Applications of Computer Vision for Remote Sensing of Infant and Adult Physiology

by

Fatema-Tuz-Zohra Khanam

B.Eng. in Electrical & Electronic Engineering

M.Sc. in Electrical & Computer Engineering

A thesis submitted for the degree of

Doctor of Philosophy

STEM



University of
South Australia

October 2023

Contents

List of Figures ..... **i**

List of Tables ..... **iv**

List of Abbreviations ..... **v**

Summary ..... **ix**

Declaration ..... **x**

Acknowledgments ..... **xi**

List of Publications ..... **xii**

CHAPTER 1: General Introduction and Motivation

1.1. Introduction and Motivation 3

1.2. Problem Statement 6

1.3. Thesis Objectives 7

1.4. Thesis Contributions 7

1.5. Thesis Structure 9

CHAPTER 2: Literature Review

2.1. Technologies for Non-contact Vital Sign Monitoring 11

2.1.1. Thermal Camera Imaging Technology 14

2.1.1.1. Body Temperature Measured with Thermal Camera 16

2.1.1.2. HR and BR Measured with Thermal Camera 18

2.1.2. Video Camera Imaging Technology 22

2.1.2.1. Motion-Based Method 22

2.1.2.2. Colour-Based Method 25

2.1.3. Combinations of Different Technologies 29

2.2. Different Aspects of Video Camera Imaging 31

2.2.1. Motion Artefacts 31

2.2.2. Illumination Variations 34

2.2.3. Both Illumination Variations and Motion Artefacts 35

2.2.4. Alternate Sensors 37

2.2.5. Different Subjects 38

2.2.6. Different Vital Signs 39

2.2.7. Multiple ROIs 39

2.2.8. Long Distance and Multiple Persons Simultaneously 40

2.2.9. Others 41
2.3. Applications of Video Camera Imaging Technology 42
2.3.1. Clinical Applications 43
2.3.2. Non-Clinical Applications 44
2.4. Technologies for Jaundice Detection 45
2.4.1. Invasive Approaches 45
2.4.2. Non-invasive Approches 46
2.4.2.1. Computer Vision-Based Approaches Using Sclera Colour 46
2.4.2.2. Computer Vision-Based Approaches Using Skin Colour 47
2.5. Research Gaps and Limitations 50
2.6. Concluding Remarks 51

CHAPTER 3: Research Methodology and Techniques

3.1. Introduction 53
3.2. Research Design 54
3.3. Data Collection 54
3.3.1. Research Ethics 54
3.3.2. Participants 55
3.3.3. Video Data Acquisition 55
3.3.4. Validation 56
3.4. Data Processing and Analysis 56
3.4.1. Image Processing Techniques 56
3.4.1.1. Image Colour Space 56
3.4.1.2. ROI Selection 59
3.4.2. Raw Signal Extraction 63
3.4.3. Signal Processing Techniques 63
3.4.3.1. Signal Decomposition 63
3.4.3.2. Spectral Analysis 65
3.4.3.3. Temporal Filtering 67
3.4.3.4. Peak Detection 67
3.4.4. Machine Learning Algorithms 68
3.4.4.1. Naive Bayes 68
3.4.4.2. K-Nearest Neighbours (KNN) 68
3.4.4.3. Support Vector Machine (SVM) 68

3.4.4.4. Decision Tree 68
3.4.4.5. Random Forest (RF) 69
3.4.4.6. Stochastic Gradient Boosting (SGB) 69
3.5. Software 69
3.6. Statistics 70
3.6.1. Pearson Correlation Coefficient 70
3.6.2. Bland-Altman Plot 70
3.6.3. Root Mean Square Error 71
3.6.4. Mean Absolute Error 72
3.6.5. Accuracy 72
3.6.6. Precision 72
3.6.7. Recall 73
3.6.8. F1 Score 73
3.7. Concluding Remarks 73

CHAPTER 4: Non-contact Neonatal Vital Signs Monitoring in NICU

4.1. Introduction and Related Work 75
4.2. Materials and Protocol 78
4.2.1. Study Design 78
4.2.1. Experimental Setup 78
4.3. Methods 79
4.3.1. System Framework Using Image Decomposition Method 79
4.3.1.1. Video Decomposition 80
4.3.1.2. ROI Selection 81
4.3.1.3. Spatial Averaging 82
4.3.1.4. Spectral Analysis and Band-pass Filtering 83
4.3.1.5. Peak detection 83
4.3.2. System Framework Using Skin Detection Method 84
4.3.2.1. Automatic ROI Selection 85
4.3.2.2. Spatial Averaging 86
4.3.2.3. Spectral Analysis and Band-pass Filtering 86
4.3.2.4. Peak detection 86
4.4. Experimental Results 87
4.4.1. Results Using Image Decomposition Method 87

4.4.2. Results Using Skin Detection Method 93
4.5. Concluding Remarks 97

CHAPTER 5: Non-contact Neonatal Vital Signs Monitoring Based on EEMD

5.1. Introduction and Related Work 99
5.2. Methods and Materials 101
5.2.1. Study Design and Experimental Set-up 101
5.2.2. System Framework Using Neural Network Method 101
5.2.2.1. Automatic ROI Selection 102
5.2.2.2. Spatial Averaging 104
5.2.2.3. Signal Decomposition 105
5.2.2.4. Spectral Analysis and Filtering 107
5.5.2.5. Peak detection 107
5.3. Experimental Results 107
5.4. Concluding Remarks 114

CHAPTER 6: Non-contact Respiration Rate Monitoring Integrating RGB-Thermal Cameras

6.1. Introduction and Related Work 116
6.2. Materials and Protocol 118
6.2.1. Ethics Protocol 118
6.2.1. Experimental Setup and Data Acquisition 118
6.3. Methods 119
6.3.1. RR Measurement System Using Thermal Sensor 119
6.3.1.1. RGB-Thermal Image Acquisition 120
6.3.1.2. RGB and Thermal Image Alignment 120
6.3.1.3. Automatic ROI Selection 121
6.3.1.4. Raw Signal Extraction 123
6.3.1.5. Spectral Analysis and Bandpass Filtering 123
6.3.1.6. Peak Detection 123
6.3.2. RR Measurement System Using RGB Sensor 124
6.3.2.1. ROI Selection 124
6.3.2.2. Raw Signal Extraction 125
6.3.2.3. Spectral Analysis and Band-pass Filtering 125

6.3.2.4. Peak Detection 126
6.4. Experimental Results 126
6.4.1. Results Using Thermal Data 126
6.4.2. Results Using RGB Data 128
6.4.3. Comparison between Thermal & RGB data 130
6.5. Discussion 130
6.6. Concluding Remarks 131

CHAPTER 7: Non-contact Jaundice Detection Based on Machine Learning
7.1. Introduction and Related Work 133
7.2. Methods and Materials 135
7.2.1. Data Collection 135
7.2.2. System Framework for Jaundice Detection 135
7.2.2.1. Automatic ROI Selection 136
7.2.2.2. Skin Colour Analysis 138
7.2.2.3. Feature Extraction 139
7.2.2.4. Jaundice Detection Based on Random Forest 139
7.3. Experimental Results 141
7.4. Discussion 148
7.5. Concluding Remarks 149

CHAPTER 8: Conclusions and Future Directions
8.1. Conclusions 151
8.2. Future Directions 152

References 155
Appendix A 172
Appendix B 175

List of Figures

2.1: Contactless measuring methods 12

2.2: Thermal image captured by thermal camera 15

2.3: The non-contact body temperature measurement framework proposed in [75]. 17

2.4: The thermal camera-based continuous body temperature measurement framework proposed in [78] 17

2.5: Block diagram for the BR estimation proposed in [62] 20

2.6: System diagram of the remote respiratory monitoring system [106] 24

2.7: Block diagram of contactless vital sign monitoring system using Microsoft Kinect sensor [108] 25

2.8: Skin reflection model of iPPG [adapted from [111]]. 27

2.9: System overview of the noise removal technique using UAV [142]. 37

2.10: Data collection using the UAV (3DR solo drone) [146] 38

2.11: The experimental setup and data acquisition of the noise artefact removal, multiple people, and long detection system [161] 41

3.1: Basic block diagram of the proposed non-contact monitoring system. 54

3.2: Visualization of the 68 facial landmark coordinates [adapted from [216]] 62

3.3: The graphical form of the Bland-Altman method 71

4.1: Experimental setup for data collection 79

4.2: System framework using image decomposition method 80

4.3: Video decomposition and ROI selection 81

4.4: System framework using skin detection technique. 84

4.5: Automatic ROI selection 85

4.6: Skin detection 86

4.7: Cardiac signal 88

4.8: Respiratory signal 89

4.9: Correlation plot for HR 91

4.10: Bland Altman plot for HR 91

4.11: Correlation plot for RR 92

4.12: Bland Altman plot for RR 92

4.13: Cardiac signal 93

4.14: Respiratory signal 94

4.15: Statistical measurement for HR. 96

4.16: Statistical measurement for RR 96

5.1: The experimental setup where the data recording was performed 101

5.2: The system framework using neural network method 102

5.3: YOLO network architecture (adapted from [214]) 103

5.4: Automatic ROI selection using the YOLO neural network 104

5.5: The spectral analysis and band-pass filtering process 107

5.6: An infant image with detected ROI is shown in (a). The corresponding ROI extracted from the original is shown in (b). 108

5.7: Raw cardiorespiratory signals for 300 frames 109

5.8: IMF components of the raw cardiorespiratory signals using the EEMD technique. 110

5.9: The frequency spectrum of decomposed IMF3, IMF4, IMF5 and IMF6 111

5.10: The filtered cardiorespiratory signals are shown in the figure 112

5.11: Statistical measurement for HR 113

5.12: Statistical measurement for RR 113

6.1: Experimental setup of the proposed system. 119

6.2: Overview of the proposed RR measurement using the thermal sensor 120

6.3: RGB and thermal image alignment 121

6.4: Overview of the proposed automatic ROI selection technique 121

6.5: Face detection and landmark selection 122

6.6: Automatic ROI selection 122

6.7: Overview of the proposed RR measurement using RGB sensor 124

6.8: ROI selection from RGB image 125

6.9: Respiratory signal obtained from thermal data 127
6.10: Statistical measurement between reference and thermal data 128
6.11: Respiratory signal obtained from RGB data 128
6.12: Statistical measurement between reference and RGB data 129
6.13: Statistical measurement between RGB and thermal data 130
7.1: Block diagram of the proposed jaundice detection system 135
7.2: Block diagram of automatic ROI selection 136
7.3: Representation of automatic ROI selection 137
7.4: Algorithm of ROI selection 138
7.5: Representation of RF algorithm 140
7.6: ROC curve for different machine learning algorithms 143
7.7: A 3D baby with light skin colour (jaundice-free). 144
7.8: A 3D baby with a yellowish-light skin colour and jaundice 144
7.9: A 3D baby with brown skin colour and jaundice-free. 144
7.10: A 3D baby that has yellowish-brown skin and jaundice 145
7.11: A normal baby without jaundice 146
7.12: A baby with a yellowish skin colour who has jaundice 146

List of Tables

2.1: Different studies extract body temperature using thermal cameras 18

2.2: Different studies to monitor vital signs using thermal cameras 21

2.3: IPPG -based methods under well controlled condition 28

2.4: Different studies to monitor vital signs using combined technology 30

2.5: IPPG-based methods concern about motion artefacts 33

2.6: IPPG-based methods concern about illumination variations 35

2.7: IPPG-based methods concern about both illumination variations and motion artefacts 37

2.8: Summary of different iPPG-based methods 41

5.1: Comparison with state-of-the-art methods for HR 114

6.1: Summary of the results of the proposed method 131

6.2: Comparison with state-of-the-art methods. 131

7.1: Comparison of the performance of different ML algorithms 142

7.2: Summary of 20 infants considered in the first dataset 145

7.3: Summary of 30 infants considered in the second dataset 147

7.4: Comparison with state-of-the-art methods 148

List of Abbreviations

ABF	Adaptive Bandpass Filter
ANFIS	Adaptive Neuro-Fuzzy Inference System
AR	Auto regressive
AUC	Area Under the Curve
B-A	Bland-Altman
BCG	Ballistocardiogram
BP	Blood Pressure
BT	Body Temperature
BSS	Blind Source Separation
CBTM	Continuous Body Temperature Measurement
CCA	Canonical Correlation Analysis
CCD	Charge-Coupled Device
CEEMD	Complete Ensemble Empirical Mode Decomposition
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CHROM	Chrominance
CMOS	Complementary Metal-oxide-semiconductor
CNN	Convolutional Neural Networks
CRF	Conditional Regression Forests
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
DEAP	Database for Emotion Analysis using Physiological Signals
DFT	Discrete Fourier Transform
DSP	Digital Signal Processing
DSLR	Digital Single-lens Reflex
DTCWT	Dual Tree Complex Wavelet Transform
ECG	Electrocardiogram
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
EMMS	Efficient Motion Magnification System
EVM	Eulerian Video Magnification
FFT	Fast Fourier Transform

FIR	Far-Infrared
FP	False Positive
FN	False Negative
GFT	Good Feature Tracking
GRD	Green/Red Difference
GRT	Green, Red, and Thermal
GUI	Graphical User Interface
HCI	Human Computer Interaction
HR	Heart Rate
HREC	Human Research Ethics Committee
HRV	Heart Rate Variability
HSV	Hue, Saturation, Value
ICA	Independent Component Analysis
IoU	Intersection over Union
IMF	Intrinsic Mode Function
iPPG	Imaging Photoplethysmography
IR	Infrared
IRT	Infrared Thermography
ITI	Infrared Thermal Imaging
JBSS	Joint Blind Source Separation
JFTD	Joint-Time-Frequency Diagram
KCF	Kernel Correlation Filter
KLT	Kanade Lucas Tomasi
KNN	K Nearest Neighbour
LARS	Least Angle Regression
LDV	Laser Doppler Velocimetry
LED	Light-emitting Diode
LoA	Limits of Agreement
LWIR	Long Wavelength Infrared
MAE	Mean Absolute Error
MEMD	Multivariate Empirical Mode Decomposition
MLR	Multiple Linear Regression
MRSPT	Motion Resistant Spectral Peak Tracking

MSE	Mean Squared Error
MSSIM	Multi-scale Structural Similarity
NICU	Neonatal Intensive Care Unit
NIR	Near-Infrared
NMS	Non-Maximum Suppression
NPD	Normalised Pixel Difference
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PLS	Partial Least Squares
POS	Plane Orthogonal to the Skin
PPG	Photoplethysmography
PSD	Power Spectral Density
PSNR	Peak Signal to Noise Ratio
RF	Random Forest
RGB	Red, Green, Blue
RGBMT	Red, Green, Blue, Magenta, and Thermal
RIP	Respiratory Inductive Plethysmography
RMSE	Root Mean Square Error
ROI	Region of Interest
ROM	Region of Measurement
ROC	Receiver Operating Characteristic
RR	Respiratory Rate
TE	Time of the Expiration
TI	Time of the Inspiration
TN	True Negative
TP	True Positive
SB	Sub Band
SCICA	Single-Channel Independent Component Analysis
SD	Standard Deviation
SDM	Supervised Descent Method
SGB	Stochastic Gradient Boosting
SNR	Signal to Noise Ratio
SV	Speckle Vibrometry

SpO_2	Blood Oxygen Saturation
SQI	Signal Quality Index
SSA	Singular Spectrum Analysis
STC	Spatio-Temporal Context
STFT	Short-Time Fourier Transform
SURF	Speeded Up Robust Feature
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
TcB	Transcutaneous Bilirubinometry
TFR	Time-Frequency Representation
TSB	Total Serum Bilirubin
UAV	Unmanned Aerial Vehicle
V–J	Viola–Jones
WHO	World Health Organization
YOLO	You Only Look Once

Summary

The interest in using non-contact technologies to monitor vital signs remotely is rapidly increasing for both clinical and non-clinical purposes. Vital signs are important parameters that indicate the condition of human health. Conventional monitoring systems necessitate direct physical contact as they rely on contact-based sensors, straps, probes, or electrodes, which may cause discomfort, stress, and pain. These systems may also result in skin infections and damage, particularly when applied to premature babies or individuals with sensitive skin. Therefore, there is a need to monitor vital signs in a non-contact manner when contact with skin is either undesirable or not feasible. On the other hand, jaundice is a common phenomenon in neonates and a significant cause of morbidity and mortality. Early detection of jaundice is still a significant challenge. Moreover, existing invasive techniques require a blood sample taken with a syringe and needle, causing stress and pain to newborns. Hence, a non-contact and non-invasive system is required to detect neonatal jaundice without any pain. Therefore, this project aims to investigate non-contact systems using camera imaging techniques to monitor heart rate and respiratory rate, as well as to detect jaundice in a non-contact manner. The research employs various signal and image processing techniques to measure cardiorespiratory signals and identify jaundice. The main contributions of this thesis involve overcoming challenges associated with developing robust non-contact image-based systems. These challenges include ensuring subject safety, managing system costs, selecting appropriate regions of interest automatically, addressing noise artefacts, and handling the complexity of dealing with different subjects, such as premature babies. Moreover, the integration of multi-camera fusion is also addressed. The experimental results demonstrate promising performance, with strong correlation, close agreement, and low error rates when compared to reference measurements from clinical instruments. Consequently, this thesis introduces a new perspective on non-contact systems for monitoring vital signs and detecting jaundice, showing significant potential for both clinical and non-clinical applications.

Declaration

This thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, are clearly acknowledged.



……

Fatema-Tuz-Zohra Khanam 03 October 2023

Acknowledgments

First and foremost, I would like to express my deep appreciation, sincere thanks, and gratitude to my principal supervisor, Professor Javaan Chahl, for his invaluable advice, continuous guidance, assistance, support, and encouragement throughout the progress of the research study. I would like to use this opportunity to thank you, not only for guiding me, but also for providing me with many opportunities to be involved in a broad range of exciting side projects.

My appreciation, thanks, and gratitude also extend to my End-user advisor, Dr. Ali Al-Naji, for his invaluable advice, continuous guidance, assistance, support, and encouragement. His advice has consistently been highly beneficial to me during my Ph.D. study.

I would also like to extend thanks to another End-user advisor, Mrs. Kim Gibson, who provided clinical support and assisted me in data collection. Moreover, I would like to thank Dr. Jasmine Huynh for her assistance in performing the ethical procedures.

I would also like to thank my colleagues in the Aerospace Lab, especially Dr. Asanka G. Perera, Dr. Titilayo Ogunwa and Timothy D. McIntyre, for their help, support, and encouragement throughout my Ph.D. study.

My gratitude also goes to the staff of STEM for providing the necessary support and facilities to advance with this research.

I would like to thank Australian Commonwealth Department of Education and Training for giving me the opportunity to do my Ph.D. and continue my research by providing a Research Training Program domestic (RTPd) Scholarship.

I must offer my heartfelt thanks to my family: my parents, especially my mother (Rokeya Akhter Khanum), my beloved husband (Dr. Kamrul Hassan), my brothers, my son (Faizan Hassan) for always believing in me and for their blessings, unconditional love, and support without which the completion of this work would have been a distant dream.

Lastly, all praise to the most beneficent **ALLAH (SWT)** for the mercy and kindness in keeping me happy and healthy throughout this long journey.

List of Publications

[I] F.-T.-Z. Khanam, A. Al-Naji & J. Chahl, "Remote Monitoring of Vital Signs in Diverse Non-Clinical and Clinical Scenarios Using Computer Vision Systems: A Review," *Applied Sciences*, vol. 9, no. 20, p. 4474, 2019.

[II] F.-T.-Z. Khanam, LA Chahl, JS Chahl, A. Al-Naji, AG Perera, D. Wang, YH Lee, … & J. Chahl, "Noncontact sensing of contagion," *Journal of Imaging*, vol. 7, no. 2, p. 28, 2021.

[III] F.-T.-Z. Khanam, A. Al-Naji, A. G. Perera, K. Gibson & J. Chahl, "Remote Vital Signs Monitoring in Neonatal Intensive Care Unit Using a Digital Camera," *International Journal of Biomedical and Biological Engineering*, vol. 16, no. 10, pp. 138–144, 2022.

[IV] F.-T.-Z. Khanam, A. Al-Naji, A. G. Perera, K. Gibson & J. Chahl, "Non-contact automatic vital signs monitoring of neonates in NICU using video camera," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 2, pp. 278–285, 2023.

[V] F.-T.-Z. Khanam, A. G. Perera, A. Al-Naji, K. Gibson & 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.

[VI] F.-T.-Z. Khanam, A. G. Perera, A. Al-Naji, K. Gibson & J. Chahl, "Contactless Vital Signs Monitoring in NICU based on Machine Vision," *presented at the 19th International Conference on Intelligent Unmanned Systems (ICIUS), Adelaide, Australia*, 2023.

[VII] F.-T.-Z. Khanam, A. G. Perera, A. Al-Naji, T. D. Mcintyre & J. Chahl, "Integrating RGB-Thermal Image Sensors for Non-contact Automatic Respiration Rate Monitoring," *Biomedical Signal Processing and Control*, 2023. (Submitted)

[VIII] F.-T.-Z. Khanam, A. G. Perera, A. Al-Naji, D. Wang & J. Chahl, "Non-invasive and Non-contact Automatic Jaundice Detection of Infants Based on Random Forest," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp.1-14, 2023.

CHAPTER 1

General Introduction and Motivation

1 General Introduction and Motivation

The vital signs are important indicators of human health condition and in some circumstances need continuous monitoring. The currently used conventional monitoring systems require direct physical contact as they use contact-based sensors, straps, probes, or electrodes, which may cause discomfort, stress, and pain. Therefore, there is a need to monitor vital signs in a non-contact manner when contact with skin is either undesirable or not possible. With the rapid growth of non-contact monitoring technology, it has become feasible to monitor vital signs unobtrusively.

Jaundice is a common phenomenon in neonates and a significant cause of morbidity and mortality. Detecting jaundice early remains a considerable challenge. Moreover, existing invasive techniques require a blood sample taken via a syringe and needle causing stress and pain to newborns. Therefore, a non-contact and non-invasive system is required to detect neonatal jaundice without pain.

Several non-contact approaches are available for remotely monitoring vital signs and detecting jaundice with sufficient accuracy and reliability. However, each technology has its pros and cons, depending on factors like subject safety, system cost, region of interest (ROI) selection, noise-artefacts, multi-camera fusion and the complexity of dealing with vulnerable subjects, such as premature babies.

To ensure the feasibility of non-contact systems in real-world scenarios, this thesis addresses the following key questions: (i) How can an efficient system be developed using camera imaging-based technologies to observe physiological and physical variations? (ii) How can an efficient system be developed using camera imaging-based technologies to detect jaundice? (iii) How can the proposed systems be applied under the specified assumptions to handle various potential applications? To address these questions, the research presented in this thesis introduces novel methods and techniques and conducts experiments to validate their effectiveness.

Section 1.1 of this chapter provides a general introduction and motivation for the research described in this thesis. Section 1.2 includes the problem statement of the proposed imaging systems, followed by the related main objectives in section 1.3. Section 1.4 outlines the associated research contributions, while Section 1.5 presents the thesis's overall structure.

1.1. Introduction and Motivation

Monitoring of human vital signs, such as respiratory rate (RR), blood oxygen saturation (SpO_2), heart rate (HR), heart rate variability (HRV) and blood pressure (BP) plays a significant role in modern clinical care of patients in hospital and at home [1]. It has great importance in both clinical and non-clinical applications such as medical diagnosis including detection of abnormal cardiorespiratory events such as bradycardia, tachycardia, tachypnoea, bradypnoea, and apnoea [2-4], training programs, fitness assessment, lie detection and stress measurement [5].

There are various instruments for measuring these vital signs, such as Electrocardiogram (ECG), pulse oximeter, nasal thermocouples, respiratory belt transducer and piezoelectric transducer [6]. These instruments require direct physical contact with the human body as they use contact-based sensors, straps, probes or electrodes [7]. These instruments may cause stress, discomfort, pain or, in the worst case, damage the skin especially in case of premature babies, aged people or burns victims who have sensitive skin [8]. Sardesai et al. [9] documented the harmful effects of adhesives on the skin, highlighting that a single removal of an adhesive can strip away 70 to 90% of the epidermis, and repeated removal or application of adhesives can cause even deeper damage. For instance, a study conducted on 37 infants found that approximately 26% of them experienced skin lesions due to various medical procedures, including the use of life-support equipment in the neonatal intensive care unit (NICU). These procedures encompassed activities like body hygiene, dressing changes, repositioning, and the application and removal of adhesive devices, sensors, electrodes, and probes [10]. Another study indicated that 80% of premature infants faced some form of skin disorder or lesion within the first month of receiving hospital care [11]. Moreover, there may be a risk of entanglement or strangulation of infants who are attached to monitor by means of wires and leads [12]. These instruments are also not appropriate for long term monitoring as they may cause discomfort, irritation and a cumulative risk of fungal and bacterial infection to the subjects [13]. About 18.4% of infection occurred in the NICU as a result of using adhesive sensors, leads and electrodes directly on the skin [11].

Cost is an important issue as the monitoring electrodes and leads are only intended and certified for a single use, followed by disposal [14]. As an illustration, the approximate daily expenses for neonatal monitoring in the NICU are around $3,500 per infant in the

United States [15], £800 in the United Kingdom [16] and ¥10,413.7 in China [17]. Placement of the sensors with self-adhesive pads leads to difficulties with wet, oily, dirty or hairy subjects which is a limitation of these technologies in emergency situations [18]. Accuracy is another issue with conventional contact methods since they are sensitive to artefacts produced by the subject's movement or from cardiac activity [19]. In addition, reduced amplitude of chest wall expansion can affect the respiratory rate input signal from the impedance lead [20]. Furthermore, it is also difficult to make many people wear such sensors and collect data from them when they are in public places such as hospitals, schools, or airports.

Numerous wearable devices such as smart watches [21, 22], smart patches [23], smart bracelets [24, 25] etc. offer many advantages for vital sign monitoring for a high number of applications [26, 27]. These devices are smaller and more user friendly than current hospital monitoring devices [24]. Despite their wearable nature, these contact-based modalities still introduce inconvenience to patients' movement and may cause discomfort skin irritation, or a dislike for wearing accessories. Moreover, users and healthcare professionals should be aware of considerations related to accuracy, battery life, user compliance, data privacy, device cost and accessibility, and the need for clinical validation in specific medical contexts [21]. Furthermore, these devices are not feasible for neonates as they mainly designed for adults.

Therefore, to minimise these limitations there is value in an alternative method where vital signs can be measured without any physical contact.

There are various non-contact methods for measuring vital signs remotely, including magnetic induction, the Doppler effect, video camera imaging, and thermal camera imaging [28]. These techniques have been shown to be effective in remotely monitoring vital signs with acceptable accuracy, reliability, and sometimes practicality [28]. However, each method mentioned above has its pros and cons and may outperform in some challenges but be inferior in others. Doppler radar is affected by motion artefacts, can have biological effects on humans, and is not suitable for monitoring multiple subjects simultaneously. Visible spectrum and thermal spectrum cameras offer the best options of those studied, thermal cameras due to their potential to measure all likely symptoms on a single camera, especially temperature, and video cameras due to their availability, cost, adaptability, and compatibility.

Another significant issue addressed in this thesis is related to jaundice detection using video camera imaging-based methods. Neonatal jaundice is a common disease found in newborn infants in their first few weeks of life [29, 30]. According to worldwide statistics, 80% of preterm infants and 60% of term newborns, and 10% of breastfed newborns were prone to neonatal jaundice [31]. Every year a total of 24 million newborns become jaundiced [32]. Jaundice or hyperbilirubinemia is observed as the yellow discolouration of the skin and sclera of the eyes due to an excess level of bilirubin [33]. Short-term excess of bilirubin is mostly harmless and self-limiting. But a high level of bilirubin in newborn infants is neurotoxic and can permanently damage the brain, which is called kernicterus. It may cause cerebral palsy, deafness or hearing loss, language difficulties, and developmental delay or can be fatal in the worst cases [34, 35]. Kernicterus can be avoided by early detection and treatment of jaundice or hyperbilirubinemia.

For detecting jaundice, there are different invasive and non-invasive approaches [33]. Total Serum Bilirubin (TSB) measurement is a common invasive technique to determine the level of hyperbilirubinemia that works based on spectrophotometry using a small blood sample taken by a syringe and needle [36]. However, the TSB measurement can be performed only by medical experts with good facilities. Moreover, as it is an invasive technique, this procedure is stressful for both infants and parents [37]. For the diagnosis of jaundice, there are some other techniques that involve the collection of blood samples. Nevertheless, all invasive procedures are uncomfortable and painful for the infant, have a risk of infections, may cause anaemia, and will create anxiety for parents [38]. Visual assessment [39] and Transcutaneous Bilirubinometry (TcB) [40] are non-invasive alternative techniques to TSB measurement. However, they are expensive and not reliable for detecting jaundice [41, 42]. Computer vision can be a potential alternative to detect jaundice in non-contact and noninvasive manner, more immediately and cost-effectively.

For non-contact vital sign monitoring and jaundice detection, camera imaging-based methods emerge as the most promising methods, therefore, in this thesis these methods have been used as the basis for extracting cardiorespiratory signals and identifying jaundice. The main goal of this thesis is to develop a safe, cost-effective, and reliable non-contact system for vital sign monitoring and jaundice detection, based on computer vision technologies. This aims to address the challenges associated with existing non-contact monitoring systems which include the subject's safety, the system cost, ROI selection,

noise-artefacts, multi-camera fusion, and the complexity of dealing vulnerable subjects such as premature babies.

1.2. Problem Statement

In recent years, significant progress has been made in camera imaging-based technologies, and numerous studies have been published in this domain. However, these existing works have various limitations that need to be overcome.

The problems are stated as follows:

- The reliability of vital sign extraction and jaundice detection depends on selecting an appropriate ROI from the subject. The ROI selection is an important issue to overcome and very few researchers have paid enough attention to this issue. Most of the existing camera-based technologies used manual processes to select the ROI. Therefore, advanced image processing techniques are required to automatically select appropriate ROIs.

- Non-contact monitoring is greatly affected by noise in video and signals can become undetectable or heavily distorted from noise resulting from skin tone and illumination variation. It is necessary to develop a noise removal technique using advanced signal processing to reduce noise artefacts.

- Researchers have mainly focused on healthy and young participants as their subjects for experiments and do not access elderly people and premature babies. More research needs to be conducted, especially considering different subjects such as premature babies, who are adversely affected by existing contact-based methods.

- Most existing non-contact systems use either thermal camera or video camera to monitor vital signs. However, video cameras do not function in complete darkness and are affected by illumination variation, whereas thermal cameras may work in complete darkness and are not affected by illumination variation. While there are different face detection algorithms in the visible band, it is challenging to detect and track faces and facial tissue in thermal bands owing to a lack of textural and geometric details. Therefore, to enhance the reliability of the non-contact system, the use of multiple spectra is required.

The research presented in this thesis addresses the first challenge by developing different automatic ROI selection algorithms for the automatic selection of appropriate ROIs. Additionally, this thesis addresses the second challenge by introducing a noise removal technique to mitigate illumination variation. We conducted experiments involving both healthy adults and premature babies in the NICU, effectively addressing the third issue. Furthermore, the fourth challenge was successfully met by integrating both thermal and video camera-based imaging techniques.

1.3. Thesis Objectives

The primary objective of this PhD thesis is to design and develop efficient non-contact camera-imaging systems to extract cardiorespiratory signals and to detect jaundice, while being safe, cost-effective, and reliable.

The aims and objectives are outlined as follows:

- To design various automatic ROI selection techniques using advanced image processing methods, enabling the automatic selection of appropriate ROIs and enhancing the reliability of the non-contact systems.
- To develop a noise removal technique using advanced signal processing methods to reduce the noise artefacts caused by illumination variation.
- To monitor vital signs of diverse subjects, including healthy individuals and infants in NICU, thereby contributing to a comprehensive understanding of the system's applicability.
- To synergistically integrate both thermal and video camera-based imaging techniques, thereby enhancing the reliability of the monitoring system.
- To detect jaundice remotely, employing a non-invasive and non-contact approach through machine learning methodologies.

1.4. Thesis Contributions

This section provides a concise overview of the original contributions made in this thesis. These contributions have been published through six peer-reviewed journal papers, one conference presentation, and another currently under review. The detailed contributions of this thesis can be summarised as follows:

- Providing a thorough review of the non-contact technologies (presented in Chapter 2) that discusses the advantages and disadvantages of each technology based on several challenges. This review was designed as a resource for researchers to assess existing gaps and challenges, thereby paving the way for further research opportunities. This contribution is based on material published in [I, II].

- Development of a new and cost-effective system (presented in Chapter 4) to remotely extract cardiorespiratory signals from infants in NICU without the need for physical contact. Moreover, ROI was selected efficiently and automatically using image decomposition and skin detection method. This contribution is based on material and results published in [III, IV].

- Proposal of an efficient non-contact imaging system (presented in Chapter 5) that measures cardiorespiratory signals from infants in NICU. The system incorporates a noise removal technique based on signal decomposition to enhance signal accuracy and an automatic ROI selection technique using neural network. This contribution is based on material and results published in [V, VI].

- Introduction of a reliable non-contact system (presented in Chapter 6) to extract respiratory rate of healthy adults by integrating RGB-thermal camera imaging to increase the reliability of the system. Moreover, ROI was selected automatically using landmark detection technique. This contribution is based on material and results submitted in [VII].

- Development of a non-invasive and non-contact jaundice detection system (presented in Chapter 7) to remotely detect neonatal jaundice using machine learning and a graphical user interface (GUI). Moreover, an advanced image processing technique was used for automatic selection of appropriate ROI. This contribution is based on material and results published in [VIII].

These contributions collectively signify a significant advance in the field of non-contact vital sign monitoring and disease detection. The integration of various innovative techniques and methodologies presented in this thesis contributes to the potential enhancement of healthcare practices and technology.

1.5. Thesis Structure

The remainder of this thesis is structured as follows:

Chapter 2 provides the foundational concepts relevant to this thesis. It offers a comprehensive review of the existing literature concerning non-contact technologies, elucidating the associated challenges and complexities.

Chapter 3 details the research methodology and provides the research design and data collection of the proposed non-contact imaging systems. It also describes the techniques, software and statistics adopted to conduct the research in the thesis.

Chapter 4 presents a new and cost-effective system to remotely extract cardiorespiratory signals from infants in NICU using video camera imaging.

Chapter 5 proposes an efficient non-contact imaging system that measures cardiorespiratory signals from infants in NICU with a noise removal method based on signal decomposition.

Chapter 6 provides a reliable non-contact monitoring system to extract respiratory rate of healthy adults through the fusion of video and thermal camera imaging, thus enhancing the reliability of the system.

Chapter 7 proposes a non-invasive and non-contact jaundice detection system to remotely detect neonatal jaundice using a user-friendly graphical interface leveraging skin colour analysis and machine learning techniques.

Finally, **Chapter 8** presents the conclusions and potential future work in this area of research.

CHAPTER 2

Literature Review

2 Literature Review

Non-contact technologies for monitoring vital signs and detecting jaundice have been attracting increasing attention in the literature. The aim of this chapter is to provide a thorough review of the non-contact technologies, discussing the pros and cons of each technology based on different influencing challenges. This provides an opportunity to assess the gaps and challenges of these technologies, enabling informed pursuit of research opportunities. The content presented in this chapter is based on material published in [I, II].

This chapter comprises six main sections. Section 2.1. presents different technologies used for measuring vital signs remotely. Sections 2.2 and 2.3 different aspects of video camera imaging and application of it in both clinical and non-clinical sector, respectively. Section 2.4. presents different technologies used for detecting jaundice remotely. Section 2.5 discusses challenges of existing studies that need to be overcome in the future, while Section 2.6 presents the concluding remarks.

2.1. Technologies for Non-Contact Vital Sign Monitoring

As presented in Figure 2.1, there are several noncontact means based on magnetic induction, Doppler effect, thermal camera imaging and video camera imaging which can be an effective alternative means of monitoring vital signs with acceptable reliability and accuracy [28]. These methods depend on the observation of physical and physiological variations including skin colour, temperature, impedance changes, head motion, arterial pulse motion, and importantly, thoracic, and abdominal motion due to the cardiorespiratory activity of both the respiratory and cardiovascular systems.

Magnetic induction-based methods can detect the impedance changes caused by blood and air volume variations due to mechanical action of the heart, diaphragm, and thorax. The basic principle is to induce eddy currents into the tissue and to calculate the re-inducted magnetic field externally, the impedance changes can be observed remotely to extract vital signs [43]. The method uses a simple arrangement based on multiple coils [44] or a single coil [45] integrated into a mattress [46], bed [47] or seat [48]. However, the method is highly susceptible to relative movements between coil and body. Moreover, electrical sensors suffer from electromagnetic interference, short life, and easy corrosion which is

not suitable for long-term medical monitoring. Furthermore, implementing a magnetic coil-based monitoring system may involve higher costs and technical complexities compared to some other sensing technologies. Besides, there may be safety concerns associated with the use of magnetic fields, especially in individuals with certain medical conditions or those using medical devices that could be affected by magnetic interference.

A smart mattress [49, 50] for vital sign monitoring is a modern and innovative solution that leverages technology to track various physiological parameters while individuals sleep. These mattresses typically incorporate sensors and other technologies such as pressure sensors, accelerometers and gyroscopes, temperature sensors, optical fibre sensors, and ballistocardiogram (BCG) sensors to monitor vital signs without the need for the user to wear any additional devices. However, it is important to note that users must remain lying down on the mattress for continuous monitoring, limiting the feasibility to tracking sleep patterns exclusively. Concerns regarding accuracy, privacy, cost, and comfort are significant factors that need careful consideration when opting for a smart mattress.

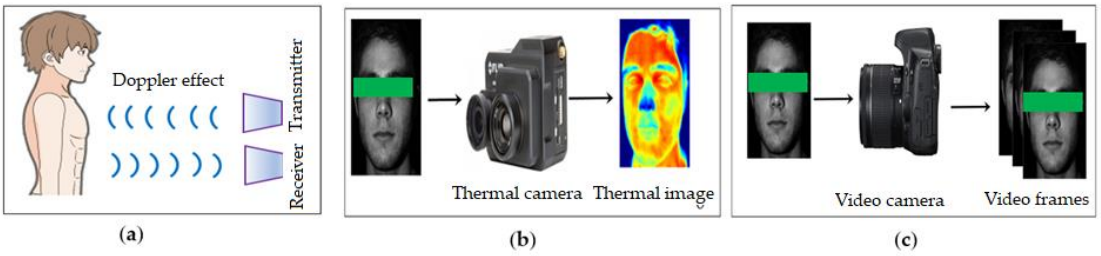


Figure 2.1: Contactless measuring methods. (a) Doppler effect, (b) Thermal camera imaging, (c)Video camera imaging.

The Doppler effect is an active noncontact method that can detect subtle chest movements due to cardiorespiratory activity. In this method, vital signs are extracted using Doppler radar [51, 52], or laser sensors [53] as well as digital signal processing (DSP) techniques [54] where phase shift between the transmitted waves and the reflected received waves from a ROI are calculated. There are three types of Doppler-based methods - Doppler with electromagnetic wave [55, 56], laser [57, 58], and ultrasonic wave [59-61]. However, Doppler approaches are susceptible to noise and motion artefacts and constrain movement of the subjects, preventing saturation sampling of the environment. Moreover, these methods need exposed ROI and specialized hardware making them costly [7]. They are also constrained to short-term monitoring and monitoring a single subject at a time.

Additionally, Doppler-based method may have biological effects on human [62] with unknown future population risks if broadly adopted.

Speckle vibrometry (SV) is an active noncontact method that can measure subtle movements of a given surface (e.g., the thoracic wall) due to cardiorespiratory activity. In this method, vital signs [63-65] are extracted by exploiting the laser speckle phenomenon, where a granular pattern will be formed and observed in a defocused camera when a laser beam is illuminated on the surface that is optically rough on the scale of the laser wavelength. This pattern is very sensitive to minute motions and thus will change according to each movement [66]. However, SV is susceptible to motion artifacts caused by movement of the subject. Speckle patterns can be sensitive to environmental conditions such as ambient light and surface reflections. SV typically provides measurements at a single point on the surface. The use of lasers in SV raises safety concerns, especially if the technique is applied for continuous or long-term monitoring. Setting up an SV system can be more complex [7].

Thermal camera imaging [67, 68] is a passive noncontact method that can detect the radiation emitted from particular parts of the human body in the infrared (IR) range of the electromagnetic spectrum to measure the physiological signal using a thermal camera [69-71].Thermal imaging based methods extract vital signs by measuring temperature changes around the nostril area [72-75] as well as heat differences due to pulsating blood flow in the main superficial arteries at different regions such as the carotid artery in the neck [76, 77] and temporal artery in the forehead [78, 79]. Thermal camera imaging techniques have advantages over other non-contact techniques as they are not affected by illumination variation and may work in complete darkness. However, thermal imaging-based approaches are susceptible to temperature variations and motion artefacts and constrain movement of the subjects. Their relatively low resolution, limits detection range and specificity to a particular subject. Besides, the thermal image is constrained in spatial resolution, and thermal cameras with even modest resolution are costly [7]. Nevertheless, thermal images give less geometric and textural information, which makes face and facial tissue detection algorithms ineffective in thermal imaging. They are also constrained to short-term monitoring and monitoring a single subject at a time.

Digital cameras offer high resolution, both spatial (number of pixels per degree), temporal (number of frames per second), intensity (number of bits per pixel) and in spectrum (at

least 3 visible channels, with hyperspectral options increasingly common), all due to consumer market demand. Furthermore, a large base of research assets exist for processing imagery, much of it free for use, for example OpenCV [80]. Flexibility with visible light optical design, offering panoramic, microscopic, and telescopic solutions in well integrated commercial product families, allows diverse measurement scenarios. Tailored fields of view allow analysis of multiple ROIs in parallel, or in series based on availability. The mass market has led to low cost [7, 28] and affordable optics that can be used in almost any conceivable application scenario [14].

Video camera imaging is a passive contactless method where video cameras are used to extract different physiological signals from several regions of the human body exploiting two principles. The first principle relies on skin colour variations due to cardiorespiratory activity, photoplethysmography (PPG). The vital signs are measured by exploiting variations in the reflectance properties of human skin from video, which causes variation in brightness values in sequences of images. The second principle depends on cyclic body motion owing to cardiorespiratory activity in techniques that can be broadly characterised as motion-based methods. The motion in the regions of the head, arterial pulse, and thoracic and abdominal region are included in this method. For noncontact physiological assessments, video camera imaging based methods seem to be a promising approach since they are robust, reliable, safe, cost-effective, suitable for long distance and long-term monitoring as well as multiple person detection simultaneously [28]. However, video camera imaging is affected by noise artefacts caused by motion artefacts and illumination variation.

Visible and thermal camera imaging are the best options among non-contact techniques since thermal cameras are capable of measuring many possible symptoms on a single camera, especially temperature, while video cameras are versatile owing to their cost, availability, compatibility, and adaptability [81].

2.1.1. Thermal Imaging Technology

Infrared thermography (IRT), also known as thermal imaging, is a promising monitoring and diagnostic technique in the medical field [82]. For example, thermal imaging can be used to assess HR and BR, endocrinological disorders (especially diabetes mellitus), vascular disorders, neurological disorders, musculoskeletal disorders, oncology, regenerative medicine, and surgery. Additionally, a popular application of thermal imaging

is screening persons with fever at airports, schools, hospitals, etc. Therefore, IRT is a passive, non-contact monitoring technique, which senses the radiation naturally emitted from an object such as human skin. Thermal imaging does not require an illuminating radiation source or a dedicated light source. These are the most significant advantages of thermal imaging over other non-contact techniques. They are not affected by illumination variation, can work in darkness, and are difficult to falsify using makeup or masks [83].

Thermal cameras respond to wavelengths starting at around 3 micrometers (µm) compared to visible wavelengths, ending at about 0.75 µm. An array of thermally sensitive elements behind a specified lens produces an image much like an optical camera, with higher radiation measured from hotter objects. This leads to a grey image, with lighter areas being warmer rather than darker. Depending on the application, this scale might be reversed for human operators. The use of "false colour" is common in thermal images and thermal video because it allows for increased ability for human operators to discern temperature. Colder temperatures are often assigned a shade of purple, blue, or green, whereas hotter temperatures can be given a shade of orange, red, or yellow [84]. Figure 2.2 shows a thermal image captured by a thermal camera where the person is covered in shades of orange and yellow, whereas other areas are blue and purple. That is because he is radiating more heat than surrounding objects. False colour is purely a remapping of grey levels and is not a feature of the underlying thermal images.

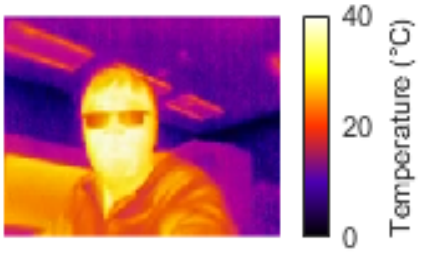


Figure 2.2: Thermal image captured by thermal camera.

