births`, 2025. Accessed: 2025-06-14.

[9] Office for National Statistics (ONS), "Birth characteristics in england and wales: 2021," 2023.

[10] Office for National Statistics (ONS), "Birth characteristics in england and wales: 2022." Statistical Bulletin, ONS Website, 2024. Released 17 May 2024.

[11] J. Wise, "Target to reduce preterm births in england to 6% will not be met without action, say peers," *BMJ*, vol. 387, no. q2488, 2024.

[12] S. Shimano, T. Yamada, K. Cho, K. Sengoku, T. Mariya, and T. Saito, "Changes in preterm and extremely preterm birth rates in japan after the introduction of obstetrical practice guidelines in 2008," *Journal of Obstetrics and Gynaecology Research*, vol. 49, no. 9, pp. 2283–2294, 2023.

[13] M. J. Osterman, B. E. Hamilton, J. A. Martin, A. K. Driscoll, and C. P. Valenzuela, "Births: final data for 2022.," *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, vol. 73, no. 2, pp. 1–56, 2024.

[14] O. Lincetto and A. Banerjee, "World prematurity day: improving survival and quality of life for millions of babies born preterm around the world," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 319, no. 5, pp. L871–L874, 2020.

[15] R. E. Behrman, A. S. Butler, *et al.*, "Mortality and acute complications in preterm infants," in *Preterm Birth: Causes, consequences, and prevention*, National Academies Press (US), 2007.

[16] F. Operto, "Preterm births and cognitive development," *Acta Scientific Neurology (ISSN: 2582-1121)*, vol. 3, no. 9, 2020.

[17] D. Morniroli, V. Tiraferri, G. Maiocco, D. U. De Rose, F. Cresi, A. Coscia, F. Mosca, and M. L. Giannì, "Beyond survival: the lasting effects of premature birth," *Frontiers in Pediatrics*, vol. 11, p. 1213243, 2023.

[18] J. Zivaljevic, M. Z. Jovandaric, S. Babic, and M. Raus, "Complications of preterm birth—the importance of care for the outcome: A narrative review," *Medicina*, vol. 60, no. 6, p. 1014, 2024.

[19] N. P. Kondamudi, L. Krata, and A. S. Wilt, *Infant apnea*. StatPearls Publishing LLC, 2023.

[20] H. Lee, C. G. Rusin, D. E. Lake, M. T. Clark, L. Guin, T. J. Smoot, A. O. Paget-Brown, B. D. Vergales, J. Kattwinkel, J. R. Moorman, *et al.*, "A new algorithm for detecting central apnea in neonates," *Physiological measurement*, vol. 33, no. 1, p. 1, 2011.

[21] J. Jorge, M. Villarroel, S. Chaichulee, A. Guazzi, S. Davis, G. Green, K. McCormick, and L. Tarassenko, "Non-contact monitoring of respiration in the neonatal intensive care unit," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 286–293, IEEE, 2017.

[22] J. Zhao, F. Gonzalez, and D. Mu, "Apnea of prematurity: from cause to treatment," *European journal of pediatrics*, vol. 170, no. 9, pp. 1097–1105, 2011.

[23] M. Bester, D. van den Heever, R. Joshi, and K. Dellimore, "A study of short cessations and temporal dynamics of breathing in preterm infants," in *2018 3rd Biennial South African Biomedical Engineering Conference (SAIBMEC)*, pp. 1–4, IEEE, 2018.

[24] H. Almutairi, G. M. Hassan, and A. Datta, "Classification of obstructive sleep apnoea from single-lead ecg signals using convolutional neural and long short term memory networks," *Biomedical Signal Processing and Control*, vol. 69, p. 102906, 2021.

[25] R. J. Martin, J. M. Abu-Shaweesh, and T. M. Baird, "Apnoea of prematurity," *Paediatric respiratory reviews*, vol. 5, pp. S377–S382, 2004.

[26] M. Seppä-Moilanen, *Sleep and Breathing in Preterm Infants: Polysomnography Studies on the Effects of Caffeine and Supplemental Oxygen*. PhD thesis, University of Helsinki, 2022.

[27] G. Erickson, N. R. Dobson, and C. E. Hunt, "Immature control of breathing and apnea of prematurity: the known and unknown," *Journal of Perinatology*, vol. 41, no. 9, pp. 2111–2123, 2021.

[28] M. A. Mohr, B. D. Vergales, H. Lee, M. T. Clark, D. E. Lake, A. C. Mennen, J. Kattwinkel, R. A. Sinkin, J. R. Moorman, K. D. Fairchild, *et al.*, "Very long apnea events in preterm infants," *Journal of Applied Physiology*, vol. 118, no. 5, pp. 558–568, 2015.

[29] J. V. Pergolizzi Jr, P. Fort, T. L. Miller, J. A. LeQuang, and R. B. Raffa, "The epidemiology of apnoea of prematurity," *Journal of clinical pharmacy and therapeutics*, vol. 47, no. 5, pp. 685–693, 2022.

[30] M. Williamson, R. Poorun, and C. Hartley, "Apnoea of prematurity and neurodevelopmental outcomes: current understanding and future prospects for research," *Frontiers in Pediatrics*, vol. 9, 2021.

[31] E. G. A. Jawdeh, H. Huang, P. M. Westgate, A. Patwardhan, H. Bada, J. A. Bauer, and P. Giannone, "Intermittent hypoxemia in preterm infants: a potential proinflammatory process," *American journal of perinatology*, vol. 38, no. 12, pp. 1313–1319, 2021.

[32] A. K. Yee, L. S. Siriwardhana, G. M. Nixson, L. M. Walter, F. Y. Wong, and R. S. Horne, "Developmental consequences of short apneas and periodic breathing in preterm infants," *Journal of Perinatology*, vol. 43, no. 11, pp. 1420–1428, 2023.

[33] X. Xu, L. Li, D. Chen, S. Chen, L. Chen, and X. Feng, "Establishment and validation of apnea risk prediction models in preterm infants: a retrospective case control study," *BMC pediatrics*, vol. 24, no. 1, p. 654, 2024.

[34] A. Dormishian, A. Schott, A. C. Aguilar, V. Jimenez, E. Bancalari, J. Tolosa, and N. Claure, "Etiology and mechanism of intermittent hypoxemia episodes in spontaneously breathing extremely premature infants," *The Journal of pediatrics*, vol. 262, p. 113623, 2023.

[35] C. F. Poets, M. P. Samuels, and D. P. Southall, "Epidemiology and pathophysiology of apnoea of prematurity," *Neonatology*, vol. 65, no. 3-4, pp. 211–219, 1994.

[36] J. Adams, I. Zabaleta, and M. Sackner, "Hypoxemic events in spontaneously breathing premature infants: etiologic basis," *Pediatric Research*, vol. 42, no. 4, pp. 463–471, 1997.

[37] S. A. Pullano, I. Mahbub, M. G. Bianco, S. Shamsir, S. K. Islam, M. S. Gaylord, V. Lorch, and A. S. Fiorillo, "Medical devices for pediatric apnea monitoring and therapy: past and new trends," *IEEE reviews in biomedical engineering*, vol. 10, pp. 199–212, 2017.

[38] H. Almutairi, G. M. Hassan, and A. Datta, "Detection of obstructive sleep apnoea by ecg signals using deep learning architectures," in *2020 28th European Signal Processing Conference (EUSIPCO)*, pp. 1382–1386, IEEE, 2021.

[39] L. K. Bawua, C. Miaskowski, X. Hu, G. W. Rodway, and M. M. Pelter, "A review of the literature on the accuracy, strengths, and limitations of visual, thoracic impedance, and electrocardiographic methods used to measure respiratory rate in hospitalized patients," *Annals of Noninvasive Electrocardiology*, vol. 26, no. 5, p. e12885, 2021.

[40] C. F. Poets, "Pulse oximetry vs. transcutaneous monitoring in neonates: practical aspects," *www. bloodgas. com, Neonatology. Copenhagen: Radiometer Medical A/S*, 2003.

[41] J. Vagedes, M. Sobh, M. O. A. Islam, and C. F. Poets, "Averaging times for pulse oximeter measurements–a review of manuscripts published in the top five sleep medicine journals," *Nature and Science of Sleep*, pp. 1131–1139, 2024.

[42] J. M. Di Fiore, P. M. MacFarlane, and R. J. Martin, "Intermittent hypoxemia in preterm infants," *Clinics in perinatology*, vol. 46, no. 3, pp. 553–565, 2019.

[43] L. Horakova and K. Roubik, "Pulse oximeter performance during rapid desaturation," *Sensors*, vol. 22, no. 11, p. 4236, 2022.

[44] M. Al-Beltagi, N. K. Saeed, A. S. Bediwy, and R. Elbeltagi, "Pulse oximetry in pediatric care: Balancing advantages and limitations," *World Journal of Clinical Pediatrics*, vol. 13, no. 3, 2024.

[45] Philips Healthcare, "Expand your view: Critical care monitoring for demanding conditions - intellivue mx700 and mx800 patient monitors." `https://www.documents.philips.com/assets/20191008/d5ee9ea660d843f79addaae0011c0492.pdf`, 2019. Accessed: June 16, 2023.

[46] V. Rafl-Huttova, V. Hlavac, J. Rafl, and M. Rozanek, "Dependence of spo2 signal noise on the pulse oximeter averaging time," *Current Directions in Biomedical Engineering*, vol. 7, no. 2, pp. 351–354, 2021.

[47] D. Bracco and S. B. Backman, "Philips monitors: Catch the wave!," *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*, vol. 59, pp. 325–326, 2012.

[48] M. T. Petterson, V. L. Begnoche, and J. M. Graybeal, "The effect of motion on pulse oximetry and its clinical significance," *Anesthesia & Analgesia*, vol. 105, no. 6, pp. S78–S84, 2007.

[49] C. McClure, S. Y. Jang, and K. Fairchild, "Alarms, oxygen saturations, and spo2 averaging time in the nicu," *Journal of neonatal-perinatal medicine*, vol. 9, no. 4, pp. 357–362, 2016.

[50] D. Bertoni and A. Isaiah, "Towards patient-centered diagnosis of pediatric obstructive sleep apnea—a review of biomedical engineering strategies," *Expert review of medical devices*, vol. 16, no. 7, pp. 617–629, 2019.

[51] D. Zou, S. Vits, C. Egea, D. Ehrsam-Tosi, F. Lavergne, M. Azpiazu, and I. Fietze, "A new approach to streamline obstructive sleep apnea therapy access using peripheral arterial tone-based home sleep test devices," *Frontiers in Sleep*, vol. 2, p. 1256078, 2023.

[52] S. Nwaneri, B. Ezenwa, A. Osuntoki, V. Ezeaka, and F. Ogunsola, “A review of infant apnea monitor design,” *Journal of Clinical Sciences*, vol. 21, no. 2, pp. 93–98, 2024.

[53] A. Nicolò, C. Massaroni, E. Schena, and M. Sacchetti, “The importance of respiratory rate monitoring: From healthcare to sport and exercise,” *Sensors*, vol. 20, no. 21, p. 6396, 2020.

[54] C. Massaroni, A. Nicolò, D. Lo Presti, M. Sacchetti, S. Silvestri, and E. Schena, “Contact-based methods for measuring respiratory rate,” *Sensors*, vol. 19, no. 4, p. 908, 2019.

[55] E. J. Ciaccio, M. Hiatt, T. Hegyi, and G. M. Drzewiecki, “Measurement and monitoring of electrocardiogram belt tension in premature infants for assessment of respiratory function,” *BioMedical Engineering OnLine*, vol. 6, pp. 1–11, 2007.

[56] D. Ostojic, S. Guglielmini, V. Moser, J.-C. Fauchère, H. U. Bucher, D. Bassler, M. Wolf, S. Kleiser, F. Scholkmann, and T. Karen, “Reducing false alarm rates in neonatal intensive care: a new machine learning approach,” *Oxygen Transport to Tissue XLI*, pp. 285–290, 2020.

[57] M. Hravnak, T. Pellathy, L. Chen, A. Dubrawski, A. Wertz, G. Clermont, and M. R. Pinsky, “A call to alarms: current state and future directions in the battle against alarm fatigue,” *Journal of electrocardiology*, vol. 51, no. 6, pp. S44–S48, 2018.

[58] R. Lloyd, R. Goulding, P. Filan, and G. Boylan, “Overcoming the practical challenges of electroencephalography for very preterm infants in the neonatal intensive care unit,” *Acta paediatrica*, vol. 104, no. 2, pp. 152–157, 2015.

[59] M. M. Baharestani, “An overview of neonatal and pediatric wound care knowledge and considerations.,” *Ostomy/wound management*, vol. 53, no. 6, pp. 34–6, 2007.

[60] O. Bonner, K. Beardsall, N. Crilly, and J. Lasenby, “There were more wires than him: the potential for wireless patient monitoring in neonatal intensive care,” *BMJ innovations*, vol. 3, no. 1, 2017.

[61] H. Pavlyshyn, I. Sarapuk, C. Casper, and N. Makieieva, “Kangaroo mother care can improve the short-term outcomes of very preterm infants,” *Journal of neonatal-perinatal medicine*, vol. 14, no. 1, pp. 21–28, 2021.

[62] B. D. Vergales, A. O. Paget-Brown, H. Lee, L. E. Guin, T. J. Smoot, C. G. Rusin, M. T. Clark, J. B. Delos, K. D. Fairchild, D. E. Lake, *et al.*, “Accurate automated apnea analysis in preterm infants,” *American journal of perinatology*, vol. 31, no. 02, pp. 157–162, 2014.

[63] A. B. Sankar, D. Kumar, K. Seethalakshmi, *et al.*, “Performance study of various adaptive filter algorithms for noise cancellation in respiratory signals,” *Signal processing: An international journal (SPIJ)*, vol. 4, no. 5, p. 267, 2010.

[64] S. G. Moxon, J. E. Lawn, K. E. Dickson, A. Simen-Kapeu, G. Gupta, A. Deorari, N. Singhal, K. New, C. Kenner, V. Bhutani, *et al.*, “Inpatient care of small and sick newborns: a multi-country analysis of health system bottlenecks and potential solutions,” *BMC pregnancy and childbirth*, vol. 15, no. Suppl 2, p. S7, 2015.

[65] S. R. Kamala, Z. Julius, E. M. Kosia, and F. Manzi, “Availability and functionality of neonatal care units in healthcare facilities in mtwara region, tanzania: The quest for quality of in-patient care for small and sick newborns,” *Plos one*, vol. 17, no. 11, p. e0269151, 2022.

[66] A. Tarus, G. Msemo, R. Kamuyu, D. Shamba, R. P. Kirby, K. M. Palamountain, E. Gicheha, M. B. Kumar, T. Powell-Jackson, C. Bohne, *et al.*, “Devices and furniture for small and sick newborn care: systematic development of a planning and costing tool,” *BMC pediatrics*, vol. 23, no. Suppl 2, p. 566, 2023.

[67] F. P. Wieringa, F. Mastik, and A. v. d. Steen, “Contactless multiple wavelength photoplethysmographic imaging: A first step toward “spo2 camera” technology,” *Annals of biomedical engineering*, vol. 33, pp. 1034–1041, 2005.

[68] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light.,” *Optics express*, vol. 16, no. 26, pp. 21434–21445, 2008.

[69] M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” *IEEE transactions on biomedical engineering*, vol. 58, no. 1, pp. 7–11, 2010.

[70] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak, “Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity,” in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 405–410, IEEE, 2011.

[71] L. A. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, J. S. Nelson, S. B. Oetomo, and W. Verkruysse, “Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—a pilot study,” *Early human development*, vol. 89, no. 12, pp. 943–948, 2013.

[72] M. Uenoyama, T. Matsui, K. Yamada, S. Suzuki, B. Takase, S. Suzuki, M. Ishihara, and M. Kawakami, “Non-contact respiratory monitoring system using a ceiling-attached microwave antenna,” *Medical and Biological Engineering and Computing*, vol. 44, pp. 835–840, 2006.

[73] M. Beelke, S. Angeli, M. Del Sette, C. Gandolfo, M. E. Cabano, P. Canovaro, L. Nobili, and F. Ferrillo, “Prevalence of patent foramen ovale in subjects with obstructive sleep apnea: a transcranial doppler ultrasound study,” *Sleep medicine*, vol. 4, no. 3, pp. 219–223, 2003.

[74] S. D. Min, D. J. Yoon, S. W. Yoon, Y. H. Yun, and M. Lee, “A study on a non-contacting respiration signal monitoring system using doppler ultrasound,” *Medical & biological engineering & computing*, vol. 45, pp. 1113–1119, 2007.

[75] J. Fei, I. Pavlidis, and J. Murthy, “Thermal vision for sleep apnea monitoring,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009: 12th International Conference, London, UK, September 20-24, 2009, Proceedings, Part II 12*, pp. 1084–1091, Springer, 2009.

[76] U. Muhammad, R. Evans, R. Saatchi, R. Kingshott, and H. Elphick, “P016 using non-invasive thermal imaging for apnoea detection,” *BMJ Open Respiratory Research*, vol. 6, pp. A9–A10, 2019.

[77] L. Cattani, D. Alinovi, G. Ferrari, R. Raheli, E. Pavlidis, C. Spagnoli, and F. Pisani, “A wire-free, non-invasive, low-cost video processing-based approach to neonatal apnoea detection,” in *2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings*, pp. 67–73, IEEE, 2014.

[78] C. Yang, G. Cheung, V. Stankovic, K. Chan, and N. Ono, “Sleep apnea detection via depth video and audio feature learning,” *IEEE Transactions on Multimedia*, vol. 19, no. 4, pp. 822–835, 2016.

[79] A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, “Real time apnoea monitoring of children using the microsoft kinect sensor: a pilot study,” *Sensors*, vol. 17, no. 2, p. 286, 2017.

[80] C. Qian, J. Li, and M. Khine, “Noninvasive vital signs monitoring in the neonatal intensive care unit,” *Journal ISSN*, vol. 2766, p. 2276, 2024.

[81] D. Vitazkova, E. Foltan, H. Kosnacova, M. Micjan, M. Donoval, A. Kuzma, M. Kopani, and E. Vavrinsky, “Advances in respiratory monitoring: a comprehensive review of wearable and remote technologies,” *Biosensors*, vol. 14, no. 2, p. 90, 2024.

[82] M. S. Ishrak, F. Cai, S. M. M. Islam, O. Borić-Lubecke, T. Wu, and V. M. Lubecke, “Doppler radar remote sensing of respiratory function,” *Frontiers in Physiology*, vol. 14, p. 1130478, 2023.

[83] S. M. M. Islam, O. Borić-Lubecke, Y. Zheng, and V. M. Lubecke, “Radar-based non-contact continuous identity authentication,” *Remote Sensing*, vol. 12, no. 14, p. 2279, 2020.

[84] L. Maurya, P. Mahapatra, and D. Chawla, “Non-contact breathing monitoring by integrating RGB and thermal imaging via RGB-thermal image registration,” *Biocybernetics and Biomedical Engineering*, vol. 41, no. 3, pp. 1107–1122, 2021.

[85] Z. Peng, D. Kommers, R.-H. Liang, X. Long, W. Cottaar, H. Niemarkt, P. Andriessen, and C. van Pul, “Continuous sensing and quantification of body motion in infants: A systematic review,” *Heliyon*, vol. 9, no. 7, 2023.

[86] S. Shu, H. Liang, Y. Zhang, Y. Zhang, and Z. Yang, “Non-contact measurement of human respiration using an infrared thermal camera and the deep learning method,” *Measurement Science and Technology*, vol. 33, no. 7, p. 075202, 2022.

[87] E. E. Geertsema, G. H. Visser, J. W. Sander, and S. N. Kalitzin, “Automated non-contact detection of central apneas using video,” *Biomedical Signal Processing and Control*, vol. 55, p. 101658, 2020.

[88] M. Villarroel, S. Chaichulee, J. Jorge, S. Davis, G. Green, C. Arteta, A. Zisserman, K. McCormick, P. Watkinson, and L. Tarassenko, “Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit,” *NPJ digital medicine*, vol. 2, no. 1, pp. 1–18, 2019.

[89] S. Ruhrberg Estévez, A. Grafton, L. Thomson, J. Warnecke, K. Beardsall, and J. Lasenby, “Continuous non-contact vital sign monitoring of neonates in intensive care units using rgb-d cameras,” *Scientific Reports*, vol. 15, no. 1, p. 16863, 2025.

[90] V. Selvaraju, N. Spicher, J. Wang, N. Ganapathy, J. M. Warnecke, S. Leonhardt, R. Swaminathan, and T. M. Deserno, “Continuous monitoring of vital signs using cameras: A systematic review,” *Sensors*, vol. 22, no. 11, p. 4097, 2022.