Thermal cameras work on the principle that all objects that have a temperature above absolute zero (−273.15 °C) emit electromagnetic energy, also known as infrared radiation or thermal radiation [85].

To accurately observe the temperature of a target using a thermal camera, the target should be an ideal black body. A black body in thermal equilibrium will absorb all incident electromagnetic radiation and emit an equal amount of energy through isotropic radiation.

The distribution in wavelength of the emitted radiation is governed by the temperature of the body; thus, as the temperature increases, the distribution shifts towards shorter wavelengths. This phenomenon is described by Planck's law. A thermal camera measures the predominant wavelength of the radiation emitted by a body. The human body is not an ideal Plank black body radiator but is a close approximation in some bands [85]. It is generally accepted that the emissivity of human skin in the IR range of 3–14 μm is 0.98 ± 0.01 [85], which, while close to that of a black body, can cause disturbances due to non-ideal radiation emission and reflections. Thus, there will always be some degree of ambiguity when measuring skin temperature through thermal imaging. In a controlled environment, this can be accounted for with calibration. In uncontrolled environments, however, it is significantly more difficult to calibrate.

This is just one of many limitations that arise when using thermal cameras for the detection of elevated body temperature. Infrared camera manufacturers, such as FLIR [86], state that such technology can only be used to detect skin temperature (as opposed to core body temperature) and must be operated in a controlled environment to do so. The temperature must be in the range of 20 to 24 °C, and the relative humidity should be in the range between 10% and 50%. The temperature measured on the skin's surface has been used as an indicator of core body temperature. Environmental, operational, and subjective factors such as convective airflow, reflective surfaces, IR contamination (from sources such as sunlight and heaters for example), the sweat produced by the subject, wearable accessories covering some part of the face (such as glasses or baseball caps), ambient temperature, humidity, and emissivity all have an effect on the acquired data. Thus, thermal imaging may be used to determine anomalies relative to other subjects; however, there are many factors that prevent determination of the absolute core body temperature of a given subject from range. It is suggested that the area medially adjacent to the inner canthus of the eye provides the most consistent measurement, as determined with mass screening [87].

2.1.1.1 Body Temperature Measured with Thermal Camera

In the literature, thermal cameras have been used by numerous researchers to extract body temperature, as shown in Table 2.1. For example, Bilodeau et al. [88] proposed a non-contact method to measure the body temperature of a moving animal or human in a laboratory setting based on IRT using a thermal camera. As shown in Figure 2.3, the ROI was selected manually and tracked automatically with a particle filter. A Raw temperature signal was measured from extracted skin pixels. To minimize camera measurement noise,

a Kalman filter was applied to the raw signal. However, in the proposed method, the ROI was selected manually.

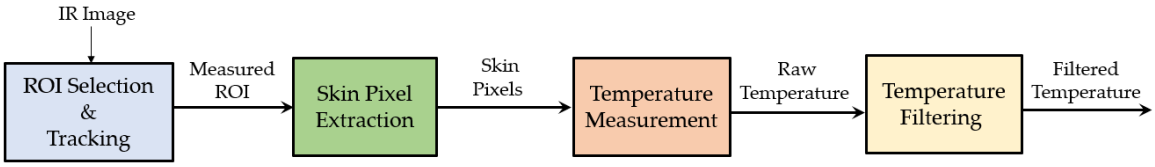


Figure 2.3: The non-contact body temperature measurement framework proposed in [88].

Aubakir et al. [89] used a long wavelength infrared (LWIR) camera to monitor body temperature based on an automatic face detection using the Viola–Jones (V–J) algorithm and ROI averaging. It is noted that this method of ROI averaging is prone to perturbances due to environmental factors. Another study by Sharma et al. [90] proposed a contactless approach to measure temperature using the Viola–Jones face detection algorithm. Authors used an algorithm to select the best frame containing all the features of the face rather than computing the temperature of the face in each frame and then calculated the temperature for the same frame. However, the Viola–Jones face detection method was sensitive to the subject's head movement.

Lin et al. [91] presented a non-contact, automatic continuous body temperature measurement (CBTM) system using a thermal camera. In the proposed method depicted in Figure 2.4, face detection was implemented using a neural network, and a kernel correlation filter (KCF) tracker was used to track the face region across each frame. Nevertheless, in the proposed method, participants were advised to sit and remain stationary, making the system not applicable for real-time situations.

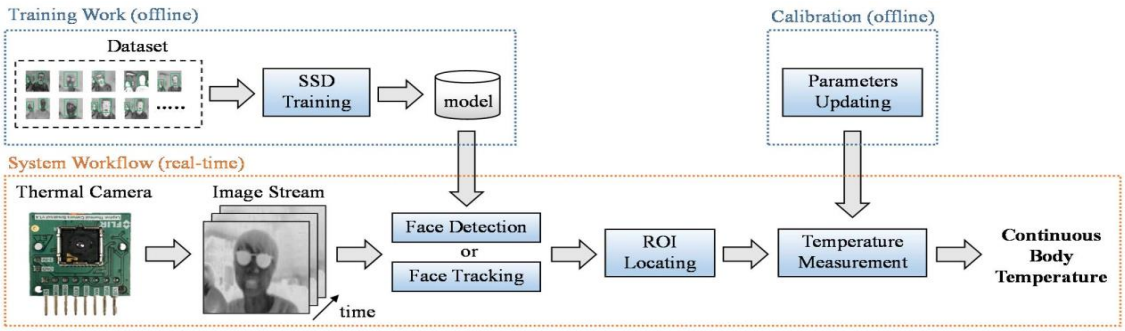


Figure 2.4: The thermal camera-based continuous body temperature measurement framework proposed in [91].

Sumriddetchkajorn et al. [92] introduced a parallel measurement technique to screen human temperatures that is applicable to large public areas by combining infrared technology, image processing techniques, and human flow management. Through image filtering, morphological operations, and particle filtering, large numbers of subjects were able to be processed efficiently. With 100% sensitivity and 36.4% specificity, however, this approach requires refinement. Using an infrared thermal camera, Silawan et al. [93] proposed a novel idea to estimate core body temperature. The method enhances the sensitivity and specificity of a fever screening system considering various environmental situations. A pilot study by Thomas et al. [94] explored the use of an IR camera for measuring the core body temperature of workers in hot work environments. The strong correlation between the measured and reference data showed potential for the infrared camera as a non-invasive alternative to the thermometer pill and rectal thermometer.

Table 2.1. Different studies extract body temperature using thermal cameras.

Ref	Sensor Used	ROI	Used Technique	Temperature Measured
Bilodeau et al. [88]	7.5–13 μm, LWIR (FLIR ThermoVision A40M)	Face	Particle filter, Kalman filter	-
Aubakir et al. [89]	8–14 μm, LWIR (FLIR Lepton 2.5)	Forehead	V-J method	34.95 °C to 37.00 °C
Sharma et al. [90]	7.5–13 μm, LWIR (FLIR X63900) NIR (CP-PLUS CP-USC-TAL2)	Face	V-J method	29.45 °C to 32.82 °C
Lin et al. [91]	8–14 μm, LWIR (FLIR Lepton 2.5) 8–14 μm, LWIR (KeySight Keysight U5855A)	Forehead	Deep- learning	27 ℃ to 37.5 ℃
Sumriddetch kajorn et al. [92]	7.5–13 μm, LWIR (FLIR ThermoVision A40M)	Face	Image filtering, particle analysis	35 °C to 40 °C
Silawan et al. [93]	8–14 μm, LWIR (Optris PI450)	Forehead, mouth, cheek	Multiple data comparison	36.0 °C to 39.5 °C
Thomas et al. [94]	7.5–14 μm, LWIR (Fluke TiS65)	Face	Linear regression	34 °C to 41 °C

2.1.1.2. HR and BR Measured with Thermal Camera

To assess a human's health condition, especially to identify a virus-infected patient, measuring body temperature alone is not sufficient. Other vital signs, such as breathing rate and heart rate, should be considered as well. Thermal imaging techniques can detect two

physiological phenomena caused by cardiorespiratory activity. The first phenomenon related to cardiac activity is the slight heat differences produced by the pulsating blood flow in the major superficial arteries in particular anatomical regions. A thermal camera can detect these heat differences to extract the pulse signal. The heat difference due to the pulsating blood flow is approximately 0.08 K, which is far less than the temperature of normal skin (310 K) from a black-body perspective [67]. The second phenomenon related to breathing activity is the small temperature differences around the nostrils of approximately 0.1 K [95] due to the heat transfer from warm exhaled and cold inhaled lung contents.

A summary of studies in literature where thermal cameras were used by researchers to monitor the breathing rate and heart rate is shown in Table 2.2. Murthy et al. [96] proposed a noncontact technique to monitor breathing rate using thermal imaging based on a statistical methodology that modelled breathing as a combination of expiration and non-expiration distributions. Fei et al. [97, 98] enhanced the thermal imaging technique by using an optical bandpass filter and Fourier analysis based on the fast Fourier transform (FFT). In these studies, the ROI selection was done manually, and there was not sufficient compensation for motion.

Sun et al. [78] presented a novel non-contact technique for measuring a cardiovascular pulse using thermal imaging. The algorithm includes motion tracking, manual ROI selection, FFT, and adaptive estimation to measure cardiac pulse. However, in this study, the subject sample size was very small. Garbey et al. [76] further elaborated the Sun et al.

method by increasing the sample size from five to 32 subjects and recording time from two minutes to five minutes. However, both methods are highly affected by the movement of subjects and the manual selection of ROI. Chekmenev et al. [99] introduced a novel non-contact method to remotely measure HR and BR based on wavelet analysis using continuous wavelet transform (CWT), which gave a better performance than FFT. However, they only considered four subjects to validate their method, and the ROI selection was manual. In another study, Fei et al. [100] introduced a new and enhanced technique to monitor BR based on automatic tracking/localization of ROI using a coalitional tracking algorithm and wavelet analysis based on CWT. Using CWT, not only the breathing rate, but also the full breathing waveform was recovered.

Shakhih et al. [101] presented an approach to measure the time of the inspiration (TI) and expiration (TE) considering three different breathing patterns that were simultaneously captured by infrared thermal imaging (ITI) and respiratory inductive plethysmography (RIP). However, the method was highly susceptible to head movement and image processing was also not sufficient.

Pereira et al. [75] presented a new robust method to remotely measure BR by using thermal imaging captured by an LWIR camera. As shown in Figure 2.5, in the proposed method, the ROI and region of measurement (ROM) were identified automatically, and then tracked using a particle filter framework. A Butterworth bandpass filter was applied to the raw signal. Finally, three estimators were calculated to compute instantaneous breathing frequencies. However, the authors only considered BR.

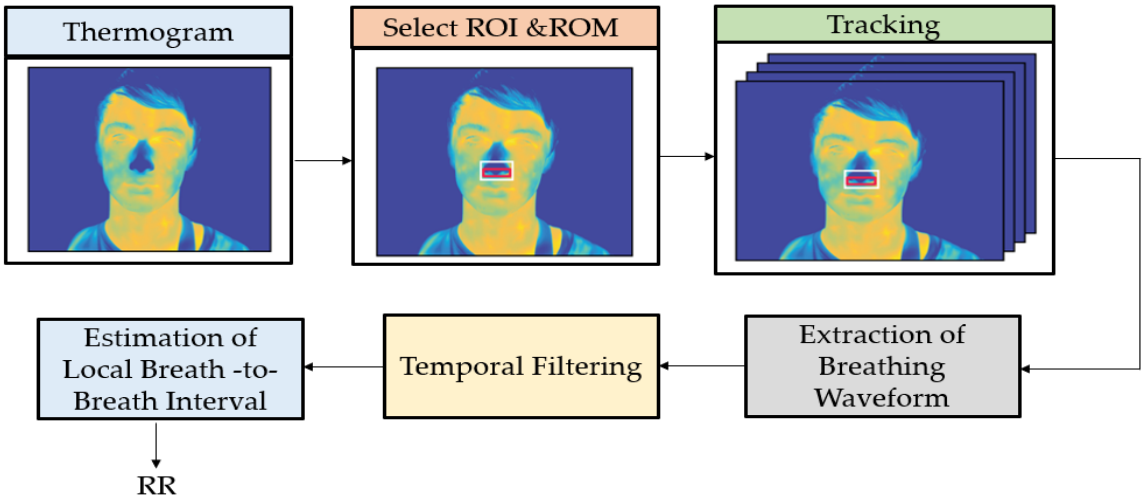


Figure 2.5: Block diagram for the BR estimation proposed in [75].

The authors [102] proposed a non-contact method to monitor cardiorespiratory signals using thermal imaging. BR was extracted incorporating the same algorithm as [75]. Feature points were detected and tracked within the manually selected ROI. Temporal filtering, principal component analysis (PCA), and peak detection were used to estimate HR. Most of the above-discussed methods used thermal information around the nostril area, which is not suitable for long term monitoring, because it can happen that the nose is not in the field of view of the thermal camera. To deal with this issue, Pereira et al. [103] presented an enhanced and robust method to estimate BR using thermal imaging, considering both the temperature difference around the nostril area and mouth and the movement of the

shoulders. In the proposed algorithm, four ROIs were detected automatically and tracked using a particle filter framework. Only five subjects were considered in this study.

Most of the works discussed above considered only adults as their subjects to monitor vital signs. However, a few works did research on infants as well. For example, Abbas et al. [72] first used infrared thermography to monitor the breathing rate of neonates in the neonatal intensive care unit (NICU) using an infrared camera and an algorithm based on CWT. However, they did not consider a tracking algorithm, and the sample size was small. Pereira et al. [104] proposed a feasible approach based on IRT to monitor the BR of preterm babies incorporating the same algorithm as [75]. However, the sample size was very small, as they considered only four infants. In [105], the authors presented a "black-box" or grid-based approach algorithm to remotely monitor the BR of neonates in the NICU using an LWIR camera. However, this method is highly affected by motion artefacts.

Table 2.2: Different studies to monitor vital signs using thermal cameras.

Ref	Sensor Used	Vital Signs	ROI	Used Technique	Result
Murthy et al. [96]	3–5 μm, MWIR	BR	Nose	Advanced statistical algorithm	Accuracy = 98.5%
Fei et al. [97, 98]	3–5 μm, MWIR (FLIR) 3–5 μm, MWIR (Indigo Systems)	BR	Nose	Optical bandpass filter	-
Sun et al. [78]	MWIR (Indigo Systems)	HR	Forehead, neck, and wrist	FFT	PCC = 0.994
Garbey et al. [76]	MWIR Indigo camera (Indigo Systems)	HR	Forehead, neck, and wrist	FFT	CAND = 88.52%
Chekmenev et al. [99]	LWIR (FLIR)	HR and BR	Face and neck	CWT	-
Fei et al. [100]	3–5 μm, MWIR (FLIR SC6000)	BR	Nose	CWT	CAND = 98.27%
Shakhih et al. [101]	7–14 μm, LWIR (Infrared Camera Incorporation 7640 P-series)	TI and TE	Nose	Mean pixel intensity	PCC = 0.796(TI), 0.961(TE)
Pereira et al. [75]	7.5–14 μm, LWIR (VarioCAM R HD head 820S/30)	BR	Nose	Particle filter framework and temporal filtering	MAE = 0.33, 0.55 and 0.96 breaths/min.
Pereira et al. [102]	2–5.5 μm, MWIR (InfraTec 9300)	HR and BR	Head and nose	Particle filter framework, temporal filtering and PCA	RMSE = 3 bpm (HR), RMSE = 0.7 breaths/min.

Pereira et al. [103]	7.5–14 μm, LWIR (VarioCAM R HD head 820S/30)	BR	Nose, mouth, shoulders	Particle filter framework and signal fusion	RMSE = 0.24,0.89 breaths/min.
Abbas et al. [72]	1–14 μm, LWIR (VarioCAM HR head)	BR	Nose	CWT	-
Pereira et al. [104]	7.5–14 μm, LWIR (VarioCAM R HD head 820S/30)	BR	Nose	Particle filter framework and temporal filtering	Relative error = 3.42%
Pereira et al. [105]	7.5–14 μm, LWIR (VarioCAM R HD head 820S/30)	BR	-	Black-box	RMSE = 4.15 ± 1.44 breaths/min.

2.1.2. Video Camera Imaging Technology

2.1.2.1. Motion-Based Method

Cardiorespiratory activity causes subtle motion in various parts of human body that can be measured from video to extract vital signs. Numerous researchers have introduced different techniques to monitor vital signs using information extracted from the motion. Nakazima et al. [106] proposed a contactless method based on optical flow analysis to extract RR from the video of whole body motion captured by a CCD (charge-coupled device) camera. Another optical flow-based respiratory monitoring system was introduced by Frigola et al. [107] using videos of a subject's chest movement recorded by video camera. Optical flow-based techniques have been found to be susceptible to motion artefacts caused by the subject's movement, ambient light, and unclear ROI. Computational complexity is a consideration for optical flow-based analysis compared to other processing options.

Several studies have been published based on blind source separation (BSS) using either principal component analysis (PCA) or independent component analyses (ICA). Balakrishnan et al. [108] proposed a novel non-contact method based on head motion using facial video recorded by a digital camera to measure heart rate. In their proposed method, the Viola-Jones face detector [109] and the Kanade Lucas Tomasi (KLT) [110] tracking algorithm were exploited for detecting the face, extracting the ROI and tracking the ROI feature points based on the good feature tracking (GFT) method. Then, using a Butterworth filter, the tracked points were temporally filtered, and PCA was exploited to remove artefacts and measure the physiological signal. Finally, the pulse rate was extracted by means of a simple peak detection algorithm followed by a FFT. On the other hand, to monitor HR, Shan et al. [111] introduced another noncontact method using head motion captured by smart phone camera based on ICA rather than PCA. These methods are susceptible to motion artefacts as they considered stationary subjects only without any

internal or external movement. To mitigate these limitations, an improved head motion-based method to extract HR using a webcam was designed by Irani et al. [112] based on discrete cosine transform (DCT) and a moving average filter considering various facial expressions and head poses. Haque et al. [113] further improved this technique by integrating both the GFT method and supervised descent method (SDM) with training against the MANHOB-HCI (Human computer interaction) database with moving objects. Lomazia et al. [114] proposed a method to track background and facial features to solve the handshaking problem with a smartphone camera. This method is further enhanced in [115] by exploiting a system with two cameras in a smartphone, where the front camera was for tracking facial features and the rear camera was for tracking background features.

Some researchers have used video magnification to extract vital signs from motion. He et al. [116] introduced a contactless method combining both the Eulerian video magnification (EVM) [117] method and 2-Gaussian curve modelling method for extracting pulse signal from a subject's wrist using a digital camera. A study by Al-Naji et al. [118] used video magnification and a frame subtraction method to remotely measure cardiac activity (heart rate, pulse width and the total cycle length) from head motion. Another study by Al-Naji et al. [119] presented a non-contact respiratory monitoring scheme for detecting and measuring respiratory rate as well as respiratory cycle timing parameters by using a video camera from the movement of the chest or blanket draped over an infant subject in various sleeping postures based on the two processing techniques as shown in Figure 2.6. The first technique was for magnifying the motion resulting from chest movement based on 8 level wavelet pyramid decomposition and 5th order temporal elliptical filtering. The second technique was to measure respiratory rate by using motion detection exploiting frame subtraction, local contrast enhancement, binarization, morphological and masked filtering, white area detection and logical matrix calculation where the mean distance between ones in the logical matrix was used to calculate RR. Although the system was an effective solution for the unclear ROI problem, it still had some limitations. They have only considered a single vital sign i.e., RR and a single infant, which limits the real applications of the system. In [120] a contactless real time monitoring system was introduced for measuring respiratory rate and detecting apnoea by exploiting a Microsoft Kinect v2 sensor from the movement of thorax and abdomen in different sleeping positions and various conditions of light including dark environments. They used an improved motion magnification scheme based on the Lanczos resampling method, wavelet pyramid

decomposition, temporal band pass filtering and image denoising to magnify the input signal. Frame subtraction was used to calculate respiratory rate. Nevertheless, the system suffered from motion artefacts, short range, and a limited number of viable ROI. Furthermore, they only considered respiratory activity and apnoea, and did not consider other vital signs. To mitigate this limitation an improved contactless system was presented to measure the HR and RR and to sense irregular cardiopulmonary functions such as bradycardia, tachycardia, bradypnea, tachypnoea, and central apnoea by using the signal from the thoracic-abdominal region based on image sequences taken by the Microsoft Kinect v2 sensor with processing that considered unclear ROI, various illumination conditions and different sleeping postures [121]. The basic block diagram of the proposed system is presented in Figure 2.7. An efficient motion magnification technique (EMMS) [122] was used to magnify the input data to make the movement apparent. Then, to calculate HR, an intensity-based technique was used, including signal decomposition, blind source separation, spectral analysis, filtering, and peak detection. To calculate RR, they used a frame subtraction technique, binarization, filtering, white area detection and binary matrix calculation. If any irregularity was detected an alarm would be sent to a carer to notify them of the irregularity. However, the detection range was limited to 4.5 m and the number of ROI was also limited. Furthermore, system failure may occur in the case of a fully covered subject or if the participant should lie on their stomach.

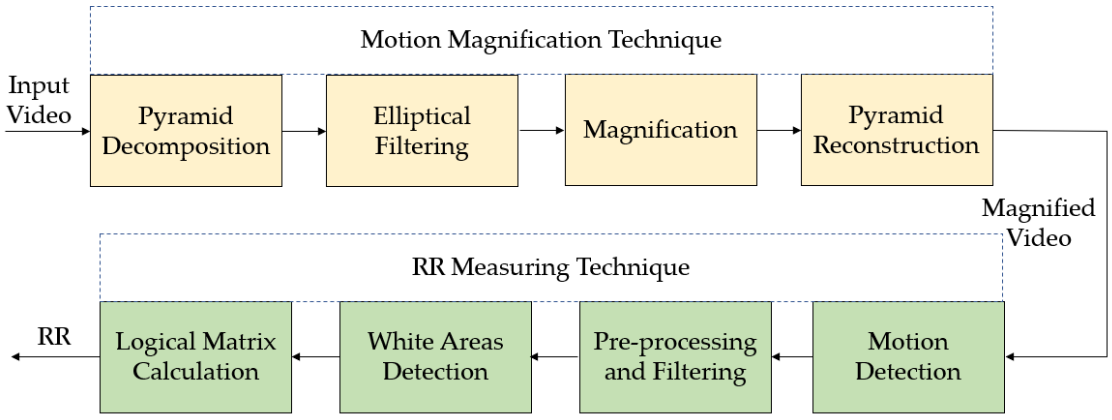


Figure 2.6: System diagram of the remote respiratory monitoring system [119].

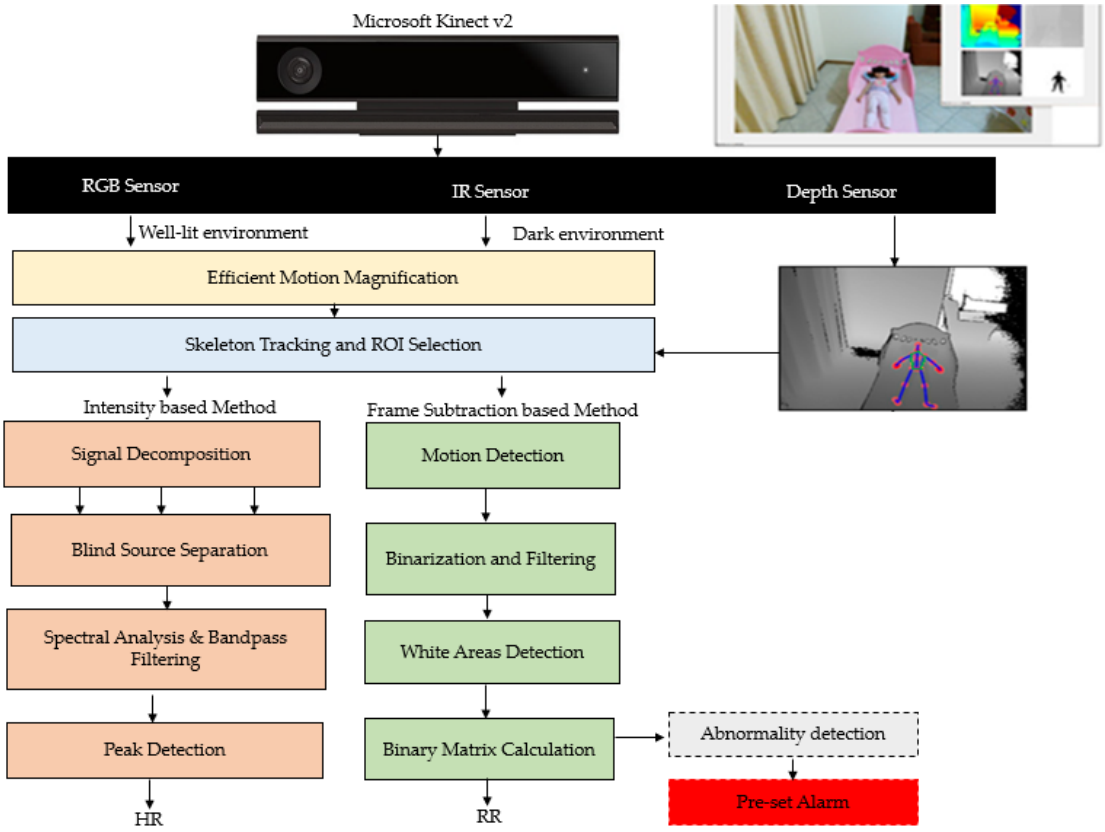


Figure 2.7: Block diagram of contactless vital sign monitoring system using Microsoft Kinect sensor [121].

The motion-based method was very beneficial in this case, since the method is robust to illumination variance, skin tone and unclear ROI. However, the main limitation of this method is the dependency on motion features to extract the physiological signals. As a result, the technique may be highly vulnerable to voluntary motion variations such as different facial expressions, walking and talking by the subject. Consequently, determining vital signs during voluntary motion may reduce the consistency and reliability of the method, and remains a research problem to be overcome.

2.1.2.2. Colour-Based Method

Photoplethysmography (PPG) is a non-invasive, low-cost, passive, optical technique, first proposed by Hertzman et. al. in 1937 [123]. It can monitor three vital signs, HR, RR and SpO_2. PPG is used extensively in biomedical, clinical, and non-clinical fields due to its simple design and relatively low cost. PPG measures variations in the optical properties of transmitted or reflected light from the human skin because of the blood volume changes during cardiorespiratory activity. Blood absorbs light more than the adjacent tissue which causes subtle optical property variations of the skin because of haemoglobin in the blood.

Generally, PPG can function either in a transmissive or a reflective mode, however, the first one is limited to regions such as ear lobe and fingertip. In the PPG technique, a dedicated light source is required to illuminate a part of the body and an optical sensor or photodetector, connected to the skin is needed to sense the optical properties of the skin. Imaging photoplethysmography (iPPG) is basically remote or non-contact PPG whereas for skin-contact a PPG sensor and illuminator replace the camera. To detect PPG signals remotely, many researchers have considered a video camera as an optical sensor. Some studies have used a dedicated light source and others considered just ambient light as a source of light. Nevertheless, a dedicated light source increases the hardware setup of experiments as well as starting to become invasive to the participant, particularly infants.

As shown in Figure 2.8, the iPPG reflection model mainly consists of three parts such as a light source, a patch of human skin containing pulsating blood and a video camera. The light source can be a dedicated light source or ambient light. When a light source illuminates human skin, subtle colour changes are observed from the videos captured by the camera. Without illumination variations and motion artefacts, colour changes denote the blood volume changes in the microvascular tissue bed under the skin because pulsatile blood flows vary each cardiac cycle. Nevertheless, illumination variations and motion artefacts can also be a reason for the intensity and spectral composition variations. So, the skin area observed by the camera has a colour variation because of motion-induced and illumination-induced intensity/specular changes and pulse-induced subtle colour changes. It is assumed that the spectral composition of the light is fixed. The variation of light intensities is dependent on the distance from light source to camera and to the skin and the specific geometry of the situation.

Based on the dichromatic model al. [124], the light reflected from the skin can be presented as follows

$$P_k(t) = I(t)\,(R_s(t) + R_d(t)) + R_n(t) \tag{2.1}$$

where $P_k(t)$ represents the RGB channels of the *k*th skin pixel; $I(t)$ denotes the illumination intensity level; $R_s(t)$ signifies the specular reflection and $R_d(t)$ represents the diffuse reflection. $I(t)$ is modulated by both specular and diffuse reflection. $R_n(t)$ signifies the quantization noise of the camera sensor.

The specularly reflected light from the skin surface is like light reflected from a mirror. As specularly reflected light does not penetrate skin tissues, it does not carry any information about the physiological signal. Additionally, the spectral composition of specular reflected light is identical to the light source al. [124]. On the other hand, diffuse reflected light penetrates the skin, it is absorbed and scattered inside skin tissues and then reflected. Useful information about the physiological signal can be extracted from the diffuse reflected light.

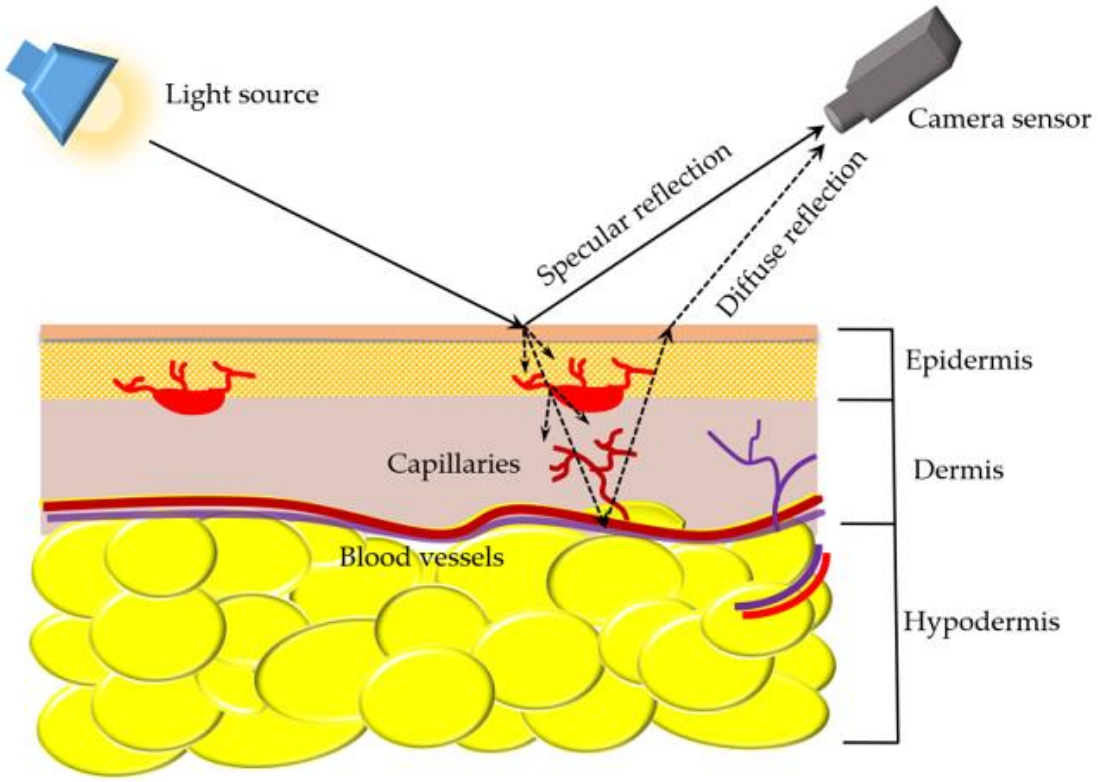


Figure 2.8: Skin reflection model of iPPG [adapted from [124]].

Even though the slight colour variations of the skin are invisible to human eyes, they can be detected by video cameras. From the frame sequences of the video captured by a camera, the PPG signal can be extracted by image and signal processing techniques.

Over the past several years, numerous techniques in iPPG have been proposed under well controlled situations such as considering stationary subjects and stable illumination as listed in Table 2.3. Verkruysse et al. [125] used normal ambient light as the light source to measure the HR from the human face using a digital camera. In their proposed method, the ROI was selected manually, and the raw PPG signal was calculated per frame by an intensity method: averaging the spatial pixel value of the red, green, blue (RGB) colour

channels. The FFT algorithm was used to calculate power spectral density of the signal to extract HR. They showed that under these circumstances, the green channel with highest signal to noise ratio (SNR) contains the strongest plethysmography signal because oxygenated haemoglobin absorbs green light more than blue and red. However, this method is strongly affected by motion artefacts caused by the subject's movement, as they did not consider any method for noise removal. Moreover, their ROI selection was manual. To overcome the limitations of the previous technique, Pho et al. [126] proposed a novel method to measure HR considering BSS using a webcam and considering three subjects at a time. Moreover, they used the OpenCV face detection algorithm which is automatic and based on the Viola and Jones [109] method. The facial ROI was defined as a rectangular bounding box. However, the proposed work only considered very small movements and did not consider illumination variations. Additionally, in every case Poh et al. considered the second component produced by ICA as the PPG signal. Moreover, they only measured HR in that study, a drawback that was minimised in [127] where they presented an enhanced method to measure RR and heart rate variability (HRV) along with HR. Another ICA-based method was introduced by Purche et al. [128] where they used two algorithms for extracting the pulse rate: a peak detection algorithm and power-spectrum analysis algorithm under both relaxed and active conditions and found that power-spectrum performed better than peak detection. However, this work is highly susceptible to noise artefacts caused by illumination variations, subject movement, facial expressions etc. Lewandoska et al. [129] proposed another method based on BSS using PCA and reported that PCA is faster than ICA in terms of computation time and it can be a good alternative if only HR is extracted. However, in this method only the stationary case and HR were considered.

Table 2.3. IPPG-based methods under well controlled condition.

Ref.	Used Sensors	Vital Signs	ROIs	Used Technique	Distance (m)	No. of participants at a time
Verkruysse et al. [125]	CCD	HR and RR	Face and forehead	Single/multichannel analysis	1-2 m	1
Pho et al. [127]	Webcam	HR, RR and HRV	Face	ICA	0.5 m	1

Purche et al. [128]	Webcam	HR	Forehead, nose, and mouth	ICA	0.5 m	1
Lewandoska et al. [129]	Webcam	HR	Face and forehead	PCA	1 m	1

2.1.3. Combinations of Different Technologies

To enhance the performance of monitoring vital signs, different technologies can be combined, as shown in Table 2.4. For example, by combining an RGB camera, a monochrome camera with colour filter, and a thermal camera, a study by Gupta et al. [130] proposed an efficient multicamera measuring system to measure HR and HRV. In the proposed method, face segmentation was done using conditional regression forests (CRF) to detect facial landmarks in real-time, and an ROI was selected. The raw signal was calculated based on a spatial average of the pixels within the ROI for each channel, i.e., red, green, blue, magenta, and thermal (RGBMT). To recover the underlying source signals, ICA was used, and a bandpass filter was applied to the selected source signal. Finally, peak detection was performed to calculate HR and HRV. The experimental results showed that the GRT (green, red, and thermal) channel combination gave the most accurate results, demonstrating that the inclusion of more spectral channels could attain more robust and accurate measurements.

Hu et al. [131] presented a dual-mode imaging technique based on visible and long-wave infrared wavelengths to extract breathing rate and pattern remotely using both an RGB and a thermal camera. After performing image registration, to identify a ROI in the RGB image, the cascade object detector based on the Viola–Jones algorithm and the screening technique using biological characteristics were used. Next, to select the corresponding regions in the thermal image, linear coordinate mapping was applied. To select the interest points from ROIs in the visible light gray images, the Shi–Tomasi corner detection method derived from the Harris–Stephens was applied. Cross-spectrum ROI tracking was attained using linear coordinate mapping. After that, the mean pixel intensity was obtained within the ROI of the thermal image. In all of the frames, the raw pixel intensities of ROIs were smoothed by the moving average filter and smoothing method. Finally, breathing rate was extracted from the smoothed signal. By adding RGB images, it was possible to detect and track the face and facial tissue in thermal images more accurately and faster. However, the above systems may not be feasible for measuring vital signs at night.

Another dual camera imaging system was proposed in [132] to measure BR and HR at night using an RGB infrared camera and thermal imager. Image registration was done based on affine transformation, face detection was performed using a pre-trained boosted Cascade classifier or fully convolutional network, and a discriminative regression-based approach was used to select landmarks on the face. IR images were used to select ROI from thermal images using linear coordinate mapping. ROI tracking in thermal images was done by spatio-temporal context learning (STC). A moving average filter was applied to the raw signal and finally, HR and BR were calculated.

Bennett et al. [68] used both a thermal and optical camera to monitor heart rate and blood perfusion based on EVM. After identifying ROI from both thermal and optical videos, the average intensity was computed within the ROI. After that, a Butterworth lowpass filter was used to filter these signals, and then normalized. The experimental results of blood perfusion showed that optical video with EVM was sensitive to skin colour and lighting conditions, while thermal video with EVM was not sensitive to these variables. However, in the proposed method, movement of the subject was not considered, and the sample size was small. Although a temperature response caused by blood occlusion was sensed by the thermal camera, it was not in the pulsatile manner of perfusion.

Table 2.4. Different studies to monitor vital signs using combined technology.

Ref	Sensor Used	Vital Signs	ROI	Used Technique	Result
Gupta et al. [130]	RGB, monochrome and thermal camera	HR and HRV	Cheeks and forehead	ICA	Error = 4.62%
Hu et al. [131]	RGB and thermal camera	BR	nose and mouth	Viola–Jones algorithm together with the screening technique	LCC = 0.971
Hu et al. [132]	RGB infrared and thermal camera	HR and BR	Mouth and nose regions	Moving average filter	LCC =0.831 (BR), LCC = 0.933 (HR),
Bennett et al. [68]	Thermal and digital camera	HR and blood perfusion	Face and arm	EVM	-

2.2. Different Aspects of Video Camera Imaging

2.2.1. Motion Artefacts

There are some researchers who took motion artefacts as a serious issue and proposed various methods to suppress the motion artefacts based on blind source separation (BSS) or model-based methods listed in Table 2.5.

Pho et al. [126] first introduced a novel method to remove motion artefacts using BSS based on ICA where a linear combination of three channels were decomposed into three independent components to remove motion artefacts. Nevertheless, the proposed work only considered very small movement and did not consider illumination variations. In [133], a motion compensated method was designed based on single-channel independent component analysis (SCICA) to extract both HR and RR using a CMOS video camera during exercise. Moreover, they implemented image stabilisation via 2-D cross correlation and used time-frequency representation (TFR) for spectral analysis. Still, the motion artefacts could not be totally removed using this method. Moreover, in the extracted components, physiological waveforms were different under different situations. Feng et al. [134] introduced an improved motion tolerant technique to measure both average and instantaneous HR via webcam considering a substantial amount of movement by the body as well as the head. After detecting a face by the V-J method they used speeded up robust feature (SURF) detection to find trackable interest points in the facial region and exploited the KLT algorithm for tracking the faces. To extract the signal, they used an adaptive bandpass filter and automatic sorting algorithm to sort ICA output components by taking a sine function as a reference signal. Nevertheless, this method is not suitable if the subject walks or does other difficult movements and the range tested was also very short with limited ROI options. Using a digital camera to remotely measure heart rate, Qi et al. [135] proposed a new technique to improve the photoplethysmography signal by combining facial sub-regions landmark localization and joint blind source separation (JBSS) methods to extract the physiological signal.

However, blind source separation-based approaches may have limitations when exposed to other incidental periodic signals, including illuminance or motion variations. To overcome this limitation, some researchers considered different techniques to suppress motion artefacts rather than using BSS. For instance, a continuous wavelet transform based method was presented in [136] to extract both instantaneous heart and respiratory rates

using a webcam with normal head movements. However, their motion consideration was limited to head movements, and they did not consider illumination variation or other movements. De Haan et al. [137] proposed a robust technique to extract HR from CCD camera video based on chrominance (CHROM) during exercise. In chrominance-based methods, the colour difference signals are combined linearly considering a standardized skin-tone. However, in this case they considered two fitness devices, a stationary bike, and a stepping device with only one subject. Their method was also affected by skin colour especially dark skin. This method was further improved in [138] by using the blood volume pulse vector and five different exercise devices. Here, blood volume pulse was considered as a signature for various reflection spectra of skin for explicitly distinguishing the physiological signal from noise caused by motion. Another study by Feng et al. [139] presented a system, robust to motion, using adaptive colour variation between red and green channels as well as an adaptive bandpass filter (ABF) to measure HR from moving subjects exploiting a webcam. Using a CCD camera, Wang et al. [140] introduced a new robust technique exploiting spatial subspace rotation (2SR), where a spatial subspace of skin-pixels was estimated and its temporal rotation was measured in the image domain to measure HR. Nevertheless, as 2SR is a completely data-driven algorithm, it might cause undetected inaccurate results because of noise or a poorly selected skin-mask. Another algorithm to measure HR by exploiting imaging of the plane orthogonal to skin (POS) was presented in [124] where normalized RGB channels were combined into two new channels which were merged by weighting to the desired signal. Later, to allow the independent reduction of various motion frequencies by exploiting sub band (SB) decomposition, Wang et al. [141] enhanced the POS method to measure continuous HR considering different fitness applications. However, this technique cannot supress motion artefacts when the motion and the pulse have the same frequency. Moreover, this method is susceptible to degradation with skin tone and illumination variation. Wu et al. introduced a time-frequency analysis method using CWT [142] and a motion resistant spectral peak tracking (MRSPT) method [143] to measure HR via Webcam considering seven motion circumstances when subjects were driving, running and engaged in fitness training. Nevertheless, these methods considered very short ranges of around 1.5m. For monitoring precise HR using a video camera, Xie et al.[144] presented another method using singular spectrum analysis (SSA) where the motion signal was measured from signal and SSA based on singular value decomposition (SVD) was employed to rectify motion artefacts during treadmill exercise. However, they only considered tread mill exercise and a very small

number of subjects. Moreover, only the first three leading decomposed components of the motion signal were removed from BVP which may affect the performance of the proposed method. McDuff et al. [145] designed a computationally efficient method based on linear transformation exploiting parameters attained from tissue-like models of the skin to extract HR and HRV accurately in the presence of various head motions using a colour camera. However, the apparatus needed to be calibrated via a known colour grid and only systematic head motions were considered. Fallet et al. [146] introduced a signal quality index (SQI) and verified the capability of SQI as a tool in order to enhance the consistency of heart rate measuring applications considering videos of a moving object. However, this SQI on videos needs to be tested with participants carrying out smoother movements, in time-varying illumination environments.

Table 2.5. IPPG-based methods concern about motion artefacts.

Ref.	Used Sensors	Vital Signs	ROIs	Used Technique	Distance (m)	No. of participants at a time	Results
Pho et al. [126]	Webcam	HR	Face	ICA	0.5 m	3	PPC=0.95, RMSE= 4.63bpm,
Yu et al. [133]	CMOS	HR and RR	Palm and face	SCICA	0.35-0.4 m	1	PCC=0.9
Feng et al. [134]	Webcam	HR	Forehead	ICA	0.75 m	1	PCC=0.99
Qi et al. [135]	Digital camera	HR	Face	JBSS	-	1	PCC=0.74
Bousefsaf et al. [136]	Webcam	HR	Face	CWT	1 m	1	PCC=1
De Haan et al. [137]	CCD	HR	Face	CHROM	-	1	PCC=1, RMSE=0.5
De Haan et al. [138]	CCD	HR	Face	PBV	-	1	PCC=0.99, RMSE=0.64
Feng et al. [139]	Webcam	HR	Cheeks	GRD	0.75m	1	PCC=0.97
Wang et al. [140]	CCD	HR	Face and forehead	2SR	1.5 m	1	PCC=0.94
Wang et al. [124]	CCD	HR	Face	POS	1.5 m	1	SNR (dB) = 5.16
Wang et al. [141]	CCD	HR	Face	Sub-band decomposition	2 m	1	SNR (dB) = 4.77
Wu et al. [142]	Webcam	HR	Face	CWT	0.5-1.5 m	1	SNR (dB)= -3.01

Wu et al. [143]	Webcam	HR	Cheeks	MRSPT	0.6-1.6 m	1	RMSE = 6.44 bpm
Xie et al. [144]	Video camera	HR	Face	SSA and SVD	-	1	PCC = 0.99, RMSE=3.99
McDuff et al. [145]	Colour camera	HR and HRV	Face	Linear transformation	-	1	MAE= 4.17 bpm, SNR (dB) = 3.13
Fallet et al. [146]	Video camera	HR	Forehead and face	SQI	-	1	MAE= 4.72 bpm

2.2.2. Illumination Variations

While most studies are concerned about the motion artefacts, other studies have considered how to mitigate illumination variations as listed in Table 2.6. For example, Chen et al. [147] proposed a new robust method to measure the pulse signal using a camera which is based on reflectance decomposition on the green channel and ensemble empirical mode decomposition (EEMD) to suppress the noise caused by illumination variation. This method was further improved in [148] where HR was evaluated by means of a multiple linear regression (MLR) model followed by poisson distribution to reduce the effects of ambient light changes. Nevertheless, EEMD has limitations of inadvertently considering periodic illuminance variations as physiological signal, especially if the frequency is near the normal cardiac frequency range particularly from 0.75 to 4 Hz. Moreover, both methods are not suitable for real time applications. To measure HR, Lee et al. [149] used a different approach based on multi order curve fitting (MOCF) to remove noise artefacts considering subjects watching television in a dark room. In the proposed approach, they subtracted the estimated brightness signal from the raw PPG signal to cancel out the noise caused by illumination variations. Another study by Tarassenko et al. [150] presented a new technique based on autoregressive (AR) modelling and pole cancellation, to cancel out the aliased frequency components caused by artificial light flicker which enhanced robustness of measurement under strong fluorescent lights. Nevertheless, the periodic illumination variations may affect the performance, as AR modelling is a spectral analysis technique. Moreover, there is a calibration issue while calculating oxygen saturation as well. To extract HR using a webcam, Cheng et al. [151] designed a robust method exploiting illumination variation by combining JBSS and EEMD and considering background images as well. However, this method is not free from motion artefacts as they only considered stationary subjects. Moreover, the illumination variation considered here was mostly controlled, which restricts the applicability of the proposed method in real time. Another method that was robust to illumination variations was proposed by Xu et al. [152] using

partial least squares (PLS) and multivariate empirical mode decomposition (MEMD) to efficiently extract HR using a webcam under varying illumination conditions. However, they only considered artificial illumination created by an LED lamp.