[91] S. L. Rossol, J. K. Yang, C. Toney-Noland, J. Bergin, C. Basavaraju, P. Kumar, and H. C. Lee, “Non-contact video-based neonatal respiratory monitoring,” *Children*, vol. 7, no. 10, p. 171, 2020.

[92] D. L. August, L. Edmonds, D. K. Brown, M. Murphy, and Y. Kandasamy, “Pressure injuries to the skin in a neonatal unit: fact or fiction,” *Journal of Neonatal Nursing*, vol. 20, no. 3, pp. 129–137, 2014.

[93] V. Boivin, M. Shahriari, G. Faure, S. Mellul, E. D. Tiassou, P. Jouvet, and R. Noumeir, “Multimodality video acquisition system for the assessment of vital distress in children,” *Sensors*, vol. 23, no. 11, p. 5293, 2023.

[94] I. Torres-Cuevas, M. Cernada, A. Nuñez, J. Escobar, J. Kuligowski, C. Chafer-Pericas, and M. Vento, “Oxygen supplementation to stabilize preterm infants in the fetal to neonatal transition: no satisfactory answer,” *Frontiers in Pediatrics*, vol. 4, p. 29, 2016.

[95] C. M. J. Tan and A. J. Lewandowski, “The transitional heart: from early embryonic and fetal development to neonatal life,” *Fetal diagnosis and therapy*, vol. 47, no. 5, pp. 373–386, 2020.

[96] D. Zhang and S. E. Lindsey, “Recasting current knowledge of human fetal circulation: The importance of computational models,” *Journal of Cardiovascular Development and Disease*, vol. 10, no. 6, p. 240, 2023.

[97] J. G. Betts, K. A. Young, J. A. Wise, E. Johnson, B. Poe, D. H. Kruse, O. Korol, J. E. Johnson, M. Womble, and P. DeSaix, *Anatomy and Physiology*. Houston, TX: OpenStax, Rice University, 2013. Access for free at `https://openstax.org/books/anatomy-and-physiology/pages/20-6-development-of-blood-vessels-and-fetal-circulation`. Licensed under CC BY.

[98] J. G. Betts, K. A. Young, J. A. Wise, E. Johnson, B. Poe, D. H. Kruse, O. Korol, J. E. Johnson, M. Womble, and P. DeSaix, *Anatomy and Physiology*, ch. 19 The Cardiovascular System: The Heart. Houston, Texas: OpenStax,Rice University, 2022.

[99] W. D. Tucker, Y. Arora, and K. Mahajan, “Anatomy, blood vessels,” in *StatPearls [internet]*, StatPearls Publishing, Treasure Island (FL), 2017.

[100] R. N. Pittman, “The circulatory system and oxygen transport,” in *Regulation of Tissue Oxygenation*, Morgan & Claypool Life Sciences, 2011.

[101] J.-H. Kim, “Heart and circulatory system,” in *Recent Advancements in Microbial Diversity*, pp. 229–254, Elsevier, 2022.

[102] A. J Patterson and L. Zhang, “Hypoxia and fetal heart development,” *Current molecular medicine*, vol. 10, no. 7, pp. 653–666, 2010.

[103] G. Acharya, S.-E. Sonesson, K. Flo, J. Räsänen, and A. Odibo, "Hemodynamic aspects of normal human feto-placental (umbilical) circulation," *Acta obstetricia et gynecologica Scandinavica*, vol. 95, no. 6, pp. 672–682, 2016.

[104] S. Felis, "Fetal oxygenation," *American Journal of Medical and Clinical Research & Reviews*, vol. 2, no. 10, pp. 1–15, 2023.

[105] I. Lara-Cantón, S. Badurdeen, J. Dekker, P. Davis, C. Roberts, A. Te Pas, and M. Vento, "Oxygen saturation and heart rate in healthy term and late preterm infants with delayed cord clamping," *Pediatric Research*, pp. 1–6, 2022.

[106] N. H. Hillman, S. G. Kallapur, and A. H. Jobe, "Physiology of transition from intrauterine to extrauterine life," *Clinics in perinatology*, vol. 39, no. 4, pp. 769–783, 2012.

[107] P. Wu, M. Gulati, C. S. Kwok, C. W. Wong, A. Narain, S. O'Brien, C. A. Chew-Graham, G. Verma, U. T. Kadam, and M. A. Mamas, "Preterm delivery and future risk of maternal cardiovascular disease: a systematic review and meta-analysis," *Journal of the American Heart Association*, vol. 7, no. 2, p. e007809, 2018.

[108] R. I. Clyman, "Ibuprofen and patent ductus arteriosus," *New England Journal of Medicine*, vol. 343, no. 10, pp. 728–730, 2000.

[109] J. M. Ryder and E. Bae, "Increasing alprostadil requirements in a neonate with cardiac anomalies and co-administration of rectal and oral acetaminophen," *The Journal of Pediatric Pharmacology and Therapeutics*, vol. 27, no. 6, pp. 573–577, 2022.

[110] J. E. Dice and J. Bhatia, "Patent ductus arteriosus: an overview," *The Journal of Pediatric Pharmacology and Therapeutics*, vol. 12, no. 3, pp. 138–146, 2007.

[111] A. Patwa and A. Shah, "Anatomy and physiology of respiratory system relevant to anaesthesia," *Indian journal of anaesthesia*, vol. 59, no. 9, p. 533, 2015.

[112] H. Schöpper, C. Bartel, and K. Szalai, "Breathing: Comparative aspects of the respiratory system," in *Comparative Medicine: Anatomy and Physiology*, pp. 179–194, Springer, 2013.

[113] J. Ward, "Physiology of breathing i," *Surgery (Oxford)*, vol. 23, no. 11, pp. 419–424, 2005.

[114] J. Petersson and R. W. Glenny, "Gas exchange in the lung," in *Seminars in Respiratory and Critical Care Medicine*, vol. 44, pp. 555–568, Thieme Medical Publishers, Inc., 2023.

[115] Nucleus Medical Media, "Anatomy and functions of the respiratory system." `https://nmal.nucleusmedicalmedia.com/anatomy-and-functions-of-the-respiratory-system/view-item?ItemID=4888`, 1969. Digital image, retrieved from Nucleus Medical Media.

[116] P. H. Burri, "Fetal and postnatal development of the lung," *Annual review of physiology*, vol. 46, no. 1, pp. 617–628, 1984.

[117] J. C. Schittny, "Development of the lung," *Cell and tissue research*, vol. 367, pp. 427–444, 2017.

[118] I. Caldeira, H. Fernandes-Silva, D. Machado-Costa, J. Correia-Pinto, and R. S. Moura, "Developmental pathways underlying lung development and congenital lung disorders," *Cells*, vol. 10, no. 11, p. 2987, 2021.

[119] J. C. Woods and J. C. Schittny, "Lung structure at preterm and term birth," *Fetal lung development-clinical correlates & future technologies. Cambridge University Press, New York*, pp. 126–140, 2016.

[120] L. S. Inselman and R. B. Mellins, "Growth and development of the lung," *The Journal of pediatrics*, vol. 98, no. 1, pp. 1–15, 1981.

[121] M. Elshazzly, A. A. Anekar, and O. Caban, *Physiology, Newborn*. StatPearls Publishing, Treasure Island (FL), 2025.

[122] S. Rehman and D. Bacha, *Embryology, pulmonary*. StatPearls Publishing, Treasure Island (FL), 2023.

[123] A. LoMauro and A. Aliverti, "Physiology masterclass: Extremes of age: newborn and infancy," *Breathe*, vol. 12, no. 1, pp. 65–68, 2016.

[124] M. J. Herring, L. F. Putney, G. Wyatt, W. E. Finkbeiner, and D. M. Hyde, "Growth of alveoli during postnatal development in humans based on stereological estimation," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 307, no. 4, pp. L338–L344, 2014.

[125] S. Garg and S. Sinha, "Non-invasive respiratory support in preterm infants: do we need more evidence," *Infant*, vol. 10, pp. 44–8, 2014.

[126] T. Gerhardt and E. Bancalari, "Chestwall compliance in full-term and premature infants," *Acta Pædiatrica*, vol. 69, no. 3, pp. 359–364, 1980.

[127] D. Trachsel, T. O. Erb, J. Hammer, and B. S. von Ungern-Sternberg, "Developmental respiratory physiology," *Pediatric Anesthesia*, vol. 32, no. 2, pp. 108–117, 2022.

[128] S. de Gelidi, A. Bardill, N. Seifnaraghi, Y. Wu, A. Demosthenous, M. Rahtu, M. Kallio, and R. Bayford, "Thoracic shape changes in newborns due to their position," *Scientific Reports*, vol. 11, no. 1, p. 4446, 2021.

[129] R. Warren, S. Horan, and P. Robertson, "Chest wall motion in preterm infants using respiratory inductive plethysmography," *European Respiratory Journal*, vol. 10, no. 10, pp. 2295–2300, 1997.

[130] J. C. Smith and S. H. Loring, "Passive mechanical properties of the chest wall," *Comprehensive Physiology*, pp. 429–442, 2011.

[131] S. Fouzas, A. Vervenioti, A. Tsintoni, T. Dassios, A. A. Karatza, and G. Dimitriou, "Diaphragmatic muscle function in term and preterm infants," *European Journal of Pediatrics*, vol. 182, no. 12, pp. 5693–5699, 2023.

[132] C. Dani, M. Fusco, S. Manti, L. Marseglia, M. Ciarcià, V. Leonardi, I. Corsini, and E. Gitto, "Effects of caffeine on diaphragmatic activity in preterm infants," *Pediatric Pulmonology*, vol. 58, no. 7, pp. 2104–2110, 2023.

[133] T. Dassios, A. Vervenioti, and G. Dimitriou, "Respiratory muscle function in the newborn: a narrative review," *Pediatric Research*, vol. 91, no. 4, pp. 795–803, 2022.

[134] M. Richardson, "Understanding the structure and function of the skin.," *Nursing times*, vol. 99, no. 31, pp. 46–48, 2003.

[135] G. V. Baranoski and A. Krishnaswamy, "Light interaction with human skin: from believable images to predictable models," in *ACM SIGGRAPH Asia 2008 courses*, pp. 1–80, Association for Computing Machinery, New York, 2008.

[136] D. Shier, J. Butler, and R. Lewis, *Hole's human anatomy & physiology*. Boston Burr Ridge: McGraw-Hill Higher Education, 2007.

[137] J. G. Betts, K. A. Young, J. A. Wise, E. Johnson, B. Poe, D. H. Kruse, O. Korol, J. E. Johnson, M. Womble, and P. DeSaix, *Anatomy and Physiology 2e*, ch. 5.1 Layers of the Skin. Houston, Texas: OpenStax,Rice University, 2022.

[138] J. Barbieri, K. Wanat, and J. Seykora, "Skin: Basic structure and function," in *Pathobiology of Human Disease*, pp. 1134–1144, Academic Press, 2014.

[139] H. Yousef, M. Alhajj, and S. Sharma, "Anatomy, skin (integument), epidermis," in *StatPearls [Internet]*, StatPearls Publishing, Treasure Island FL, 2017.

[140] G. S. Barsh, "What controls variation in human skin color?," *PLoS biology*, vol. 1, no. 1, p. e27, 2003.

[141] T. M. Brown and K. Krishnamurthy, “Histology, dermis,” in *StatPearls [Internet]*, StatPearls Publishing, 2022.

[142] T. Oranges, V. Dini, and M. Romanelli, “Skin physiology of the neonate and infant: clinical implications,” *Advances in wound care*, vol. 4, no. 10, pp. 587–595, 2015.

[143] L. E. Eichenfield and C. A. Hardaway, “Neonatal dermatology,” *Current opinion in pediatrics*, vol. 11, no. 5, pp. 471–474, 1999.

[144] W. Q. Mok, W. Wang, and S. Y. Liaw, “Vital signs monitoring to detect patient deterioration: An integrative literature review,” *International journal of nursing practice*, vol. 21, pp. 91–98, 2015.

[145] C. Lockwood, T. Conroy-Hiller, and T. Page, “Vital signs,” *JBI reports*, vol. 2, no. 6, pp. 207–230, 2004.

[146] N. Kumar, G. Akangire, B. Sullivan, K. Fairchild, and V. Sampath, “Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront,” *Pediatric Research*, vol. 87, no. 2, pp. 210–220, 2020.

[147] I. J. Brekke, L. H. Puntervoll, P. B. Pedersen, J. Kellett, and M. Brabrand, “The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review,” *PloS one*, vol. 14, no. 1, p. e0210875, 2019.

[148] E. Phillipos, A. L. Solevåg, K. Aziz, S. van Os, G. Pichler, M. O’Reilly, P.-Y. Cheung, and G. M. Schmölzer, “Oxygen saturation and heart rate ranges in very preterm infants requiring respiratory support at birth,” *The Journal of Pediatrics*, vol. 182, pp. 41–46, 2017.

[149] M. Chiera, F. Cerritelli, A. Casini, N. Barsotti, D. Boschiero, F. Cavigioli, C. G. Corti, and A. Manzotti, “Heart rate variability in the perinatal period: a critical and conceptual review,” *Frontiers in Neuroscience*, vol. 14, p. 561186, 2020.

[150] E. Nerdrum Aagaard, A. L. Solevåg, and O. D. Saugstad, “Significance of neonatal heart rate in the delivery room—a review,” *Children*, vol. 10, no. 9, p. 1551, 2023.

[151] W. Daw, R. Kaur, M. Delaney, and H. Elphick, “Respiratory rate is an early predictor of clinical deterioration in children,” *Pediatric Pulmonology*, vol. 55, no. 8, pp. 2041–2049, 2020.

[152] S. A. Kayser, R. Williamson, G. Siefert, D. Roberts, and A. Murray, “Respiratory rate monitoring and early detection of deterioration practices,” *British Journal of Nursing*, vol. 32, no. 13, pp. 620–627, 2023.

[153] E. M. Chiocca, *Advanced pediatric assessment*. Springer Publishing Company, 2024.

[154] A. Fenner, U. Schalk, H. Hoenicke, A. Wendenburg, and T. Roehling, “Periodic breathing in premature and neonatal babies: incidence, breathing pattern, respiratory gas tensions, response to changes in the composition of ambient air,” *Pediatric Research*, vol. 7, no. 4, pp. 174–183, 1973.

[155] S. M. Barlow, “Oral and respiratory control for preterm feeding,” *Current opinion in otolaryngology & head and neck surgery*, vol. 17, no. 3, pp. 179–186, 2009.

[156] I. Wheatley, “Respiratory rate 3: How to take an accurate measurement,” *Nursing Times*, vol. 114, no. 7, pp. 21–22, 2018.

[157] N. Keshvani, K. Berger, A. Gupta, S. DePaola, O. K. Nguyen, and A. N. Makam, “Improving respiratory rate accuracy in the hospital: A quality improvement initiative,” *Journal of hospital medicine*, vol. 14, no. 11, pp. 673–677, 2019.

[158] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, “Respiratory rate: the neglected vital sign,” *Medical Journal of Australia*, vol. 188, no. 11, pp. 657–659, 2008.

[159] J. Hogan, “Why don’t nurses monitor the respiratory rates of patients?,” *British Journal of nursing*, vol. 15, no. 9, pp. 489–492, 2006.

[160] J. H. Lee, L. A. Nathanson, R. C. Burke, B. W. Anthony, N. I. Shapiro, and A. S. Dagan, “Assessment of respiratory rate monitoring in the emergency department,” *JACEP Open*, vol. 5, no. 3, p. e13154, 2024.

[161] A. I. Paganelli, A. G. Mondejar, A. C. da Silva, G. Silva-Calpa, M. F. Teixeira, F. Carvalho, A. Raposo, and M. Endler, “Real-time data analysis in health monitoring systems: A comprehensive systematic literature review,” *Journal of Biomedical Informatics*, vol. 127, p. 104009, 2022.

[162] R. Seth, M. Nagpal, R. Swami, and M. Kaur, “Fast and precise: Ai medical devices at the intersection of speed and accuracy,” *Current Pharmacology Reports*, vol. 12, no. 1, p. 10, 2026.

[163] T. Adjei, R. Purdy, J. Jorge, E. Adams, M. Buckle, R. E. Fry, G. Green, C. Patel, R. Rogers, R. Slater, *et al.*, “New method to measure interbreath intervals in infants for the assessment of apnoea and respiration,” *BMJ open respiratory research*, vol. 8, no. 1, p. e001042, 2021.

[164] J. Jorge, M. Villarroel, S. Chaichulee, G. Green, K. McCormick, and L. Tarassenko, “Assessment of signal processing methods for measuring the respiratory rate in the neonatal intensive care unit,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2335–2346, 2019.

[165] B. B. Hafen and S. Sharma, “Oxygen saturation,” in *StatPearls [Internet]*, StatPearls Publishing: Treasure Island, FL, 2018.

[166] M. Chang, “Optimal oxygen saturation in premature infants,” *Korean journal of pediatrics*, vol. 54, no. 9, p. 359, 2011.

[167] D. G. Sweet, V. Carnielli, G. Greisen, M. Hallman, E. Ozek, A. Te Pas, R. Plavka, C. C. Roehr, O. D. Saugstad, U. Simeoni, *et al.*, “European consensus guidelines on the management of respiratory distress syndrome–2019 update,” *Neonatology*, vol. 115, no. 4, pp. 432–450, 2019.

[168] J. J. Cummings, R. A. Polin, C. on Fetus, Newborn, K. L. Watterberg, B. Poindexter, J. J. Cummings, W. E. Benitz, E. C. Eichenwald, B. B. Poindexter, D. L. Stewart, *et al.*, “Oxygen targeting in extremely low birth weight infants,” *Pediatrics*, vol. 138, no. 2, p. e20161576, 2016.

[169] B. Schmidt and H. Kirpalani, “Safe and sound oxygen therapy for extremely preterm infants: a literature review,” *Pediatric Medicine*, vol. 5, 2022.

[170] V. Manja, S. Lakshminrusimha, and D. J. Cook, “Oxygen saturation target range for extremely preterm infants: a systematic review and meta-analysis,” *JAMA pediatrics*, vol. 169, no. 4, pp. 332–340, 2015.

[171] F. O’Brien, D. Clapham, K. Krysiak, H. Batchelor, P. Field, G. Caivano, M. Pertile, A. Nunn, and C. Tuleu, “Making medicines baby size: the challenges in bridging the formulation gap in neonatal medicine,” *International journal of molecular sciences*, vol. 20, no. 11, p. 2688, 2019.

[172] V. Dumpa and I. D. Avulakunta, “Umbilical artery catheterization,” in *StatPearls [Internet]*, StatPearls Publishing, Treasure Island (FL), 2020.

[173] E. Senechal, D. Radeschi, L. Tao, S. Lv, E. Jeanne, R. Kearney, W. Shalish, and G. S. Anna, “The use of wireless sensors in the neonatal intensive care unit: a study protocol,” *PeerJ*, vol. 11, p. e15578, 2023.

[174] A. T. Reisner, G. D. Clifford, and R. G. Mark, “The physiological basis of the electrocardiogram,” *Advanced methods and tools for ECG data analysis*, vol. 1, p. 25, 2006.

[175] Y. Sattar and L. Chhabra, “Electrocardiogram,” in *StatPearls [Internet]*, StatPearls Publishing, 2022.

[176] F. Monitillo, M. Leone, C. Rizzo, A. Passantino, and M. Iacoviello, “Ventricular repolarization measures for arrhythmic risk stratification,” *World journal of cardiology*, vol. 8, no. 1, p. 57, 2016.

[177] J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm,” *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[178] A. Peterkova and M. Stremy, “The raw ecg signal processing and the detection of qrs complex,” in *IEEE European modelling symposium*, pp. 80–85, 2015.

[179] R. E. Klabunde, *Cardiovascular physiology concepts, 3e*. Lippincott Williams & Wilkins, a Wolters Kluwer business, 2021.

[180] P. De Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. O’Malley, “Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea,” *IEEE transactions on biomedical engineering*, vol. 50, no. 6, pp. 686–696, 2003.

[181] L. Bacharova, E. Triantafyllou, C. Vazaios, I. Tomeckova, I. Paranicova, and R. Tkacova, “The effect of obstructive sleep apnea on qrs complex morphology,” *Journal of electrocardiology*, vol. 48, no. 2, pp. 164–170, 2015.

[182] A. Zarei and B. M. Asl, “Automatic classification of apnea and normal subjects using new features extracted from hrv and ecg-derived respiration signals,” *Biomedical Signal Processing and Control*, vol. 59, p. 101927, 2020.