Table 2.6. IPPG-based methods concern about illumination variations.

Ref.	Used Sensors	Vital Signs	ROIs	Used Technique	Distance (m)	No. of participants at a time	Results
Chen et al. [147]	Digital camera	HR	Brow area	EEMD	0.07-0.09 m	1	PCC=0.91
Lin et al. [148]	Digital camera	HR	Brow area	EEMD + MLR	0.10-0.25 m	1	PCC=0.96
Lee et al. [149]	Digital Camera	HR	Cheek	MOCF	1 m	1	RMSE = 1.8 bpm
Tarassenko et al. [150]	Digital camera	HR, RR, SpO_2	Forehead and cheek	AR modelling and pole cancellation	1 m	1	MAE = 3 bpm
Cheng et al. [151]	Webcam	HR	Face	JBSS + EEMD	0.50 m	1	PCC=0.91
Xu et al. [152]	Webcam	HR	Face	PLS + MEMD	0.50 m	1	PCC=0.81

Most of the above-mentioned works are concerned about the elimination of the effect of illumination variations and to keep the participants still, which means that they did not consider motion artefacts. Nevertheless, both illumination variations and motion artefacts are important in realistic applications. So, these approaches initially considered to remove illumination variations should be further improved by considering motion artefacts.

2.2.3. Both Illumination Variations and Motion Artefacts

Some studies are concerned about either motion artefacts or illumination variations. However, only a few studies have considered both illumination variation and motion artefacts as an issue. Table 2.7 summarizes the studies concerned about both illumination and motion. Li et al. [153] introduced a normalized least mean square (NLMS) adaptive filtering scheme to suppress the effect of both the subject's motion and illumination variations considering realistic situations, for example watching movies or playing games by using an iSight camera to measure cardiac pulse. Nevertheless, this method is seriously affected by large head movements. Using a monochrome camera, Kumar et al. [154] designed a robust method to extract HR and HRV based on a weighted average and considering different skin tones, illumination variation, and several motion scenarios. Al-Naji et al. [155] presented a new noise elimination technique using both complete ensemble

empirical mode decomposition with adaptive noise (CEEMDAN) and canonical correlation analysis (CCA) to eliminate noise artefacts caused by movement of subjects and camera, variations of illumination and variation of skin tone. As shown in Figure 2.9, first data was collected using an unmanned aerial vehicle (UAV) from 3m distance at various times of day with various illumination levels. Then, an improved video magnification technique [156] was used to magnify skin colour variation. After that, an efficient face detection method [157] was used to detect facial ROI as it is more effective with inclined or angled face. Raw iPPG signal was extracted using spatial averaging within the facial ROI of green channel. To supress the noise caused by illumination variations, CEEMDAN was applied as it shows better performance than EMD and EEMD by decreasing noise from the intrinsic mode functions (IMFs) with more physical meaning. By using CEEMDAN, the iPPG signal was decomposed into eight IMFs and 5, 6 and 7th IMFs were selected as their frequency band fall within 0.2-4 Hz, corresponding to 12-240 beats/min. Then, to reduce motion artefacts CCA was used on chosen IMFs as CCA generates components derived from their uncorrelated signals rather than independence components used in ICA and gives better performance than ICA. Furthermore, spectral analysis and filtering were done using FFT, and two Butterworth bandpass filters respectively. Finally, HR and RR were extracted using a peak detection algorithm. To obtain experimental results, they considered 15 subjects with different skin tones in four different scenarios such as without movement, different facial expressions, talking and different illumination levels. Experimental results show that the proposed method (with and without magnification) achieved better performance than ICA and PCA for all four scenarios. However, the CCA (1.22s) needs higher computational time than ICA (0.86s) and PCA (0.79s). Moreover, in the proposed method, they did not consider higher level of movements such as walking or exercising. In addition, this method is also constrained to limited distance and single subject detection.

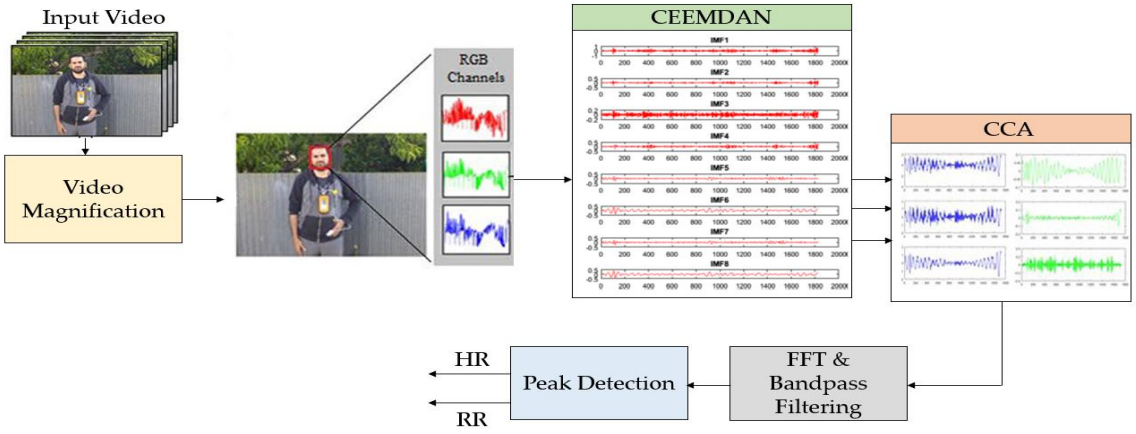


Figure 2.9: System overview of the noise removal technique using UAV [155].

Table 2.7. IPPG-based methods concern about both illumination variations and motion artefacts.

Ref.	Used Sensors	Vital Signs	ROIs	Used Technique	Distance (m)	No. of participants at a time	Results
Li et al. [153]	iSight camera	HR	Face	NLMS adaptive filtering	0.35-0.50 m	1	RMSE = 1.8 bpm
Kumar et al. [154]	Monochrome camera	HR and HRV	Face	Weighted average	0.5 m	1	SNR (dB) = 6.5
Al-Naji et al. [155]	Hovering UAV	HR and RR	Face	CEEMDAN + CCA	3 m	1	PCC= 0.99 RMSE = 0.65 bpm

2.2.4. Alternate Sensors

Rather than using digital cameras or webcams, some researchers have used other sensors to capture PPG data as shown in Table 2.8. For example, Kwon et al. [158] extracted HR exploiting the built-in camera of a smartphone and introduced FaceBEAT which is an application in an iPhone for measuring HR remotely. Al-Naji et al. [159] presented a robust method to monitor cardiorespiratory signal remotely from the video taken by a hovering UAV as shown in Figure 2.10. They used an improved video magnification technique, an intensity-based method and advanced signal processing methods such as signal decomposition based on complete EEMD and blind source separation based on ICA to eliminate noise artefacts. Nevertheless, only slow and small movements were considered, they did not address low lighting either. To extract vital signs Mcduff et. al [160] used

DSLR camera with five bands (red, green, blue, cyan and orange) considering only stationary subjects both at rest and under cognitive stress and reported that CGO is better than RGB. However, this method is susceptible to noise artefacts and is not suitable for real time application. In [161], a comparison between CMOS camera and a webcam was done to extract HR under cycling exercise and reported HR values were independent of the measurement method. However, the proposed method is applicable to very limited rnage of 0.2-0.35m. Blanik et al. [73] combined both a CCD camera and a thermal camera to monitor a broad range of physiological signs. Using a Kinect device, Bernacchia et al. [162] measured HR and RR by means of spatial averaging and ICA. Smilkstein et al. [163] and Gambi et al. [164] used Microsoft Kinect to extract HR, exploiting the EVM technique based on RGB signals.



Figure 2.10: Data collection using the UAV (3DR solo drone) [159].

2.2.5. Different Subjects

Most of the work considered adult, healthy subjects, and limited number of babies to measure vital signs. However, there are some researchers who considered infants in NICU as described in Table 2.8. Scalise et al. [165] proposed a system to measure heart rate using a web camera in a NICU using an intensity-based method and ICA. However, this method is susceptible to noise artefacts such as motion and illumination variations and range is very short at only 20cm. Another work [18] by Arts monitored HR of newborn infants in the NICU using a digital camera over up to 1m range exploiting ambient light. Still, this method is not free from motion artefacts, illumination variation and the skin tone problem. To suppress illumination variation and moderate motion artefacts, Cobos-Torres et al.

[166] designed a computationally efficient method based on numerical analysis techniques and filtering using a digital camera to measure HR and RR of preterm infants in NICU. Nevertheless, the proposed method is affected by strong motion, poor lighting, and shadow. Gibson et al. [167] introduced a remote monitoring system to monitor HR and RR of infants in NICU using digital camera based on video magnification and compared their results with ECG data. Their system was also able to detect an apnoea event in a clinical setting.

2.2.6. Different Vital Signs

Almost all studies aim to extract heart rate and respiratory rate. Few researchers consider other vital signs such as heart rate variability and oxygen saturation as shown in Table 2.8. To extract blood oxygen saturation (SpO_2) along with HR and RR, Tarassenko et al. [150] presented a new technique based on autoregressive (AR) modelling and pole cancellation in order to cancel out the aliased frequency components caused by artificial light flicker which enhanced the robustness of the method under strong fluorescent lights. By means of a webcam, another robust method to calculate SpO_2 and average HR was introduced by Bal [168] using a noise removal algorithm based on dual tree complex wavelet transform (DTCWT) to rectify the artefacts caused by movement and artificial lighting. However, the range was very short, at only 50cm.

2.2.7. Multiple ROIs

Almost all of the works, considered limited ROI particularly face for the vital sign extraction. Some studies used face and cheeks [18], [169], face and forehead [129], cheeks and forehead [150], nose, forehead and mouth [128], as ROIs to measure physiological signals. Another study by Datcu et al. [170] considered 10 different parts of the face as ROIs. To extract vital signs, Yu et al. [133] and Feng et al. [171] used both the face and the palm. Another study by Bernacchia et al. [162] considered the neck, thorax and abdominal area to calculate HR and RR. Zhao et al. [172] extracted physiological signals from face, arm and hand using a webcam considering two subjects simultaneously under stationary conditions. However, this method is susceptible to motion artefacts and was only tested over short ranges. Al-Naji et al. [173] measured cardiopulmonary signals from various regions of the human body such as the face, palm, wrist, arm, neck, leg, forehead, head and chest under stationary scenarios using a digital camera. They proposed a noise elimination technique based on EEMD and ICA. They used various methods such as intensity, frame subtraction and feature tracking to extract cardiopulmonary signals by

considering skin colour variation, chest motion and head motion. Nevertheless, the proposed methods suffer from various issues such as limited range, a single subject at a time and noise artefacts.

2.2.8. Long Distance and Multiple Persons Simultaneously

Most of the previous methods discussed above have some limitations. Firstly, most of the works considered very short ranges, the highest distance considered was only 3m. Secondly, except for Pho et al. (three subjects) and Zhao et al. (two subjects) all other methods measured vital signs for one subject at a time. In [174] a robust contactless method was proposed to calculate HR and RR using both iPPG and head motion from the videos taken by both a digital camera and a hovering UAV considering long ranges of up to 60m, multiple subjects in groups of up to six people simultaneously under both stationary and non-stationary circumstances using a video magnification technique. To eliminate noise artefacts caused by movement of subjects and camera, variations of illumination and variation of skin tone, they used a noise elimination technique using both CEEMDAN and CCA. Furthermore, FFT was used for spectral analysis and two Butterworth bandpass filters were used for filtering. Finally, HR and RR were extracted by calculating number of peaks using MATLAB built-in function '*findpeaks*'. Moreover, they introduced a graphical user interface (GUI) that facilitates a user for loading video data, selecting the magnification type, and executing the proposed system and configurations.

The experimental setup and data acquisition of the proposed system is presented in Figure 2.11 where they considered three groups of people consisting of 15, 20 and 10 subjects with different skin tone respectively under three assumptions: noise artefacts, multiple detection, and long distance. The proposed method achieved better performance than ICA and PCA for both stationary and nonstationary scenarios with multiple subject detection and long distance respectively. However, the system used limited ROI, observing only the face, and will be affected by unclear ROI.

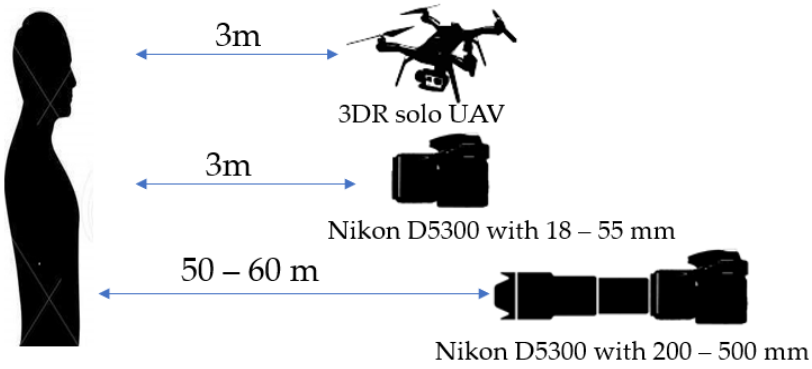


Figure 2.11: The experimental setup and data acquisition of the noise artefact removal, multiple people, and long detection system [174].

2.2.9. Others

Based on machine learning, Hsu et al. [175] presented a novel method to extract HR using a video camera where the PPG signal was recovered from either ICA or chrominance-based methods and may be enhanced by utilizing the mid-level PPG-based features exploiting support vector regression (SVR). However, they only considered the stationary case. To integrate colour-based and motion-based approaches, Weidi et al. [176] introduced two fusion techniques, mean method and the ratio of variations method, to extract HR using an industrial camera. They reported that a colour-based method is more reliable than a motion-based method. However, the accuracy of the fusion-based method was not satisfactory, and this method was confined to rehabilitation only.

Table 2.8. Summary of different iPPG-based methods.

Ref.	Used Sensors	Vital Signs	ROIs	Used Technique	Distance (m)	No. of participants at a time	Results
Kwon et al. [158]	Smartphone	HR	Face	ICA	0.3 m	1	MAE = 1.47 bpm
Al-Naji et al. [159]	UAV	HR and RR	Face	CEEMD+ICA	3 m	1	PCC = 0.99, RMSE= 0.7 bpm
Mcduff et. al [160]	Digital camera with five bands	HR, RR, and HRV	Face	ICA	3 m	1	PCC=0.92
Sun et al. [161]	CMOS, Webcam	HR	Face	TFR	0.2-0.35 m	1	PCC = 0.85
Bernacchia et al. [162]	Microsoft Kinect	HR and RR	Neck, thorax and	ICA	-	1	PCC=0.91

			abdominal area				
Smilkstein et al. [163]	Microsoft Kinect	HR	Face	EVM	-	1	-
Gambi et al. [164]	Microsoft Kinect	HR	Forehead, cheeks, neck,	EVM	-	1	RMSE = 2.2 bpm
Scalise et al. [165]	Webcam	HR	Forehead	ICA	20 cm	1	PCC = 0.94
Arts et al. [18]	Digital camera	HR	Face and Cheek	JFTD	1 m	1	-
Cobos-Torres et al. [166]	Digital camera	HR	Abdominal area	Stack FIFO	50 cm	1	PCC=0.94
Gibson et al. [167]	Digital camera	HR and RR	Face and chest	EVM	1-2 m	1	Mean bias = 4.5 bpm
Bal [168]	Webcam	HR and SpO_2	Face	DTCWT	50 cm	1	PCC=0.92, RMSE=2.05 bpm
Tarassenko et al. [150]	Digital camera	HR, RR, SpO_2	Forehead and cheek	AR modelling and pole cancellation	1 m	1	MAE = 3 bpm
Foud et al. [169]	Webcam	HR	Face and cheeks	ICA	1-2 m	1	RMSE = 2.7 bpm
Datcu et al. [170]	Video camera	HR	10 different parts of face	ICA	-	1	RMSE = 1.47 bpm
Feng et al. [171]	Webcam	HR	Face and palm	Block division	0.75 m	1	PCC=0.96
Zhao et al. [172]	Webcam	HR	Face, arm, hand	POSCC	1.5 m	2	SNR (dB) = 4.5
Al-Naji et al. [173]	Digital camera	HR and RR	Face, palm, wrist, arm, neck, leg, forehead, head and chest	EEMD + ICA	2 m	1	PCC = 0.96, RMSE=3.52
Al-Naji et al. [174]	Digital camera, UAV	HR and RR	Face and Forehead	CEEMDAN+CCA	60 m	6	PCC=0.99, RMSE= 0.89 bpm
Hsu et al. [175]	Video camera	HR	Face	SVR	-	1	PCC=0.88, RMSE = 5.48 bpm
Weidi et al. [176]	Industrial camera	HR	Forehead	ICA	2 m	1	RMSE= 4.83 bpm

2.3. Applications of Video Camera Imaging Technology

As camera imaging methods overcome various limitations of contact sensors especially related to sensitive skin damage or infection, such contactless vital sign monitoring

methods have promised to be realistic and attractive to numerous potential applications in both clinical and nonclinical scenarios.

2.3.1. Clinical Applications

Because of its contactless sensing mode, camera imaging can offer a comfortable method to monitor elderly people, infants, and patients with chronic pain, burnt skin, under dialysis, during and after surgery.

Neonatal Monitoring

Neonates and infants have very delicate and sensitive skin, so non-contact methods would be the preferred approach to monitoring them. Several studies used iPPG methods to monitor neonates. For example, Klaessens et al. [177] introduced a baby-friendly non-contact approach to monitor several vital signs like HR, RR, skin temperature, and SpO_2 of infants and neonates in the NICU with gestational ages of 24 to 39 weeks. To measure HR and RR, RGB colour magnification and IR-thermography were used respectively. Another study by Villario et al. [178] applied iPPG technique to continuously monitor HR, RR, and SpO_2 of neonates and infants in the NICU. They also detected bradycardia while monitoring participants continuously for long periods. Other studies [18], [165-167] also proved the promise of iPPG in neonatal monitoring.

Assessing Patients with Chronic Pain

Rubins et al. [179] used imaging photoplethysmography to assess patients with chronic pain particularly neuropathic pain using a monochrome camera. Another study by Zaproudina et al. [180] also employed the iPPG method to find new biomarkers of migraine patients using a monochrome camera.

Critical Patient Monitoring

Rasche et al. [181] used the iPPG technique in patients after heart surgery using a mobile camera under critical care conditions. In their proposed method, ROI was selected manually and optical flow techniques were used to detect motion artefacts. Moreover, high pass filter and Fourier spectra were employed to calculate HR. Nevertheless, their method is affected by patient's movements and illumination variations.

Arrythmia Detection

An abnormal heart rhythm is called an arrhythmia. Amelard et al. [182] applied iPPG technique to detect arrhythmia using a monochrome camera based on a signal fusion technique. The signal fusion technique was used to extract the blood pulse signal using prior information. They formulated the problem based on a Bayesian problem and modeled a novel probabilistic pulsatility model which combined spectral and spatial priors derived from physiology of the blood pulse signal.

Anesthesia Monitoring

Rubins et al. [183] introduced the concept that a PPG imaging system might be employed to continuously monitor regional anesthesia by monitoring skin microcirculation. Using both RGB camera and a near-infrared camera, Trumpp et al. [184] applied camera-based PPG to monitor patients of cardiovascular disease during surgery in an intraoperative environment. It can help anesthetists to take measurements in response to any cardiovascular events and to administer appropriate medication.

Monitoring Dialysis Patient

Villarroel et al. [185] extracted HR, RR and SpO_2 of patients going through haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford, UK based on iPPG using a high-quality 5 megapixel camera. Tarassenko et al. [150] also applied iPPG to monitor HR, RR, and SpO_2 of hemodialysis patients in the Oxford Kidney Unit.

Burn Care

Thatcher et al. [186] introduced that PPG imaging can be used to identify the correct depth of burn excision as there was significantly less blood circulation in burnt tissue. Consequently, it can help surgeons by giving an idea of where and how much to resect. Thus, iPPG helps to improve burn care as well.

2.3.2. Non-Clinical Applications

In nonclinical sectors such as home health care [162], fitness monitoring[133, 137, 138, 141, 144, 151], sleep monitoring [163], polygraph [5], living skin detection [187], stress monitoring [188], driver monitoring [189], security [188], war zone or natural calamity [174], animal research [190], the video camera imaging may play an important future role by monitoring vital signs.

2.4. Technologies for Jaundice Detection

Jaundice or hyperbilirubinemia is indicated by a yellow discolouration of the skin and sclera of the eyes due to an excess level of bilirubin [33]. Generally, jaundice is noticeable when serum bilirubin level exceeds 2.0 mg/dl in the blood [191]. Bilirubin is a water-soluble tetrapyrrolic yellowish pigment that is present in the blood and whose excess accumulation in the skin results in neonatal jaundice symptoms [192]. Bilirubin is created due to the breakdown of old red blood cells. In the human body, new red blood cells are produced, and old ones are broken down continuously. In an adult, the red blood cells survive for about 120 days; however, in a newborn infant, they survive for a significantly shorter time. Hence, newborns have higher than average quantities of red blood cells, which leads to excess bilirubin level due to the breakdown of more red blood cells [192]. Normally, the damaged blood cells that produce bilirubin are metabolised by the liver for excretion. Later, bilirubin is secreted through urine and bile [34]. Short-term excess of bilirubin is mostly harmless and self-limiting. But a high level of bilirubin in newborn infants is neurotoxic and can permanently damage the brain, which is called kernicterus. It may cause cerebral palsy, deafness or hearing loss, language difficulties, and developmental delay or can be fatal in the worst cases [34, 35].

Kernicterus can be avoided by early detection and treatment of jaundice or hyperbilirubinemia. High bilirubin can be controlled through phototherapy, a method that requires exposure of the affected newborn infant to particular wavelengths of blue light that transform bilirubin into a harmless, excretable form [193, 194]. On the other hand, excess bilirubin may need to be removed through exchange transfusions (ETs) in cases of extremely high levels [195]. However, ET is an invasive technique where a calculated amount of the newborn's blood is replaced by a donor's blood to decrease the excess level of bilirubin. Therefore, phototherapy has been the most popular method for the treatment of jaundice or hyperbilirubinemia [30, 196].

For detecting jaundice, there are different invasive and non-invasive approaches [33] as follows.

2.4.1. Invasive Approaches

Total serum bilirubin (TSB) measurement is a common invasive technique to determine the level of hyperbilirubinemia that works based on spectrophotometry using a small blood

sample taken by a syringe and needle [36]. However, the TSB measurement can be performed only by medical experts with good facilities. Moreover, as it is an invasive technique, this procedure is stressful for both infants and parents [37]. For the diagnosis of jaundice, there are some other techniques that involve the collection of blood samples. Nevertheless, all invasive procedures are uncomfortable and painful for the infant, have a risk of infections, may cause anaemia, and will create anxiety for parents [38].

2.4.2. Non-Invasive Approaches

Visual assessment is the simplest, non-invasive approach to identify hyperbilirubinemia in newborns [39]. However, as this process is operator-dependent and may be affected by the skin colour of the infant, this process is not reliable for detecting jaundice [41, 42]. On the other hand, transcutaneous bilirubinometry (TcB) is a non-invasive alternative technique to TSB measurement, which can reduce blood sampling by approximately 30% [40]. Still, TcB techniques are expensive as they require products that have a huge recurrent cost. Moreover, The TcB test can be performed on babies who are 35 weeks or more of gestation age and 24 hours after birth [197]. Therefore, in the case of infants who are born before 35 weeks of gestation as well as in the first 24 hours after birth, TcB is ineffective.

With the development of new advanced technologies such as smartphones and digital cameras, there is the possibility of a more efficient, cost-effective, and portable system to monitor jaundice. These new technologies may perform equivalently to existing monitoring tools, while being affordable and available [198]. Most of these methods have used image processing and colour analysis for jaundice detection [199].

In the literature, jaundice detection systems were proposed based on computer vision and image processing techniques [200]. Different systems used different objects, such as skin colour [201], sclera colour [202], urine strips [203] and stool [204] for the colour analysis process. However, using urine strips and stool is not feasible as it is difficult to collect urine and stool on demand. Moreover, these systems are not reliable and accurate as the colour of urine and stool may be affected by food [200].

2.4.2.1. Computer Vision-Based Approaches Using Sclera Colour

Some researchers have considered sclera colour to detect jaundice. For example, Laddi et al. [202] proposed a non-invasive, low cost and instant jaundice detection method using the images of sclera of the eyes captured by a 3CCD colour camera. The authors used

multiple image-processing techniques, such as image segmentation, image morphology, and colour analysis. PCA was used to differentiate between jaundice and non-jaundice patients. Finally, an adaptive neuro-fuzzy inference system (ANFIS) algorithm was incorporated to measure jaundice and its degree of severity. However, the eyes of the babies remain closed most of the time, and to take the image of sclera, the head needed to be in a fixed position which is difficult to attain with infants. Hence, the proposed approach is more appropriate for adults rather than infants. Moreover, the accuracy of the proposed method depends on the training data to a great extent. Miah et al. [205] introduced another non-invasive system for jaundice detection and bilirubin level quantification by capturing the image of the sclera using a webcam and a controllable LED flashlight. They selected ROI automatically using colour thresholding and bounding box algorithms. Furthermore, to estimate the bilirubin level, different machine learning-based regressions such as decision tree regression, linear regression, k-NN regression, SVM regression, and RF regression were used. Among all, RF regression gave the best results, with an absolute mean error of 0.29. The proposed technique was able to identify jaundice and deliver results within 5 minutes. Furthermore, to operate the system, no trained person was required. However, the system required the subject to wear specially made goggles and a controlled lighting system, and it was only suitable for adults.

For detecting neonatal jaundice, using sclera colour is not feasible as it is difficult to capture the image of an infant's eyes due to the amount of time, they spend sleeping and their unpredictable movements. Furthermore, for capturing the image of the sclera, participants need to maintain a controlled posture which is not possible for infants. Moreover, some infants' sclera may have a blue hue as they are thinner than adult sclera.

2.4.2.2. Computer Vision-Based Approaches Using Skin Colour

Skin colour is one of the common indicators that has been used for detecting jaundice, especially in the case of neonates [206]. For instance, Aune et al. [207] presented a physics-based system to measure bilirubin from digital images of newborn babies using a smartphone and a colour calibration card. In the proposed system, the authors undertook a combination of colour analysis of digital photographs and a physics-based model of transportation of light in the skin. A library of simulated reflectance spectra of skin was created by means of a mathematical model using diffusion theory. However, they mainly considered Caucasian-term infants as participants in their study and used manual ROI selection.

Hashim et al. [201, 208] developed a diagnosis and treatment process for neonatal jaundice based on infants' skin colour analysis using a digital camera and MATLAB graphical user interface. The system proposed in [201] has two key sections: skin colour analysis and the practical circuit. The first one included manual ROI selection and colour range selection. On the other hand, the practical circuit was designed employing an Arduino Uno Microcontroller that was linked with the MATLAB GUI panel. However, in the proposed method [201], the ROI was selected manually, and the sample size was very small, as they considered only two infants. Moreover, they only considered controlled lighting. In [208], the authors considered 20 infants with varying skin tones, different lighting conditions and camera distance. In the proposed system, different image processing techniques were considered, such as skin detection based on colour space transformation and threshold value, automatic ROI selection using a bounding box algorithm, skin colour analysis in RGB and YC_bC_r colour spaces. The values of the B and Cb channels were compared with the predefined threshold value in order to determine whether the neonate had jaundice or not. The system provided the treatment of jaundice as well by employing an Arduino Uno microcontroller which controlled a blue LED light that provided phototherapy. However, the selected ROI included background pixels which may affect the results. Moreover, they only considered values of the B and C_b channel to detect jaundice.

Kawano et al. [209] proposed a non-contact and non-invasive neonatal jaundice detection method using a video camera and 8-colour calibration card based on skin colour analysis. They first detected the face automatically by using a cascaded object detector and then extracted the face skin region by removing the feature map, which was a combination of a mouth map and eye map from the detected face region. Two image features were used to detect jaundice: the mean value of skin colour information and eigenvalues of the variance-covariance matrix. However, the proposed skin detection technique could not extract skin regions clearly, as eyes and eyebrows might not be separated using an eye map. As a result, there were some failure cases, especially in the case of premature babies. Therefore, local areas, such as the forehead and cheek in the face, need to be detected to get better results. Moreover, the calibration card needed to be discarded after every use and changed to a new one before every use to prevent the possibility of infection of the infant, which requires an additional consumable expense.

Machine learning has great potential in medical applications, especially in detecting jaundice and measuring the level of bilirubin. For example, Rong et al. [210] introduced

an automated image-based (AIB) system that used an automated smartphone and a colour calibration card to detect jaundice. The images of skin taken by the smartphone had been uploaded to a cloud server that used machine learning-based software to deliver an immediate and individualized estimated bilirubin index. The results were compared with TSB and TcB for both term and preterm babies and showed a good correlation and strong relevance between AIB and TSB. However, the specificity and sensitivity of the proposed scheme were perceived to be not good enough. Moreover, they used manual ROI selection.

Chakraborty et al. [211] proposed a non-invasive system for detecting neonatal jaundice based on the yellowness of skin using a Raspberry Pi microcontroller and Pi camera. They used different image processing techniques and Convolutional Neural Networks (CNN) to detect jaundice. However, CNN has high computational complexity and is thus time-consuming. Castro-Ramos et al. [212] introduced the detection of jaundice in adults or newborns using digital images from a smartphone. They considered RGB attributes as well as diffuse reflectance spectra acquired from a spectrometer as the factor to distinguish whether subjects had jaundice or not. A SVM was used to distinguish between sick and healthy subjects. However, they only considered Mexican patients as their subject.

Juliastuti et al.[213] estimated the risk zone of jaundice in newborn infants based on skin colour analysis using a digital camera. First, the image of a baby was filtered and colour-corrected using a median filter and a colour card, respectively. Then, baby skin was obtained by segmenting the image using a K-Means segmentation method. Finally, bilirubin level was estimated using multivariable linear regression. The proposed method was simple and had good correlation with TSB; however, the error percentage and accuracy of the system were only 17% and 67%, respectively. Moreover, they selected the ROI manually.

Greef et al. [214] developed BiliCam, a new non-invasive and low-cost system for monitoring jaundice in newborns using a smartphone camera (iPhone 4S) based on skin colour. The algorithm includes colour balancing, feature extraction and different machine learning-based regressions such as least angle regression (LARS), LARS-Lasso Elastic Net, k-NN regression, SVR, and RF regression. However, in the proposed system, a colour calibration card was required to be placed on the neonate's forehead and sternum before capturing the images. Moreover, they used manual ROI selection. The technique may also be constrained by skin tone especially darker complexion, body position and unclear ROI.

In another study, Mansor et al. [215] proposed a non-invasive method of detecting the jaundice of newborns using PCA)based feature extraction and k-NN algorithm. They used the Intel Open CV cascade Classifier developed by Viola-Jones [216] to detect the face. However, the detected face also contains background pixels which may affect the results of jaundice detection.

2.5. Research Gaps and Limitations

During the past few years, many advances have been made using camera imaging-based methods and numerous studies have been published in this field. However, the existing studies have several limitations that need to be overcome in the future.

- Most of the current studies focus on either motion artefacts or illumination variations. However, in real time scenarios both motion and illumination play an important role in degrading accuracy and usability. Only a few researchers have given solutions to both motion artefacts and illumination variations. Therefore, it can be a further topic of research for new researchers to develop a method which can deal with both motion and illumination artefacts.
- Still researchers are mainly interested to extract HR and RR. Nevertheless, blood oxygen also plays an important role in assessing a human's health condition. Blood glucose [217-219] is another important vital sign which helps to identify and maintain the welfare of diabetes patients. Therefore, more research needs to be done to monitor blood oxygen saturation and blood glucose.
- All of the studies except Pho et al. (3 subjects), Zhao et al. (2 subjects) and Al-Naji et al. (6 subjects), considered a single subject at a time when monitoring vital signs. Detecting multiple persons simultaneously is a major issue of current studies that need to be overcome in future.
- Distance between camera and subject is another challenge for current researchers. All the researchers except Al-Naji et al. (60 m) considered very small distances while monitoring vital signs.
- The number of ROI and selecting ROI are important issues to overcome as very few researchers have paid attention to these problems well enough for industrial or commercial use. Most of the studies used manual processes while some used automatic methods to select from a very limited number of ROI in a limited

scenario. However, advanced techniques are required to automatically select multiple ROIs.

- Researchers mainly considered healthy and young participants as their subjects to do the experiments and ignored patients, elderly people, and premature babies. More research needs to be done considering different subjects such as premature babies and elderly people.
- Most of the existing works considered privately owned databases. Only a few used publicly available databases like MAHNOB-HCI (Human computer interaction) or DEAP (Database for Emotion Analysis using Physiological Signals). Lack of publicly available datasets taken under a realistic situation is another challenge to deal with in future.
- To validate proposed methods, researchers mainly use a pulse oximeter as a ground truth. However, ECG was used by very few researchers as a ground truth. There are indications that no commercial instrumentation is truly accurate and most are simply accepted as accurate [167].
- Future research could also include multi-camera fusion and new non-visible light sensors to tackle the visible light camera's shortcomings.

2.6. Concluding Remarks

In conclusion, each study mentioned above, however, had its advantages and disadvantages based on different assumptions, leading to several issues to be considered. This chapter gives a thorough review of recent works in non-contact methods and identifies the research gaps and challenges of the existing studies and provides some direction for future research.

CHAPTER 3

Research Methodology and Techniques

3 Research Methodology and Techniques

This chapter outlines the methodology and techniques utilized to conduct the research presented in this thesis, addressing the relevant issues of the existing non-contact imaging measurement technologies in terms of the subject's safety, the system cost, the ROI selection, noise artefacts, the complexity of dealing different subjects such as premature babies and multi-camera fusion.

This chapter is structured in seven main sections. Section 3.1 presents a general introduction of the research methodology and related techniques used in this thesis. Section 3.2 describes the research design of the proposed non-contact monitoring system. Section 3.3 presents the data collection procedures, including details of the research ethics, participants, video data acquisition and validation. Section 3.4 presents the image and signal processing techniques and machine learning algorithms used in data processing and analysis. Section 3.5 presents the software and related tools that were used. Section 3.6 describes the statistics used for result validation purposes. Finally, Section 3.7 presents the concluding remarks.

3.1. Introduction

Most of the reviewed non-contact technologies described in the previous chapter are promising for extracting vital signs and detecting jaundice. However, each non-contact technology has pros and cons based on different assumptions, resulting in several challenges that need to be taken into account. These challenges can be summarized as issues related to possible biological effects, the subject's safety, the system cost, the ROI selection, noise artefacts, the complexity of dealing with different subjects such as premature babies and multi-camera fusion. Computer system-based camera imaging stands out as a highly potential and feasible research design for extracting vital signs and detecting jaundice employing different image and signal processing techniques, in which the above challenges are the main considerations.

This chapter presents the methodology and techniques used in this thesis to address the research questions by providing descriptions of the study design, data collection procedures, data processing and analysis, software tools used, and the statistical approaches applied.

3.2. Research Design

A key task of the research presented in this thesis was to utilize image and signal processing techniques to extract and measure vital signs from the physical and physiological variations caused by cardiorespiratory activity, using a computer system based on camera imaging. Figure 3.1 shows the basic framework, which includes data acquisition, ROI selection, raw signal extraction, signal decomposition, spectral analysis, temporal filtering, and peak detection.

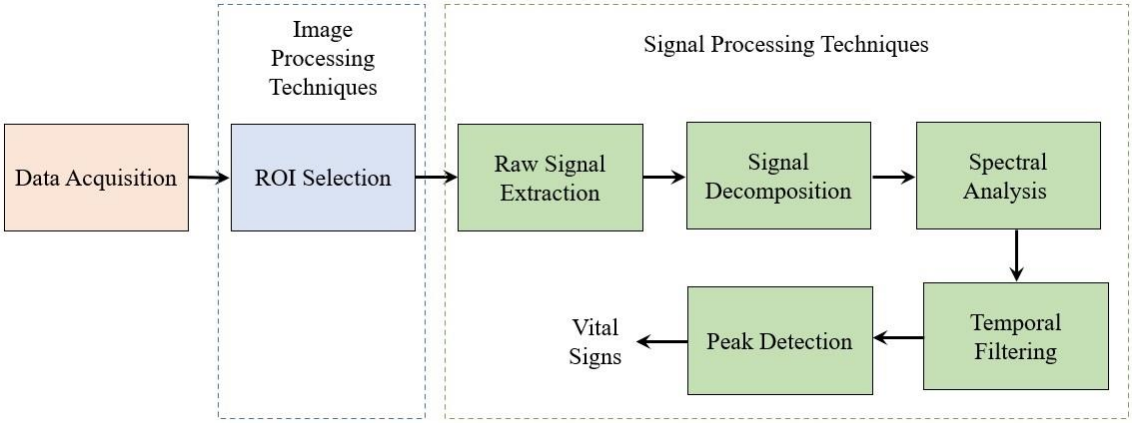


Figure 3.1: Basic block diagram of the proposed non-contact monitoring system.

3.3. Data Collection

3.3.1. Research Ethics

The research procedure outlined in this thesis strictly adhered to the ethical principles laid out in the Declaration of Helsinki (Finland 1964). Moreover, all procedures involving human subjects received approval from the University of South Australia (UniSA), Human Research Ethics Committee (HREC) (protocol number: 0000034901, Protocol number:202973.) and Southern Adelaide Local Network Research Committee (HREC/17/SAC/340; SSA/17/SAC/341).

Prior to the commencement of the experiment, all participants (or their legal guardians, in the case of infants) were thoroughly informed about the research procedure. Subsequently, written informed consent was obtained from all adult participants and guardians of infants, ensuring that they were fully aware of the study's objectives and agreed to participate.

To ensure asset protection and maintain the privacy of participants, all video data collected during the research were securely stored on a computer at UniSA and these video data will not be made available for public use.

3.3.2. Participants

For vital sign monitoring, a group of 18 participants with different ages and different skin tones participated in the research experiments. The participants were divided into two groups according to their ages:

- the first group: 8 infants (males and females, <1 year of age).
- the second group: 10 adults (males and females, over 18 years of age).

For jaundice detection, two different datasets were used in the experiment. The first data was obtained by creating 20 images of 3D babies with varying skin tones. These images were used as an alternative to real babies. The second data set contains 30 images of babies taken from the internet; 20 of these images are for healthy infants without jaundice, and the other 10 images are for infants with jaundice. Collecting our own jaundice dataset from real patients was challenging and time-consuming. Therefore, some suitable images from the internet were collected for this study.

3.3.3. Video Data Acquisition

Various camera sensors were used in the research presented in this thesis to capture video data, and are summarized as follows:

- A digital camera (Nikon D610 digital single-lens reflex [DSLR]) with a resolution 1920×1080 and a frame rate of 30fps was used.
- Another digital camera (Nikon D5300 DSLR) with a resolution 1920×1080 and a frame rate of 30fps was used.
- A FLIR A400 camera which had both thermal and RGB sensor options. The thermal sensor had a resolution of 320 × 240 and a frame rate of 30fps. The RGB camera had a resolution of 640 × 420 and a frame rate of 30fps.

3.3.4. Validation

For validation purposes, different equipment and tools were used as reference measurement methods in this thesis to gather physiological data, as follows:

- ECG was used as the reference for validation purposes. For extracting HR and RR, the difference in electrical impedance and chest wall movement is measured by the impedance lead of the ECG, respectively. ECG is constrained to patient movement and cardiac activity [220]; however, it was used for validation purposes to avoid interruption.

- A Go Direct® Respiration Belt (Vernier Software & Technology, Beaverton, Oregon, United States) was also employed as the reference measurement in the study [221]. During respiration, the force exerted by the abdomen or chest is measured by the respiration belt with a resolution of 0.01 N, range of 0-50N and response time of 50 ms.

3.4. Data Processing and Analysis

The focus in this thesis was to achieve a safe, cost-effective, fast, and reliable non-contact monitoring system using camera imaging-based technologies that can be employed to detect cardiorespiratory activity in both clinical and non-clinical environments, using the developed techniques for image and signal processing and analysis.

3.4.1. Image Processing Techniques

Different image and video processing techniques were applied in this thesis to analyse and extract useful features of interest from video data. These techniques included a description of procedures for colour space transformation and ROI selection.

3.4.1.1. Image Colour Space

Colour space is a specific representation that can provide an informative colour representation of digital colour images. Colour is a visual perception created by light; hence, any incident spectrum can be perceived by the human eye as a combination of red, green, and blue, which are known as the three ‘primary colours’. Most digital cameras capture images in the RGB colour space, which is considered a fundamental and commonly used colour space. However, the RGB colour space is sometimes not suitable to analyse

colour images because of the high correlation among its components and difficulty to evaluate the similarity of the components from their distance in colour space [222, 223].

RGB Colour Space

The R, G, and B components of a colour image can be presented by the brightness or intensity values of each pixel in the image using the following equations [222, 224]:

$$R = \int_{\lambda 1}^{\lambda 2} E(\lambda) F_r(\lambda)\, d\lambda \tag{3.1}$$

$$G = \int_{\lambda 1}^{\lambda 2} E(\lambda) F_g(\lambda)\, d\lambda \tag{3.2}$$

$$B = \int_{\lambda 1}^{\lambda 2} E(\lambda) F_b(\lambda)\, d\lambda \tag{3.3}$$

where F_r, F_g and F_b are the colour filters on the incoming light $E(\lambda)$, and λ is the wavelength and $\lambda 1 = 300$ and $\lambda 2 = 830$nm [224]. From the RGB representation, other types of colour representations (such as YIQ and YC_bC_r colour spaces) can be derived by using different transformations to separate the independent characteristics from the components, thereby providing important parameters to analyse digital colour images. Selecting the best colour space is still a challenge in colour image processing applications [222].

YIQ Colour Space

YIQ colour space is a linear transformation of the RGB colour space that is used to separate the chrominance information from the luminance information. The YIQ is obtained from the RGB colour space as follows [225]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.2989 & 0.5866 & 0.1144 \\ 0.5959 & -0.2741 & -0.3218 \\ 0.2113 & -0.5227 & 0.3113 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.4}$$

where Y is the luminance component and I and Q are the chrominance components. The YIQ colour space was adopted by the National Television System Committee.

YC_bC_r Colour Space

YC_bC_r is another colour representation with one luminance component (Y) and two chrominance components (C_b and C_r). The YC_bC_r is derived from the RGB components using a linear transformation as follows [225]:

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -39.797 & -74.203 & 112 \\ 112 & -93.786 & -18.2214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (3.5)$$

where Y has a range from 16 to 235 and C_b and C_r have a range from 16 to 240. The colour space transformation from RGB into YIQ and YC_bC_r can be implemented using MATLAB's built–in functions called '`rgb2ntsc`' and '`rgb2ycbcr`'.

LAB Colour Space

The LAB colour space is a three-dimensional colour model that was developed to mimic human vision and provide a more perceptually uniform colour representation [226]. It is widely used in various fields such as colour science, image processing, and computer vision.

The LAB colour space consists of three components: L, A, and B. Here's a brief description of each component:

L (Lightness): It represents the brightness of the colour. The L component ranges from 0 (darkest black) to 100 (brightest white). Values in between represent various shades of gray.

A (Green-Red): It represents the colour along the green-red axis. The A component ranges from negative to positive values. Negative values represent green shades, and positive values represent red shades. The absolute value indicates the intensity of the colour.

B (Blue-Yellow): Represents the colour along the blue-yellow axis. Like the A component, the B component ranges from negative to positive values. Negative values represent blue shades, and positive values represent yellow shades. The absolute value indicates the intensity of the colour.