[183] B. A. Krbec, X. Zhang, I. Chityat, A. Brady-Mine, E. Linton, D. Copeland, B. W. Anthony, E. R. Edelman, and J. M. Davis, “Emerging innovations in neonatal monitoring: a comprehensive review of progress and potential for non-contact technologies,” *Frontiers in Pediatrics*, vol. 12, p. 1442753, 2024.

[184] R. A. Marnani, R. Jaros, J. Pavlicek, R. Martinek, and R. V. Kahankova, “Advancements and challenges in non-invasive electrocardiography for prenatal, intrapartum, and postnatal care: A comprehensive review,” *IEEE Access*, vol. 12, pp. 44730–44747, 2024.

[185] K. Brockmeier, R. Nazal, and N. Sreeram, “The electrocardiogram of the neonate and infant,” *Journal of Electrocardiology*, vol. 49, no. 6, pp. 814–816, 2016.

[186] M. Chu, T. Nguyen, V. Pandey, Y. Zhou, H. N. Pham, R. Bar-Yoseph, S. Radom-Aizik, R. Jain, D. M. Cooper, and M. Khine, “Respiration rate and volume measurements using wearable strain sensors,” *NPJ digital medicine*, vol. 2, no. 1, p. 8, 2019.

[187] F. Q. AL-Khalidi, R. Saatchi, D. Burke, H. Elphick, and S. Tan, “Respiration rate monitoring methods: A review,” *Pediatric pulmonology*, vol. 46, no. 6, pp. 523–529, 2011.

[188] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, P. J. Watkinson, and J. Alastruey, “An impedance pneumography signal quality index: Design, assessment and application to respiratory rate monitoring,” *Biomedical Signal Processing and Control*, vol. 65, p. 102339, 2021.

[189] V.-P. Seppä, J. Hyttinen, and J. Viik, “Impedance pneumography for ambulatory assessment of obstructive respiratory diseases,” in *Recent Advances in Ambient Assisted Living–Bridging Assistive Technologies, e-Health and Personalized Health Care*, pp. 45–62, IOS Press, 2015.

[190] A. F. Pacela, “Impedance pneumography—a survey of instrumentation techniques,” *Medical and biological engineering*, vol. 4, pp. 1–15, 1966.

[191] M. Gaiduk, L. Weber, Á. S. Alarcón, R. Seepold, N. M. Madrid, S. Orcioni, and M. Conti, “Design of a sleep apnoea detection system for a home environment,” *Procedia computer science*, vol. 192, pp. 3225–3234, 2021.

[192] J. Gracia-Tabuenca, V.-P. Seppä, M. Jauhiainen, A. Kotaniemi-Syrjänen, K. Malmström, A. Pelkonen, M. Mäkelä, J. Viik, and L. P. Malmberg, “Tidal breathing flow volume profiles during sleep in wheezing infants measured by impedance pneumography,” *Journal of Applied Physiology*, vol. 126, no. 5, pp. 1409–1418, 2019.

[193] K. Goyal, D. A. Borkholder, and S. W. Day, “Dependence of skin-electrode contact impedance on material and skin hydration,” *Sensors*, vol. 22, no. 21, p. 8510, 2022.

[194] N. Marjanovic, O. Mimoz, and J. Guenezan, “An easy and accurate respiratory rate monitor is necessary,” *Journal of clinical monitoring and computing*, vol. 34, pp. 221–222, 2020.

[195] R. T. Brouillette, A. S. Morrow, D. E. Weese-Mayer, and C. E. Hunt, "Comparison of respiratory inductive plethysmography and thoracic impedance for apnea monitoring," *The Journal of pediatrics*, vol. 111, no. 3, pp. 377–383, 1987.

[196] D. W. Ball, "The electromagnetic spectrum: a history," *Spectroscopy*, vol. 22, no. 3, p. 14, 2007.

[197] R. Morris, "Spectrophotometry," *Current Protocols Essential Laboratory Techniques*, vol. 11, no. 1, pp. 2–1, 2015.

[198] J. G. Webster, *Design of pulse oximeters*. Crc Press, 1997.

[199] I. Oshina and J. Spigulis, "Beer–lambert law for optical tissue diagnostics: current state of the art and the main limitations," *Journal of biomedical optics*, vol. 26, no. 10, pp. 100901–100901, 2021.

[200] P. D. Mannheimer, "The light–tissue interaction of pulse oximetry," *Anesthesia & Analgesia*, vol. 105, no. 6, pp. S10–S17, 2007.

[201] A. Krishnaswamy and G. V. Baranoski, "A study on skin optics," *Natural Phenomena Simulation Group, School of Computer Science, University of Waterloo, Canada, Technical Report*, vol. 1, pp. 1–17, 2004.

[202] P. van Beest, G. Wietasch, T. Scheeren, P. Spronk, and M. Kuiper, "Clinical review: use of venous oxygen saturations as a goal-a yet unfinished puzzle," *Critical care*, vol. 15, pp. 1–9, 2011.

[203] H. Christopherson and E. Crimi, "Mixed venous oxygen saturation," *ADVANCED ANESTHESIA*, p. 7, 2023.

[204] A. N. Krauss, P. W. Post, S. Waldman, and P. A. Auld, "Skin reflectance in the newborn infant," *Pediatric Research*, vol. 10, no. 9, pp. 776–778, 1976.

[205] V. Harpin and N. Rutter, "Barrier properties of the newborn infant's skin," *The Journal of pediatrics*, vol. 102, no. 3, pp. 419–425, 1983.

[206] M. Elgendi, R. Fletcher, Y. Liang, N. Howard, N. H. Lovell, D. Abbott, K. Lim, and R. Ward, "The use of photoplethysmography for assessing hypertension," *NPJ digital medicine*, vol. 2, no. 1, pp. 1–11, 2019.

[207] H. Ugnell and P. Öberg, "The time-variable photoplethysmographic signal; dependence of the heart synchronous signal on wavelength and sample volume," *Medical engineering & physics*, vol. 17, no. 8, pp. 571–578, 1995.

[208] Y. Sun and N. Thakor, "Photoplethysmography revisited: from contact to noncontact, from point to imaging," *IEEE transactions on biomedical engineering*, vol. 63, no. 3, pp. 463–477, 2015.

[209] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors—past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, 2014.

[210] Y. Mendelson, "Pulse oximetry: theory and applications for noninvasive monitoring," *Clinical chemistry*, vol. 38, no. 9, pp. 1601–1607, 1992.

[211] S. Hu, V. Azorin-Peris, J. Zheng, *et al.*, "Opto-physiological modeling applied to photoplethysmographic cardiovascular assessment," *Journal of healthcare engineering*, vol. 4, pp. 505–528, 2013.

[212] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, p. R1, 2007.

[213] J. Park, H. S. Seok, S.-S. Kim, and H. Shin, "Photoplethysmogram analysis and applications: an integrative review," *Frontiers in physiology*, vol. 12, p. 808451, 2022.

[214] M. Nitzan, A. Romem, and R. Koppel, "Pulse oximetry: fundamentals and technology update," *Medical Devices: Evidence and Research*, pp. 231–239, 2014.

[215] J. E. Sinex, “Pulse oximetry: principles and limitations,” *The American journal of emergency medicine*, vol. 17, no. 1, pp. 59–66, 1999.

[216] T. Duke, R. Subhi, D. Peel, and B. Frey, “Pulse oximetry: technology to reduce child mortality in developing countries,” *Annals of tropical paediatrics*, vol. 29, no. 3, pp. 165–175, 2009.

[217] T. Y. Abay and P. A. Kyriacou, “Reflectance photoplethysmography as noninvasive monitoring of tissue blood perfusion,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2187–2195, 2015.

[218] E. D. Chan, M. M. Chan, and M. M. Chan, “Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations,” *Respiratory medicine*, vol. 107, no. 6, pp. 789–799, 2013.

[219] Y. Sun, *Imaging photoplethysmography: Towards effective physiological measurements*. PhD thesis, Loughborough University, 2011.

[220] H. Mansor, M. H. A. Shukor, S. S. Meskam, N. Q. A. M. Rusli, and N. S. Zamery, “Body temperature measurement for remote health monitoring system,” in *2013 IEEE International conference on smart instrumentation, measurement and applications (ICSIMA)*, pp. 1–5, IEEE, 2013.

[221] A. Dassel, R. Graaff, M. Aardema, W. Zijlstra, and J. Aarnoudse, “Effect of location of the sensor on reflectance pulse oximetry,” *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 104, no. 8, pp. 910–916, 1997.

[222] M. G. Joo, D. H. Lim, K.-K. Park, J. Baek, J. M. Choi, and H. W. Baac, “Reflection-boosted wearable ring-type pulse oximeters for spo2 measurement with high sensitivity and low power consumption,” *Biosensors*, vol. 13, no. 7, p. 711, 2023.

[223] S. Stockwell, T. Kwok, S. Morgan, D. Sharkey, and B. Hayes-Gill, “Forehead monitoring of heart rate in neonatal intensive care,” *Frontiers in Physiology*, vol. 14, p. 531, 2023.

[224] M. Grubb, J. Carpenter, J. A. Crowe, J. Teoh, N. Marlow, C. Ward, C. Mann, D. Sharkey, and B. R. Hayes-Gill, “Forehead reflectance photoplethysmography to monitor heart rate: preliminary results from neonatal patients,” *Physiological measurement*, vol. 35, no. 5, p. 881, 2014.

[225] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, “A review on wearable photoplethysmography sensors and their potential future applications in health care,” *International journal of biosensors & bioelectronics*, vol. 4, no. 4, p. 195, 2018.

[226] A. C. Kevat, D. V. Bullen, P. G. Davis, and C. O. F. Kamlin, “A systematic review of novel technology for monitoring infant and newborn heart rate,” *Acta Paediatrica*, vol. 106, no. 5, pp. 710–720, 2017.

[227] F. Peng, W. Wang, and H. Liu, “Development of a reflective ppg signal sensor,” in *2014 7th International Conference on Biomedical Engineering and Informatics*, pp. 612–616, IEEE, 2014.

[228] C. R. Amirtharaj, L. C. Palmeri, G. Gradwohl, Y. Adar, M. Nitzan, D. Gruber, A. Blaufox, D. B. Meyer, E. N. Kwon, and R. Koppel, “Photoplethysmographic assessment of pulse transit time correlates with echocardiographic measurement of stroke volume in preterm infants with patent ductus arteriosus,” *Journal of perinatology*, vol. 38, no. 9, pp. 1220–1226, 2018.

[229] J. Moyle, “Uses and abuses of pulse oximetry.,” *Archives of disease in childhood*, vol. 74, no. 1, p. 77, 1996.

[230] S. Fouzas, K. N. Priftis, and M. B. Anthracopoulos, “Pulse oximetry in pediatric practice,” *Pediatrics*, vol. 128, no. 4, pp. 740–752, 2011.

[231] M. J. Tobin and A. Jubran, “Pulse oximetry, racial bias and statistical bias,” *Annals of intensive care*, vol. 12, no. 1, pp. 1–2, 2022.

[232] D. L. J. Bunker, R. Kumar, A. Martin, and S. P. Pegg, “Thermal injuries caused by medical instruments: A case report of burns caused by a pulse oximeter,” *Journal of burn care & research*, vol. 35, no. 2, pp. e132–e134, 2014.

[233] M. Villarroel, J. Jorge, C. Pugh, and L. Tarassenko, "Non-contact vital sign monitoring in the clinic," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 278–285, IEEE, 2017.

[234] W. Wang, *Robust and automatic remote photoplethysmography*. PhD thesis, Eindhoven University of Technology, 2017.

[235] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. Clifton, and C. Pugh, "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," *Physiological measurement*, vol. 35, no. 5, p. 807, 2014.

[236] M. Villarroel, A. Guazzi, J. Jorge, S. Davis, P. Watkinson, G. Green, A. Shenvi, K. McCormick, and L. Tarassenko, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthcare technology letters*, vol. 1, no. 3, pp. 87–91, 2014.

[237] V. Blazek, T. Wu, and D. Hoelscher, "Near-infrared ccd imaging: Possibilities for noninvasive and contactless 2d mapping of dermal venous hemodynamics," in *Optical Diagnostics of Biological Fluids V*, vol. 3923, pp. 2–9, SPIE, 2000.

[238] K. S. Tan, R. Saatchi, H. Elphick, and D. Burke, "Real-time vision based respiration monitoring system," in *2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010)*, pp. 770–774, IEEE, 2010.

[239] Y.-W. Bai, W.-T. Li, and C.-H. Yeh, "Design and implementation of an embedded monitor system for body breath detection by using image processing methods," in *2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, pp. 193–194, IEEE, 2010.

[240] J. Jorge, M. Villarroel, H. Tomlinson, O. Gibson, J. L. Darbyshire, J. Ede, M. Harford, J. D. Young, L. Tarassenko, and P. Watkinson, "Non-contact physiological monitoring of post-operative patients in the intensive care unit," *NPJ digital medicine*, vol. 5, no. 1, p. 4, 2022.

[241] C. H. Antink, S. Lyra, M. Paul, X. Yu, and S. Leonhardt, "A broader look: Camera-based vital sign estimation across the spectrum," *Yearbook of medical informatics*, vol. 28, no. 1, p. 102, 2019.

[242] A. Caroppo, A. Manni, G. Rescio, P. Siciliano, and A. Leone, "Vital signs estimation in elderly using camera-based photoplethysmography," *Multimedia Tools and Applications*, vol. 83, no. 24, pp. 65363–65386, 2024.

[243] P. S. Addison and J. N. Watson, "Secondary transform decoupling of shifted nonstationary signal modulation components: application to photoplethysmography," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 01, pp. 43–57, 2004.

[244] P. H. Charlton, T. Bonnici, L. Tarassenko, J. Alastruey, D. A. Clifton, R. Beale, and P. J. Watkinson, "Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants," *Physiological measurement*, vol. 38, no. 5, p. 669, 2017.

[245] S. Zhuang, F. Li, Z. Zhuang, W. Rao, A. N. Joseph Raj, and V. Rajangam, "Improved ecg-derived respiration using empirical wavelet transform and kernel principal component analysis," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, p. 1360414, 2021.

[246] P. H. Charlton, T. Bonnici, L. Tarassenko, D. A. Clifton, R. Beale, and P. J. Watkinson, "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram," *Physiological measurement*, vol. 37, no. 4, p. 610, 2016.

[247] X. Bao, A. K. Abdala, and E. N. Kamavuako, "Estimation of the respiratory rate from localised ecg at different auscultation sites," *Sensors*, vol. 21, no. 1, p. 78, 2020.

[248] C. Varon, J. Morales, J. Lázaro, M. Orini, M. Deviaene, S. Kontaxis, D. Testelmans, B. Buyse, P. Borzée, L. Sörnmo, *et al.*, "A comparative study of ecg-derived respiration in ambulatory monitoring using the single-lead ecg," *Scientific reports*, vol. 10, no. 1, p. 5704, 2020.

[249] M. Schmidt, A. Schumann, J. Mueller, K.-J. Baer, and G. Rose, "Ecg derived respiration: comparison of time-domain approaches and application to altered breathing patterns of patients with schizophrenia," *Physiological measurement*, vol. 38, no. 4, p. 601, 2017.

[250] E. Helfenbein, R. Firoozabadi, S. Chien, E. Carlson, and S. Babaeizadeh, "Development of three methods for extracting respiration from the surface ecg: A review," *Journal of electrocardiology*, vol. 47, no. 6, pp. 819–825, 2014.

[251] H.-b. Wang, C.-w. Yen, J.-t. Liang, Q. Wang, G.-z. Liu, and R. Song, "A robust electrode configuration for bioimpedance measurement of respiration," *Journal of healthcare engineering*, vol. 5, no. 3, pp. 313–328, 2014.

[252] C. S. Zandvoort, F. Usman, S. Robinson, O. Fatunla, E. Adams, K. T. Pattinson, S. F. Farmer, and C. Hartley, "Cortical motor activity modulates respiration and reduces apnoea in neonates," *bioRxiv*, pp. 2025–04, 2025.

[253] H. Jung, S. Mabrouk, and O. T. Inan, "Impedance pneumography: Assessment of dual-frequency calibration approaches," in *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–4, IEEE, 2021.

[254] H. Liu, F. Chen, V. Hartmann, S. G. Khalid, S. Hughes, and D. Zheng, "Comparison of different modulations of photoplethysmography in extracting respiratory rate: From a physiological perspective," *Physiological Measurement*, vol. 41, no. 9, p. 094001, 2020.

[255] M. A. Pimentel, A. E. Johnson, P. H. Charlton, D. Birrenkott, P. J. Watkinson, L. Tarassenko, and D. A. Clifton, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2016.

[256] E. Mejia-Mejia, J. Allen, K. Budidha, C. El-Hajj, P. A. Kyriacou, and P. H. Charlton, "Photoplethysmography signal processing and synthesis," in *Photoplethysmography*, pp. 69–146, Elsevier, 2022.

[257] V. Hartmann, H. Liu, F. Chen, W. Hong, S. Hughes, and D. Zheng, "Toward accurate extraction of respiratory frequency from the photoplethysmogram: Effect of measurement site," *Frontiers in Physiology*, vol. 10, p. 732, 2019.

[258] F. Massie, S. Vits, A. Khachatryan, B. Van Pee, J. Verbraecken, and J. Bergmann, "Central sleep apnea detection by means of finger photoplethysmography," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 11, pp. 126–136, 2023.

[259] S. D. Xavier, C. A. Eckley, A. C. Duprat, L. H. de Souza Fontes, T. Navarro-Rodriguez, J. Patrocínio, D. Tridente, and G. Lorenzi-Filho, "Temporal association between respiratory events and reflux in patients with obstructive sleep apnea and laryngopharyngeal reflux," *Journal of Clinical Sleep Medicine*, vol. 15, no. 10, pp. 1397–1402, 2019.

[260] G. Korompili, L. Kokkalas, S. A. Mitilineos, N.-A. Tatlas, and S. M. Potirakis, "Detecting apnea/hypopnea events time location from sound recordings for patients with severe or moderate sleep apnea syndrome," *Applied Sciences*, vol. 11, no. 15, p. 6888, 2021.

[261] A. Azarbarzin, G. Labarca, Y. Kwon, and A. Wellman, "Physiologic consequences of upper airway obstruction in sleep apnea," *Chest*, vol. 166, no. 5, pp. 1209–1217, 2024.

[262] N. Blekic, I. Bold, T. Mettay, and M. Bruyneel, "Impact of desaturation patterns versus apnea–hypopnea index in the development of cardiovascular comorbidities in obstructive sleep apnea patients," *Nature and Science of Sleep*, pp. 1457–1468, 2022.

[263] Y. Zhang, Y. Shi, Y. Su, Z. Cao, C. Li, Y. Xie, X. Niu, Y. Yuan, L. Ma, S. Zhu, *et al.*, "Detection and severity assessment of obstructive sleep apnea according to deep learning of single-lead electrocardiogram signals," *Journal of Sleep Research*, vol. 34, no. 1, p. e14285, 2025.

[264] M. Bsoul, H. Minn, and L. Tamil, "Apnea medassist: real-time sleep apnea monitor using single-lead ecg," *IEEE transactions on information technology in biomedicine*, vol. 15, no. 3, pp. 416–427, 2010.

[265] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Transactions on information technology in biomedicine*, vol. 16, no. 3, pp. 469–477, 2012.

[266] C. Song, K. Liu, X. Zhang, L. Chen, and X. Xian, "An obstructive sleep apnea detection approach using a discriminative hidden markov model from ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1532–1542, 2015.

[267] I. García-López, R. X. A. Pramono, and E. Rodriguez-Villegas, "Artifacts classification and apnea events detection in neck photoplethysmography signals," *Medical & Biological Engineering & Computing*, vol. 60, no. 12, pp. 3539–3554, 2022.

[268] S. Haghighat, M. Joghatayi, J. Issa, S. Azimian, J. Brinz, A. Ashkan, A. Chaurasia, Z. Rahimian, and L. Sangalli, "Diagnostic accuracy of artificial intelligence for obstructive sleep apnea detection: a systematic review," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, p. 278, 2025.

[269] E. Urtnasan, J.-U. Park, E.-Y. Joo, and K.-J. Lee, "Automated detection of obstructive sleep apnea events from a single-lead electrocardiogram using a convolutional neural network," *Journal of medical systems*, vol. 42, pp. 1–8, 2018.

[270] S. Biswal, H. Sun, B. Goparaju, M. B. Westover, J. Sun, and M. T. Bianchi, "Expert-level sleep scoring with deep neural networks," *Journal of the American Medical Informatics Association*, vol. 25, no. 12, pp. 1643–1650, 2018.

[271] R. V. Sharan, S. Berkovsky, H. Xiong, and E. Coiera, "End-to-end sleep apnea detection using single-lead ecg signal and 1-d residual neural networks," *Journal of Medical and Biological Engineering*, vol. 41, no. 5, pp. 758–766, 2021.