The LAB colour space has several advantages over other colour models, such as RGB or CMYK. One significant advantage is its perceptual uniformity, which means that a given numerical difference in LAB values corresponds to a roughly equal perceived difference in colour by the human eye. This makes it suitable for colour analysis, image editing, and colour management tasks.

3.4.1.2. ROI Selection

ROI selection is a basic image processing technique to automatically detect and locate a ROI in digital images. To develop any real-time system, a robust, fast, and efficient detection method is required for ROI analysis in different scenarios. Four ROI detection methods have been used in this thesis based on image/video decomposition, skin detection, YOLO, and facial landmark detection.

Image/Video Decomposition

Image decomposition is a process where an image is divided into several smaller tiles or grids. Image decomposition is often used as a preprocessing step to identify potential ROIs that may contain human faces and limbs. This method is explained in detail in chapter 4.

Skin Detection

Skin detection is an important preprocessing step in various computer vision applications, especially in tasks related to human detection, face recognition, and gesture recognition. The main goal of skin detection is to identify skin-coloured pixels or regions within an image or video. It's worth noting that skin detection is often used as a preprocessing step to identify potential ROIs that may contain human faces and limbs. After skin detection, more sophisticated techniques, such as face detection, object detection, or segmentation, can be applied to further analyse and process the regions of interest for specific computer vision tasks.

Several computer vision approaches have been developed for skin detection, and these methods can broadly be categorized into colour-based and machine learning-based techniques [227].

Colour-based Approaches: These methods rely on the observation that skin pixels typically exhibit specific colour characteristics in different colour spaces, such as RGB, HSV (Hue, Saturation, Value), YC_bC_r, LAB etc. The colour-based methods set predefined thresholds or ranges for each colour channel to determine whether a pixel is likely to be skin-coloured or not. For example, skin tones in the YC_bC_r colour space tend to have specific ranges for C_b and C_r values, which can be used to identify skin pixels.

Machine Learning-based Approaches: These methods involve training machine learning models, such as Support Vector Machines, Random Forests, Decision Trees, or Neural Networks, on a labelled dataset containing skin and non-skin samples. The model learns to

classify pixels based on their feature representations and makes predictions on unseen data to identify skin pixels. The features used can include colour information, texture, and spatial information. The colour-based technique for skin detection was used in this thesis. The skin detection technique for selecting an accurate ROI is explained in detail in chapter 4.

YOLO (You Only Look Once)

The YOLO algorithm [228] was inspired by the human ability to quickly process visual information and understand the context and semantics of their surroundings by just glancing at them. The main idea behind YOLO is to reframe object detection as a regression task rather than a traditional two-step process of region proposal and classification. In YOLO, the algorithm takes an image as input and directly predicts bounding boxes along with their corresponding class labels and probabilities. This means that the algorithm "sees" the entire image at once and processes it in a single pass, hence the name "You Only Look Once." The bounding box output consists of the coordinates that define the location of the object in the image, the class label representing the detected object type (e.g., car, dog, person), and the confidence score indicating the certainty of the presence of that object within the bounding box.

The YOLO algorithm's key advantages are its ability to capture contextual information in images without relying on spatial segmentation, and its enhanced generalization capability compared to earlier object detection approaches. This makes YOLO faster and more accurate for real-time object detection tasks, making it a popular choice in various computer vision applications.

In the YOLO algorithm [228], the input image is divided into an S×S grid, and each grid cell is responsible to detect objects that fall within its spatial region. For each grid cell, the algorithm predicts multiple bounding boxes and confidence scores. The output tensor has dimensions of S×S× (B×5 + C), where:

S is the number of grid cells along each dimension of the image.

B is the number of bounding boxes that each grid cell can predict.

5 represents the five elements of each bounding box prediction (x, y, width, height, and confidence score).

C is the number of classes being considered for object detection.

During training and evaluation, an essential parameter used for object detection is the Intersection over Union (IoU). IoU is a metric that measures the overlap between the predicted bounding box and the ground truth (actual) bounding box. It is defined as the ratio of the area of intersection of the predicted box and the ground truth box divided by the area of their union i.e.

$$IoU = \frac{A_{intersection}}{A_{union}} \tag{3.6}$$

An IoU of 1 indicates that the predicted bounding box perfectly matches the ground truth box, while lower values indicate less accurate detections.

In the YOLO algorithm, the confidence score (P) represents the algorithm's certainty that an object is present within a given bounding box. It can be calculated as follows.

$$P = \ Pr_{object} \times IoU \tag{3.7}$$

where, Pr_{object} is the probability of the presence of the object. When an object is detected, the confidence score will be positive (usually ranging from 0 to 1), indicating the probability of the object's presence. However, if no object is detected within a specific grid cell, the confidence value will be set to 0.

Non-Maximum Suppression (NMS) is a technique used in the post-processing stage of object detection to filter out duplicate or redundant bounding box predictions. After the YOLO algorithm generates multiple bounding box predictions for each detected object, NMS is applied to retain only the most accurate bounding box for each object while removing any overlapping or redundant boxes. This process helps improve the accuracy and clarity of the final object detection results. NMS is a critical step in the YOLO algorithm to obtain reliable and non-redundant bounding box predictions during inference. Chapter 5 gives a more detailed explanation about the ROI selection using YOLO.

Facial Landmark Detection

In computer science and computer vision, landmark detection is the process of identifying and locating significant landmarks or key points in an image. These landmarks serve as distinctive reference points for various applications and tasks. Originally, landmark detection was used in navigation and robotics to find key landmarks in an environment,

enabling robots or autonomous systems to navigate and create maps. However, landmark detection methods have been extended and widely applied to various other fields, including facial recognition, medical imaging, and object recognition [229].

In facial recognition and analysis, landmark detection is used to identify key points on a face, such as the corners of the eyes, the tip of the nose, the mouth corners, etc. These facial landmarks play a crucial role in tasks like facial expression analysis, emotion recognition, and face alignment. Detecting facial landmarks typically involves two main steps:

Step 1: Localizing the face in the image.

Step 2: Detecting the key facial structures on the face ROI.

In recent years, deep learning algorithms, especially CNNs, have shown remarkable performance in facial landmark detection due to their ability to learn complex spatial representations from the data. Tools like dlib's facial landmark detector [230], which utilizes a pre-trained CNN, have become popular choices for this task. The pre-trained model can detect the location of 68 (x, y) coordinates that correspond to the facial structures on the face. As shown in Figure 3.2, the 68 facial landmark points can be visualised as coordinates on the face, representing locations such as the tip of the nose, corners of the eyes, mouth corners, and so on.

In this thesis, Dlib’s facial landmark detector [230] was used to detect 68 facial landmarks on the detected face. Chapter 6 and 7 give more detail explanation about the ROI selection using facial landmark detection.

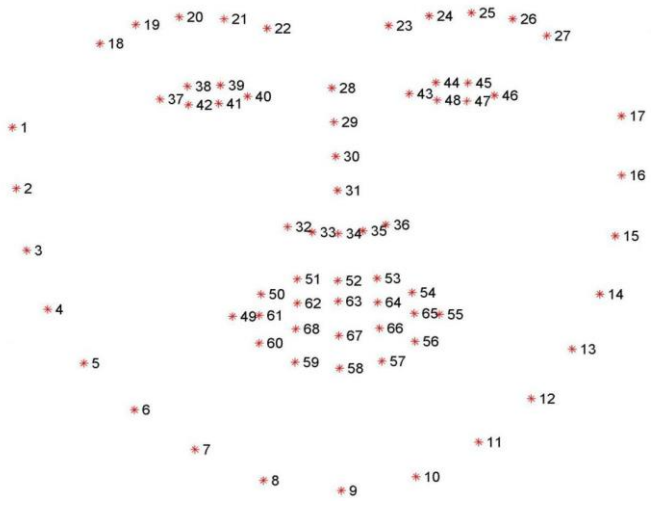


Figure 3.2: Visualization of the 68 facial landmark coordinates [adapted from [230]].

3.4.2. Raw Signal Extraction

The raw signal was extracted from the selected ROI by calculating the spatial average of pixel value in ROI for each frame. This is called intensity-based method or spatial averaging.

$$i_{R(t)}, i_{G(t)} i_{B(t)} = \frac{\sum_{x,y \in ROI} I(x,y,t)}{|ROI|} \quad (3.8)$$

where $i_{R(t)}, i_{G(t)} i_{B(t)}$ are three source signals from R, G and B components respectively, $I(x,y,t)$ is the brightness pixel value at image location (x,y) at time (t), and $|ROI|$ is the size of the selected ROI.

3.4.3. Signal Processing Techniques

The proposed systems in this thesis employ a variety of signal processing techniques for extracting relevant features from video signals and eliminating noise artefacts in the image spatial domain. These techniques included signal decomposition techniques, spectral analysis methods, filtering methods and peak detection.

3.4.3.1. Signal Decomposition

Signal decomposition techniques are widely used for biomedical signal processing tasks to decompose a temporal signal into a collection of modes of interest, while the most interesting modes need to be selected to represent the original signal. The most effective signal decomposition techniques include empirical mode decomposition (EMD), EEMD and complete ensemble empirical mode decomposition (CEEMD).

Empirical Mode Decomposition (EMD)

EMD is a multiresolution signal decomposition technique for a complex and multi-component signal representation developed by Huang et al. [231]. It is commonly used to remove noise artefacts from biomedical signals. The EMD uses the local temporal and structural characteristics of a non-linear and non-stationary signal and adaptively decomposes it into a set of stationary modes in different time scales, called intrinsic mode functions (IMFs). IMF is commonly used to remove noise artefacts from biomedical signals, such as the removal of noise artefacts from ECG data [232], the removal of noise artefacts from electromyogram data [233], the removal of muscle artefacts from

electroencephalogram data [234], the removal of tissue artefacts from respiratory signals [235], and the removal of illumination variations from photoplethysmography signal [236].

The decomposition of the original signal, $x(t)$, into a set of IMFs must occur under two assumptions, as follows: (i) the number of extrema of $x(t)$ is either equal to the number of zero-crossing or differs at most by one and (ii) the mean values of the envelopes defined by local maxima and local minima are equal to zero. The signal decomposition based on EMD is defined in the following steps:

1. Identify all peaks (maxima and minima) of $x(t)$.
2. Generate the upper and lower envelopes of the peaks via the cubic spline interpolation.
3. Calculate the mean value, $m(t)$, of the upper and lower envelopes point by point.
4. Extract the detail signal, $d(t)$, by subtracting the mean value $m(t)$ from $x(t)$:

$$d(t) = x(t) - m(t) \tag{3.9}$$

5. Check the properties of $d(t)$:
   a. If $d(t)$ meets the above two assumptions and becomes a zero-mean process, then it would be the first IMF component of $x(t)$, named IMF_1, and replace $x(t)$ with the residue:

$$r(t) = x(t) - d(t) \tag{3.10}$$

   b. Otherwise, go to step (1) and replace $x(t)$ with $d(t)$.
6. Repeat Steps 1 to 5 to obtain the IMF_2, IMF_3, …., IMF_N, where N is the number of alterations. The process is stopped when $r(t)$ becomes a monotonic function and no further IMF can be extracted.

As a result of the EMD process, the original signal, $x(t)$, can be recovered by the following expression:

$$x(t) = \sum_{i=1}^{N} IMF_i + r_N(t) \tag{3.11}$$

where N is the number of IMFs and $r_N(t)$ is the residue of the signal $x(t)$.

Ensemble Empirical Mode Decomposition (EEMD)

EEMD is a noise-assisted signal decomposition technique proposed by Wu and Huang [237], aiming at eliminating the mode-mixing problem caused by the original EMD because of intermittency. It is based on adding white Gaussian noise into the original signal, $x(t)$, in a controlled manner to obtain the noisy signal, $x_m(t)$, as follows:

$$x_m(t) = x(t) + \omega_m(t), \qquad m = 1,2,\dots,L \tag{3.12}$$

where $\omega_m(t)$ is the mth added white noise and L is the ensemble number of the EEMD technique. The acquired original signal, $x_m(t)$, is then decomposed into a set of IMFs using EMD, and can be recovered as follows:

$$x_m(t) = \sum_{i=1}^{N} IMF_{i,m} + r_{N,m}(t), \tag{3.13}$$

$$x(t) = \frac{1}{L}\sum_{i=1}^{N}\sum_{m=1}^{L} IMF_{i,m} + \frac{1}{L}\sum_{m=1}^{L} r_{N,m}(t) \tag{3.14}$$

3.4.3.2. Spectral Analysis

Spectral analysis methods are being used increasingly in biomedical signal analysis for frequency spectral evaluation and to show the dominant characteristics of biomedical signals. Spectral analysis methods can identify dominant frequencies in each time series using different computational methods, such as discrete Fourier transform.

Discrete Fourier Transform (DFT)

DFT is one of the most commonly used methods for spectral analysis to extract the relevant spectral content from the temporal biomedical signal. Fast Fourier transform (FFT) is an efficient signal processing method for computing DFT with minimum computational effort. This is because the FFT method reduces the computational complexity to $O\ (N\ log\ N)$, instead of the extensive $O(N^2)$ used by DFT [238]. In addition, round-off error can essentially also be reduced over the direct method by a factor of $(\log_2 N)/N$), where N is the number of data samples in the discrete time periodic signal.

To transform a discrete time periodic signal with N samples from time domain to frequency domain, assume $x(n)$ represents a sequence of sampled signal $x(t)$ with sampling frequency f_s. The DFT of $x(n)$, $X(k)$, is expressed as follows [239]:

$$X(k) = \sum_{n=0}^{N-1} x(n)\, e^{-j\left(\frac{2\pi nk}{N}\right)} \qquad k = 0,1,2. \ldots., N-1$$

$$= \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi nk}{N}\right) - j \sum_{n=0}^{N-1} x(n) \sin\left(\frac{2\pi nk}{N}\right) \tag{3.15}$$

The inverse DFT (iDFT) is given by:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)\, e^{j\left(\frac{2\pi nk}{N}\right)} \qquad n = 0,1,2. \ldots., N-1$$

$$= \frac{1}{N}\left\{\sum_{k=0}^{N-1} X(k) \cos\left(\frac{2\pi nk}{N}\right) + j \sum_{k=0}^{N-1} X(k) \sin\left(\frac{2\pi nk}{N}\right)\right\} \tag{3.16}$$

The FFT method can be implemented either using decimation in time or decimation in frequency. To implement the FFT method on Eq. (3.15) based on decimation in time, it is necessary to separate $x(n)$ into odd and even numbered points, $e(n)$ and $d(n)$, respectively, with $(N/2)$-point power sequences as follows [240]:

$$x(n) = \begin{cases} e(n) = x(2n) \\ d(n) = x(2n+1), \end{cases} \qquad n = 0,1,2 \ldots, \left(\frac{N}{2}\right) - 1, \tag{3.17}$$

Assuming $W = \exp\left(\frac{-2\pi j}{N}\right)$ in Eq. (3.15), it can be rewritten as follows:

$$X(k) = \sum_{n=0}^{\left(\frac{N}{2}\right)-1} \{e(n)\, W^{2nk} + \; d(n)\, W^{(2n+1)k}\},$$

$$= \sum_{n=odd}^{\left(\frac{N}{2}\right)-1} x(2n)\, W^{2nk} + \sum_{n=even}^{\left(\frac{N}{2}\right)-1} x(2n+1)\, W^{(2n+1)k} \tag{3.18}$$

To implement the FFT method on Eq. (3.15) based on decimation in frequency, it is necessary to separate $x(n)$ into two equal parts, $g(n)$ and $h(n)$, respectively, where $g(n)$ has the first $(N/2)$ samples and $h(n)$ has the last $(N/2)$ samples, as follows [240]:

$$x(n) = \begin{cases} g(n) \\ h(n) = x\left(n + \frac{N}{2}\right), \end{cases} \qquad n = 0,1,2, \ldots, \left(\frac{N}{2}\right) - 1, \tag{3.19}$$

The N points FFT of $x(n)$ as depicted in Eq. (3.18) can be rewritten as follows:

$$X(k) = \sum_{n=0}^{(\frac{N}{2})-1} \{g(n)\, W^{nk} + \; h(n)\, W^{(n+\frac{N}{2})k}\}$$

$$= \sum_{n=0}^{(\frac{N}{2})-1} x(n)\, W^{nk} + \sum_{n=0}^{(\frac{N}{2})-1} x(n+\frac{N}{2})\, W^{(n+\frac{N}{2})k} \tag{3.20}$$

3.4.3.3. Temporal Filtering

Digital filters are extensively used in biomedical signal and image processing applications due to their ability to detect specific frequency bands of interest and remove unwanted noise artefacts. These filters play a crucial role in enhancing the quality and reliability of biomedical data, which is essential for accurate analysis and diagnosis in various medical fields.

Biomedical signals, such as electrocardiograms (ECG), electromyograms (EMG), and electroencephalograms (EEG), often contain valuable information in specific frequency ranges. However, they can also be contaminated with various types of noise, such as electrical interference or muscle artefacts. Digital filters can help to extract the relevant frequency components while suppressing noise, resulting in cleaner and more informative signals [241].

Overall, digital filters are valuable tools in biomedical signal and image processing, allowing researchers and healthcare practitioners to extract meaningful information and make accurate diagnoses from complex biomedical data.

3.4.3.4. Peak Detection

Peak detection is a common task for time-series biomedical signal analysis. It is used to extract peaks of interest from a periodic signal that can be informative in biomedical applications. This study used the MATLAB built-in function '`findpeaks`' as the peak detection method to reveal the peaks and their locations within a signal. This function can also be used to determine the time between consecutive peaks with substantial prominence and can accurately detect peaks that are close to each other. Many physiological signs parameters can be easily measured when the periodicity peaks, their locations and number of peaks are identified from the relevant signal.

3.4.4 Machine Learning Algorithms

3.4.4.1. Naive Bayes

Naive Bayes [242] is a probabilistic classifier based on Bayes' theorem. It assumes that features are conditionally independent given the class label. It is computationally efficient, especially for high-dimension data, and it can handle both categorical and numerical features. It performs well with small training datasets and is relatively robust to irrelevant features. However, Naive Bayes assumes independence between features, which may not hold, leading to suboptimal performance.

3.4.4.2 K-Nearest Neighbours (KNN)

KNN [243] is a non-parametric, instance-based learning algorithm that classifies new instances based on the majority class of its k nearest neighbours in feature space. It is simple to understand and implement, and it can handle both classification and regression tasks. It is a lazy learning algorithm, which means it does not require explicit training. The training phase only involves storing the training data. KNN can be sensitive to the choice of the k value, the distance metric, and feature scaling. It can also be computationally expensive, especially with large datasets.

3.4.4.3 Support Vector Machine (SVM)

SVM [244] is a supervised learning algorithm that finds an optimal hyperplane to separate the data into different classes while maximizing the margin between the classes. SVM can handle both linear and non-linear classification tasks through the use of different kernel functions. It is effective in high-dimensional spaces and can handle datasets with a small number of samples. SVM can be sensitive to the choice of the kernel and the regularisation parameter, and it may not perform well on noisy datasets or those with overlapping classes.

3.4.4.4 Decision Tree

Decision Tree [245] is a flowchart-like model that makes predictions by partitioning the feature space based on feature values and learning simple decision rules. Decision Trees are interpretable and can handle both categorical and numerical features. They can capture complex relationships and interactions between features. They are prone to over-fitting, especially with deep or complex trees, and they may not generalize well to unseen data. They are computationally efficient during both training and prediction.

3.4.4.5 Random Forest (RF)

RF [245] is an ensemble learning method that combines multiple decision trees to make predictions. Each tree is trained on a random subset of the training data and a random subset of features. Random Forest reduces over-fitting by aggregating the predictions of multiple trees, improving generalization, and increasing model robustness. It can handle high-dimensional data, measure variable importance, and handle both classification and regression tasks. Random Forests can be slower to train and make predictions compared to single decision trees due to the ensemble approach.

3.4.4.6. Stochastic Gradient Boosting (SGB)

SGB [246] is an ensemble learning method that combines weak learners, typically decision trees, through sequential model building. It iteratively fits weak learners to the negative gradient of the loss function, minimizing the loss and improving predictions at each step. It tends to produce more accurate predictions compared to single decision trees or Random Forests and can handle complex relationships in the data. However, SGB can be computationally expensive and sensitive to hyperparameter tuning. It requires careful optimization and can be prone to overfitting if the number of boosting iterations is too high.

3.5. Software

In this thesis, the following software and tools were used:

- MATLAB program: Image Processing Toolbox, Signal Processing Toolbox, Computer Vision Toolbox.
- Python 3.9
- Movavi Video Suit
- Adobe Video Editor
- Microsoft Visual Object Tagging Tool (VoTT)

3.6. Statistics

This section briefly describes the statistical methods applied in the research presented in this thesis.

3.6.1. Pearson Correlation Coefficient

The Pearson correlation coefficient (PCC) is a statistical measure of linear association that is used to specify the level of correlation (or similarity) between two datasets. The PCC is defined as the covariance of the two datasets divided by their standard deviations. Let dataset $x = [x_1, x_2, \ldots., x_n]$ and $y = [y_1, y_2, \ldots., y_n]$ containing n values, then the PCC can be expressed as:

$$PCC = \frac{cov\ (x, y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{n} (x_i - \mu_y)^2} \sqrt{\sum_{i=1}^{n} (y_i - \mu_y)^2}} \tag{3.21}$$

where $cov\ (x, y)$ is the covariance; σ_x and σ_x are the standard deviations of dataset x and y, respectively; μ_x and μ_y are the mean values of all the samples of datasets x and y respectively; and x_i and y_i are the samples. The PCC ranges from +1 to -1, where the value +1 indicates that the relationship between datasets has a perfect positive linear correlation, -1 indicates a perfect negative linear correlation, and 0 indicates no correlation between the datasets.

3.6.2. Bland-Altman Plot

The Bland-Altman (B-A) plot, also known as a Tukey mean-difference plot, is a popular method used in biomedicine and clinical research to assess the agreement between two measurement techniques or instruments [247, 248]. It provides a graphical representation (Figure 3.3) of the differences between the measurements, allowing for a visual assessment of the agreement and potential biases between the two methods.

It is used to qualitatively assess the agreement between two datasets in a graphical form as shown in Figure 3.3, where the differences between measurements are plotted on the y-axis against the mean value of these measurements on the x-axis. Using this statistical method, the limits of agreement (LoA) and overall measurement variability can be estimated by determining the standard deviation (SD) of the differences and calculating a confidence interval within 95% of the differences between measurements, which calculated by a mean difference (bias) is ±1.96 SD.

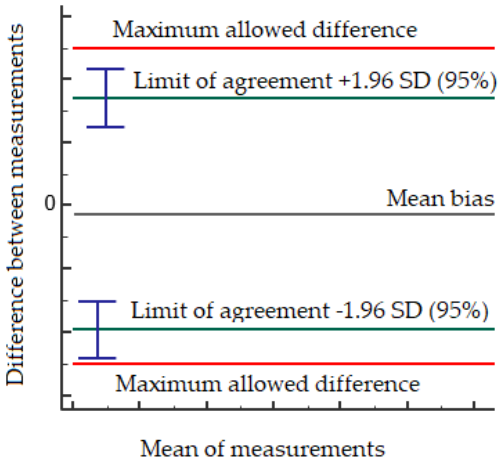


Figure 3.3: The graphical form of the Bland-Altman method.

The symmetry between methods will be in good agreement if the 95% confidence interval for the mean difference falls within the range of ±1.96 SD of the differences. The 95% confidence interval values above and below the bias demonstrate the influence of random variations between the measurements.

By analyzing the Bland-Altman plot, researchers can gain insights into the comparability and interchangeability of the two measurement techniques, helping to determine the appropriateness of using one method as a substitute for the other in specific clinical or research settings. The Bland-Altman plot is a powerful and widely used tool for assessing measurement agreement, and it is commonly employed in various biomedical and clinical research studies.

3.6.3. Root Mean Square Error

The root mean square error (RMSE) is a statistical measure of precision characterizing the difference between the predicted and observed values for both systematic and non-systematic errors. The RMSE is defined as the square root of the mean square error between two datasets, x and y, as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3.22}$$

where the best values of the RMSE are close to zero.

3.6.4. Mean Absolute Error

The mean absolute error (MAE) is another statistical measure that is used to evaluate the accuracy between the predicted and observed values. The MAE is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(x_i - y_i)| \tag{3.23}$$

where the best values of the MAE are close to zero.

3.6.5. Accuracy

Accuracy is a common metric used to evaluate the overall performance of a classification model. It denotes the ratio of correctly classified instances to the total number of instances evaluated.

$$Accuracy = \frac{TP+TN}{\text{TP+TN+FP+FN}} \tag{3.24}$$

Where TP represents True Positive, TN presents True Negative, FP represents False Positive, and FN presents False Negative (FN).

TP (True Positive): It denotes the number of cases that are actually positive and are correctly predicted as positive by the model. In other words, these are the cases where the model predicted the positive class correctly.

TN (True Negative): It presents the number of cases that are actually negative and are correctly predicted as negative by the model. These are the cases where the model predicted the negative class correctly.

FP (False Positive): It signifies the number of cases that are actually negative but are incorrectly predicted as positive by the model. These are the cases where the model made a false positive error.

FN (False Negative): It denotes the number of cases that are actually positive but are incorrectly predicted as negative by the model. These are the cases where the model made a false negative error.

3.6.6. Precision

Precision measures the proportion of predicted positive instances that are actually positive. It focuses on minimizing false positives.

$$Precision = \frac{TP}{\text{TP+FP}} \tag{3.25}$$

3.6.7. Recall

Recall estimates the ratio of actual positive cases that are correctly recognised by the model. It focuses on identifying all positive instances, minimizing false negatives.

$$Recall = \frac{TP}{\text{TP+FN}} \tag{3.26}$$

3.6.8. F1 Score

The F1 score is a widely used metric in classification tasks that combines precision and recall into a single value, providing a balance between the two measures. It is particularly useful in scenarios where there is an imbalanced class distribution or when both false positives and false negatives have significant consequences.

$$F1\ Score = \frac{2\times(Precision\times Recall)}{(Precision+Recall)} \tag{3.27}$$

3.7. Concluding Remarks

This thesis investigated how to potentially obtain clinical and biomedical information from video data captured by various camera sensors, and how non-contact vital signs measurements can be improved when different challenges need to be considered during the measurements. Many image and signal processing techniques have been developed and applied to video data to address these challenges and the ways this may help obtain an efficient and accurate non-contact system. This chapter has outlined the methodology and techniques used in the research presented in this thesis to address the research questions by providing descriptions of the study's design, data collection procedures, data processing and analysis, software used, and statistical methods applied.

CHAPTER 4

Non-contact Neonatal Vital Signs Monitoring in NICU

4 Non-contact Neonatal Vital Signs Monitoring in NICU

Conventional contact-based vital signs monitoring sensors, such as pulse oximeters or ECG may cause discomfort, skin damage, and infections, particularly in neonates with fragile, sensitive skin. Therefore, non-contact monitoring of the vital sign is desired in both clinical and non-clinical settings to overcome these issues. Camera-based vital signs monitoring is a recent technology for these applications with many positive attributes. However, there are still limited camera-based studies on neonates in a clinical setting. Therefore, in this chapter, the heart rate and respiratory rate of infants at the NICU in Flinders Medical Centre were remotely monitored using a digital camera applying colour and motion-based computational methods. The content of this chapter is based on results published in [III, IV].

The outline of this chapter is as follows. Section 4.1 presents a general introduction to the proposed system and related work. Materials and protocol, with study design, experimental set up for data collection in the NICU, are explained in section 4.2. Section 4.3 provides the methods of the proposed non-contact system including different image and signal processing techniques. Then, the experimental results are represented and discussed in section 4.5 and section 4.6 presents the concluding remarks for the chapter.

4.1. Introduction and Related Work

A delivery between 37 and 42 weeks of gestation is defined as term pregnancy by the World Health Organization (WHO) [249]. Preterm birth is defined as any birth prior to 37 weeks of gestation, and it is a priority health issue worldwide. It is projected that more than one in ten of the infants in the world are born prematurely [250]. As preterm infants are not completely developed and have a tendency to have medical conditions that need specialist care, they are often placed into the NICU immediately after birth [251].

In NICU, vital signs including respiratory rate and heart rate of neonates, are monitored continuously by various monitoring sensors, such as ECG, pulse oximeters, nasal thermocouples, piezoelectric transducers, and respiratory belt transducers [6]. These monitoring sensors are attached to the neonates with wires, straps, electrodes or probes, which may cause painful stimuli, discomfort, skin irritation, and even skin damage when the adhesive sensors are removed as neonates have fragile, sensitive skin [252]. These

complications can lead to injuries such as strangulation or entanglement, bacterial and fungal infection risk, and sleep disruption for the neonates [253]. Moreover, bonding of parent-infant, especially at the time of kangaroo mother care, is impaired due to obtrusiveness of the wires [254]. As explained earlier in section 2.1, in recent years, non-contact remote vital sign monitoring techniques, including dopplers [55, 255], thermal/infrared imaging [62, 68] and video camera imaging [118, 151, 256], have been effective in solving the issues of contact vital sign monitoring approaches due to their unobtrusiveness and absence of physical contact.

Among non-contact techniques, video camera imaging is the most popular because of the high resolution and low cost of digital cameras [28]. Video camera imaging technology can be categorized into two main categories: motion-based methods and colour-based methods. From section 2.2, it can be noticed that in most of the studies, participants were adults; therefore, the investigation into the feasibility of non-contact monitoring of neonates in a clinical setting is limited.

Some studies have used video camera imaging technology for monitoring vital signs in the NICU. For instance, Scalise *et al*. [165] first proposed a non-contact HR monitoring system for newborns using a webcam in combination with a non-ambient green light. In this work, the ROI was selected manually. ICA and PSD were also used to extract HR. However, the infants were recorded only in supine positions in this work. The system was limited by short effective range, motion artefacts, light reflections, and unclear ROI due to occlusion of the ROI. Aarts *et al*. [18] first used ambient light to monitor the vital signs of 19 infants in a NICU and investigated different scenes with challenging conditions. In their work, for tracking the global motion of the infant, a manually chosen ROI was used as a template. Fourier analysis and a joint-time-frequency diagram (JFTD) were also considered. Nevertheless, the system was affected by infant motion and low ambient light.

A camera-based scheme was presented by Klaessens *et al*. [177] to monitor the RR and HR of seven neonates in NICU using RGB colour magnification and IRT. However, this system was also highly affected by motion. Another study by Gibson *et al*. [167] considered colour magnification to measure HR and motion magnification to measure RR based on EVM. A method comparison study between the digital camera and the unit's ECG monitor was performed on 10 neonates in the NICU. However, the computational cost of EVM is high. Using two LWIR and three CMOS cameras, Paul *et al*. [257] performed another

clinical study for monitoring the vital sign of 19 neonates. After a manual ROI selection, a kernelized correlation filter (KCF) was incorporated for tracking the ROI. The key issues pointed out by the authors were artefacts from motion, light sources, medical devices, and the detection and tracking of ROI to extract the signal.

An efficient and robust technique was proposed by Cobos-Torres *et al*. [166], considering multichannel analysis based on the least square method to lessen the effect of the moderate motion artefacts and changing light conditions. This work used a moving average filter, two narrowband filters, and a short-time Fourier transform (STFT). It was reported that, the proposed system needed 75% less CPU usage than ICA. Chen *et al*. [258] introduced another robust HR monitoring framework by resolving motion artifact challenges and considered nine hospitalized neonates. However, this study was conducted in a personal area with no noisy hospital environment and an unchanged lighting environment.

A study by Mestha *et al*. [259] introduced a continuous HR monitoring system where they recorded videos of eight newborns for 30 minutes each using a webcam. Another continuous monitoring scheme was presented by Villarroel *et al*. [178], where two neonates were monitored for 40 hours using a video camera. Nevertheless, this approach was constrained to unclear ROI, illumination variation, and subject movement. However, the ROI was selected manually in all the studies discussed above.

In the literature there are different algorithms such as Viola-Jones algorithm [109], Liao's method [260], Chen's method [157] to detect ROI automatically. However, all of these existing algorithms work for adults only. In [168] and [136], the Viola-Jones algorithm was used to automatically detect the face of adult participants. Then, a skin detection algorithm [261] based on skin colour was applied on the detected face to detect and gather skin pixels that encompass the physiological signal. In this skin detection algorithm, the face image was converted from RGB colour space to YCbCr space. The skin mask was generated from YCbCr colour space by setting a threshold on the two or three channels of YCbCr space. However, the proposed skin detection algorithm was used to detect skin after selecting the ROI and worked for adults only.

To address some of these challenges, this chapter aims to monitor the vital signs of neonates in NICU using a digital camera in normal neonatal scenarios that have affected previous results. Rather than selecting ROI manually, an efficient ROI selection method based on image decomposition and skin detection was designed. The skin detection algorithm was

done based on skin colour as skin colour has been recognized to be a useful and robust cue for human-related image analysis, such as, automatic detection of skin in images, face detection, localization and tracking, hand detection and tracking, image filtering, people retrieval in Internet and databases [262-264].

4.2. Materials and Protocol

This section explains the research design and the materials used in the study.

4.2.1. Study Design

A single-center cross-sectional experimental study was performed at Flinders Medical Centre Neonatal Intensive Care Unit, a tertiary neonatal provision in Adelaide, South Australia. This study was approved by the Southern Adelaide Local Network Research Committee (HREC/17/SAC/340; SSA/17/SAC/341) and the University of South Australia Human Research Ethics Committee (Protocol number 0000034901). Prior to filming, the experimental procedure was thoroughly explained to the guardian of neonates and a written consent was also obtained from them.

Eight neonates were filmed while being monitored using the regular ECG monitor of the unit. Seven neonates were preterm (gestational age was less than 37 weeks), and one neonate was greater than 37 weeks. Neonates who had abnormalities that may make them distinguishable in publications or who were not monitored by ECG or who were expected to be discharged during data collection were excluded in this study.

In this work, ECG was considered as the reference for validation purposes for all neonates. For extracting HR and RR, the difference in electrical impedance and chest wall movement were measured by the impedance lead of the ECG, respectively. ECG was constrained to patient movement and cardiac activity [220]; however, it was used for validation purpose to avoid interruption to the neonates.

4.2.2. Experimental Set-Up

This study used two digital cameras mounted on tripods to record videos, as depicted in Figure 4.1. A Nikon D610 with resolution 1920×1080 and frame rate 30fps was used to capture the video of the neonate. On the other hand, a Nikon D5300 was used to record the video of the ECG monitor. Two cameras were synchronized manually. Videos from each camera started at the same time, for synchronizing datasets from the proposed method and

the ECG monitor. The videos were saved in .MOV format. The distance between the camera and the subject was 1-2 meters. 10-minute-long videos of each neonate were recorded. For the experiment, 10 second videos were considered when neonates remained stable, i.e., not moving. For each neonate, five samples were considered for data analysis. Moreover, HR from the screen was extracted manually by taking average of numerical values of HR displayed on the screen.



Figure 4.1: Experimental setup for data collection.

4.3. Methods

4.3.1. System Framework Using Image Decomposition Method

As shown in Figure 4.2, the overall system includes numerous techniques: video decomposition into tiles, ROI selection, spatial averaging, spectral analysis, band-pass filtering, and peak detection.

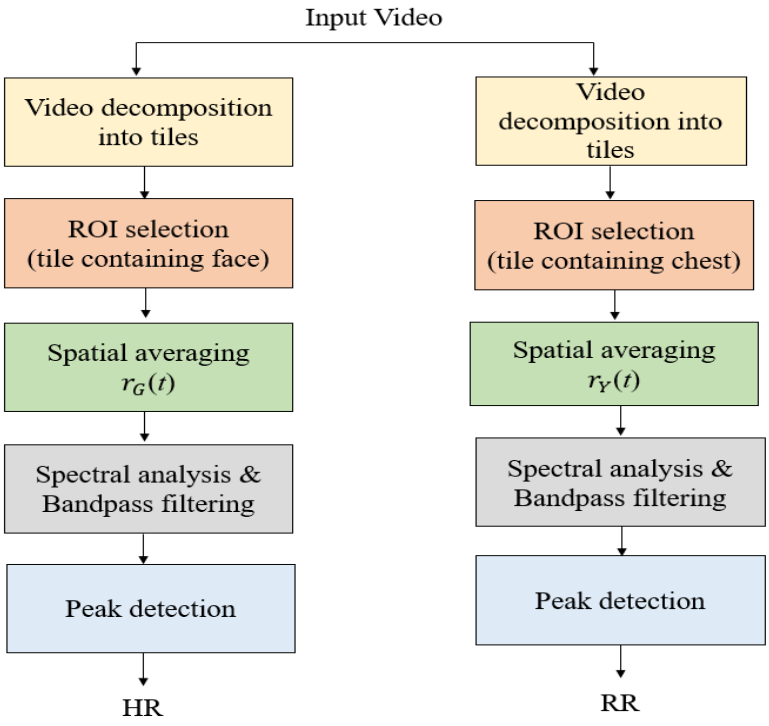


Figure 4.2: System framework using image decomposition method.

4.3.1.1. Video Decomposition

The first frame of the raw video was first decomposed into tile images. To create a tile video from ROI, each frame was divided into small tiles corresponding to the grid in Figure 4.3(a). The tile size $d \times d$ was selected heuristically. Creating of tiles was started from the (1,1) image coordinate on the raw image frame. When the image width or height are not integer multiples of d, the tiles do not cover the whole raw image frame. In such cases, only the area covered by the tiles on the raw image frame was considered (i.e., the red grid in Figure 4.3(a) shows the selected tiles and each grid cell represents a tile video).

The total number of tiles is given by

$$N_T = \lfloor W/d \rfloor \times \lfloor H/d \rfloor \quad (4.1)$$

W and H are the width and height of the raw image frame.

Each ROI tile video was constructed with the same number of frames and at the same frame rate as the raw video.

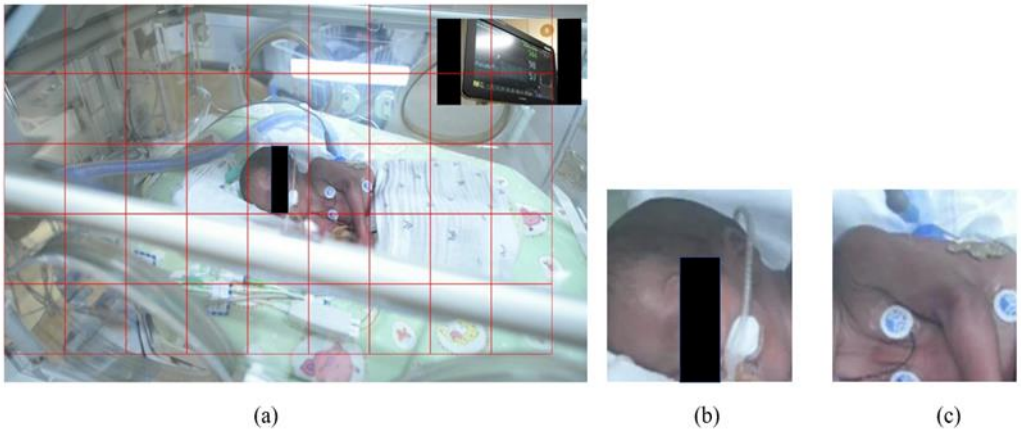


Figure 4.3: Video decomposition and ROI selection. (a) Image decomposed into tiles. (b) Tile containing face considered as an ROI for HR (c) Tile containing chest considered as an ROI for RR.

4.3.1.2. ROI Selection

Each ROI was selected from the decomposed image. The tile which contained the face region was selected as ROI to calculate HR and the tile containing the chest region was chosen as ROI to calculate RR, as shown in Figure 4.3. The same process was repeated for all frames.

In the standard manual ROI selection method, the face or chest area is selected by drawing a bounding box around it. The area covered by the bounding box is used to extract the vital signs. Here, instead of manually defining bounding boxes, the tiles containing the ROI are selected. This was done by visual inspection. The selected tiles for Figure 4.3(a) are shown in Figure 4.3(b) and (c). The main advantage of this ROI selection is that there is no need to adjust ROI bounding boxes whenever the baby changes pose. Normally, the baby moves only sideways, and therefore the face and chest areas always remain within the initially selected tiles. Since the tile size is adjustable, a safe tile size can be configured to accommodate the baby's movements.

This visual tile selection can be replaced with a baby detection machine learning model. However, it was found that this approach is challenging as (i) it is difficult to collect a sufficiently large dataset of babies, (ii) the available images contain complex equipment that occludes the view of the child, and (iii) higher numbers of fault detections were noticed with available human detectors when toys and baby like shapes on the bedsheet were present.

4.3.1.3. Spatial Averaging

In this study, colour based, and motion-based computational approaches were used for extracting raw cardiac and respiratory signals, respectively. Light absorbed by haemoglobin in blood is more than the adjacent tissue and causes a colour variation in human skin. This colour variation is not noticeable with naked eyes but can be revealed by a digital camera. The colour variation reflects the blood volume variation in the microvascular tissue bed under the skin caused by changes in pulsatile blood flow during each cardiac cycle. Hence, using this principle, the cardiac signal can be measured. First, the green channel of RGB colour space was chosen. Then, by calculating the average of the brightness pixels of the selected ROI, a raw cardiac signal was measured as depicted in (4.2). This was done to reduce camera noise enclosed in a single pixel and thus improve the signal to noise ratio.

$$r_G(t) = \frac{\sum_{i,j \in ROI} R(i,j,t)}{|ROI|} \tag{4.2}$$

where $R(i,j,t)$ represents the brightness pixel value at image location (i,j) at a time (t), and $|ROI|$ is the size of the selected ROI. Compared to the red and blue channels the green channel has the strongest cardiac frequency band [125]. Therefore, the $r_G(t)$ signal was chosen to measure the cardiac signal.

During respiration, a difference in the movement of the thoracoabdominal wall is caused by inhalation and exhalation. In the recorded videos, spatial differences in intensity values reflect the difference in the movement of the thoracoabdominal wall caused by respiration. Therefore, the respiratory signal can be measured based on this principle. It is necessary to isolate the intensity information from the colour information as the digital camera recorded the video in the RGB colour space. Therefore, the RGB colour space was converted to the YIQ colour space using MATLAB built-in function, called 'rgb2ntsc'. First, the Y channel of the YIQ colour space was chosen. Then, by taking the mean of the pixels' intensity within the selected ROI, the raw respiratory signal was extracted as follows:

$$r_Y(t) = \frac{\sum_{i,j \in ROI} R(i,j,t)}{|ROI|} \tag{4.3}$$

where $R(i,j,t)$ represents the intensity value of a pixel in Y channel at image location (i,j) at a time (t), and $|ROI|$ represents the size of the chosen ROI.

4.3.1.4. Spectral Analysis and Band-pass Filtering

The raw time series cardiorespiratory signals were transformed from the time domain to the frequency domain based on spectral analysis technique using FFT. Then, to acquire the frequency band of interest, two ideal band-pass filters were implemented at 1.5 to 3 Hz corresponding to the pulse range (90–180 beats/min), as well as 0.3 to 1.67 Hz, which corresponds to the breathing range (18–100 breaths/min). Though the frequency bands overlapped, frequency contents are properly separated using the ideal band-pass filters because different methods were incorporated to extract HR and RR. Colour-based method was used for HR extraction and motion-based method was used for RR. Moreover, two separate ideal band-pass filters were used to extract HR and RR, respectively.

After that, the inverse FFT was implemented on the filtered signals to attain the time series cardiac and respiratory signals.

The FFT and its inverse are powerful tools for analysing signals in the frequency domain. Filtering in the frequency domain through the FFT allows for a more intuitive and efficient manipulation of signal components, and it provides a way to focus on specific frequency ranges. Moreover, while filtering in the frequency domain, especially with a well-designed band-pass filter, the temporal information of the signal is generally preserved. This means that the timing relationships between different components of the signal are maintained, which is crucial in applications where the temporal aspect is significant.

To extract HR and RR, we need to acquire the frequency band of interest and bandpass filter was used to get that frequency band. In literature, it is very common to use FFT to get frequency domain signal from time domain. As our raw cardiorespiratory signal was in time domain, we used FFT to get frequency domain signal. Moreover, we used peak detection method and calculated the number of peaks to extract the HR and RR. To calculate the peaks, we need time domain signal, therefore, we used inverse FFT to convert frequency domain signal into time domain signal. As mentioned above, the temporal information of the signal is preserved while filtering in the frequency domain. Therefore, it is acceptable to use time-signal which was obtained by inverse FFT after filtering.

4.3.1.5. Peak Detection

The periodicity of peaks, peak locations, and the number of peaks of the attained signals were determined by using the peak detection method using the MATLAB built-in function

'findpeaks'. After identifying the peaks and their locations (locs), the total cycle length (C_T) between two peaks can be calculated using the following equation:

$$C_T = mean\big(diff(locs)\big) \tag{4.4}$$

the number of peaks (p) can be found as follows:

$$p = \frac{t}{C_T} \tag{4.5}$$

where t signifies the length of video recording in seconds.