[272] J. Tao, J. Huang, B. Miao, and L. Yang, "A multimodal dataset for training deep learning models aimed at detecting and analyzing sleep apnea," *Scientific Data*, vol. 12, no. 1, p. 1263, 2025.

[273] H. Fayyaz, N. S. D'Souza, and R. Beheshti, "Multimodal sleep apnea detection with missing or noisy modalities," *Proceedings of machine learning research*, vol. 252, pp. https–proceedings, 2024.

[274] T. Paul, O. Hassan, K. Alaboud, H. Islam, M. K. Z. Rana, S. K. Islam, and A. S. Mosa, "Ecg and spo2 signal-based real-time sleep apnea detection using feed-forward artificial neural network," in *AMIA Annual Symposium Proceedings*, vol. 2022, p. 379, American Medical Informatics Association, 2022.

[275] R. Lazazzera, M. Deviaene, C. Varon, B. Buyse, D. Testelmans, P. Laguna, E. Gil, and G. Carrault, "Detection and classification of sleep apnea and hypopnea using ppg and spo $_2$ signals," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 5, pp. 1496–1506, 2020.

[276] J. R. Williamson, D. W. Bliss, D. W. Browne, P. Indic, E. Bloch-Salisbury, and D. Paydarfar, "Using physiological signals to predict apnea in preterm infants," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 1098–1102, IEEE, 2011.

[277] J. R. Williamson, D. W. Bliss, D. W. Browne, P. Indic, E. Bloch-Salisbury, and D. Paydarfar, "Individualized apnea prediction in preterm infants using cardio-respiratory and movement signals," in *2013 IEEE International Conference on Body Sensor Networks*, pp. 1–6, IEEE, 2013.

[278] R. D. Shirwaikar, D. U. Acharya, K. Makkithaya, M. Surulivelrajan, and L. E. S. Lewis, "Machine learning techniques for neonatal apnea prediction," *Journal of Artificial Intelligence*, vol. 9, no. 1-3, 2016.

[279] G. Varisco, Z. Peng, D. Kommers, Z. Zhan, W. Cottaar, P. Andriessen, X. Long, and C. van Pul, "Central apnea detection in premature infants using machine learning," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107155, 2022.

[280] G. Varisco, Z. Peng, D. Kommers, E. J. Cottaar, P. Andriessen, X. Long, and C. van Pul, "Detecting central apneas using multichannel signals in premature infants," *Physiological Measurement*, 2024.

[281] R. D. Shirwaikar, D. Acharya, K. Makkithaya, M. Surulivelrajan, and S. Srivastava, "Optimizing neural networks for medical data sets: A case study on neonatal apnea prediction," *Artificial intelligence in medicine*, vol. 98, pp. 59–76, 2019.

[282] N. Vahabi, R. Yerworth, M. Miedema, A. Van Kaam, R. Bayford, and A. Demosthenous, "Deep analysis of eit dataset to classify apnea and non-apnea cases in neonatal patients," *IEEE Access*, vol. 9, pp. 25131–25139, 2021.

[283] A. P. Marshall, K. Lim, S. K. Ali, T. J. Gale, and P. A. Dargaville, "Physiological instability after respiratory pauses in preterm infants," *Pediatric Pulmonology*, vol. 54, no. 11, pp. 1712–1721, 2019.

[284] A. Janulionis, V. Sutova, V. Langiene, E. Virsilas, V. Drejeriene, A. Liubsys, and A. Valiulis, "Regional differences in lung ventilation during the early transition period in late preterm and term neonates assessed by electrical impedance tomography," *Children*, vol. 11, no. 11, p. 1314, 2024.

[285] A. Marshall, O. J. Ladlow, C. Bannink, K. Lim, S. K. Ali, T. J. Gale, and P. A. Dargaville, "Apnoea-triggered increase in fraction of inspired oxygen in preterm infants: a randomised cross-over study," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 109, no. 1, pp. 81–86, 2024.

[286] D. E. Elder, A. J. Campbell, and D. Galletly, "Current definitions for neonatal apnoea: are they evidence based?," *Journal of paediatrics and child health*, vol. 49, no. 9, pp. E388–E396, 2013.

[287] L. Pinilla, C. L. Chai-Coetzer, and D. J. Eckert, "Diagnostic modalities in sleep disordered breathing: Current and emerging technology and its potential to transform diagnostics," *Respirology*, vol. 30, no. 4, pp. 286–302, 2025.

[288] J. Abad, A. Muñoz-Ferrer, M. Á. Cervantes, C. Esquinas, A. Marin, C. Martínez, J. Morera, and J. Ruiz, "Automatic video analysis for obstructive sleep apnea diagnosis," *Sleep*, vol. 39, no. 8, pp. 1507–1515, 2016.

[289] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pp. 45–57, 2015.

[290] S. M. M. Islam, N. Molinaro, S. Silvestri, E. Schena, and C. Massaroni, "Respiratory feature extraction for contactless breathing pattern recognition using a single digital camera," *IEEE Transactions on Human-Machine Systems*, 2023.

[291] M. Gwak, K. Vatanparvar, L. Zhu, J. Kuang, and A. Gao, "Contactless monitoring of respiratory rate and breathing absence from head movements using an rgb camera," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1–4, IEEE, 2023.

[292] S. Akbarian, N. Montazeri Ghahjaverestan, A. Yadollahi, and B. Taati, "Distinguishing obstructive versus central apneas in infrared video of sleep using deep learning: Validation study," *Journal of Medical Internet Research*, vol. 22, no. 5, p. e17252, 2020.

[293] I. R. Lorato, *Video respiration monitoring: Towards remote apnea detection in the clinic*. PhD thesis, Eindhoven University of Technology, 2021.

[294] J. Zhang, Z. Tang, J. Gao, L. Lin, Z. Liu, H. Wu, F. Liu, R. Yao, *et al.*, "Automatic detection of obstructive sleep apnea events using a deep cnn-lstm model," *Computational intelligence and neuroscience*, vol. 2021, 2021.

[295] D. Dey, S. Chaudhuri, and S. Munshi, "Obstructive sleep apnoea detection using convolutional neural network based deep learning framework," *Biomedical engineering letters*, vol. 8, no. 1, pp. 95–100, 2018.

[296] X. Jiang, Y. Ren, H. Wu, Y. Li, and F. Liu, "Convolutional neural network based on photoplethysmography signals for sleep apnea syndrome detection," *Frontiers in Neuroscience*, vol. 17, 2023.

[297] M. Krichen, "Convolutional neural networks: A survey," *Computers*, vol. 12, no. 8, p. 151, 2023.

[298] L. Borawar and R. Kaur, "Resnet: Solving vanishing gradient in deep networks," in *Proceedings of International Conference on Recent Trends in Computing: ICRTC 2022*, pp. 235–247, Springer, 2023.

[299] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11976–11986, June 2022.

[300] K. Helm, A. Epp, M. Gerstlauer, M. Kaspar, L. C. Hinske, M. L. Conrad, and F. B. Fahlbusch, "Oxycardiorespirograms in neonatal monitoring—current gaps and future potential of artificial intelligence: a mini-review," *Frontiers in Pediatrics*, vol. 13, p. 1680074, 2025.

[301] F. Del Pup and M. Atzori, "Applications of self-supervised learning to biomedical signals: where are we now," *Authorea Preprints*, 2023.

[302] A. Husain, L. Knake, B. Sullivan, J. Barry, K. Beam, E. Holmes, T. Hooven, R. McAdams, A. Moreira, W. Shalish, *et al.*, "Ai models in clinical neonatology: a review of modeling approaches and a consensus proposal for standardized reporting of model performance," *Pediatric Research*, vol. 98, no. 2, pp. 412–422, 2025.

[303] D. Pais, S. Brás, and R. Sebastião, "Overcoming the small dataset challenge in healthcare," in *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*, pp. 497–502, IEEE, 2024.

[304] H. Huang, Y. Shi, Y. Hong, L. Zhu, M. Li, and Y. Zhang, "A nomogram for predicting neonatal apnea: a retrospective analysis based on the mimic database," *Frontiers in Pediatrics*, vol. 12, p. 1357972, 2024.

[305] G. Varisco, Z. Peng, H. Niemarkt, P. Andriessen, C. van Pul, and X. Long, "Improving central apnea detection in preterm infants: Neural network strategies and clinical impact," *Computer Methods and Programs in Biomedicine*, p. 109038, 2025.

[306] V. W. Lumumba, D. Kiprotich, M. Lemasulani Mpaine, N. Grace Makena, and M. Daniel Kavita, "Comparative analysis of cross-validation techniques: Loocv, k-folds cross-validation, and repeated k-folds cross-validation in machine learning models," *K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models (June 01, 2024)*, 2024.

[307] T. Abedin, H. Xu, and S. Uddin, "The impact of k selection in k-fold cross-validation on bias and variance in supervised learning models," *Scientific Reports*, 2026.

[308] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, 2023.

[309] M. A. Hasan, F. Li, P. Gouverneur, A. Piet, and M. Grzegorzek, "A comprehensive survey and comparative analysis of time series data augmentation in medical wearable computing," *Plos one*, vol. 20, no. 3, p. e0315343, 2025.

[310] I. Annaki, M. Rahmoune, and M. Bourhaleb, "Overview of data augmentation techniques in time series analysis," *Int. J. Adv. Comput. Sci. Appl*, vol. 15, no. 1, pp. 1201–1211, 2024.

[311] P. Wicaksono, P. Samuel, I. N. Alam, and S. M. Isa, "Dealing with imbalanced sleep apnea data using dcgan," *Traitement du Signal*, vol. 39, no. 5, p. 1527, 2022.

[312] S. R. Pandey, A. Saeed, and H. Lee, "Pedsleepmae: Generative model for multimodal pediatric sleep signals," in *2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 1–8, IEEE, 2024.

[313] F. Di Martino and F. Delmastro, "Challenges and limitations in the synthetic generation of mhealth sensor data," *arXiv e-prints*, pp. arXiv–2505, 2025.

[314] F.-T.-Z. Khanam, A. G. Perera, A. Al-Naji, K. Gibson, and J. Chahl, "Non-contact automatic vital signs monitoring of infants in a neonatal intensive care unit based on neural networks," *Journal of Imaging*, vol. 7, no. 8, p. 122, 2021.