HR and RR per minute can be calculated using the following equation.

$$C_M = \frac{60p}{t} \tag{4.6}$$

where C_M symbolizes the measured value.

4.3.2. System Framework Using Skin Detection Method

As shown in Figure 4.4, the overall system includes numerous techniques: Automatic ROI selection, spatial averaging, spectral analysis, band-pass filtering, and peak detection.

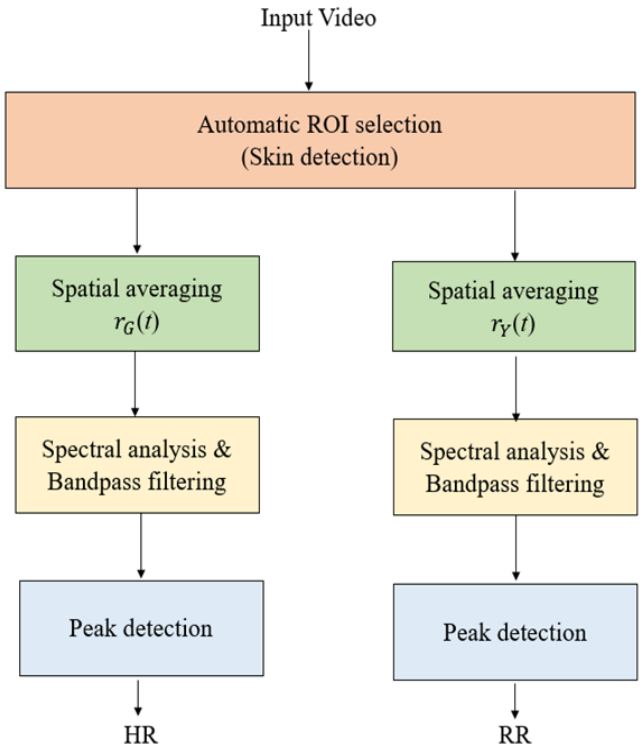


Figure 4.4: System framework using skin detection technique.

4.3.2.1. Automatic ROI Selection

After converting the videos into frames, the first frame was considered to select ROI as shown in Figure 4.5(a). First, the image was cropped to avoid nurses, attendees or other babies presented in that image as shown in Figure 4.5(b). Then, the skin detection algorithm based on skin colour was applied in the cropped image to select ROI automatically. The algorithm returned the bounding box of the skin portion, especially the face, when the body was covered with sheets or blankets. Finally, ROI was detected automatically based on skin colour (Figure 4.5(c)). This process was repeated for all frames.

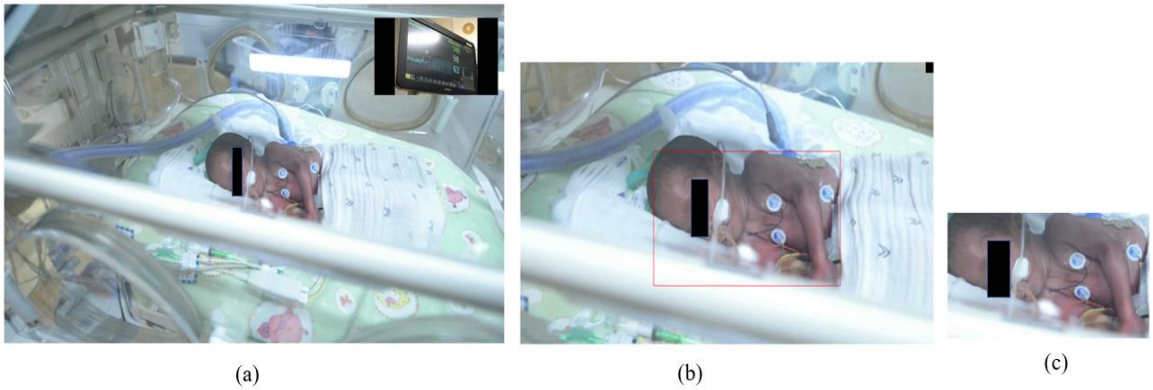


Figure 4.5: Automatic ROI selection. (a) Original image, (b) cropped image, (c) detected ROI.

In the skin detection algorithm, first, RGB image was converted to LAB colour space as this colour space is useful for image segmentation as shown in Figure 4.6(a). Unlike the RGB and CMYK colour models, LAB is designed to approximate human vision. LAB is useful for predicting small differences in colour which helps to segment images. LAB colour space was used for following reasons

- to segment skin regions from background pixels.

- for representing the colours of each image patch, ignoring the luminance channel as it is affected by illumination changes.

- due to its ability of limited variance due to sensor sensitivity.

- to remove shadows and other objects, it is closer to human vision and better represents what is able to be perceived by individuals.

A (red-green) and B (yellow-blue) component of LAB colour space were chosen (Figure 4.6(b) and 4.6(c)) and then converted them into binary images (Figure 4.6(d) and 4.6(e)). Image binarization was done based on global image threshold using Otsu's method [265]. After that, binary images are multiplied (Figure 4.6(f)) and AND logic is performed on the image, so a matrix is obtained having 1 and 0. 1 indicates the presence of skin. Thus, ROI can be detected automatically based on skin colour.

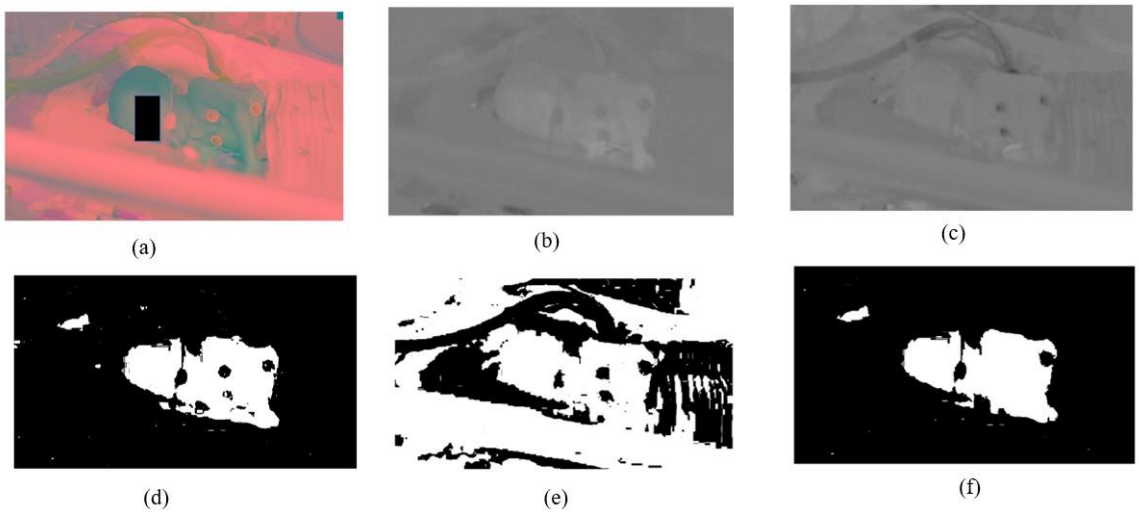


Figure 4.6: Skin detection. (a) LAB colour space, (b) A channel of LAB colour space, (c) B channel of LAB colour space, (d) binary image of A channel, (e) binary image of B channel, (f) multiplied image of binary images.

In a hospital setting, it is very challenging to record clear video of babies. Most of the time, their body is occluded with medical instruments, blankets, or clothes. The proposed skin detection method can successfully work even when ROI is unclear due to the infant being covered by a blanket, breathing mask etc.

The proposed skin detection method was tested on 40 videos (8 infants, each infant 5 videos), and each consisting of 300 images. Among the 40 videos i.e., 12,000 images, the ROI selection method succeeded with 35 videos i.e., 10500 images, hence, the accuracy rate with this data set is 87.5%. One infant was receiving phototherapy producing a blue light, and our skin detection method failed to work in this case.

The same technique described in section 4.3.1 was used for spatial averaging, signal decomposition, spectral analysis, band-pass filtering, and peak detection to extract HR and RR.

4.4. Experimental Results

To implement the proposed algorithm and compute the statistical results, MATLAB 2020a was used.

4.4.1. Results Using Image Decomposition Method

Figure 4.3 shows the video decomposition and ROI selection process. As shown in Figure 4.3(a), the image with size 1920×1080 was divided into tiles with size 200×200. The total number of tiles was 45. Among 45 tiles, the 23rd number tile contained the face, which was considered as ROI for HR calculation (Figure 4.3(b)), and 24th number tile contained the chest and was chosen as ROI for RR calculation (Figure 4.3(c)). The same process was repeated for all frames.

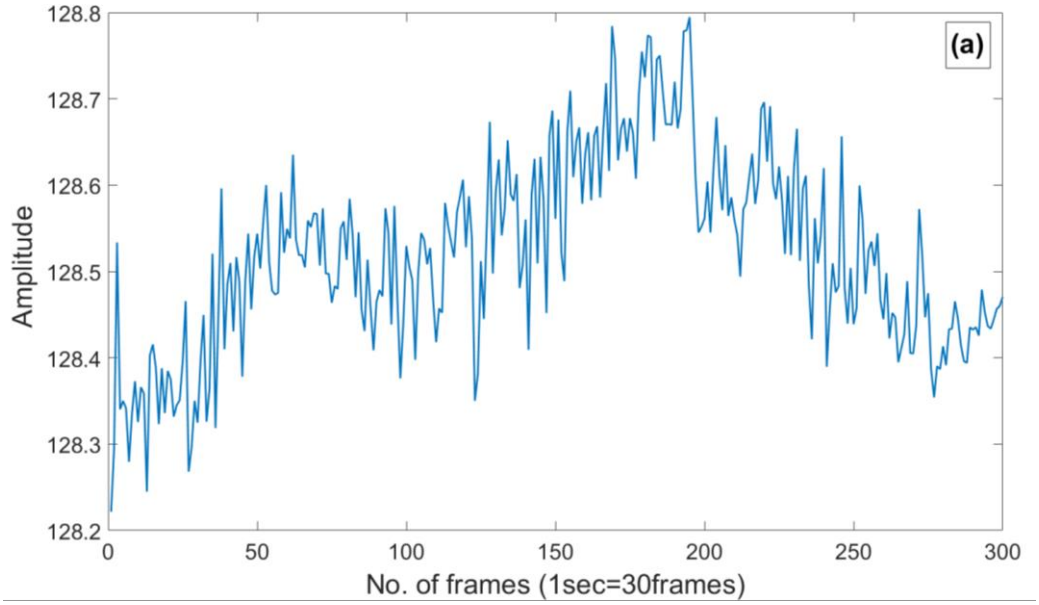
The time series cardiac signal was obtained by calculating the mean of the pixel values over the selected ROI, as depicted in Figure 4.7(a). FFT, band-pass filtering and inverse FFT were implemented to achieve the filtered cardiac signal and shown in Figure 4.7(b). The HR was measured by calculating the number of peaks from the filtered cardiac signal. In Figure 4.7(a) and (b), the labels of x-axes are different i.e. number of frames and time (sec), respectively. We have used 10s duration video, and the camera frame rate was 30 fps. Therefore, there were 300 frames (Figure 4.7(a)) which were equal to10s (Figure 4.7(b)). Likewise, Figures 4.8(a) and 4.8(b) represent the raw respiratory signal and filtered respiratory signal, respectively. RR could also be calculated by computing the number of peaks from the filtered respiratory signal.

In Figures 4.8 and 4.14, both subplots have a different scaling in y-axes. The difference in scaling between the raw respiratory signal and the filtered respiratory signal may be due to the processing steps applied to the signal. The FFT operation and subsequent filtering might introduce scaling factors. The band-pass filter might attenuate or amplify certain frequency components, affecting the amplitude of the filtered signal.

Each ROI has an intensity. When the signal changes, the intensity changes. We take the average intensity, which is a normalized value between 0 and 1. There are no units as this is just a normalized amplitude. If there is no signal, the image intensity remains constant - for example, 0.5685 in Figure 4.8 (a). When the signals in the ROI change, this average intensity changes resulting in the unfiltered signal shown in the Figure 4.8 (a). If the

average amplitude value is subtracted from all other y values, it will give a scale similar to the filtered signal (Figure 4.8(b)).

In Figure 4.8, the filtered respiratory signal (lower plot) looks quite different from the upper unfiltered plot, in particular for the first and last breath stroke. Differences between the filtered and unfiltered respiratory signals, especially around the first and last breath strokes, can be attributed to the band-pass filtering process. The band-pass filter used in the algorithm has an ideal rectangular frequency response. Ideal filters have sharp transitions between the passband and stopband, which can introduce artifacts, particularly ringing effects. The abrupt changes in the frequency response can cause distortions in the time domain, leading to differences in the filtered signal. Moreover, the characteristics of the original respiratory signal may also contribute to differences. Sudden changes or peaks in the raw signal may be affected differently by the filtering process.



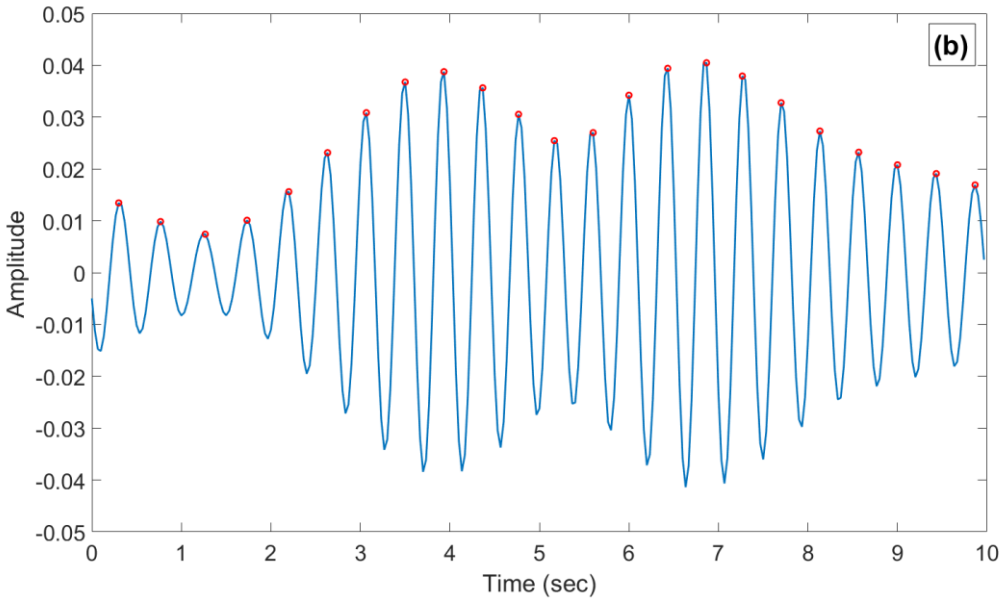
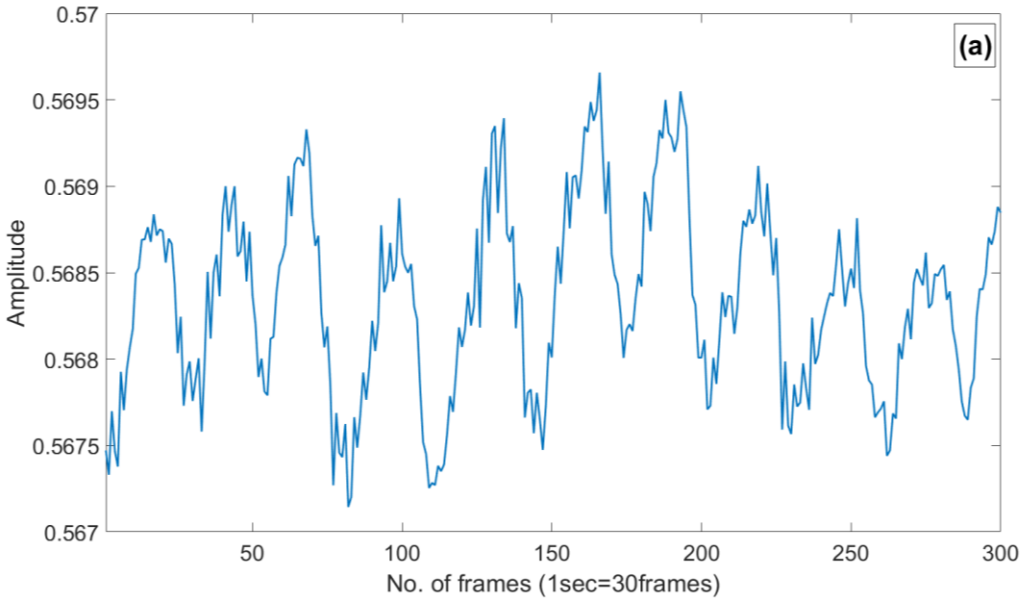


Figure 4.7: Cardiac signal. (a) Raw cardiac signal, (b) filtered cardiac signal.



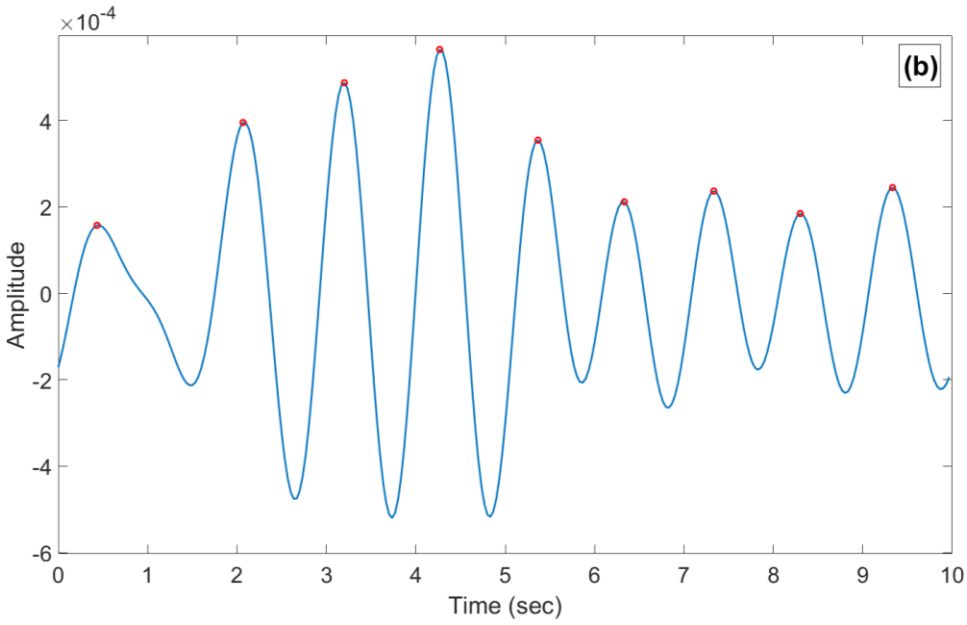


Figure 4.8: Respiratory signal. (a) Raw respiratory signal, (b) filtered respiratory signal.

Different statistical techniques, for example, the MAE, RMSE, PCC, linear regression, and Bland Altman plot, were considered for the evaluation purpose. A total sample size of n=40 was considered. Figure 4.9 presents a correlation plot for HR, and Figure 4.10 represents a Bland Altman plot for HR. A strong correlation between the ECG data and measured camera-based data with a PCC of 0.9794 can be observed in Figure 4.9. It can be seen from Figure 4.10 that the lower and upper limit of agreements were -4.9 to + 6.1 beats/min, the mean bias was 0.6 beats/min, and the reproducibility coefficient (RPC) was 5.5 beats/min (3.7%). RMSE and MAE for HR were 2.84 beats/min and 2.18 beats/min, respectively.

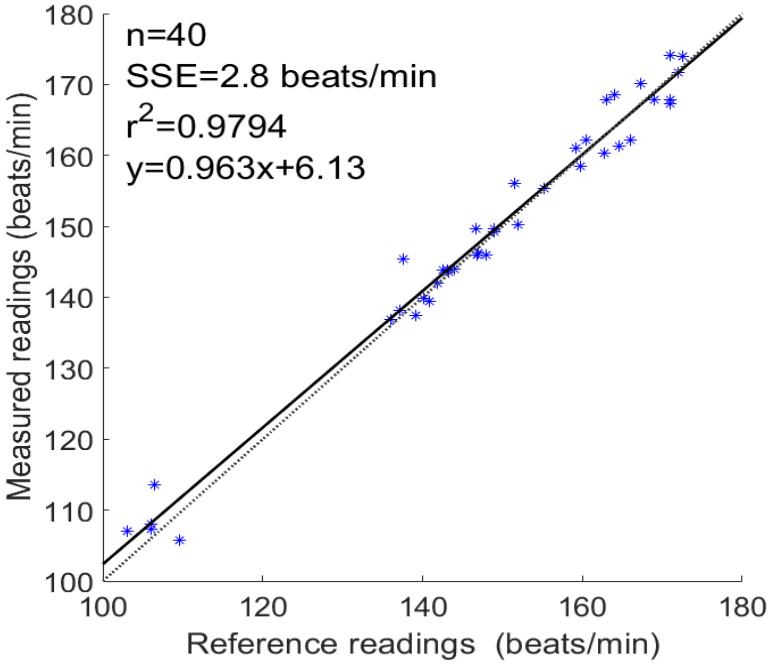


Figure 4.9: Correlation plot for HR.

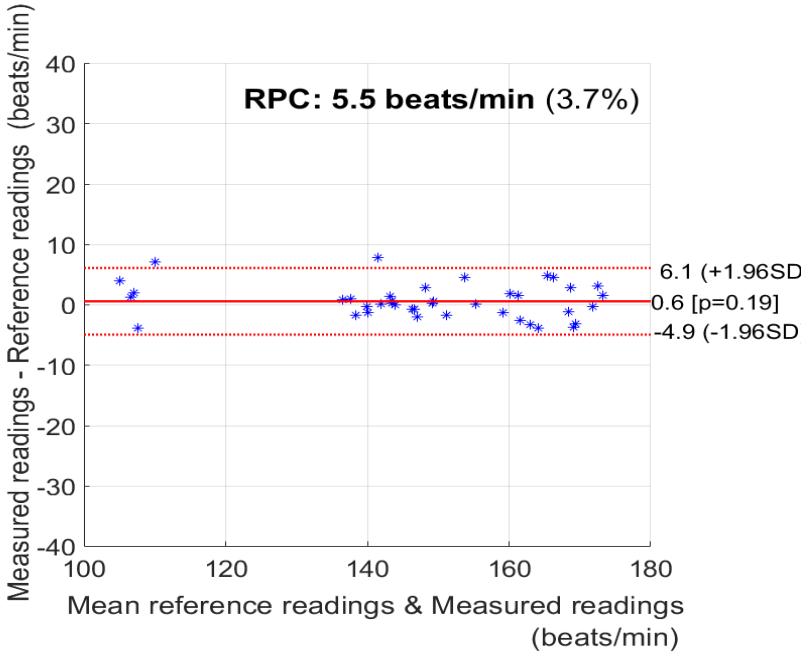


Figure 4.10: Bland Altman plot for HR.

Figures 4.11 and 4.12 show a correlation plot and a Bland Altman plot for RR, respectively. A good correlation exists between the ECG data and measured camera-based data with a

PCC of 0.9412 (Figure 4.11). The Bland Altman plot in Figure 4.12 represents RPC 5.4 breaths/min (11%), mean bias 1 breaths/min, and the lower and upper limit of agreement -4.4 to +6.4 breaths/min. RMSE was 2.91 breaths/min, and MAE was 2.17 breaths/min for RR.

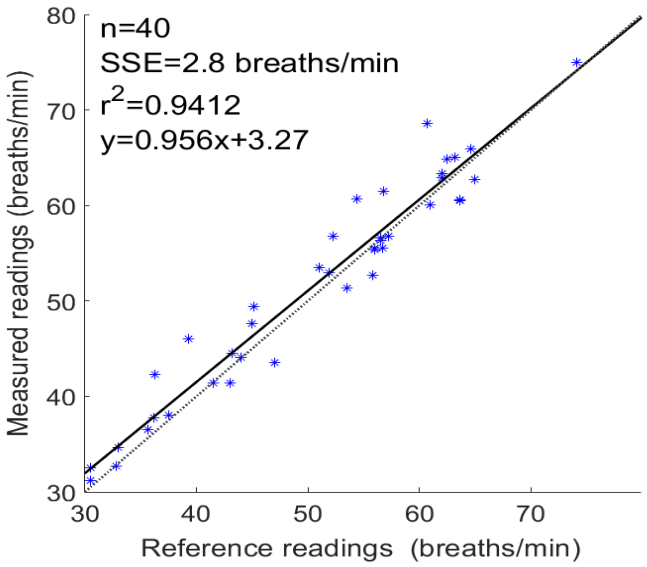


Figure 4.11: Correlation plot for RR.

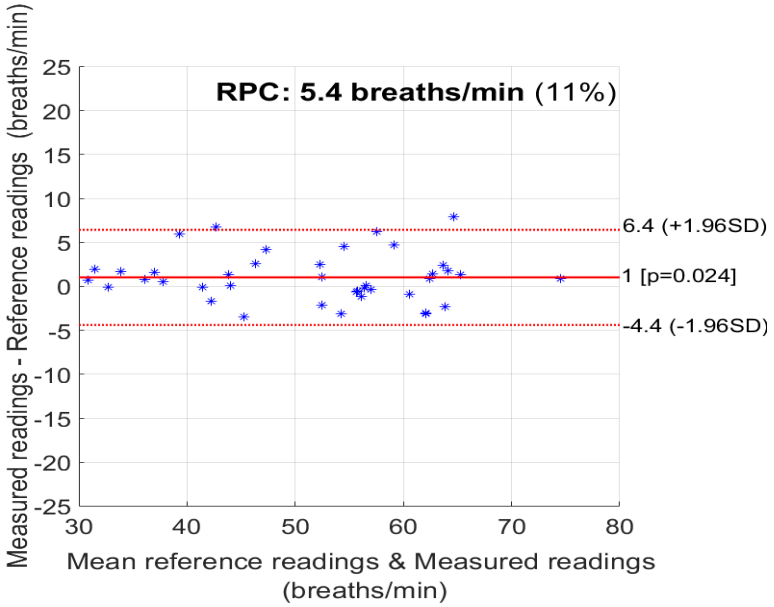
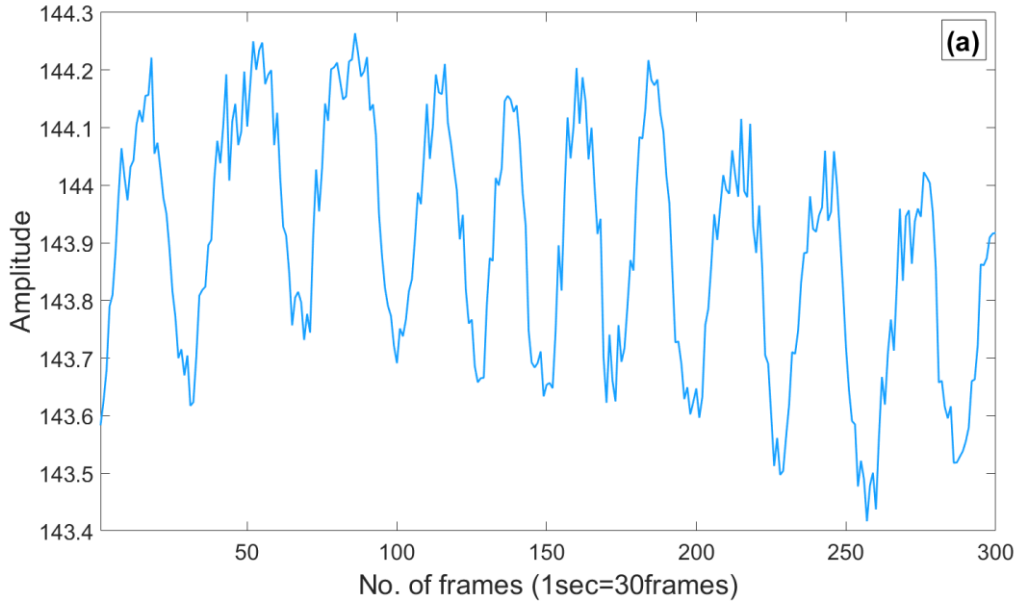


Figure 4.12: Bland Altman plot for RR.

4.4.2. Results Using Skin Detection Method

As shown in Figure 4.5(a), the first frame of the video was considered to detect ROI. Figure 4.5(b) represents the cropped image and skin detection algorithm was applied in the cropped image to detect ROI automatically. Figure 4.5(c) shows the detected ROI.

After performing spatial averaging over the detected ROI, the raw cardiac signal was obtained, as shown in Figure 4.13(a). Figure 4.13(b) shows the filtered cardiac signal that was achieved after FFT, band-pass filtering and inverse FFT. From the filtered signal, the HR was calculated by determining the number of peaks. Similarly, Figures 4.14(a) and 4.14(b) present the raw respiratory signal and filtered respiratory signal, respectively. By calculating the number of peaks, RR could also be calculated from the filtered respiratory signal.



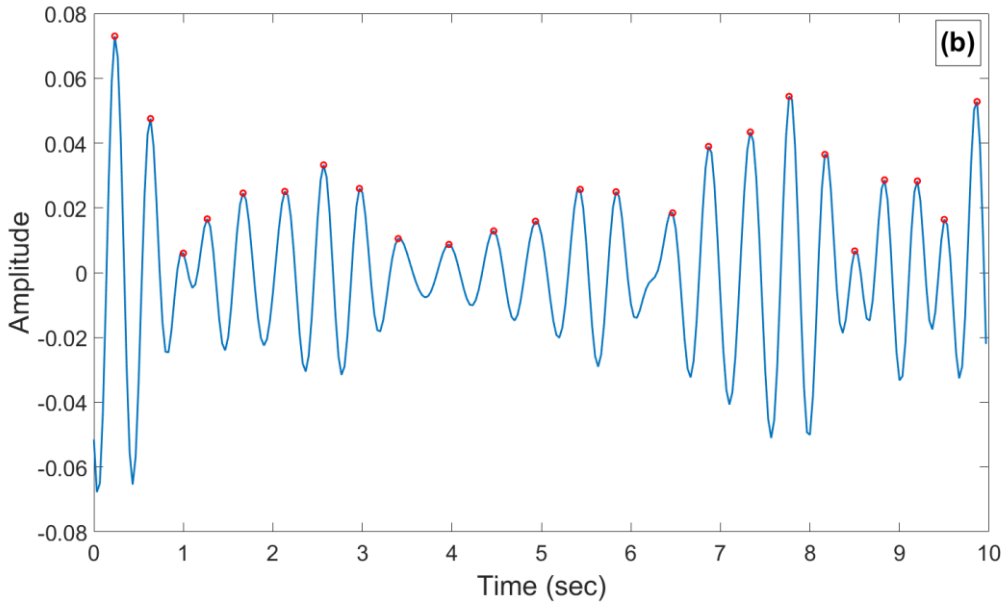
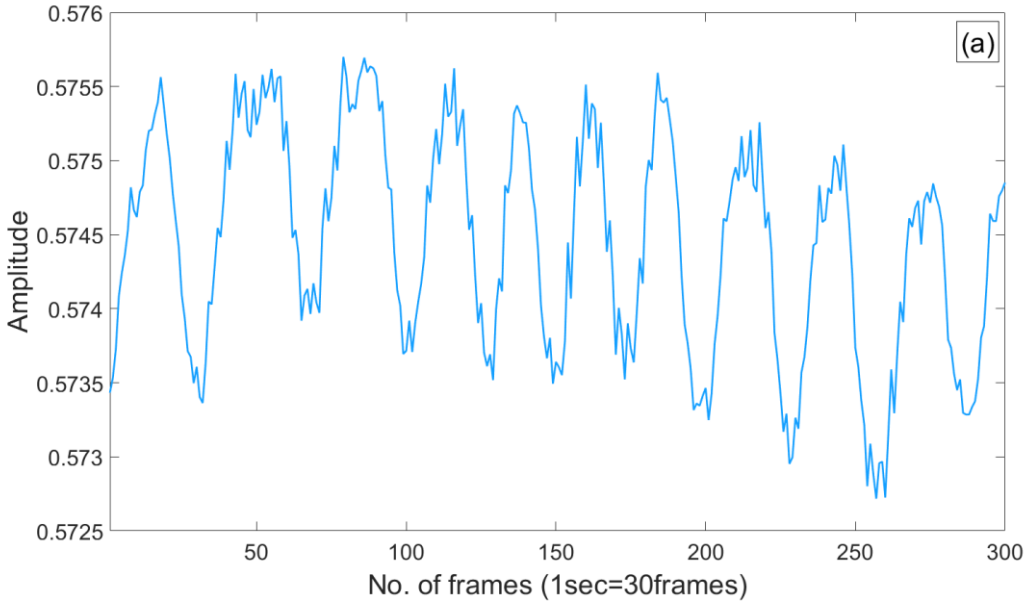


Figure 4.13: Cardiac signal. (a) Raw cardiac signal, (b) filtered cardiac signal.



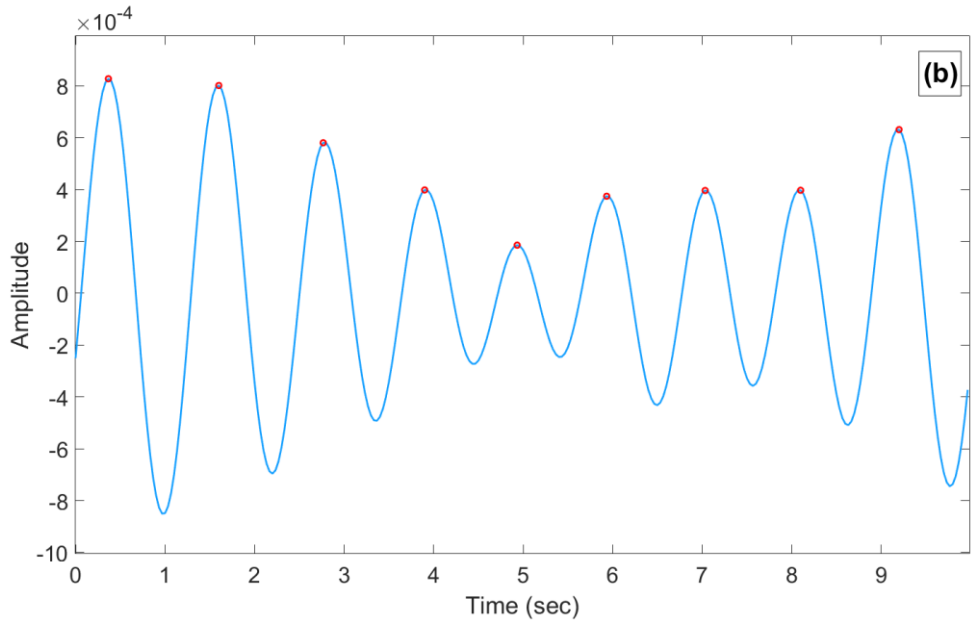


Figure 4.14: Respiratory signal. (a) Raw respiratory signal, (b) filtered respiratory signal.

To evaluate the proposed non-contact system, statistical methods based on the PCC, linear regression, Bland Altman plot, RMSE and MAE were taken into consideration with a total sample size of n=35. Figure 4.15(a) shows a strong correlation between the reference and measured data with PCC of 0.9836. As shown in the Bland Altman Plot in Figure 4.15(b), RPC) was 5.1 beats/min (3.4%), the mean bias was 0.69 beats/min, the lower and upper limit of agreement were -4.4 to +5.8 beats/min. The Bland-Altman plot comprises 35 points, as a result of considering 7 infants and collecting 5 samples per infant. For HR, RMSE was 2.66 beats/min, and MAE was 2.18 beats/min.

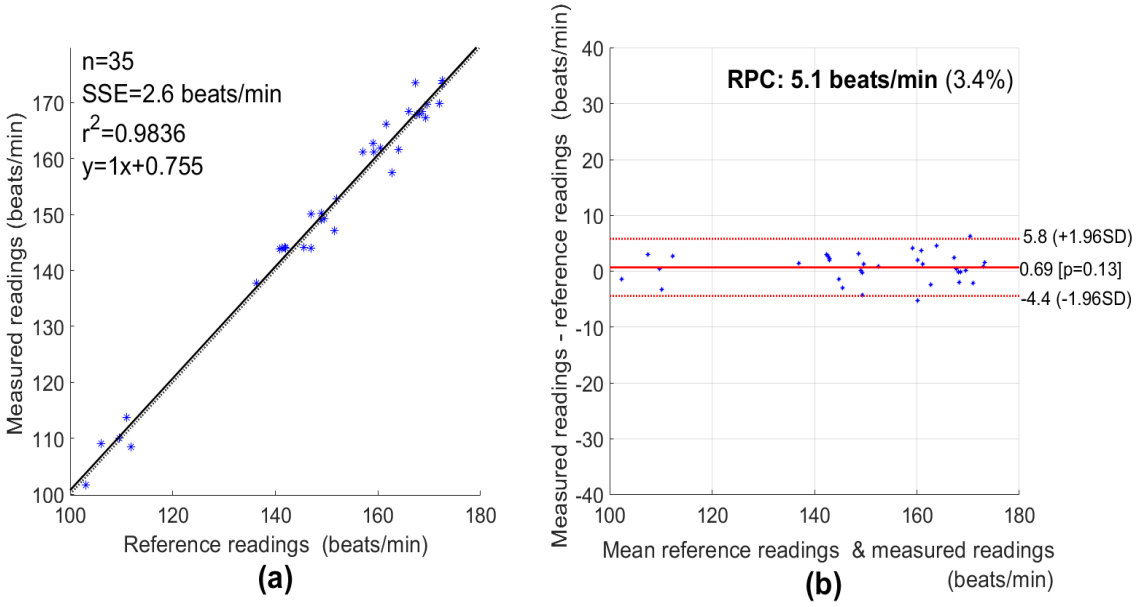


Figure 4.15: Statistical measurement for HR. (a) Correlation Plot, (b) Bland Altman plot.

Figure 4.16 represents the statistical measurement for RR. As shown in Figure 4.16(a), a strong correlation exists between the reference and measured data with PCC of 0.9426. Figure 4.16(b) represents the Bland Altman plot with mean bias 0.28 breaths/min, the lower and upper limit of agreement -5.5 to +6.1 breaths/min and RPC 5.8 breaths/min. For RR, RMSE and MAE were 2.93 and 2.45 breaths/min, respectively.

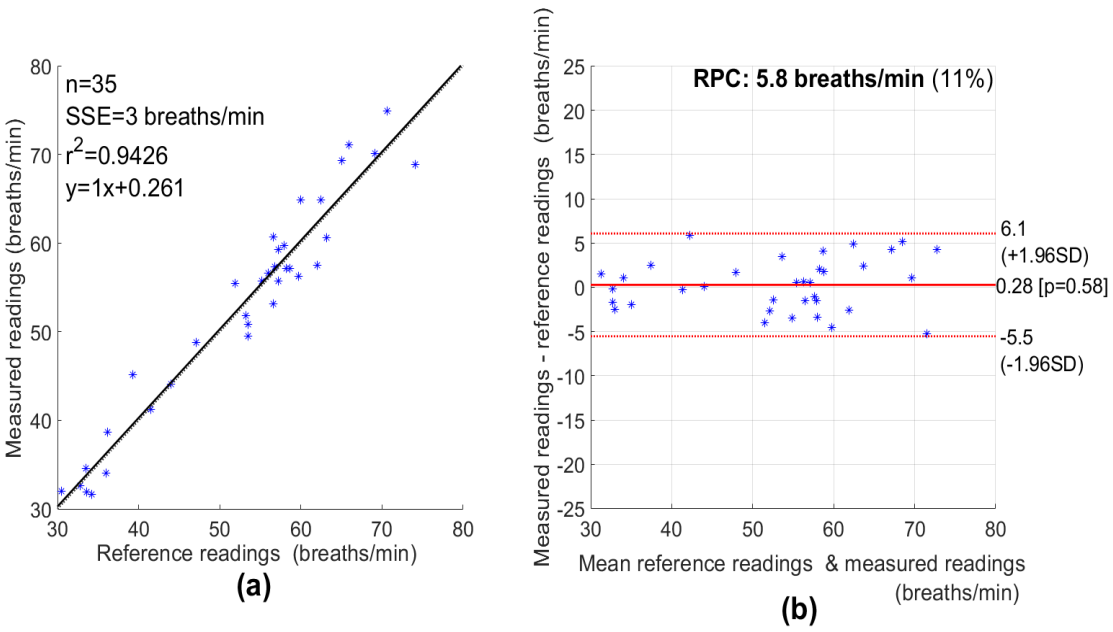


Figure 4.16: Statistical measurement for RR. (a) Correlation Plot, (b) Bland Altman plot.

4.5. Concluding Remarks

In this chapter, the HR and RR of infants in NICU were remotely monitored using a digital camera to overcome the limitation of conventional contact-based monitoring sensors. Both colour and motion-based methods have been used to measure HR and RR, respectively, which helps to measure RR even if the infant is covered with linen (i.e., ROI is unclear). The ROI was efficiently selected incorporating an image decomposition method rather than using manual ROI selection. Furthermore, an automatic ROI selection method has been used based on skin detection. Different signal processing techniques such as spatial averaging, spectral analysis, band-pass filtering, and peak detection were used to extract the vital signs.

A close correlation and small error rate between the measured data and ECG data demonstrated that video camera imaging technique may replace conventional contact-based monitoring in NICU and has potential applications in other areas such as remote home health monitoring.

CHAPTER 5

Non-contact Neonatal Vital Signs Monitoring Based on EEMD

5 Non-contact Neonatal Vital Signs Monitoring Based on EEMD

Infants with fragile skin are patients who would benefit from non-contact vital sign monitoring due to the avoidance of potentially harmful adhesive electrodes and cables. However, studies on non-contact vital signs monitoring of infants in the NICU are still limited. Moreover, most existing imaging technologies for vital sign detection are susceptible to noise artefacts caused by illumination variations. Therefore, in this chapter, A single-center study was conducted to remotely monitor HR and RR of infants in NICU using a digital camera. Moreover, a noise removal method based on a noise-assisted signal decomposition technique was used to improve the cardiorespiratory signal. In addition, the ROI was automatically selected using a neural network. The content of this chapter is based on materials and results published in [V, VI].

This chapter is structured into four key sections. Section 5.1 introduces the proposed system and describes the related work. Section 5.2 discusses methods and materials, including study design, the experimental setup for camera imaging-based monitoring in the NICU and the system framework. Section 5.3 contains the experimental results of the proposed system. Section 5.4 presents concluding remarks.

5.1. Introduction and Related Work

Continuous monitoring of vital signs is usually conducted in NICU using different monitoring equipment involving adhesive transducers or electrodes to be directly attached to the skin. Preterm infant skin is very sensitive and fragile, particularly for those born before 29 weeks of gestational age, when the bond between the dermis and attached sensor may be stronger than that between the epidermis and dermis [266]. As a result, the skin may be damaged. There is also a risk of enabling or introducing an infection [267].

As explained in section 4.1, some studies have considered contactless vital signs monitoring of infants using video cameras in a hospital environment to minimize the issues of existing contact-based methods. From section 4.1, it can be noticed that most of the studies are affected by noise artefacts caused by motion artefacts and illumination variation. Moreover, manual ROI selection is another issue with most of these studies.

The respiration of 30 preterm infants was monitored in NICU using a 3-CCD digital camera [268]. In the proposed method, a simple colour-based skin detector was used to segment the ROI. However, colour-based image segmentation is less consistent in a clinical setting. More robust image segmentation is required in order to develop a continuous monitoring system over longer periods of time.

Recent advances in deep learning research have yielded an outline to embed visual features within CNNs to produce highly precise classifiers in challenging segmentation scenarios [269]. In [270], a multi-task convolutional neural network model was presented to identify the presence of a subject and divide skin regions of the subject to estimate vital signs. This enabled continuous vital signs monitoring that can be performed automatically in challenging clinical situations. The proposed multi-task model had a shared core network with two branches: a patient detection branch employed by means of global average pooling and a skin segmentation branch using a convolutional network. This multi-task CNN model was used in [271, 272] to monitor the respiration of five neonatal patients and the vital signs (HR and RR) of 30 preterm infants in the NICU, respectively. The CNN model was extended in [273] by adding a body part detection branch to detect the body parts that are important to estimate the HR, such as the torso, head, and diaper area of the subject. The body part detection branch was executed by means of the Faster R-CNN network. However, computational cost and runtime of CNN are high.

In this chapter, using visible light video cameras, both HR and RR of neonates in the NICU were remotely monitored using a colour and motion-based method, respectively. A baby detection model was trained using the YOLO V3 weights [274] to detect ROI automatically. The YOLO weights were originally trained on the MS COCO dataset [275], which has 80 classes of objects present in it. YOLO works well with multi-scale detections and has shown better accuracy and speed than similar detection models 1000 times faster than R-CNN [276] and 100 times faster than Fast R-CNN [277]). Furthermore, a signal decomposition technique was used to minimize the noise effect using EEMD.

5.2. Methods and Materials

5.2.1. Study Design and Experimental Set-up

The study design and experimental set-up is the same as Chapter 4.2.1 and 4.2.2. The research in this chapter was conducted using seven infants. Figure 5.1 shows the experimental set up with a schematic diagram.



Figure 5.1: The experimental setup where the data recording was performed. A schematic of the setup is shown to the right of the image.

5.2.2. System Framework Using Neural Network Method

The overall system includes several image and signal processing techniques, such as automatic ROI selection, spatial averaging, signal decomposition, spectral analysis, band-pass filtering, and peak detection, as shown in Figure 5.2.

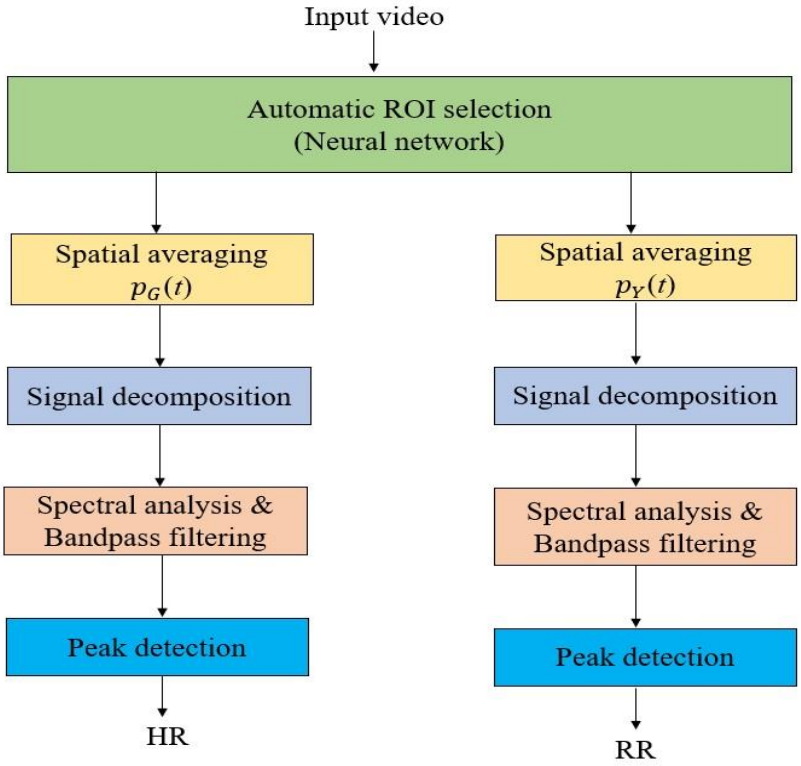


Figure 5.2: The system framework using neural network method.

5.2.2.1. Automatic ROI Selection

In a hospital setting, it is very challenging to record clear video of neonatal infants. Most of the time, their body is occluded with medical instruments or clothes. Datasets for deep learning and testing of algorithms must cover many such variations to be useful in a practical setting.

The collected images were primarily captured from hospital settings and depicted various poses of babies in different environments. The collected images were selected mainly to cover different sleeping positions of babies from different angles while they were fully visible or occluded.

Standard image processing techniques cannot be used for detecting babies in such complex images. There are popular people detector models (YOLO and Mask-RCNN, etc.) available for detecting people of any age, including babies. However, these detectors fail to detect babies in a complex hospital setting as they were trained to detect people in day-to-day situations. Such detectors were trained with very few or no baby images of interest. There are also detectors to detect skeletons and faces [278]. They work well in general situations

but fail to detect babies in a hospital setting. This is mainly due to the lack of similar images in the training dataset, occlusions, and lighting conditions.

Therefore, a baby detector was trained using a small dataset gathered from the internet. A total of 473 images were collected for training. The model was trained using the original YOLO V3 weights [274]. The network architecture is illustrated in Figure 5.3. The YOLO weights were originally trained on the MS COCO dataset [275], which has 80 classes of objects. The YOLO neural network segments the image into regions and predicts bounding boxes and probabilities for each region, as shown in Figure 5.4. These bounding boxes are weighted by the predicted probabilities. The proposed detector can detect the face of the baby receiving phototherapy producing a blue light source (Figure 5.4(b)).

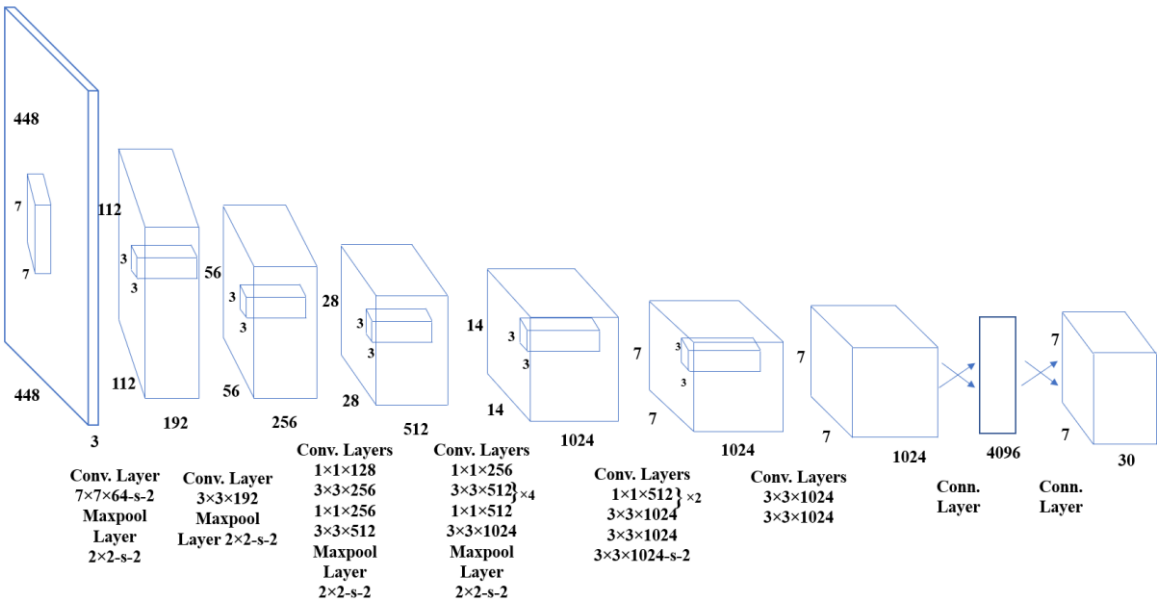


Figure 5.3: YOLO network architecture (adapted from [228]). The YOLO network has 24 layers followed by two fully connected layers.



(a) (b)

Figure 5.4: Automatic ROI selection using the YOLO neural network. The detected ROIs were shown in green bounding boxes. (a) Infant under normal light, (b) infant under blue light.

5.2.2.2. Spatial Averaging

Colour-based and motion-based methods were used to extract raw cardiac and respiratory signals, respectively. A colour variation in human skin is observed, as haemoglobin in blood absorbs illumination more than surrounding tissue, which may not be visible with the naked eye but can be detected by a video camera [5]. The colour changes reflect the blood volume changes in the microvascular tissue bed under the skin due to variation in pulsatile blood flow during each cardiac cycle. The cardiac signal can be extracted using this principle. Using the green channel of RGB colour space, a raw cardiac signal was extracted by averaging the brightness pixel values of the selected ROI as given in equation (5.1). The purpose of this was to minimize noise and improve the signal to noise ratio.

$$p_G(t) = \frac{\sum_{i,j \in ROI} P(i,j,t)}{|ROI|} \tag{5.1}$$

where $|ROI|$ represents the size of the detected ROI and $P(i,j,t)$ is the brightness value of a pixel at image location (i,j) at a time (t). To estimate the cardiac signal, the $p_G(t)$ signal was chosen as the green channel has the strongest cardiac frequency band compared to the other channels (red and blue) [125]. It was found from literature [126, 279] as well as from our experiments that green channel intensity images provide the best estimate for the heart rate. In addition, the red and blue channels were skipped to reduce impacts resulting from skin tone changes and reduce processing time.

Respiratory activity causes cyclic motion in specific body regions such as the head, nostril area, thoracic and abdominal region. In the video recording, spatial variations of intensity values directly indicate this motion. Therefore, the respiratory signal can be measured using this principle. As the video camera captured the video in the RGB colour space, the intensity data must be separated from the colour data. Therefore, the RGB colour space was changed to the YIQ colour space. From the Y channel of the YIQ colour space, the raw respiratory signal was measured by averaging the intensity values of the pixels within the selected ROI, as follows:

$$p_y(t) = \frac{\sum_{i,j \in ROI} P(i,j,t)}{|ROI|} \tag{5.2}$$

where $|ROI|$ presents the size of the detected ROI and $P(i,j,t)$ represents the intensity value of a pixel at image location (i,j) at a time (t).

5.2.2.3. Signal Decomposition

As explained earlier in section 3.4.3.1, signal decomposition techniques are applied in biomedical signal processing to separate a temporal signal into a collection of modes of interest. The most appropriate modes are then selected to represent the original signal. The most common signal decomposition techniques are EMD, EEMD, and CEEMD.

EMD is a multiresolution signal decomposition technique for a complex and multi-component signal representation developed by Huang et al. [231]. It is commonly used to remove noise artefacts from biomedical signals. The EMD uses the local temporal and structural characteristics of a non-linear and non-stationary signal and adaptively decomposes it into a set of stationary modes in different time scales, called intrinsic mode functions (IMFs). IMF is commonly used to remove noise artefacts from biomedical signals, such as the removal of noise artefacts from ECG data [232], the removal of noise artefacts from electromyogram data [233], the removal of muscle artefacts from electroencephalogram data [234], the removal of tissue artefacts from respiratory signals [235], and the removal of illumination variations from photoplethysmography signal [236].

In order to define a meaningful instantaneous frequency, the corresponding function should be symmetric (symmetry of the upper and lower envelopes) with respect to zero and have the same numbers of zero crossings and extrema. This type of function represents the oscillation mode imbedded in the data and is called an “intrinsic mode function”. Therefore, an IMF can be formally defined as follows.

The decomposition of the original signal, $x(t)$, into a set of IMFs must occur under two assumptions, as follows: (i) the number of extrema of $x(t)$ is either equal to the number of zero-crossing or differs at most by one and (ii) the mean values of the envelopes defined by local maxima and local minima are equal to zero.

The signal decomposition based on EMD is defined in the following steps:

1. Identify all peaks (maximum and minimum) of $x(t)$.

2. Generate the lower and upper envelopes of the peaks through cubic spline interpolation.

3. Calculate the mean value, $m(t)$, of the lower and upper envelopes point by point.

4. Extract the detail signal, $d(t)$, by subtracting the mean value $m(t)$ from $x(t)$:

$$d(t) = x(t) - m(t) \tag{5.3}$$

5. Verify the properties of $d(t)$:

a. If $d(t)$ meets the above two assumptions and becomes a zero-mean process, then it would be the first IMF component of $x(t)$, named IMF1, and replace $x(t)$ with the residue:

$$r(t) = x(t) - d(t) \tag{5.4}$$

b. Otherwise, go to step (1) and replace $x(t)$ with $d(t)$.

6. Repeat Steps 1 to 5 to obtain the IMF1, IMF2, IMFN, where N is the number of alterations. The process is stopped when $r(t)$ becomes a monotonic function, and no further IMF can be extracted.

As a result of the EMD process, $x(t)$ can be recovered by the following expression:

$$x(t) = \sum_{i=1}^{N} IMF_i + r_N(t) \tag{5.5}$$

where N is the number of IMFs and $r_N(t)$ is the residue of the signal $x(t)$.

In this work, EEMD was used, which is a noise-assisted signal decomposition technique proposed by Wu et al. [237], aimed at eliminating the mode-mixing problem caused by the original EMD because of intermittency. It is based on adding white Gaussian noise into the original signal, $x(t)$, in a controlled manner to obtain the noisy signal, $x_m(t)$, as follows:

$$x_m(t) = x(t) + \omega_m(t), \quad m = 1,2,\dots,L \tag{5.6}$$

where $\omega_m(t)$ is the mth added white noise and L is the ensemble number of the EEMD technique. The acquired original signal, $x_m(t)$, is then decomposed into a set of IMFs using EMD, and can be recovered as follows:

$$x_m(t) = \sum_{i=1}^{N} IMF_{i,m} + r_{N,m}(t), \tag{5.7}$$

$$x(t) = \frac{1}{L}\sum_{i=1}^{N}\sum_{m=1}^{L} IMF_{i,m} + \frac{1}{L}\sum_{m=1}^{L} r_{N,m}(t) \tag{5.8}$$

5.2.2.4. Spectral Analysis and Filtering

A spectral analysis technique based on the FFT was used to transform the time series cardiac and respiratory signals from the time domain to the frequency domain (see Figure 5.5). After that, two ideal band-pass filters were applied at 1.5 to 3 Hz, which correspond to the heart pulse range (90–180 beats/min), and 0.3 to 1.5 Hz, corresponding to the breathing range (18–90 breaths/min), to obtain the frequency band of interest. The inverse FFT was then applied to the filtered signals to obtain the time series cardiorespiratory signals.

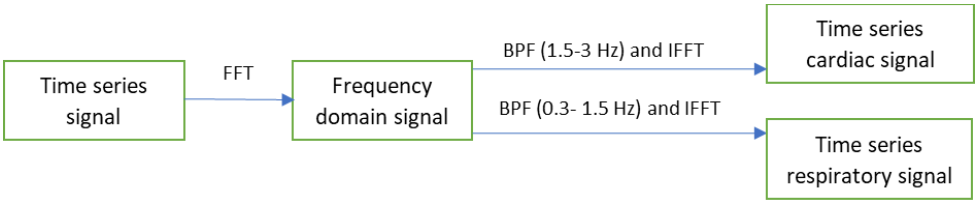


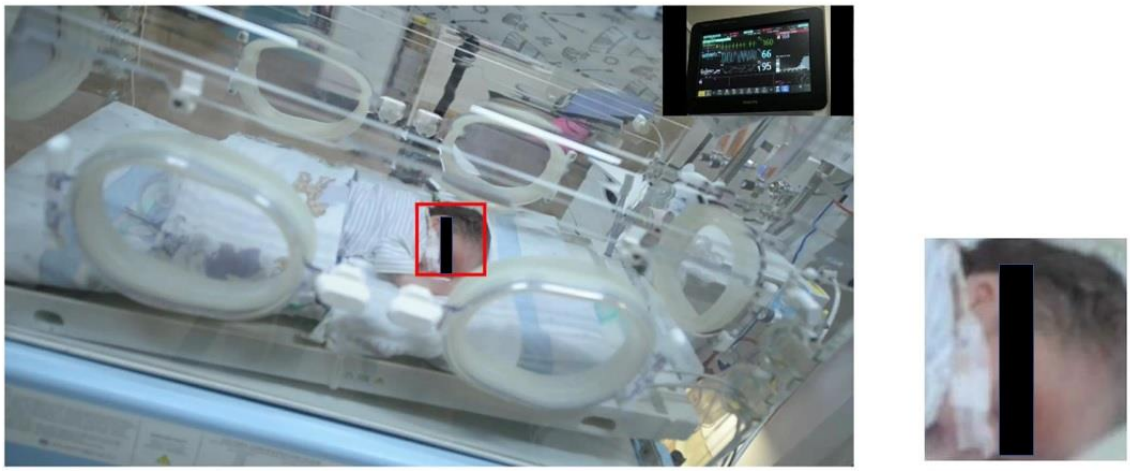
Figure 5.5: The spectral analysis and band-pass filtering process.

5.2.2.5. Peak Detection

The processing of time series cardiorespiratory signals includes calculating peaks and the distance between the consecutive peaks. MATLAB's built-in 'findpeaks' function was used for peak detection. The process is same as Chapter 4, section 4.3.1.5.

5.3. Experimental Results

MATLAB 2020a was used to implement our algorithm and calculate the statistical results. As shown in Figure 5.6(a), the first frame of the video was considered to detect ROI using the YOLO neural network. Figure 5.6(b) shows the detected ROI.



(a) (b)

Figure 5.6: An infant image with detected ROI is shown in (a). The corresponding ROI extracted from the original is shown in (b).

The raw cardiac and respiratory signals were obtained after performing spatial averaging over the detected ROI, as shown in Figure 5.7(a) and 5.7(b), respectively. Then, signal decomposition based on EEMD was applied to the raw cardiorespiratory signals. The window length of the EEMD was 10 s. The raw cardiac and respiratory signals were decomposed into IMF1, IMF2, …, IMF7 as presented in Figure 5.8(a) and 5.8(b). To select the best IMF that should be used for calculating the HR and RR, the frequency spectral analysis of the decomposed IMFs was performed using FFT, as shown in Figure 5.9. Figure 5.9 shows the spectrum of IMFs 3-6, which have the best frequency bands of interest that fall within the cardiorespiratory range, while other spectra fall outside this range. The IMF3 (Figures 5.8(a) and 5.9(a)) and IMF4 (Figures 5.8(b) and 5.9(b)) were selected for calculating HR and RR, respectively, as their highest frequencies are close to the frequency of normal HR and RR for infants.

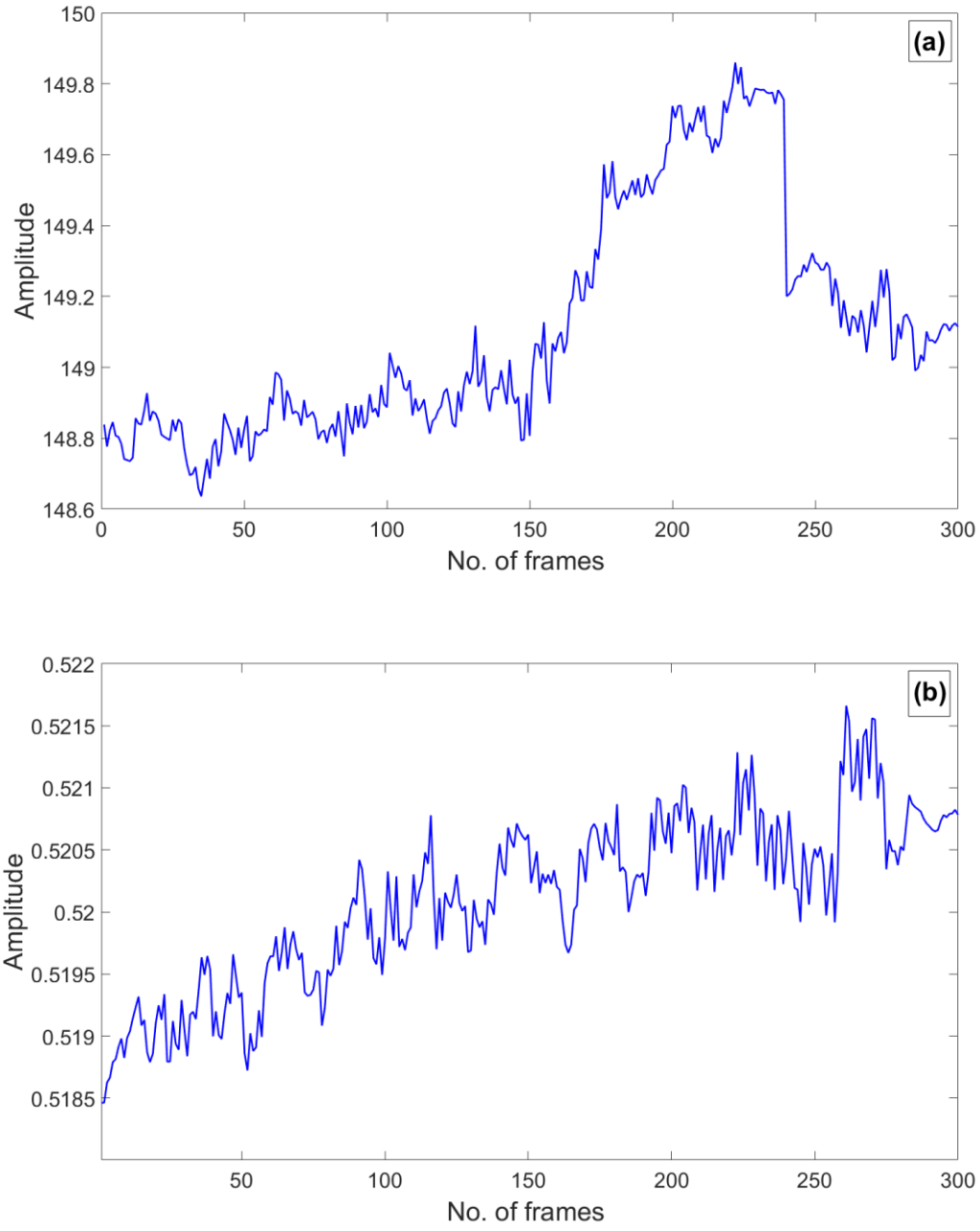


Figure 5.7: Raw cardiorespiratory signals for 300 frames are shown in the figure. (a) Raw cardiac signal, (b) raw respiratory signal.

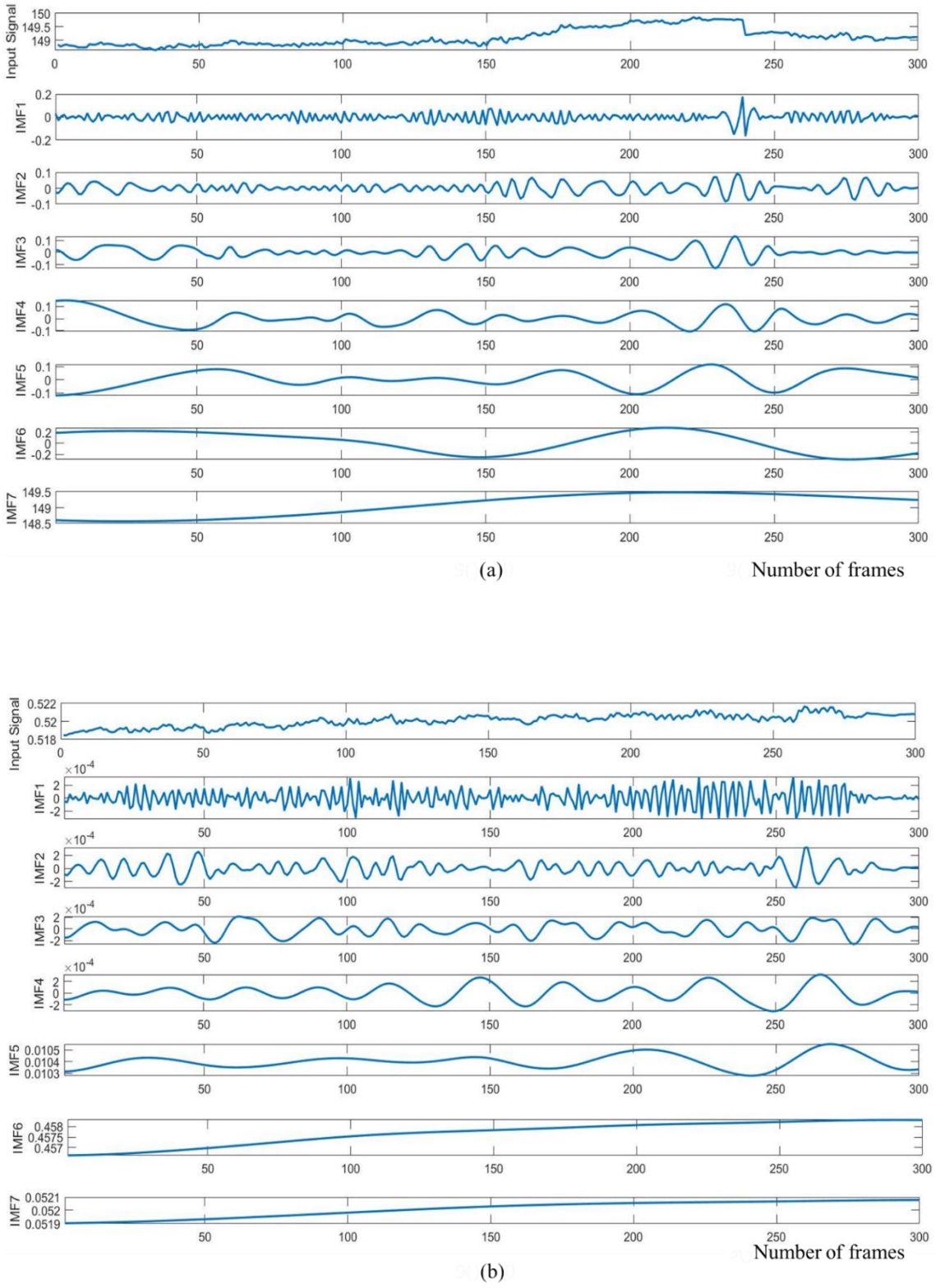


Figure 5.8: IMF components of the raw cardiorespiratory signals using the EEMD technique. (a) Cardiac signal, (b) respiratory signal.

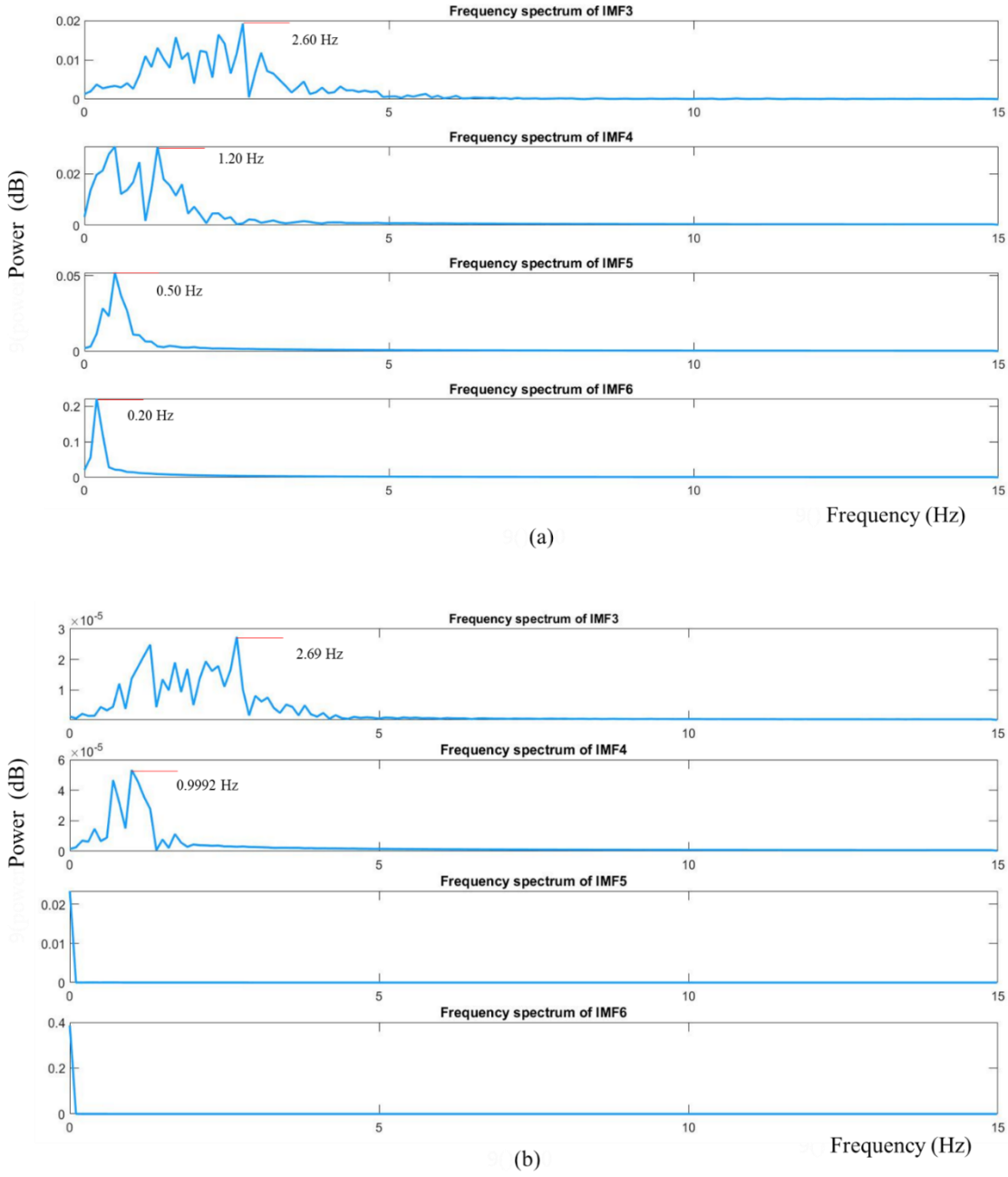


Figure 5.9: The frequency spectrum of decomposed IMF3, IMF4, IMF5 and IMF6. (a) Cardiac signal (b) respiratory signal.

Figure 5.10(a) demonstrates the filtered cardiac signal achieved after FFT, band-pass filtering and inverse FFT. The window length of the FFT was 10 s. From the filtered signal, the HR was calculated by calculating the number of peaks. Similarly, Figure 10(b) presents the filtered respiratory signal. By calculating the number of peaks, RR could also be calculated from the filtered respiratory signal.

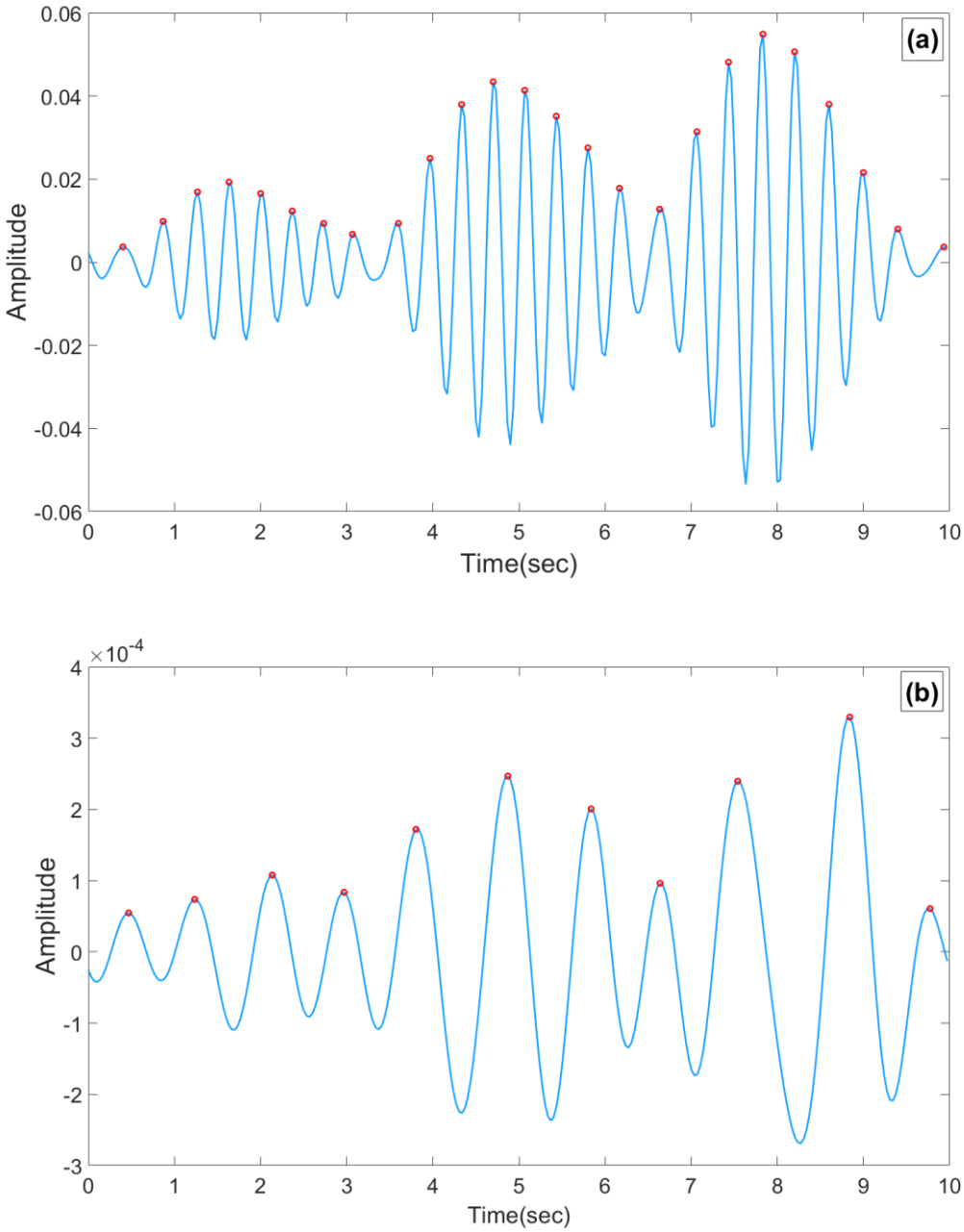


Figure 5.10: The filtered cardiorespiratory signals are shown in the figure. The red colour markers indicate the peak locations of the filtered signal. (a) Filtered cardiac signal, (b) filtered respiratory signal.

To evaluate the proposed non-contact system, statistical methods including the PCC, linear regression, Bland–Altman plot, RMSE, and MAE were taken into consideration, with a total sample size of n = 35. Figure 5.11(a) shows a strong correlation between the reference and measured data with PCC of 0.9864. As shown in the Bland–Altman Plot in Figure 5.11(b), the RPC was 4.3 beats/min (3%), the mean bias was 0.44 beats/min, the lower and

upper limits of agreement were -3.9 and +4.8 beats/min. For HR, the RMSE was 2.22 beats/min, and the MAE was 1.80 beats/min.

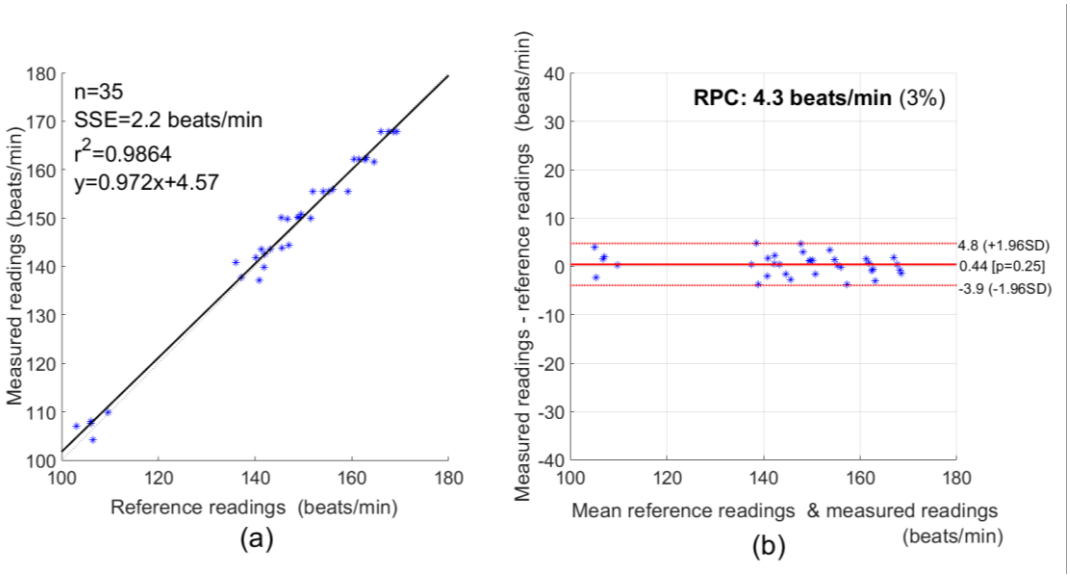


Figure 5.11: Statistical measurement for HR. (a) Correlation Plot, (b) Bland–Altman plot.

Figure 5.12 represents the statistical measurement for RR. As shown in Figure 5.12(a), a strong correlation exists between the reference and measured data with a PCC of 0.9453. Figure 5.12(b) represents the Bland–Altman plot with a mean bias of 0.71 breaths/min, a lower and upper limit of agreement of -4.5 and +5.9 breaths/min, and an RPC of 5.2 breaths/min. For RR, the RMSE and MAE were 2.69 and 2.13 breaths/min, respectively.

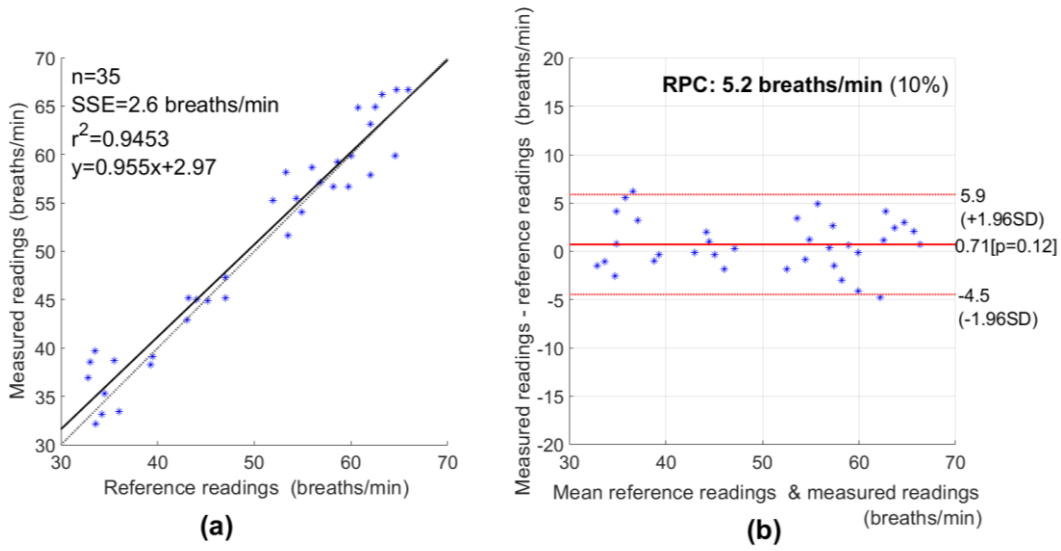


Figure 5.12: Statistical measurement for RR. (a) Correlation Plot, (b) Bland–Altman plot.

In Table 5.1, the Bland–Altman data for HR was compared to some of the state-of–the–art methods. The proposed method showed better results than other methods for all three performance measures compared in the table.

Table 5.1: Comparison with state-of-the-art methods for HR.

Methods	Lower Limit	Upper Limit	Mean Bias
Scalise et al.[165]	-9.79	7.99	0.90
Aarts et al. [18]	-5	+5.5	-
Gibson et al. [167]	-8.3	17.4	4.5
Proposed method	**-3.9**	**+4.8**	**0.44**

It could be concluded that the proposed system, for both HR and RR, showed a strong correlation with the reference method and a lower error rate.

5.4. Concluding Remarks

In this chapter, the HR and RR of infants in the NICU were measured by using video camera imaging. An automatic ROI selection based on a convolutional neural network was used instead of using manual ROI selection. Moreover, to minimize noise artefacts, a signal decomposition technique based on EEMD was also considered. The experimental results showed a strong correlation with PCC values of 0.9864 and 0.9453 for HR and RR, respectively, and a lower error rate with RMSE values of 2.23 beats/min and 2.69 breaths/min and MAE values of 1.80 beats/min and 2.13 breaths/min between measured data and reference data. A Bland–Altman analysis of the data also presented a close correlation between measured data and reference data for both HR and RR. The proposed system outperformed the conventional measurement methods.

As a result, it can be concluded that this non-contact method has valuable potential as a non-contact, economical and easily deployable monitoring system for use in clinical environments. Still, it also shows a potential application for remote, home health monitoring.

CHAPTER **6**

Non-contact Respiration Rate Monitoring Integrating RGB-Thermal Cameras

6 Non-contact Respiration Rate Monitoring Integrating RGB-Thermal Cameras

Thermal camera imaging techniques have advantages over other non-contact techniques as they are not affected by illumination variation and may work in complete darkness. However, though there are different face detection algorithms in the visible band, it is difficult to detect and track face and facial tissue in the thermal band owing to a lack of textural and geometric details. Therefore, in this chapter, a non-contact respiration rate monitoring system is proposed that combines RGB and thermal imaging systems based on RGB-thermal image alignment. In the proposed method, the ROI was detected automatically using an advanced image processing algorithm. The content of this chapter is based on materials and results submitted as a paper to a journal [VII].

This chapter is organised as follows. Section 6.1 introduces the proposed system and describes the related work. Section 6.2 provides materials and protocol, including the ethics protocols, experimental setup, and data acquisition process. Section 6.3 explains the methods involved in the proposed RR measurement system using thermal and RGB sensors. The experimental results and discussion are presented in sections 6.4 and 6.5, respectively. Finally, section 6.6 describes the concluding remarks.

6.1. Introduction and Related Work

Respiration rate, along with heart rate and body temperature, is an important indicator of human health condition [14, 280]. Different diseases, including respiratory infections [281], central nervous diseases [282], cerebral circulatory disorders [283] and sleep disorders [119], may cause abnormal respiration dynamics, which can be identified either by unusual respiratory sound, atypical breathing waveform, or abnormal respiration rate. Moreover, the RR is important in other sectors, such as sports studies [284], stress measurement [5], influenza detection at airports [285], security inspections [286] and quarantine centres [287].

Conventional RR monitoring devices such as ECG, respiratory belt transducer and piezoelectric respiratory sensor [6] necessitate direct physical contact with the human body which may cause stress, discomfort, pain, skin damage and infection [7]. In contrast, thermal camera imaging techniques have advantages over other non-contact techniques as

these techniques are not constrained to illumination variation and may function in complete darkness. However, though there are different face detection algorithms in the visible band, it is difficult to identify and track face and facial tissue in the thermal band owing to a lack of textural and geometric details. Due to the red (R), green (G), and blue (B) channels, the visible images also known as RGB images, present richer information about the participant's face than the thermal images. They contain more geometric and textural details and boast a higher spatial resolution when compared to their thermal counterparts, which primarily capture radiated heat from the face [288]. Besides, the thermal images are constrained in their spatial resolution, and thermal cameras with even modest resolution are expensive. In practice, visible images are typically preferred to select, track, and classify face and facial tissue more precisely than thermal imaging. Therefore, multi-spectral face and facial tissue detection may offer the potential to trace and track the ROI accurately and faster in thermal images, leveraging the richer information provided by visible images.

In the literature, thermal and video camera imaging have been combined to enhance the reliability and performance of non-contact vital signs monitoring systems. For example, Hu et al. [131] proposed a dual-mode imaging system combining a thermal imager and an RGB camera to monitor respiratory rate and pattern remotely using affine transformation. The cascade object detector using the Viola–Jones algorithm along with a screening technique based on biological properties were applied to locate three ROIs in the RGB image. However, it was not possible to detect ROI if the ROI was occluded. Using a thermal imager and an infrared camera, another dual-camera imaging system was developed [132] to monitor RR and HR at night remotely. A custom-built cascade face classifier, comprising the conventional Adaboost model and fully convolutional network, was applied to identify faces in infrared images. However, in these systems a fixed dual imaging arrangement was used where the two cameras were permanently connected to each other. The registration parameters are constrained to the camera setup and the subject-to-camera distance [289].

Negishi et al. [290] introduced a stable infection screening system to monitor HR and RR in a non-contact manner using an RGB camera and a thermal camera. Haar feature-based cascade classifiers and the mean shift algorithm were used to detect and track the face in RGB images. Dlib and coordinate conversion were incorporated to locate the nasal area in the thermal images. Scebba et al. [291] also used a cross-spectrum ROI selection scheme

to monitor RR and apnoea by fusing far-infrared (FIR) and near-infrared (NIR) cameras. Yang et al. [292] monitored BT, HR and RR in a non-contact manner, incorporating an RGB and a thermal camera for individuals with and without face masks. To estimate vital signs, a RetinaFace detector based on CNN was utilised to detect the face, and three ROIs were traced using facial landmarks. However, the CNN algorithm is complex and computationally expensive.

In this chapter, RR was measured in a non-invasive and non-contact manner by combining RGB and thermal imaging systems. First, RGB and thermal images were aligned. Then, the ROI was detected automatically from RGB images, and a linear coordinate mapping was done to select ROI from thermal images. Instead of using manual ROI selection, an automatic ROI selection algorithm was designed, including face detection, face landmark detection and ROI, i.e., nostril area selection. Finally, RR was extracted from the selected ROI using signal processing techniques.

6.2. Materials and Protocol

6.2.1. Ethics Protocol

The research procedure illustrated in this chapter followed the ethical tenets of the Declaration of Helsinki (Finland 1964). This study was approved by the University of South Australia Human Research Ethics Committee (Adelaide, South Australia, Protocol number: 202973). After providing a complete explanation of the research procedure, written consent from the participants was obtained before commencing the experiment.

6.2.2. Experimental Setup and Data Acquisition

Figure 6.1 shows the experimental setup of the proposed system. A FLIR A400 camera with both thermal and RGB sensor options was used. Thermal videos were acquired from the thermal sensor with a resolution of 320 × 240 and a frame rate of 30fps. The RGB videos were acquired using a visible light camera with a resolution of 640 × 420 and a frame rate of 30fps. The camera was mounted on a tripod during data collection. The camera was connected to a PC, and videos were acquired using the FLIR Research Studio Software.