[315] J. H. Klaessens, M. Van Den Born, A. Van Der Veen, J. Sikkens-Van De Kraats, F. A. van den Dungen, and R. M. Verdaasdonk, "Development of a baby friendly non-contact method for measuring vital signs: first results of clinical measurements in an open incubator at a neonatal intensive care unit," in *Advanced biomedical and clinical diagnostic systems XII*, vol. 8935, pp. 257–263, SPIE, 2014.

[316] The British Association of Perinatal Medicine (BAPM), "Categories of Care 2011: A BAPM Framework for Practice," tech. rep., British Association of Perinatal Medicine,London, 2011.

[317] A. Hijazi, A. Al-Masri, and N. Rawashdeh, "On the use of bayer sensor color cameras in digital image correlation," in *2022 11th international symposium on signal, image, video and communications (ISIVC)*, pp. 1–7, IEEE, 2022.

[318] S. Chaichulee, *Non-contact vital sign monitoring of pre-term infants*. PhD thesis, University of Oxford, 2018.

[319] A. R. Guazzi, M. Villarroel, J. Jorge, J. Daly, M. C. Frise, P. A. Robbins, and L. Tarassenko, "Non-contact measurement of oxygen saturation with an rgb camera," *Biomedical optics express*, vol. 6, no. 9, pp. 3320–3338, 2015.

[320] R. Saatchi, H. Elphick, J. Rowson, M. Wesseler, J. Marris, S. Shortland, and L. Thomas, "Development of a new prototype paediatric central sleep apnoea monitor," *Technologies*, vol. 12, no. 7, p. 116, 2024.

[321] W. Lu, M. M. Nystrom, P. J. Parikh, D. R. Fooshee, J. P. Hubenschmidt, J. D. Bradley, and D. A. Low, "A semi-automatic method for peak and valley detection in free-breathing respiratory waveforms," *Medical physics*, vol. 33, no. 10, pp. 3634–3636, 2006.

[322] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter," *Physiological measurement*, vol. 29, no. 1, p. 15, 2007.

[323] B. Hjorth, "Eeg analysis based on time domain properties," *Electroencephalography and clinical neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[324] S. Khandelwal, N. Salankar, S. Mian Qaisar, J. Upadhyay, and P. Pławiak, "Ecg based apnea detection by multirate processing hybrid of wavelet-empirical decomposition hjorth features extraction and neural networks," *Plos one*, vol. 18, no. 11, p. e0293610, 2023.

[325] B. Mulgrew, P. Grant, and J. Thompson, *Digital signal processing: concepts and applications*. Bloomsbury Publishing, 2002.

[326] M. A. Mansournia, R. Waters, M. Nazemipour, M. Bland, and D. G. Altman, "Bland-altman methods for comparing methods of measurement and response to criticisms," *Global Epidemiology*, vol. 3, p. 100045, 2021.

[327] G. Sırtbaş-Işık, D. Porsnok, B. N. Yardımcı-Lokmanoğlu, and A. Mutlu, "Sleep characteristics, early spontaneous movements, and developmental functioning in preterm infants in the early postnatal period," *Sleep Medicine*, vol. 114, pp. 151–158, 2024.

[328] F. Giganti, G. Cioni, E. Biagioni, M. T. Puliti, A. Boldrini, and P. Salzarulo, "Activity patterns assessed throughout 24-hour recordings in preterm and near term infants," *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, vol. 38, no. 2, pp. 133–142, 2001.

[329] J. Honda, M. Murakawa, and S. Inoue, "Effect of averaging time and respiratory pause time on the measurement of acoustic respiration rate monitoring," *JA Clinical Reports*, vol. 9, no. 1, p. 61, 2023.

[330] R. J. Martin, K. Wang, Ö. Köroğlu, J. Di Fiore, and P. Kc, "Intermittent hypoxic episodes in preterm infants: do they matter?," *Neonatology*, vol. 100, no. 3, pp. 303–310, 2011.

[331] K. D. Fairchild, V. P. Nagraj, B. A. Sullivan, J. R. Moorman, and D. E. Lake, "Oxygen desaturations in the early neonatal period predict development of bronchopulmonary dysplasia," *Pediatric Research*, vol. 85, no. 7, pp. 987–993, 2019.

[332] R. Joshi, C. van Pul, L. Atallah, L. Feijs, S. Van Huffel, and P. Andriessen, “Pattern discovery in critical alarms originating from neonates under intensive care,” *Physiological Measurement*, vol. 37, no. 4, pp. 564–579, 2016.

[333] J. L. Fleiss, “Measuring nominal scale agreement among many raters.,” *Psychological bulletin*, vol. 76, no. 5, p. 378, 1971.

[334] D. G. Altman, *Practical statistics for medical research*. CRC Press, 1990.

[335] P. H. Charlton, D. A. Birrenkott, T. Bonnici, M. A. Pimentel, A. E. Johnson, J. Alastruey, L. Tarassenko, P. J. Watkinson, R. Beale, and D. A. Clifton, “Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review,” *IEEE reviews in biomedical engineering*, vol. 11, pp. 2–20, 2017.

[336] A. M. Ponsiglione, M. Russo, M. G. Petrellese, A. Letizia, V. Tufano, C. Ricciardi, A. Tedesco, F. Amato, and M. Romano, “Comparison of techniques for respiratory rate extraction from electrocardiogram and photoplethysmogram,” *Sensors*, vol. 25, no. 16, p. 5136, 2025.

[337] R. Joshi, D. Kommers, X. Long, L. Feijs, S. Van Huffel, C. van Pul, and P. Andriessen, “Cardiorespiratory coupling in preterm infants,” *Journal of Applied Physiology*, vol. 126, no. 1, pp. 202–213, 2019.

[338] J. G Rolant and K. Dhanalakshmi, “Apnoea detection using ecg signal based on machine learning classifiers and its performances.,” *Journal of Medical Engineering & Technology*, pp. 1–11, 2024.

[339] J. Lázaro, A. Alcaine, D. Romero, E. Gil, P. Laguna, E. Pueyo, and R. Bailón, “Electrocardiogram derived respiratory rate from qrs slopes and r-wave angle,” *Annals of biomedical engineering*, vol. 42, pp. 2072–2083, 2014.

[340] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero, “Derivation of respiratory signals from multi-lead ecgs,” *Computers in cardiology*, vol. 12, no. 1985, pp. 113–116, 1985.

[341] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, “The apnea-ecg database,” in *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*, pp. 255–258, IEEE, 2000.

[342] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization.,” *Journal of machine learning research*, vol. 13, no. 2, 2012.

[343] J. Carter, J. Jorge, B. Venugopal, O. Gibson, and L. Tarassenko, “Deep learning-enabled sleep staging from vital signs and activity measured using a near-infrared video camera,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5940–5949, 2023.

[344] M. A. Talukder, M. Khalid, M. Kazi, N. J. Muna, M. Nur-e Alam, S. Halder, and N. Sultana, “A hybrid cardiovascular arrhythmia disease detection using convnext-x models on electrocardiogram signals,” *Scientific Reports*, vol. 14, no. 1, pp. 1–20, 2024.

[345] J. Zhu, Y. Feng, Q. Liu, H. Xu, Y. Miao, Z. Lin, J. Li, H. Liu, Y. Xu, and F. Li, “An improved convnext with multimodal transformer for physiological signal classification,” *IEEE Access*, vol. 12, pp. 11217–11229, 2024.

[346] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “Blazepose: On-device real-time body pose tracking,” *arXiv preprint arXiv:2006.10204*, 2020.

[347] M. Villarroel, *Non-contact vital sign monitoring in the clinic*. PhD thesis, University of Oxford, 2017.

[348] M. S. Sotoodeh, O. Ossmy, G. Donati, J. Hall, H. Rowan, and G. S. Forrester, “Automatic pose estimation in newborn infants: Lessons from the baby grow study,” *Behavior Research Methods*, vol. 58, no. 3, p. 82, 2026.

[349] S. Ahani, N. Niknafs, P. M. Lavoie, L. Holsti, and G. A. Dumont, “Video-based respiratory rate estimation for infants in the nicu,” *IEEE Journal of Translational Engineering in Health and Medicine*, 2024.

[350] J. Jorge, *Non-contact monitoring of respiration in the neonatal intensive care unit*. PhD thesis, University of Oxford, 2018.

[351] W. Zong, T. Heldt, G. Moody, and R. Mark, "An open-source algorithm to detect onset of arterial blood pressure pulses," in *Computers in Cardiology, 2003*, pp. 259–262, IEEE, 2003.

[352] M. Alnaggar, A. I. Siam, M. Handosa, T. Medhat, and M. Rashad, "Video-based real-time monitoring for heart rate and respiration rate," *Expert Systems with Applications*, vol. 225, p. 120135, 2023.

[353] M. Gwak, K. Vatanparvar, J. Kuang, and A. Gao, "Motion-based respiratory rate estimation with motion artifact removal using video of face and upper body," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1961–1967, IEEE, 2022.

[354] D. G. Kyrollos, K. Greenwood, J. Harrold, and J. R. Green, "Detection of false alarms in the nicu using pressure sensitive mat," in *2021 IEEE sensors applications symposium (SAS)*, pp. 1–5, IEEE, 2021.

[355] N. Koolen, O. Decroupet, A. Dereymaeker, K. Jansen, J. Vervisch, V. Matic, B. Vanrumste, G. Naulaers, S. Van Huffel, and M. De Vos, "Automated respiration detection from neonatal video data," in *International Conference on Pattern Recognition Applications and Methods*, vol. 2, pp. 164–169, SciTePress, 2015.

[356] D. Welsch and M. Neuhäuser, "Wilcoxon-signed-rank test," in *International Encyclopedia of Statistical Science*, pp. 2898–2900, Springer, 2025.

[357] K. Lim, H. Jiang, A. P. Marshall, B. Salmon, T. J. Gale, and P. A. Dargaville, "Predicting apnoeic events in preterm infants," *Frontiers in Pediatrics*, vol. 8, p. 570, 2020.

[358] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, pmlr, 2015.

[359] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," *Multimedia tools and applications*, vol. 79, no. 19, pp. 12777–12815, 2020.

[360] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[361] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[362] N. Akhtar and U. Ragavendran, "Interpretation of intelligence in cnn-pooling processes: a methodological survey," *Neural computing and applications*, vol. 32, no. 3, pp. 879–898, 2020.

[363] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[364] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[365] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, "Backpropagation and the brain," *Nature Reviews Neuroscience*, vol. 21, no. 6, pp. 335–346, 2020.