A Go Direct® Respiration Belt (Vernier Software & Technology, Beaverton, Oregon, United States) was considered as the reference method to validate the performance of the

proposed non-contact monitoring system. During respiration, the force exerted by the abdomen or chest is measured by the respiration belt with a resolution of 0.01 N, range of 0-50N and response time of 50 ms.

Videos of 10 adult participants (both male and female) with different ages and various skin tones were recorded. During video recording, the participants were seated on a chair and fitted with a respiratory belt on their chest to monitor their breathing. The 90 second videos were subsequently segmented into 30 second windows for analysis. The distance between the participant and the camera was approximately 75 cm.



Figure 6.1: Experimental setup of the proposed system.

6.3. Methods

6.3.1. RR Measurement System Using Thermal Sensor

For detecting respiration rate, the main principle of the thermal imaging technique is that the temperature around the mouth and nostril area varies as a result of the respiration process. Generally, the nostril ROI is chosen over consecutive frames, and the respiratory signal is extracted by averaging the pixel value within the selected ROI. Nevertheless, thermal images give less geometric and textural information than RGB images, which makes face and facial tissue detection algorithms ineffective in thermal imaging. To minimise this issue, the RGB image is combined with the thermal image to detect face and facial tissue precisely [69]. The general system framework of the proposed RR measurement system is shown in Figure 6.2.

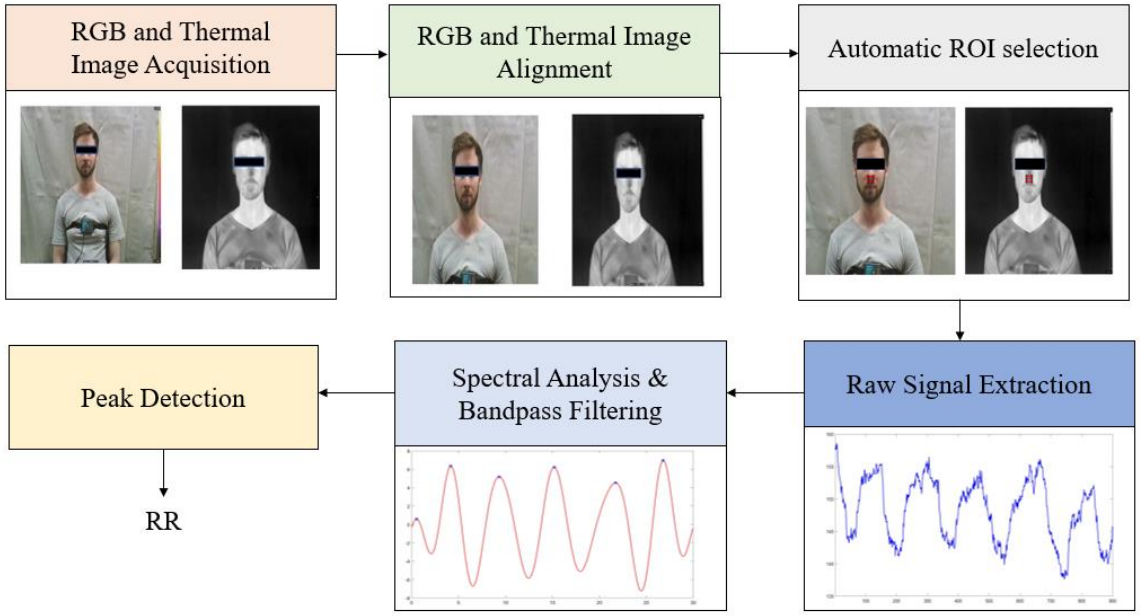


Figure 6.2: Overview of the proposed RR measurement using the thermal sensor.

6.3.1.1. RGB-Thermal Image Acquisition

First, thermal and RGB videos were recorded simultaneously using a thermal and an RGB camera. Then, the videos were converted into frames for further processing.

6.3.1.2. RGB and Thermal Image Alignment

To enhance the robustness of camera-based RR monitoring, it is necessary to align the images in the same coordinate in the multi-spectral camera-based system so that complementary features can be used. As there was a variation in camera parameters such as pixel size, focal length, etc., the acquired RGB and thermal images expose significant scale differences. Therefore, image alignment was done by using the first frame of recorded videos. First, the thermal images were scaled down, and then the RGB images were cropped using the same scale as the thermal images. This process is repeated for all frames. Figures 6.3 (a) and (b) depict the original RGB and thermal images, respectively, while Figures 6.3 (c) and (d) illustrate the aligned RGB and thermal images, respectively.

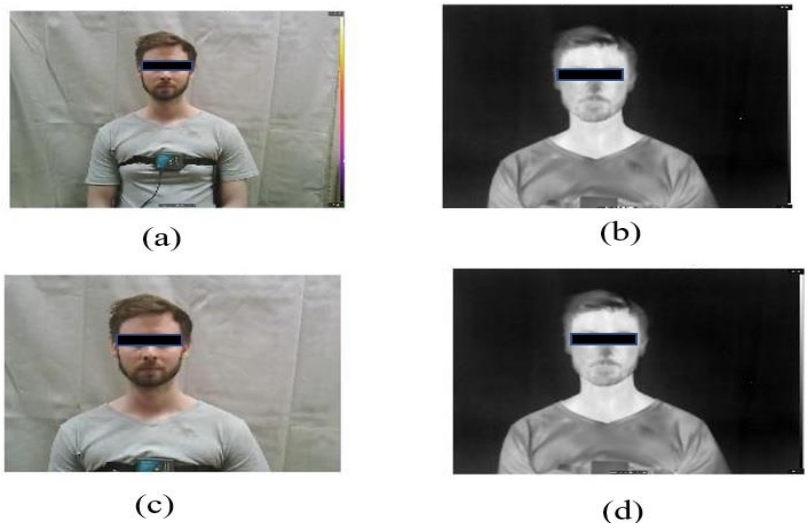


Figure 6.3: RGB and thermal image alignment. (a) Original RGB image, (b) Original thermal image, (c) aligned RGB image, (d) aligned thermal image.

6.3.1.3. Automatic ROI Selection

Figure 6.4 shows the ROI selection algorithm divided into four parts: face detection, face landmark detection, ROI selection from RGB and ROI selection from thermal.

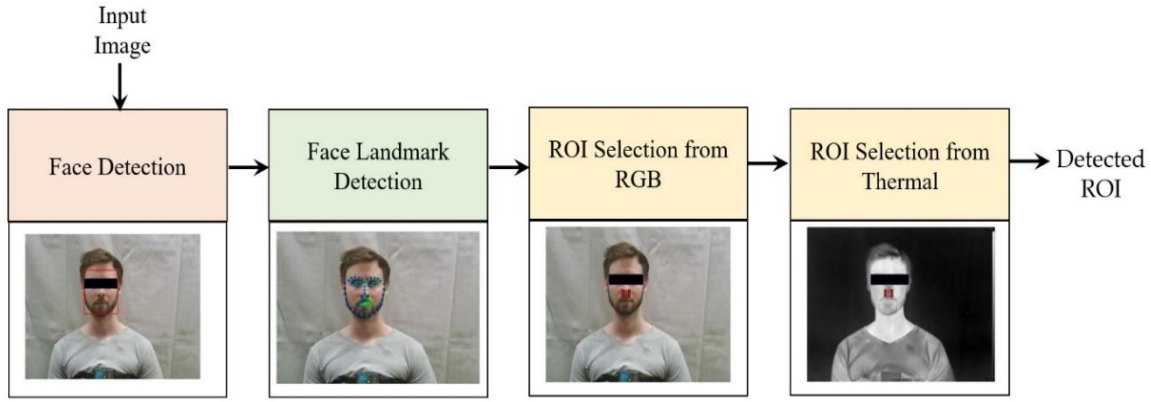


Figure 6.4: Overview of the proposed automatic ROI selection technique.

In this study, face detection was accomplished using the HOG (Histogram of Oriented Gradients) and a linear SVM (Support Vector Machine) based face detector from the Dlib library [293]. In this computing environment this face detector was faster than CNN-based methods. Since the participants were in close proximity to the camera, their faces were clearly visible in the images, making both CNN and HOG-based methods accurate for face detection. While CNNs are considered state-of-the-art for face detection, the decision to use a HOG-based detector in this study was motivated by the need for processing speed.

It's worth noting that HOG-based methods are not as invariant to variations in viewing angle and rotation as CNNs. The detected face is depicted in Figure 6.5(a).

Next, the 68 point facial landmarks of the detected face were identified, as shown in Figure 6.5(b). The landmarks encompassed the eyes, eyebrows, nose bridge, lower nose, jawline, outer lip, and inner lip. The ROI was defined as the nostril, and landmark points 32 (left nostril) and 36 (right nostril) were utilised to delineate the bounding box around the nostril area, as shown in Figure 6.6(a).

After that, the corresponding locations of the nostril area could be automatically located in the thermal images by linear coordinate mapping and drawing the bounding box around the nostril area, as shown in Figure 6.6(b).

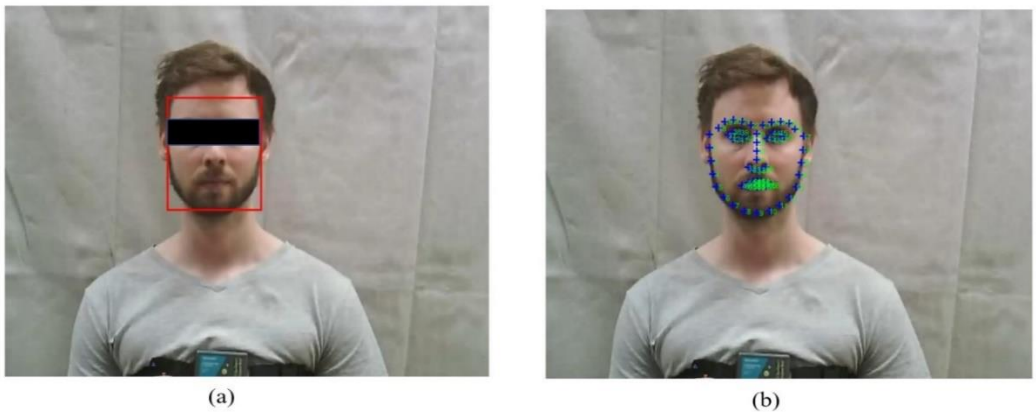


Figure 6.5: Face detection and landmark selection. (a) face detection and (b) landmarks selection.



Figure 6.6: Automatic ROI selection. (a) ROI selection from RGB image, (b) ROI selection from thermal image.

6.3.1.4. Raw Signal Extraction

During respiration, i.e., inspiration and expiration, oxygenated air is inhaled into the human body, and carbon dioxide-enriched air is exhaled from the body through the mouth or nostril area [14]. The inhaled air temperature is nearer to ambient temperature, while the temperature of exhaled air is nearer to body temperature. Therefore, the temperature of exhaled air is higher than the inhaled air, which creates a temperature variation around the mouth or nostril area. In the thermal imaging technique, this temperature variation is used to measure the RR.

Thermal variation around the nostril area was used to calculate RR. After selecting the nostril area, the raw respiratory signal was extracted by averaging the pixel intensity within the selected ROI using the following equation.

$$r_T(t) = \frac{\sum_{i,j \in ROI} R(i,j,t)}{|ROI|} \tag{6.1}$$

where $R(i,j,t)$ represents the pixel value of the thermal image at image location (i,j) at a time (t), and $|ROI|$ is the size of the chosen ROI. In brief, the raw respiratory signal was obtained by the average temperature of ROI for all frames.

6.3.1.5. Spectral Analysis and Bandpass Filtering

A spectral analysis method utilizing the FFT was employed to transform the time-varying respiratory signal from the time domain into the frequency domain. Subsequently, an ideal bandpass filter was employed within the frequency range of 0.15 to 0.5 Hz, corresponding to the typical breathing rate range of 9 to 30 breaths per minute, in order to capture the desired frequency band. Finally, to retrieve the time-domain respiratory signal, the inverse FFT was utilized on the filtered respiratory data.

6.3.1.6. Peak Detection

The processing of time series respiratory signals involves calculating peaks and the detection of the distance between the consecutive peaks within the signal. The MATLAB built-in function 'findpeaks' was used to detect peaks. After detecting the peaks and their locations (locs), the total length of the cycle (C_L) between two consecutive peaks can be calculated using the following equation.

$$C_L = mean\big(diff(locs)\big) \tag{6.2}$$

number of peaks (M) can be found using

$$M = \frac{t}{C_L} \tag{6.3}$$

where t represents the length of the video recording in seconds.

Using the following equation, the respiratory rate per minute could be measured:

$$C_v = \frac{60M}{t} \tag{6.4}$$

where C_v represents the measured value.

6.3.2. RR Measurement System Using RGB Sensor

Figure 6.7 represents the overview of the proposed RR measurement using an RGB sensor which includes four main parts: ROI selection, raw signal extraction, spectral analysis & bandpass filtering, and peak detection.

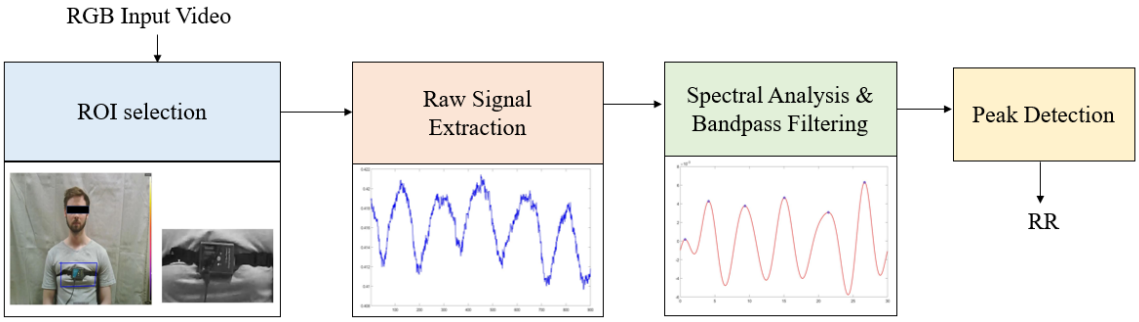
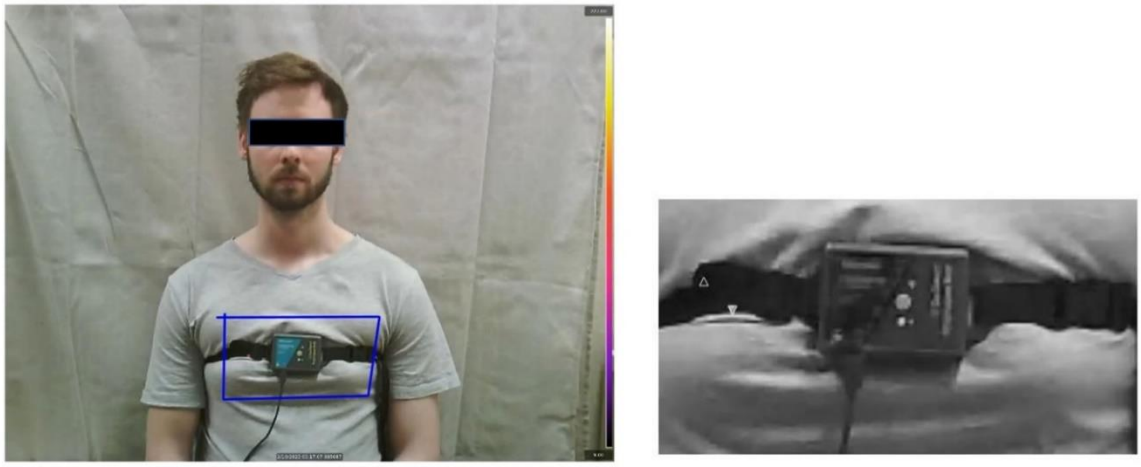


Figure 6.7: Overview of the proposed RR measurement using an RGB sensor.

6.3.2.1. ROI Selection

After collecting the RGB videos, videos were converted into frames. The first frame of the raw video was used for ROI selection. The chest ROI was manually selected from the RGB image using MATLAB's built-in function 'ginput' (Figure 6.8). The process was repeated for all frames.



(a) (b)

Figure 6.8: ROI selection from RGB image. (a) ROI selection from the image. (b) selected ROI.

6.3.2.2. Raw Signal Extraction

During respiration, inhalation and exhalation cause variations in the thoracoabdominal wall position. In the video recording, spatial differences in intensity values directly show the variation in the thoracoabdominal wall movement caused by respiration. Thus, respiratory signals can be extracted using this principle. Using the built-in function of MATLAB, called 'rgb2ntsc' the RGB colour space was converted to the YIQ colour space to separate intensity information from colour information, and the Y channel of the YIQ was selected. Then, to extract the raw respiratory signal, the mean of pixel intensity values was calculated within the chosen ROI using the following equation,

$$r_R(t) = \frac{\sum_{i,j \in ROI} R(i,j,t)}{|ROI|} \tag{6.5}$$

where $R(i,j,t)$ represents the intensity value of a pixel at image location (i,j) at a time (t), and $|ROI|$ represents the size of the selected ROI.

6.3.2.3. Spectral Analysis and Bandpass Filtering

The time series respiratory signal underwent a transformation from the time domain to the frequency domain through a spectral analysis technique utilising the FFT. Subsequently, to isolate the desired frequency range corresponding to the breathing rate (9-30 breaths/min), an ideal bandpass filter was employed with a cutoff at 0.15 to 0.5 Hz. Finally, the inverse FFT was used to convert the filtered respiratory signals back into the time domain, resulting in the reconstructed time series respiratory signal.

6.3.2.4. Peak Detection

The periodicity of peaks, the number of peaks, and peak locations from the filtered respiratory signals were obtained by using the MATLAB built-in function 'findpeaks'. The respiratory rate per minute can be calculated using equations (6.2-6.4).

6.4. Experimental Results

This section represents the experimental results of the proposed non-contact system for data collected from both thermal and RGB cameras. The experimental results of the proposed system were evaluated and validated against the measurements achieved from the reference method. Our algorithm was implemented, and the statistical results were calculated using Python 3.9 and MATLAB 2020a. To determine the level of correlation, limits of agreement (LoA), and error rate, the PCC, Bland-Altman analysis, MAE, and RMSE were performed on the data collected with a total sample size of n=30 for both thermal and RGB data.

6.4.1. Results Using Thermal Data

The raw respiratory signal from thermal data was obtained by employing spatial averaging over the selected ROI, as demonstrated in Figure 6.9(a). Subsequently, the filtered respiratory signal, depicted in Figure 6.9(b), was obtained through FFT, bandpass filtering, and inverse FFT. The RR was extracted by identifying the number of peaks from the filtered respiratory signal.

Figure 6.10 represents the statistical measurement between thermal and reference data. Figure 6.10(a) exhibits a close correlation between the reference and measured thermal data with a PCC of 0.9119. In Figure 6.10(b), the Bland-Altman plot shows that the reproducibility coefficient (RPC) was 2.5 breaths/min (15%), the mean bias was 1.5 breaths/min, the lower and upper limits of agreement were -0.9 to +4.0 breaths/min. RMSE was 1.97 breaths/min, and MAE 1.60 was breaths/min between the reference and measured thermal data.

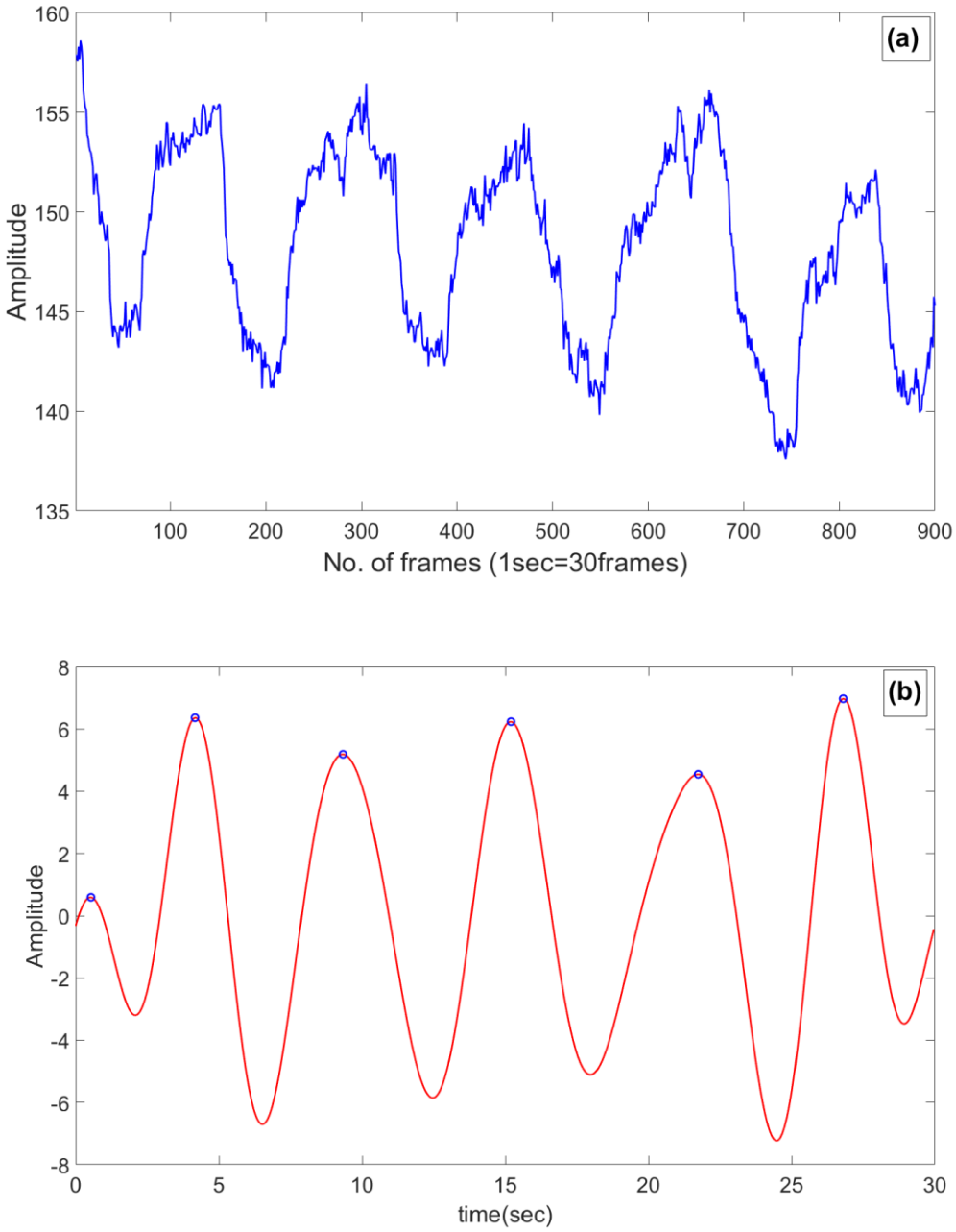


Figure 6.9: Respiratory signal obtained from thermal data. (a) Raw respiratory signal, (b) filtered respiratory signal.

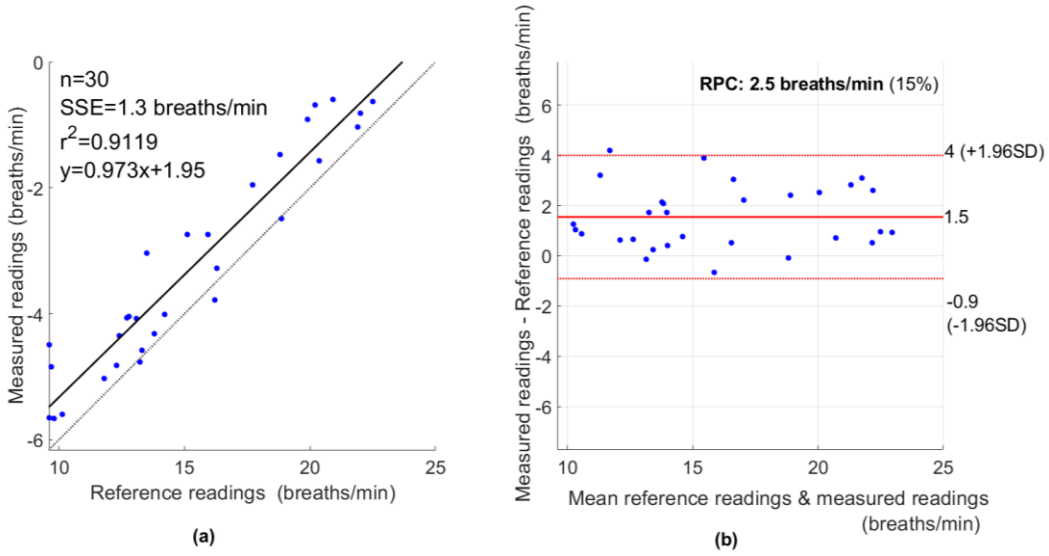
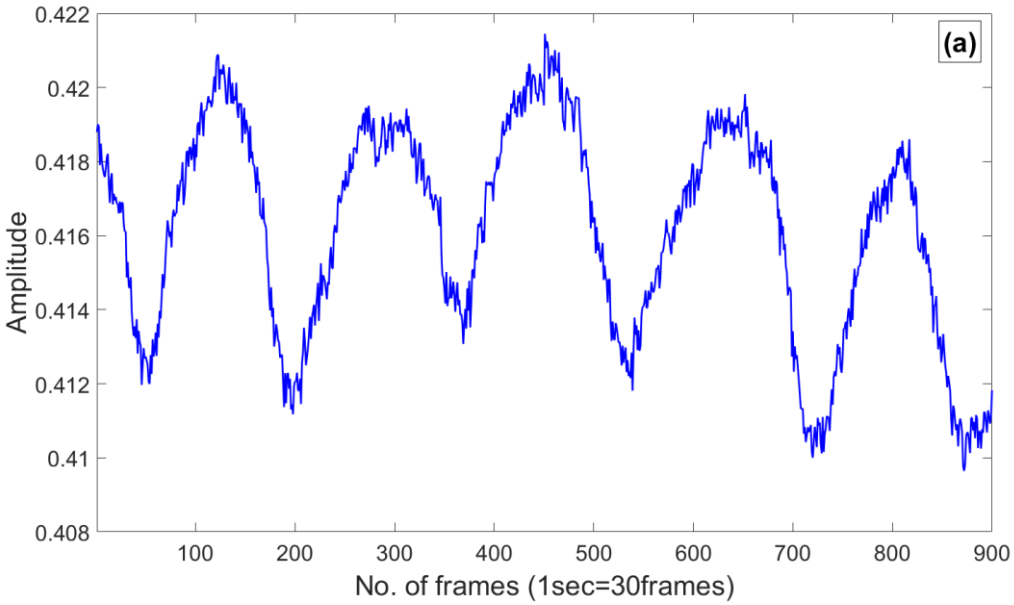


Figure 6.10: Statistical measurement between reference and thermal data, (a) the correlation plot and (b) the Bland-Altman plot.

6.4.2. Results Using RGB Data

Figures 6.11(a) and 6.11(b) represent the raw respiratory signal and filtered respiratory signal from RGB data, respectively. RR was extracted by computing the number of peaks from the filtered respiratory signal.



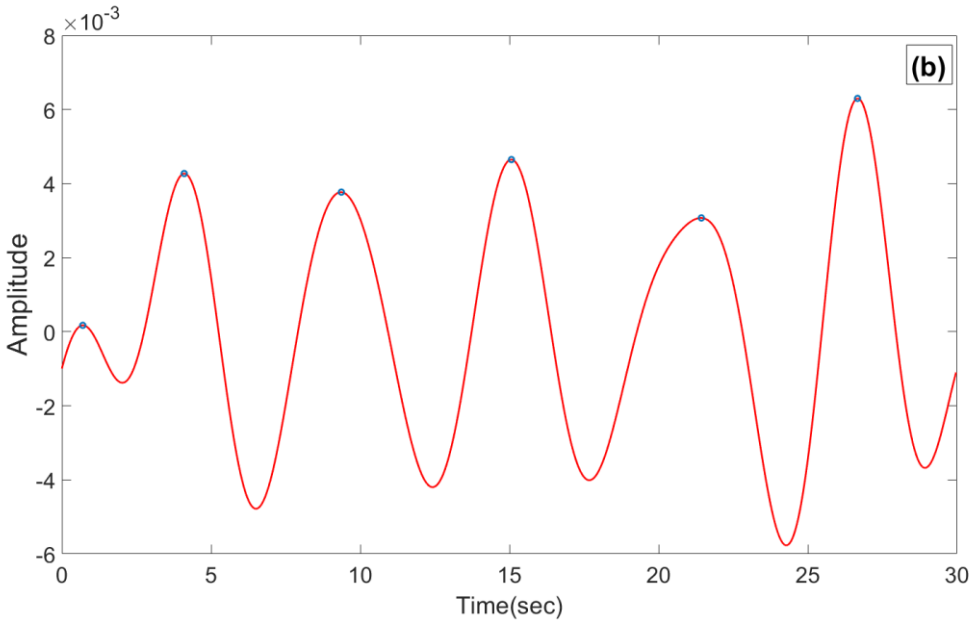


Figure 6.11: Respiratory signal obtained from RGB data, (a) raw respiratory signal, (b) filtered respiratory signal.

Figure 6.12 presents the statistical measurement between RGB and reference data. Figure 6.12(a) shows a close correlation between the reference and measured RGB data with a PCC of 0.9340. Figure 6.12(b) denotes the Bland Altman plot with a mean bias of 1.5 breaths/min, the lower and upper limits of agreement -0.54 to 3.60 breaths/min, and RPC of 2.1 breaths/min. RMSE and MAE were 1.87 and 1.61 breaths/min, respectively, between the reference and measured RGB data.

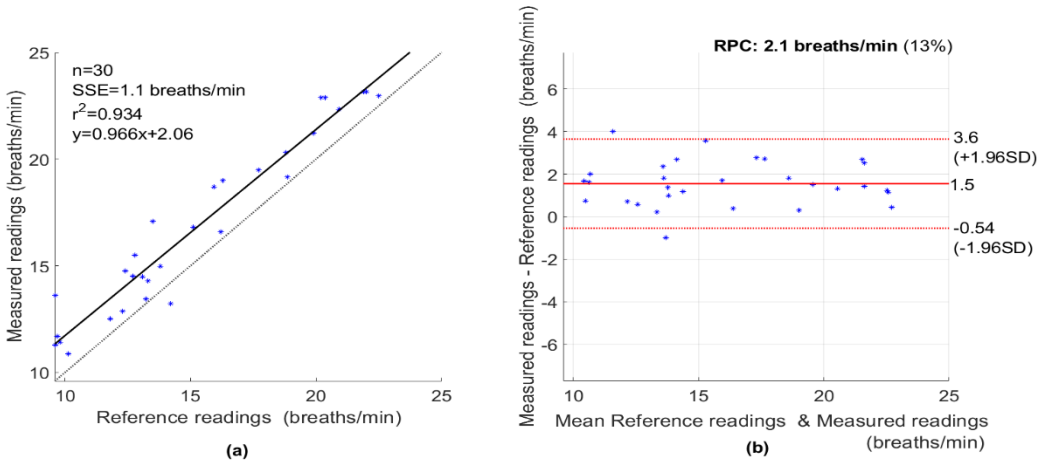


Figure 6.12: Statistical measurement between reference and RGB data, (a) the correlation plot, (b) the Bland-Altman plot.

6.4.3. Comparison between Thermal & RGB data

Figure 6.13 denotes the statistical measurement between RGB and thermal data. Figure 6.13(a) demonstrates a strong correlation between RGB and thermal data with a PCC of 0.9514. The Bland Altman Plot in Figure 6.13(b) represents that the RPC was 1.8 breaths/min (11%), the mean bias was -0.0033 breaths/min, the lower and upper limit of agreement were -1.8 to +1.8 breaths/min between RGB and thermal data which means there is a strong correlation between them. RMSE was 0.91 breaths/min, and MAE was 0.73 breaths/min between the RGB and thermal data.

It can be concluded that the proposed non-contact monitoring system represented a good correlation with the reference method and a lower error rate for both thermal and RGB data.

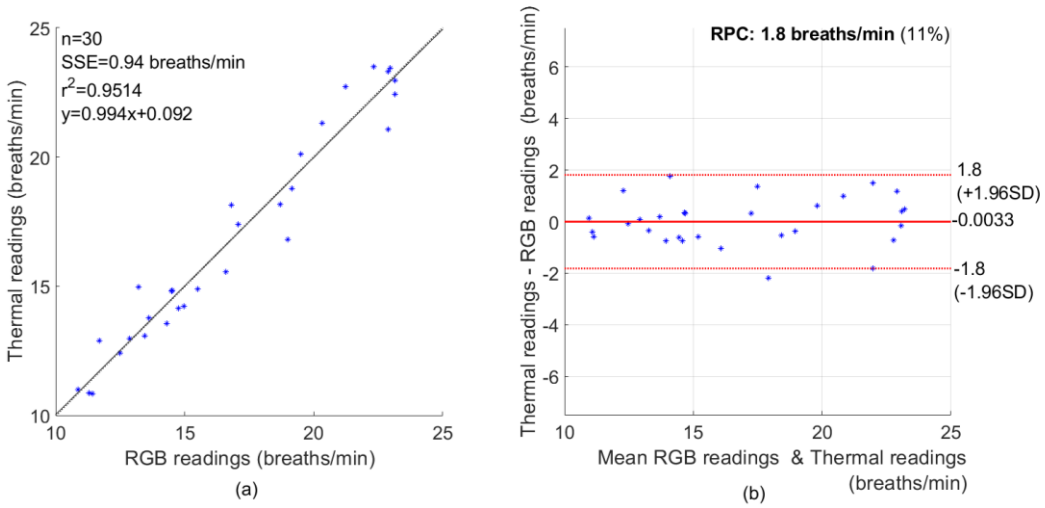


Figure 6.13: Statistical measurement between RGB and thermal data, (a) the correlation plot, (b) the Bland Altman plot.

6.5. Discussion

Table 6.1 summarizes the results of the proposed method for both RGB and thermal data. RGB data gives better performance than thermal with PCC 0.93, RMSE 1.87, LoA -0.54 to 3.6 breaths/min when compared with the reference individually. On the other hand, the level of correlation, limits of agreement, and error rate between RGB & Thermal is better than Thermal & reference and RGB & reference with PCC 0.95, RMSE 0.91, MAE 0.73 and LoA -1.8 to +1.8 breaths/min.

Table 6.1: Summary of the results of the proposed method.

Data	PCC	LoA (breaths/min)	MAE (breaths/min)	RMSE (breaths/min)
Thermal & reference	0.91	-0.9 to 4	1.60	1.97
RGB & reference	0.93	-0.54 to 3.6	1.61	1.87
RGB & Thermal	0.9514	-1.8 to +1.8	0.73	0.91

Table 6.2 shows the comparison of the proposed method with state-of-art methods for thermal data. Our proposed system outperformed other state-of-the-art systems with PCC 0.91, RMSE 1.97, MAE 1.60 and LoA -0.9 to 4.0.

Table 2: Comparison with state-of-the-art methods.

Ref	CC	LoA (breaths/min)	MAE (breaths/min)	RMSE (breaths/min)
Hu et al. [132]	0.83	-3.134 to 1.865	-	-
Negishi et al [290]	0.77	-4.53 to 5.26	-	2.52
Yang et al.[292]	0.83	-4.63 to +4.40	1.83	-
Proposed method	**0.91**	**-0.9 to + 4.0**	**1.60**	**1.97**

6.6. Concluding Remarks

In this chapter, the RR of 10 healthy participants was remotely measured by integrating thermal and RGB camera imaging techniques to overcome the limitation of the contact-based method and increase the reliability of the non-contact method. After RGB-thermal image alignment, the ROI was selected automatically from the RGB image and then the corresponding ROI was selected from the thermal image using linear coordinate mapping. The experimental results demonstrated that the proposed non-contact monitoring system had a lower error rate and good correlation with the reference method. In brief, the proposed method may be a promising alternative to a contact-based technique in the medical field and other sectors, such as quarantine centers, rehabilitation centers, sports studies, or for screening in settings like shopping malls, schools, hospitals, or airports.

CHAPTER 7

Non-contact Jaundice Detection Based on Machine Learning

7 Non-contact Jaundice Detection Based on Machine Learning

Jaundice, often indicated by a yellow discolouration of the skin and sclera of the eyes, is a common phenomenon in neonates in their first few weeks of life caused by elevated levels of bilirubin. Neonatal jaundice remains a significant cause of morbidity and mortality, and early jaundice detection is still a significant challenge. There are several invasive and non-invasive techniques for detecting neonatal jaundice. However, existing invasive techniques are stressful, disruptive, and painful for the newborn, requiring skilled personnel and time. On the other hand, existing non-invasive devices are expensive. Therefore, in this chapter the characteristics of a non-invasive and non-contact neonatal jaundice detection system were investigated based on skin colour analysis and machine learning using a graphical user interface. The content of this chapter is based on materials and results published in [VIII].

This chapter is divided into five main sections. Section 7.1 provides an introduction to the proposed system and describes the related work. Methods and materials, including data collection and the system framework, are explained in section 7.2. Then, in section 7.3, the experimental results are represented. Section 7.4 discusses the experimental results of the proposed systems. Finally, section 7.5 presents concluding remarks.

7.1. Introduction and Related Work

Neonatal jaundice is a common disease found in newborn infants in their first few weeks of life [29, 30]. Every year total of 24 million newborns become jaundiced [32]. Jaundice or hyperbilirubinemia is indicated by the yellow discolouration of the skin and sclera of the eyes due to an excess level of bilirubin [33]. Short-term excess of bilirubin is mostly harmless and self-limiting. But a high level of bilirubin in newborn infants is neurotoxic and can permanently damage the brain, which is called kernicterus. It may cause cerebral palsy, deafness or hearing loss, language difficulties, and developmental delay or can be fatal in the worst cases [34, 35]. Kernicterus can be avoided by early detection and treatment of jaundice or hyperbilirubinemia.

As explained earlier in Section 2.4, for detecting jaundice, there are different invasive and non-invasive approaches [33]. TSB measurement is a common invasive technique which

requires a blood sample to be taken by a syringe and needle and may cause stress for both infants and parents. Moreover, all invasive procedures are uncomfortable and painful for the infant, have a risk of infections, may cause anaemia, and will create anxiety for parents [38]. Visual assessment [39] and TcB [40] are non-invasive alternative techniques to TSB measurement. However, they are expensive and not reliable for detecting jaundice [41, 42]. Computer vision is a potential alternative to detect jaundice free from the need for contact or invasive procedures, yet quickly and cost effectively.

Historically skin colour is a common indicator that has been used for detecting jaundice, especially in the case of neonates. Machine learning techniques have the potential to automate the use of this indicator for medical applications.

From section 2.4.2.2, it can be seen that most of the non-invasive systems in the literature used manual ROI selection. Few of them used automatic ROI selection techniques; however, the existing ROI selection techniques detected the whole face which included areas such as mouth, eyes, nostril, clothes, and background, which may affect the results of jaundice detection.

Jaundice detection could be detected in real-time without operator intervention, and the entire procedure becomes inefficient when the ROI selection is not automated. Moreover, the reliability of jaundice detection depends on selecting the correct skin ROI of the infant. Selecting the entire face might not be the best ROI as it can include undesirable areas such as the mouth, eyes, nostril, clothes, and background. Therefore, in this study, the ROI was automatically selected from the infant's forehead.

In this chapter, a non-invasive and non-contact neonatal jaundice detection system based on skin colour analysis and machine learning was proposed to minimise the limitations of existing invasive methods. First, an automatic ROI selection technique was designed, including face detection, face landmark detection and ROI i.e., forehead selection. Then, skin colour was analysed in both RGB and YC_bC_r colour spaces for the selected area. Finally, jaundice was detected using a machine learning algorithm based on RF. Moreover, the conceptual system could notify whether the neonates require phototherapy or not based on the outcome, demonstrated through a MATLAB GUI panel.

7.2. Methods and Materials

The proposed method consists of simulated and real image data collection for the experiments, analysing facial region for the custom ROI, image feature extraction and jaundice detection.

7.2.1. Data Collection

Two different datasets were used to test the functionality of the proposed system. The first data was obtained by creating 20 images of 3D babies with varying skin tones. These images were used as an alternative to real babies for investigating four different scenarios. The first scenario is for images that have typical light skin representing infants without jaundice. On the other hand, the second scenario is for images that have yellowish light skin representing infants with jaundice. The third scenario is for images that have normal brown skin that shows brown skin infants without jaundice, and the fourth scenario is for images that have a yellowish-brown skin colour that shows brown skinned infants with jaundice.

The second data set contains 30 images of babies taken from the internet [294]; 20 of these images are for healthy infants without jaundice, and the other 10 images are for infants with jaundice. Collecting our own jaundice dataset from real patients was challenging and time-consuming. Therefore, suitable images from the internet were selected for this study.

7.2.2. System Framework for Jaundice Detection

As shown in Figure 7.1, the overall system includes various techniques such as automatic ROI selection, skin colour analysis, feature extraction and jaundice detection.

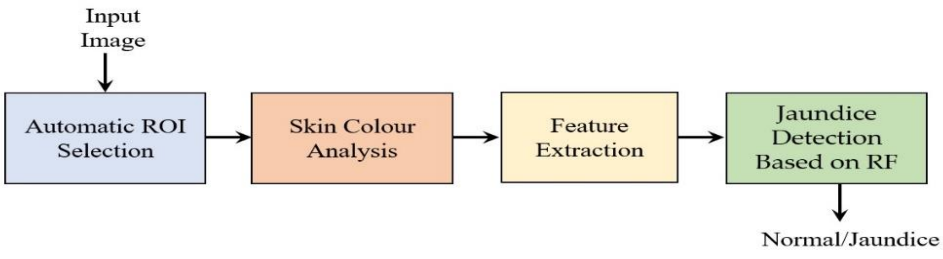


Figure 7.1: Block diagram of the proposed jaundice detection system.

7.2.2.1. Automatic ROI Selection

The reliability of jaundice detection depends on selecting the correct skin ROI of the infant. Selecting the entire face might not be the best ROI as it can include undesirable areas such as the mouth, eyes, nostrils, clothes, and background. Therefore, the ROI from the infant's forehead was selected. As shown in Figure 7.2, the ROI selection algorithm consists of three parts, namely face detection, face landmarks detection and ROI, i.e., forehead selection.

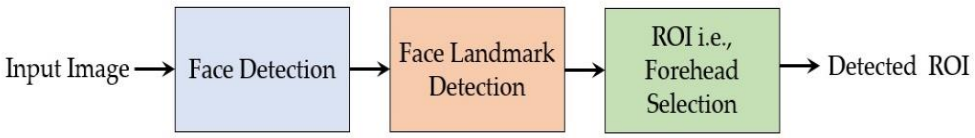


Figure 7.2: Block diagram of automatic ROI selection.

For face detection, Dlib library's [293] HOG and linear SVM based face detector was used. This face detector is faster compared to a CNN-based method. As the babies are in proximity to the camera, their faces are clearly visible in the images. Therefore, both CNN and HOG-based methods detect the faces accurately. CNN is state-of-the-art in face detection, but considering the processing speed, the HOG-based detector was used here. By nature of how the HOG descriptor performs, it is not invariant to variations in viewing angle and rotation.

Next, 68 facial landmarks were extracted from the detected face. The landmarks are the jawline, eyebrows, eyes, nose bridge, lower nose, outer lip, and inner lip. Different ROI selections from the face were explored with, and it was concluded that the forehead ROI was more reliable for the experiments. The geometry of the forehead helps to obtain a consistent skin ROI. The other areas of the face are more susceptible to shadows and inconsistent lighting.

In order to analyse the skin of the baby, a suitable area should be selected. The face is the most visible part of the body of infants in a healthcare setting. Commonly used face detection techniques include detecting the whole face as bounding box or segmenting the face area. These techniques do not provide an ROI with a skin area suitable for jaundice detection. As the colour difference between jaundice positive skin and normal skin can be subtle, a carefully selected skin area helps improve the accuracy. Based on the shape of the face, the forehead is the most suitable area for a skin analysis task. To accurately select a

ROI from the forehead, facial landmarks were used. The landmark-based ROI selection avoids distracting and colour contrasting areas such as the eyes, nostrils, mouth, and the background.

The ROI bounding box was created using the eyebrow landmarks. The mid-points of the eyebrows were projected upwards towards the forehead by a constant pixel distance *h* that was calculated by scaling down the distance between the eyes. The bounding box was oriented roughly at the middle of the forehead. The jaw line, the raised eyebrow points, and the bounding box for three infant images are shown in Figure 7.3. The ROI selection algorithm is presented in Figure 7.4.

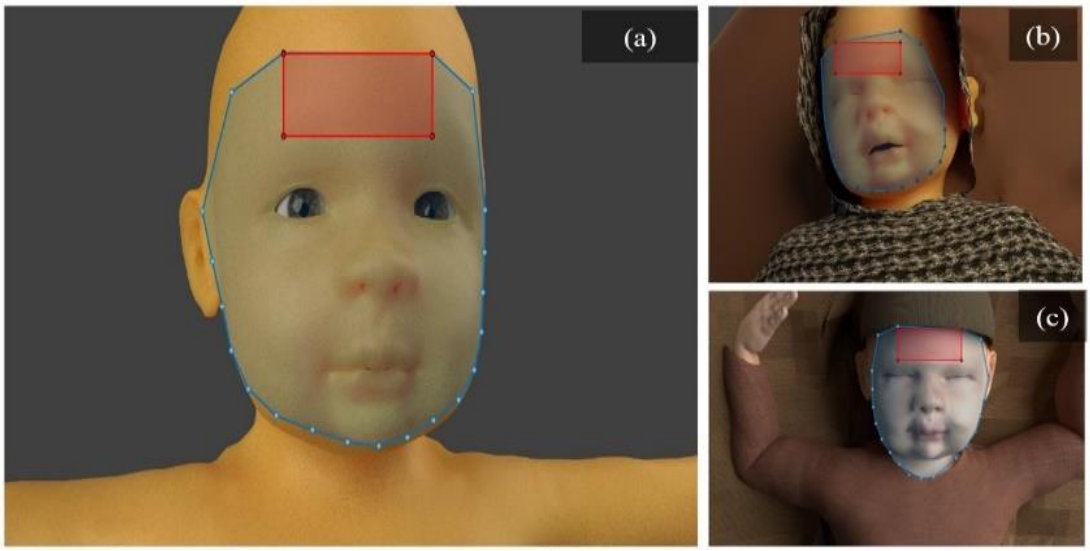


Figure 7.3: Representation of automatic ROI selection. (a) The infant is in a straight position. (b) The infant with a tilted head. (c) The infant wearing a cap.

Algorithm 1: ROI selection

Input: image_frame
Output: ROI_pixels

```
Function face_detector(image_frame):
    return face_bounding_box
Function face_landmark_detector(image_frame,
  face_bounding_box):
    return face_landmarks
Function Main:
    face_detector(image_frame)
    face_landmark_detector(image_frame, face_bounding_box)
    select two middle forehead landmarks (f1,f2)
    set the ROI width and height w.r.t. (f1,f2)
    select the ROI pixels from image_frame
    return ROI_pixels
```

Figure 7.4: Algorithm of ROI selection.

7.2.2.2. Skin Colour Analysis

After selecting ROI, the skin colour analysis was required. In numerous human-related image processing applications, skin colour is a very informative parameter [81, 262, 263, 295]. In skin colour detection, skin-coloured regions and pixels are identified in each image using three main colour spaces: RGB (Red, Green and Blue), YC_bC_r (Luminance, Chrominance) and HSV (Hue, Saturation and Value). The RGB representation of the brightness value variations can be obtained using the following equations [224, 296]

$$R = \int_{\lambda 1}^{\lambda 2} E(\lambda) F_r(\lambda)\, d\lambda \tag{7.1}$$

$$G = \int_{\lambda 1}^{\lambda 2} E(\lambda) F_g(\lambda)\, d\lambda \tag{7.2}$$

$$B = \int_{\lambda 1}^{\lambda 2} E(\lambda) F_b(\lambda)\, d\lambda \tag{7.3}$$

where F_r, F_g and F_b are the colour filters on the incoming light $E(\lambda)$, and λ is the wavelength and $\lambda 1 = 300$ and $\lambda 2 = 830$nm [224]. The colour representation of the YC_bC_r colour space can be obtained from RGB colour space using the following equation [297]

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -39.797 & -74.203 & 112 \\ 112 & -93.786 & -18.2214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (7.4)$$

An excess level of bilirubin creates a yellow discolouration of the skin. The original RGB colour space was transformed into the YC_bC_r colour spaces in order to better observe this subtle discolouration.

7.2.2.3. Feature Extraction

We calculated the mean values for each colour channel of RGB and YC_bC_r colour spaces, resulting in 6 features.

The mean value of the brightness pixel for R, G, B channels within the selected ROI can be calculated as follows.

$$F_R, F_G, F_B = \frac{\sum_{x,y \in ROI} I(x,y)}{|ROI|} \quad (7.5)$$

where $|ROI|$ is the size of the selected ROI, and $I\ (x,\ y)$ is the brightness pixel value at image location (x,y) for R,G,B channels. The value of Eq. (7.5) ranges between 0 and 255.

Within the selected ROI, the mean value of the brightness pixel for Y, C_b , C_r channels can be measured as follows

$$F_Y, F_{C_b,} F_{C_r} = \frac{\sum_{x,y \in ROI} I(x,y)}{|ROI|} \quad (7.6)$$

where $|ROI|$ is the size of the selected ROI and $I\ (x,\ y)$ is the brightness pixel value at image location (x,y) Y, C_b , C_r for channels. The value of Eq. (7.6) ranges between 16 and 240.

Therefore, the feature vector of the proposed system using 6 features is as follows.

$$F = [F_R, F_G, F_B, F_Y, F_{C_b,} F_{C_r}] \quad (7.7)$$

7.2.2.4. Jaundice Detection Based on Random Forest

Random forest (RF), developed by Leo Breiman [298], is one of the highest reported performance classification methods [205]. It can deal with continuous, binary, and categorical data. RF is a simple, fast, flexible, efficient and robust model which has numerous applications in different sectors such as banking, the stock market, e-commerce and medicine [299]. The RF algorithm is often considered for selecting an optimum and

precise combination of the components and symptoms of a particular disease in the medical field [300-302].

RF is classified as a supervised machine learning algorithm that is commonly used in regression and classification problems [299]. It develops decision trees on various samples and takes their average for regression and majority vote in case of classification, and thus the problem of overfitting is solved. It can deal with the data set containing categorical variables, as in the case of classification as well as continuous variables as in the case of regression. It achieves better outcomes for classification problems [298]. Moreover, it can handle missing values. As the average answers given by many trees are taken, it is very stable and accurate [299].

As shown in Figure 7.5, the following steps are involved in the RF algorithm [303]:

Step 1: Let a data set contains k number of records. In RF, n number of random records will be considered from that data set.

Step 2: For each sample, individual decision trees will be created.

Step 3: An output will be produced by each decision tree.

Step 4: Final output will be considered using majority voting for classification or averaging for regression.

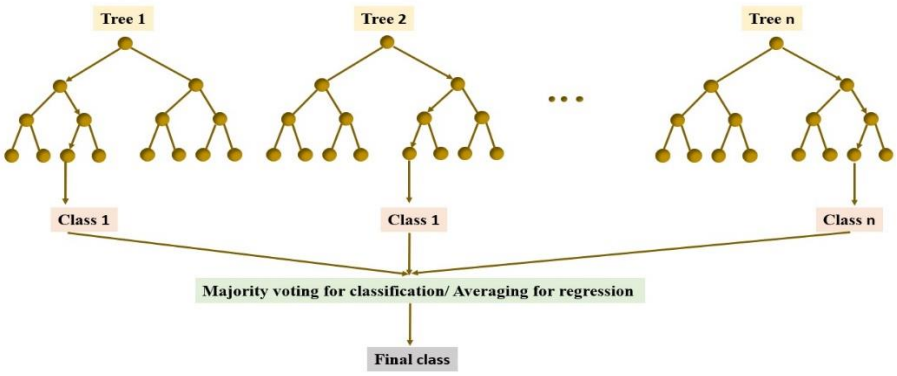


Figure 7.5: Representation of RF algorithm.

To classify jaundice or normal, RF classification based on majority voting was used. A dataset consisting of 512 images and two class labels: Jaundice and Normal was considered. These images were used to train our model. As there will always be 30% of

the data that is not considered by the decision tree, it was not required to separate data into train and test. After training, infant images were fed to the model for diagnosis.

If the output is jaundice, then the neonate requires UV therapy, and if the output is normal, then the neonate does not need any UV therapy.

7.3. Experimental Results

Python 3.9 and MATLAB 2020a were used to implement our algorithm. To evaluate our automatic ROI selection technique at different scenarios, such as the infant in a straight position without a cap (Figure 7.3a), the infant with a tilted head (Figure 7.3b), and the infant in a straight position wearing cap (Figure 7.3c). It can be seen from Figure 7.3 that ROI was successfully selected automatically using our proposed technique. The red bounding boxes in Figure 7.3 represent the selected ROI in different cases.

A 10-fold cross-validation was performed to evaluate the ML algorithm selected for our jaundice detection system. A dataset consisting of 511 images and two class labels: Normal and Jaundice was considered. In the case of Stochastic Gradient Boosting, decision tree and random forest, number of trees=100. In KNN, number of neighbours, k=5.

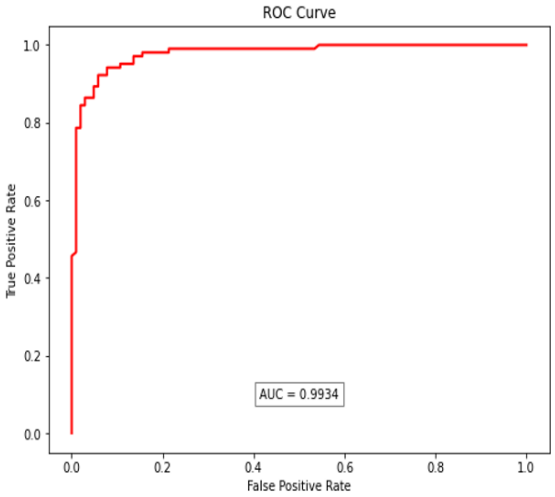
The original dataset was randomly divided into 10 subsets of approximately equal size. Each subset is referred to as a fold. For each fold, the model was trained on the remaining 9 folds (the training set) and evaluated on the held-out fold (the validation set). The performance metric (e.g., accuracy, precision, recall, and F1 score) was calculated for each fold based on the predictions made by the model on the validation set. The process was repeated 10 times, with each fold serving as the validation set exactly once. The performance metrics obtained from each fold were averaged to get an overall assessment of the model's performance. The performance of the ML algorithm that had shown the best cross validation performance was used in the proposed system.

The performance of different ML algorithms in terms of average accuracy, precision, recall, and F1-score are presented in Table 7.1. It can be seen from the table, that the performance of Random Forest shows the best cross validation performance with accuracy=0.9922, precision=0.9974, recall=0.9920, and F1-score=0.9946. Therefore, in the proposed system, random forest algorithm was used.

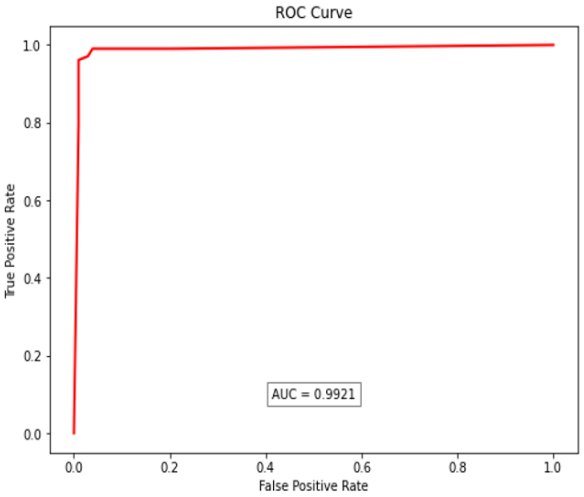
Table 7.1: Comparison of the performance of different ML algorithms.

ML Algorithms	Average Accuracy	Average Precision	Average Recall	Average F1 Score
Naïve Bayes	0.9120	0.9917	0.8876	0.9359
KNN	0.9882	0.9844	1	0.9921
SVM	0.9863	0.9895	0.9920	0.9907
SGB	0.9882	0.9947	0.9894	0.9919
Decision Tree	0.9903	0.9974	0.9894	0.9933
Random Forest	**0.9922**	**0.9974**	**0.9920**	**0.9946**

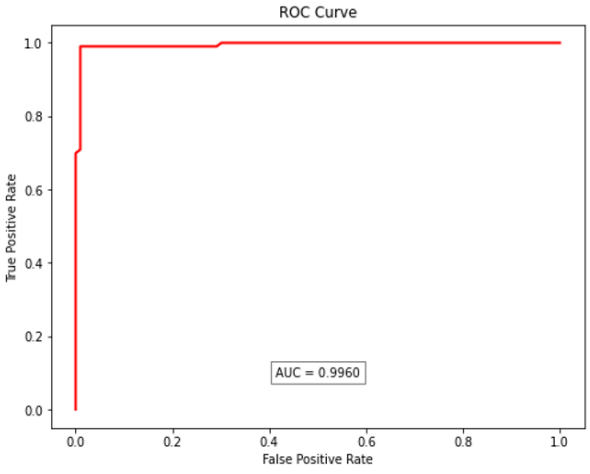
The performance of different ML algorithms is presented in Figure 7.6 in terms of ROC (Receiver operating characteristics) curve. It can be seen from the figure, that the performance of Random Forest shows the best performance with AUC=0.9984.



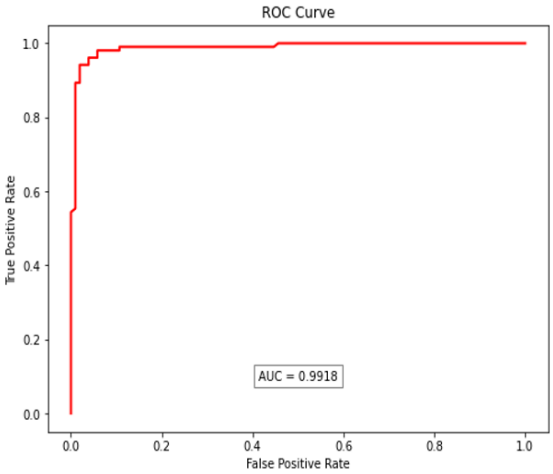
(a) Naïve Bayes



(b) KNN



(c) SVM



(d) SGB

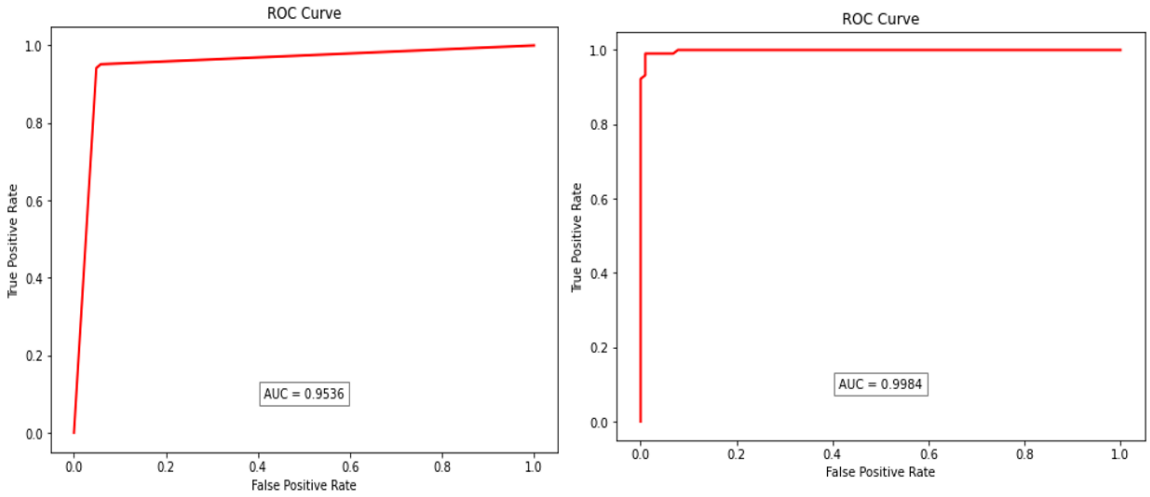


Figure 7.6: ROC curve for different machine learning algorithms. (a) Naïve Bayes, (b)KNN, (c) SVM, (d) SGB, (d) Decision tree, (f) RF.

The proposed system can be used to detect jaundice as well as to decide if the baby needs UV therapy or not using the RF algorithm based on skin colour analysis. A GUI panel using MATLAB software was designed for jaundice detection. The proposed MATLAB GUI panel (Figures 7.7-7.12) allows the operator, such as a nurse, doctor, or guardian, to open the image of the baby. Then, the algorithm was executed to identify jaundice based on skin colour and to determine whether the neonate needed UV therapy or not. The RGB histogram and their brightness values, skin state i.e., jaundice or normal, can be easily displayed through the proposed GUI panel.

Two different datasets were used to test the functionality of the proposed system. The first dataset was obtained by creating 20 images of 3D babies with varying skin tones. These images were used as an alternative to real babies to examine four different scenarios (Figure 7.7-7.10). Figure 7.7 represents a baby without jaundice in the GUI main panel of the proposed jaundice detection system. It can be seen from Figure 7.7 that the baby does not require UV therapy as the colour of his skin did not include jaundice yellowing. The GUI main panel of the proposed system for a baby with jaundice is shown in Figure 7.8, and it requires UV phototherapy as his skin is yellow. Figures 7.9 and 7.10 represent the infants with darker skin tone with and without jaundice, respectively. The results for all 20 images of 3D babies are shown in Table 7.2.

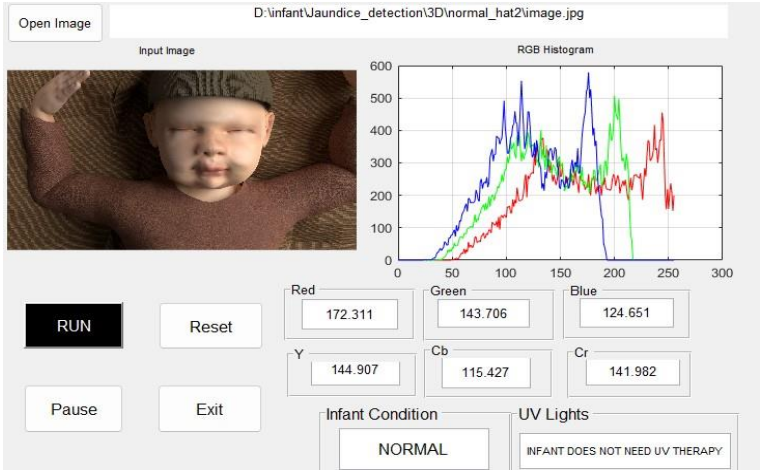


Figure 7.7: A 3D baby with light skin colour (jaundice-free).

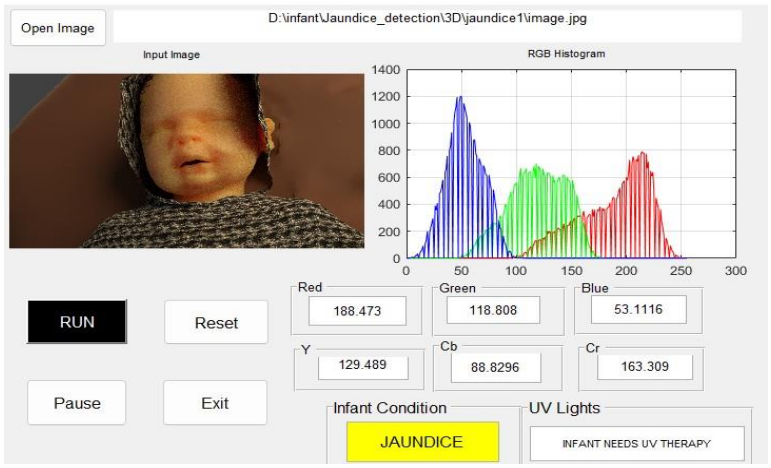


Figure 7.8: A 3D baby with a yellowish-light skin colour and jaundice.

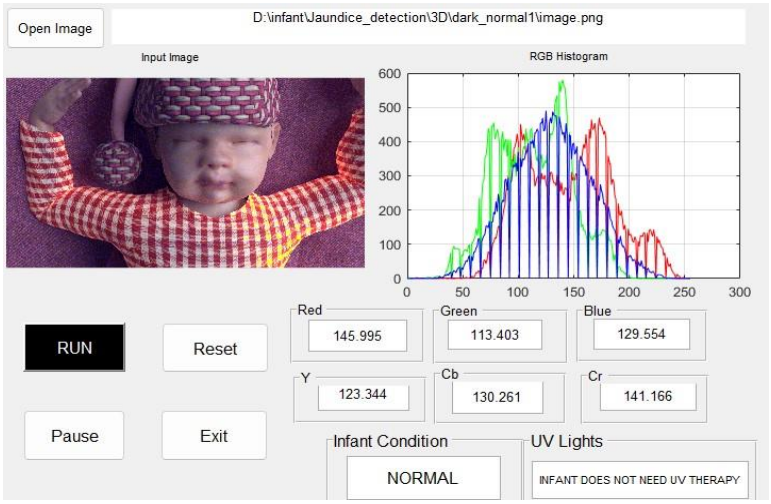


Figure 7.9: A 3D baby with brown skin colour and jaundice-free.

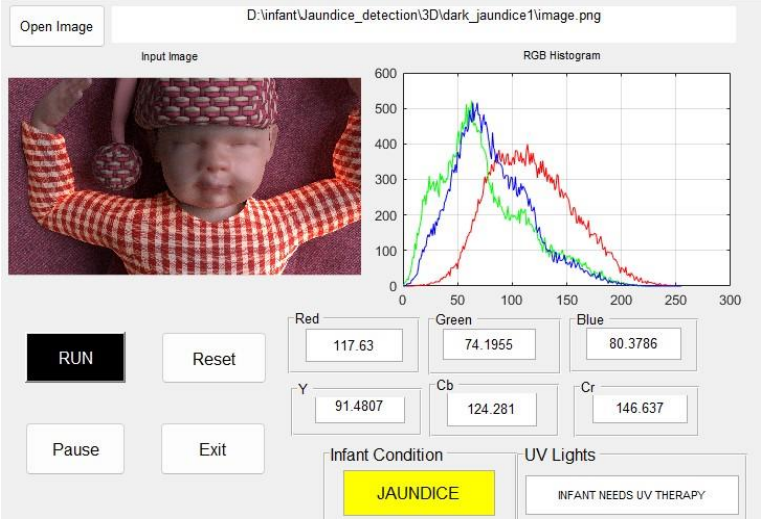


Figure 7.10: A 3D baby that has yellowish-brown skin and jaundice.

Table 7.2. Summary of 20 infants considered in the first dataset.

Infant	Red	Green	Blue	Y	C_b	C_r	Infant condition	UV Therapy
1	152.84	112.25	122.56	128.88	123.74	141.4	Normal	Does not need
2	175.64	129.19	141.6	140.11	126.57	147.50	Normal	Does not need
3	180.93	133.39	146.74	144.07	126.81	147.9	Normal	Does not need
4	199.67	167.62	145.77	166.04	113.71	143.62	Normal	Does not need
5	171.11	142.3	123.12	143.72	115.32	142.12	Normal	Does not need
6	203.02	170.5	148.13	168.59	113.42	143.86	Normal	Does not need
7	147.44	118.40	116.49	124.95	122.98	140.96	Normal	Does not need
8	145.99	113.40	129.5	123.34	130.26	141.16	Normal	Does not need
9.	172.31	143.70	124.65	144.90	115.42	141.98	Normal	Does not need
10.	193.59	153.81	148.06	157.42	119.87	146.09	Normal	Does not need
11.	117.63	74.19	80.37	91.48	124.28	146.63	Jaundice	Needs
12.	112.31	91.35	91.54	99.84	124.97	137.22	Jaundice	Needs
13.	188.47	118.80	53.11	129.48	88.82	163.31	Jaundice	Needs
14.	154.00	119.58	54.30	121.15	94.22	147.82	Jaundice	Needs
15.	140.75	105.36	53.94	110.54	100.17	147.21	Jaundice	Needs
16.	186.28	124.7	48.13	131.4	85.24	160.5	Jaundice	Needs
17.	193.48	144.45	75.3	145.88	90.23	154.48	Jaundice	Needs
18.	189.59	132.84	38.98	135.47	78.36	159.63	Jaundice	Needs
19.	205.05	131.17	58.28	140.4	85.03	165.5	Jaundice	Needs
20.	189.95	136.24	33.66	136.75	74.98	158.9	Jaundice	Needs

The second data set comprises 30 images of babies (20 of these images are for healthy babies without jaundice, and the other 10 images are for babies with jaundice) downloaded from the internet. These images were captured by different cameras, with various skin tones, illumination conditions, and camera–infant distance. As represented in Table 7.3, the proposed jaundice detection system successfully identified all the babies without jaundice as 'Normal' and not needing UV therapy. Additionally, all the images of the babies who had jaundice were detected as 'Jaundice' and needing UV therapy. Figure 7.11 represents a baby who does not have jaundice, and as a result, does not need UV therapy. Figure 7.12 shows a baby who has jaundice and needs UV therapy.

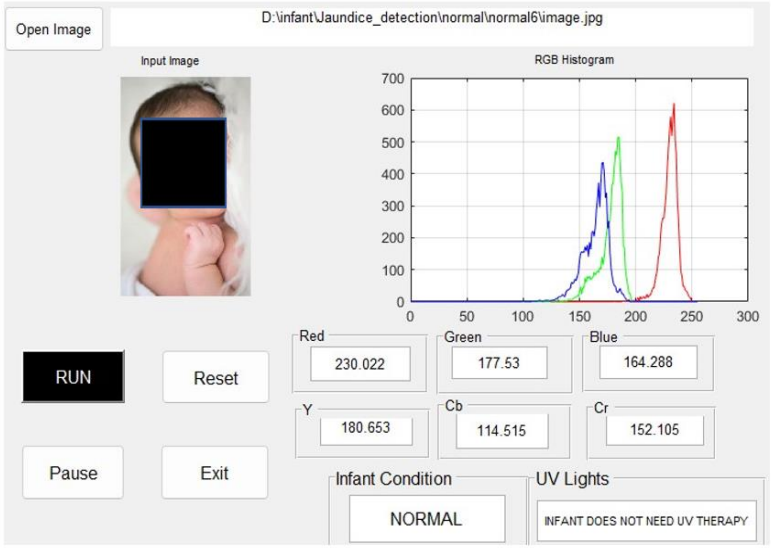


Figure 7.11: A normal baby without jaundice.

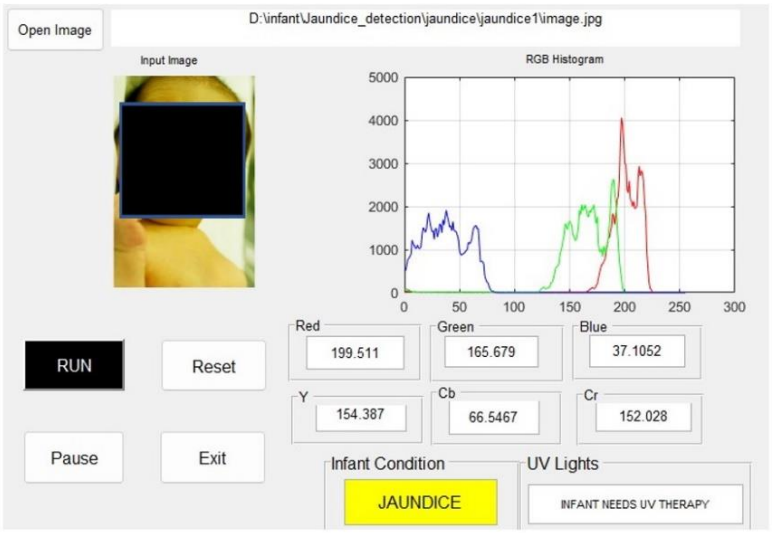


Figure 7.12: A baby with a yellowish skin colour who has jaundice.

Table 7.3: Summary of 30 infants considered in the second dataset.

Infant	Red	Green	Blue	Y	C_b	C_r	Infant condition	UV Therapy
1.	199.5	165.67	37.10	154.38	66.54	37.10	Jaundice	Needs
2.	187.39	140.11	81.4	142.72	95.22	152.94	Jaundice	Needs
3.	144.17	104.17	52.22	110.66	99.3	149.27	Jaundice	Needs
4.	226.45	165.62	112.87	168.71	95.74	158.48	Jaundice	Needs
5.	141.24	82.75	58.46	99.72	108.60	155.38	Jaundice	Needs
6.	212.37	127.56	46.40	139.39	79.77	171.04	Jaundice	Needs
7.	214.11	134.76	88.25	147.60	95.72	166.10	Jaundice	Needs
8.	222.68	147.03	97.77	156.87	95.07	164.75	Jaundice	Needs
9.	191.00	132.32	75.45	139.15	94.28	157.85	Jaundice	Needs
10.	114.9	77.26	38.24	88.18	105.41	147.35	Jaundice	Needs
11.	209	153	116	158	103	155	Normal	Does not need
12.	229.97	194.98	171.73	190.17	112.52	145.17	Normal	Does not need
13.	243.89	189.17	157.49	189.40	105.94	154.29	Normal	Does not need
14.	230.02	177.53	164.28	180.65	114.51	152.10	Normal	Does not need
15.	212.46	177.35	170.40	176.67	119.78	143.97	Normal	Does not need
16.	232.37	206.21	194.91	198.72	119.14	140.21	Normal	Does not need
17.	197	167	156	166	119	142	Normal	Does not need
18.	242.72	225.85	220.57	213.80	123.21	135.72	Normal	Does not need
19.	224.52	186.06	155.34	182.65	108.86	147.13	Normal	Does not need
20.	192.22	137.46	116.53	146.07	110.84	153.84	Normal	Does not need
21.	212.52	153.52	119.74	159.72	104.51	156.33	Normal	Does not need
22.	238.21	200.51	188.17	196.68	117.13	145.41	Normal	Does not need
23.	242.76	219.26	210.94	209.50	120.80	139.07	Normal	Does not need
24.	223.17	195.58	186.15	190.15	119.86	140.83	Normal	Does not need
25.	204.05	138.62	123.67	150.39	111.58	157.81	Normal	Does not need
26.	205.10	160.10	144.00	163.474	114.4	148.82	Normal	Does not need
27.	233	218	211	206	123	136	Normal	Does not need
28.	226.25	205.78	206.58	198.05	125.32	136.97	Normal	Does not need
29.	239.55	185.57	169.12	187.60	112.85	152.87	Normal	Does not need
30.	244.78	194.99	180.06	194.79	113.99	151.02	Normal	Does not need

The experimental result presents that the proposed jaundice detection system can be a potential non-invasive technique for jaundice detection in clinical applications.

7.4. Discussion

There are several invasive and non-invasive techniques for detecting neonatal jaundice. However, existing invasive techniques are stressful, disruptive, and painful for the newborn, requiring skilled personnel and time. On the other hand, existing non-invasive devices are expensive. Therefore, a non-invasive and non-contact neonatal jaundice detection system was proposed using a graphical user interface.

As shown in Table 7.4, most non-invasive systems in the literature used manual ROI selection [201, 207, 210, 213]. Fewer used automatic ROI selection techniques [208, 209, 215]; however, the existing ROI selection techniques detected the whole face which included areas such as mouth, eyes, nostrils, clothes, and background, which may affect the results of jaundice detection.

Jaundice detection is a real-time requirement, and the entire procedure becomes ineffective when the ROI selection is not automated. Moreover, the reliability of jaundice detection depends on selecting the correct skin ROI of the infant. Selecting the entire face might not be the best ROI as it can include undesirable areas such as the mouth, eyes, nostrils, clothes, and background. In the proposed system, an automatic ROI selection technique was developed and forehead region was considered as ROI.

From table 7.4, it can be seen that the proposed system compare favourable to the state-of-the art methods except Hashim et al.'s [201, 208] methods. However, they did skin colour analysis in RGB and YC_bC_r colour spaces to detect jaundice. The values of the B and C_b channels were compared with the predefined threshold value to determine whether the neonate had jaundice or not. In the proposed method, a machine learning algorithm based on RF was incorporated to classify jaundice or normal and determine whether the neonate requires treatment or not.

Table 7.4: Comparison with state-of-the-art methods.

Ref	ROI Selection	Jaundice detection	Detection rate of the system
Hashim et al. [201]	Manual	skin colour analysis in RGB and YC_bC_r colour spaces	100%

Rong et al. [210]	Manual	machine learning-based software	96%
Juliastuti et al.[213]	Manual	multivariable linear regression	90%
Hashim et al. [208]	Automatic	skin colour analysis in RGB and YC_bC_r colour spaces	100%
Mansor et al. [215]	Automatic	k-NN	95.6%
Proposed method	**Automatic**	**Machine learning based on RF**	**100%**

As the proposed system is based on skin colour analysis, the system may have had limitations for babies with different skin tones and under different lighting conditions. To minimise these challenges, variations of these two parameters were used to train the random forest, demonstrating that the proposed system could detect jaundice across a wide range of skin tones and lighting conditions.

7.5. Concluding Remarks

In conclusion, a non-invasive and easy-to-use conceptual system was developed to monitor neonatal jaundice by taking images of babies and using a graphical user interface. In the proposed system, first, an automatic ROI selection technique was designed, including face detection, face landmark detection and ROI i.e., forehead selection. Then, the skin colour of the selected forehead region was analysed in both RGB and YC_bC_r colour spaces. Finally, jaundice has been detected using a machine learning algorithm based on RF. Moreover, the system can notify whether the neonates require phototherapy or not.

The experimental results show that the proposed jaundice detection technique can be a potential alternative to the existing procedures that involve blood samples using syringes and needles. It lessens the stress, trauma, and pain involved in conventional invasive methods. Moreover, it gives the result within a few minutes without the involvement of any highly qualified personnel. It is also a cost-effective, affordable, and easy-to-use tool that is appropriate for monitoring an infant in the comfort of their own residence. Moreover, it is advantageous in emergency situations where the availability of specialized disease diagnosis is missing.

CHAPTER **8**

Conclusions and Future Directions

8 Conclusions and Future Directions

The primary aim of this thesis was to investigate the feasibility of monitoring cardiorespiratory signals and detecting jaundice without the need for physical contact, utilizing camera imaging technologies. The goal was driven by the challenges faced by existing non-contact systems, which include the subject's safety, the system cost, the ROI selection, noise-artefacts, the complexity of dealing with different subjects such as premature babies and multi-camera fusion. Throughout the thesis, detailed experiments were conducted, and the findings were presented in separate chapters. These experiments facilitated the proposed data processing, classification systems, and validation methods.

In this concluding chapter, the main outcomes from the investigation are succinctly summarized, with a focus on the research questions addressed throughout this thesis. Additionally, potential directions for future research in this domain are suggested, aiming to further advance the field. This chapter not only consolidates the findings of this study but also sets the stage for continued exploration and innovation in the realm of non-contact monitoring and disease detection.

8.1. Conclusions

The research presented in this thesis addresses several challenges to achieve reliable vital sign monitoring (HR and RR) and jaundice detection systems using camera imaging technologies. The baseline of these non-contact systems relies on observation of the physical and physiological variations caused by cardiorespiratory activity, including skin colour variations, head motion, and thoracic and abdominal motion. These variations, however, can be significantly distorted by noise artefacts caused by skin tone and illumination variations. Moreover, accurate ROI selection in an automatic manner is another challenge for reliable vital sign extraction and jaundice detection. This thesis has addressed these challenges as follows.

As detailed in Chapter 4 and 5, non-contact vital sign monitoring systems using digital cameras were proposed to remotely monitor HR and RR of neonates in NICU without any physical contact. Different ROI selection algorithms were developed using advanced image processing technique such as image decomposition, skin detection, and neural network for

automatic ROI selection. Additionally, a noise removal technique was used based on signal decomposition to reduce the noise artefacts caused by illumination variation. Experimental results from the proposed non-contact systems showed close correlation (PCC), strong agreement (±1.96 SD), and acceptable error rates (MAE and RMSE) when compared with the reference methods like ECG, thereby suggesting the reliability of the proposed systems in providing physiological assessments.

As detailed in Chapter 6, a non-contact monitoring system was introduced to monitor RR of healthy adults by integrating RGB and thermal imaging systems through RGB-thermal image alignment. An automatic ROI selection algorithm, including face detection, face landmarks detection, ROI selection from RGB and thermal images, was used to select ROI automatically. The performance of the proposed system i.e., close correlation (PCC), strong agreement (±1.96 SD), and acceptable error rates (MAE and RMSE) in comparison with a reference method, indicated its potential as a promising alternative to contact-based techniques across clinical and non-clinical settings.

As detailed in Chapters 7, the characteristics of a non-invasive and non-contact neonatal jaundice detection system were investigated through skin colour analysis and machine learning facilitated by a user-friendly graphical interface. The ROI was automatically selected from the image of an infant captured by a digital camera. Moreover, a machine learning algorithm based on random forest was incorporated to classify jaundice or normal and determine whether the neonate requires treatment or not. The experimental result demonstrated that the proposed jaundice detection system has the potential as a non-invasive technique in clinical application.

To conclude, the objectives of this thesis outlined in Chapter 1 were successfully achieved. Throughout the subsequent chapters in this thesis, various non-contact systems were proposed and developed to address the problem statement and the existing gaps in the literature. As a result, the proposed systems and methods introduced and described in preceding chapters may provide potential solutions for future applications in monitoring vital signs and detecting jaundice.

8.2. Future Directions

This thesis has made a significant contribution to the development of non-contact vital sign monitoring systems suitable for both clinical and non-clinical settings. Additionally, the

detection of jaundice has been achieved in a non-contact manner using machine learning. However, certain aspects remain unaddressed, and further efforts are needed to fully achieve the overall goal. The following paragraphs present potential possibilities for future research to continue and extend the research presented in this thesis.

One potential future research direction identified during this study is the estimation of another important vital sign, such as blood oxygen saturation level (SpO2). This can be estimated by analyzing the cardiorespiratory signals captured by a video camera imaging system at two different wavelengths utilizing ac/dc component analysis, as an alternative to the intensity-based methods employed in this thesis.

Another potential research direction involves using advanced signal processing techniques to reduce motion artefacts caused by camera and subject movement, thereby enhancing the reliability of the non-contact systems.

Moreover, a promising area for expanding the research presented in this thesis is the development of mobile software capable of extracting cardiorespiratory signals and detecting jaundice in real-life scenarios. Such applications could encourage users to incorporate these technologies into their daily lives, thereby increasing the practicality and usability of the proposed systems.

Furthermore, the non-contact vital sign monitoring systems proposed in this thesis may have potential applications in the field of psychophysiology, enabling the estimation of the psychological state of subjects, including stress, pain, annoyance, fatigue, and more. Moreover, it has promising applications in other sectors such as quarantine centers, rehabilitation centers, sport studies, and for screening purposes in locations like shopping malls, schools, hospitals, or airports, particularly relevant during the COVID-19 pandemic.

In this thesis, EEMD was used to reduce illumination variations. The time required for signal separation using EEMD can vary based on factors such as the signal length or number of frames. However, EEMD is too fast to deal with PPG signals (good for real-time applications), the only slow thing is the process of converting frames to signal. It is possible to develop a real-time algorithm, it may need to streamline and optimize the processing steps. This could involve parallelization, algorithmic improvements, and hardware acceleration. Moreover, to do online (real-time) signal processing several algorithms such as deep learning, optical flow, adaptive filtering etc. could be used.

To address the challenge of sparse datasets and the need for sufficient data for future deep learning studies requires a combination of strategic approaches and methodologies. For example, augmenting existing datasets by applying transformations such as rotation, scaling, cropping, or flipping. Moreover, generating synthetic data to supplement the sparse dataset. Furthermore, collaborating with other research groups or organizations to share datasets, fostering a collective effort to overcome data scarcity.

The occlusion of the ROI in vital signs monitoring, whether caused by external factors such as a nurse giving care or a patient reading a newspaper, can significantly impact the output of the monitoring system. In such scenarios, the algorithm or monitoring system needs to incorporate adaptive mechanisms to maintain accuracy and continuity. For example, implementing algorithms that dynamically adjust the ROI based on the visible and unobstructed areas or employing algorithms that capable of tracking multiple ROIs simultaneously. Moreover, incorporating data fusion techniques that combine information from multiple sources, such as different facial regions or alternative anatomical locations. By incorporating these adaptations, the monitoring system can enhance its resilience to occlusions and maintain accurate vital signs monitoring even in dynamic and obstructed scenarios.

Data fusion, also known as sensor fusion, involves combining information from multiple sensors to enhance the overall accuracy, reliability, and completeness of the data. This process is particularly valuable in scenarios where individual sensors may provide incomplete or noisy data. By effectively implementing data fusion techniques, systems can benefit from the collective strengths of diverse sensors, resulting in more accurate and reliable information for decision-making in complex and dynamic environments.

In summary, these potential future research directions offer exciting opportunities to build upon the achievements of this thesis, advancing the development of non-contact systems and expanding their applications to benefit both clinical and non-clinical scene.

References

[1] C. Brüser, C. H. Antink, T. Wartzek, M. Walter, and S. Leonhardt, "Ambient and unobtrusive cardiorespiratory monitoring techniques," *IEEE reviews in biomedical engineering*, vol. 8, pp. 30-43, 2015.

[2] C. Rotariu and V. Manta, "Wireless system for remote monitoring of oxygen saturation and heart rate," *presented at the Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on, Wroclaw, Poland*, 2012 pp. 193–196.

[3] T. I. Papon, I. Ahmad, N. Saquib, and A. Rahman, "Non-invasive heart rate measuring smartphone applications using on-board cameras: A short survey," *presented at the Networking Systems and Security (NSysS), 2015 International Conference on, Dhaka, Bangladesh*, 2015, pp. 1–6.

[4] A. Alonso, J. Sasin, and N. Bottini, "Signs of respiratory disease," *apoptosis*, vol. 85, no. 8, pp. 721-726, 2006.

[5] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, "Video-Based Heart Rate Measurement: Recent Advances and Future Prospects," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, pp. 3600–3615, 2018.

[6] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote measurements of heart and respiration rates for telemedicine," *PloS one*, vol. 8, no. 10, p. e71384, 2013.

[7] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomedical signal processing and control*, vol. 13, pp. 102-112, 2014.

[8] P. V. Rouast, M. T. Adam, R. Chiong, D. Cornforth, and E. Lux, "Remote heart rate measurement using low-cost RGB face video: a technical literature review," *Frontiers of Computer Science*, vol. 12, no. 5, pp. 858-872, 2018.

[9] S. R. Sardesai, M. K. Kornacka, W. Walas, and R. Ramanathan, "Iatrogenic skin injury in the neonatal intensive care unit," *Journal of Maternal-Fetal and Neonatal Medicine*, vol. 24, no. 2, pp. 197-203, 2011.

[10] F. C. Fontenele and M. V. L. M. L. Cardoso, "Skin lesions in newborns in the hospital setting: type, size and affected area," *Revista da Escola de Enfermagem da USP*, vol. 45, no. 1, pp. 130-137, 2011.

[11] M. T. Migoto, S. N. D. H. de Souza, and E. G. Rossetto, "Skin lesions of newborns in a neonatal unit: descriptive study," *Online Brazilian Journal of Nursing*, vol. 12, no. 2, 2013.

[12] S. Phansalkar *et al.*, "A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems," *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 493-501, 2010.

[13] 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, 2016.

[14] A. Al-Naji, K. Gibson, and J. Chahl, "Remote sensing of physiological signs using a machine vision system," *Journal of medical engineering & technology*, vol. 41, no. 5, pp. 396-405, 2017.

[15] J. Muraskas and K. Parsi, "The cost of saving the tiniest lives: NICUs versus prevention," *Virtual Mentor*, vol. 10, no. 10, p. 655, 2008.

[16] R. Oliver and R. Lamont, "Infection and antibiotics in the aetiology, prediction and prevention of preterm birth," *Journal of Obstetrics & Gynaecology*, vol. 33, no. 8, pp. 768-775, 2013.

[17] C.-H. Wang *et al.*, "Analysis of in-hospital neonatal death in the tertiary neonatal intensive care unit in china: A multicenter retrospective study," *Chinese medical journal*, vol. 129, no. 22, p. 2652, 2016.

[18] L. A. Aarts *et al.*, "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.
[19] S. Zaunseder, A. Henning, D. Wedekind, A. Trumpp, and H. Malberg, "Unobtrusive acquisition of cardiorespiratory signals," *Somnologie*, vol. 21, no. 2, pp. 93-100, 2017.
[20] P. H. Charlton *et al.*, "Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants," *Physiological measurement*, vol. 38, no. 5, p. 669, 2017.
[21] B. Reeder and A. David, "Health at hand: A systematic review of smart watch uses for health and wellness," *Journal of biomedical informatics*, vol. 63, pp. 269-276, 2016.
[22] K. J. Burdick, M. Gupta, A. Sangari, and J. J. Schlesinger, "Improved patient monitoring with a novel multisensory smartwatch application," *Journal of Medical Systems*, vol. 46, no. 12, p. 83, 2022.
[23] N. Selvaraj, "Long-term remote monitoring of vital signs using a wireless patch sensor," in *2014 IEEE Healthcare Innovation Conference (HIC), Seattle,WA,USA*, 2014, pp. 83-86.
[24] M. Weenk, S. J. Bredie, M. Koeneman, G. Hesselink, H. van Goor, and T. H. van de Belt, "Continuous monitoring of vital signs in the general ward using wearable devices: randomized controlled trial," *Journal of medical Internet research*, vol. 22, no. 6, p. e15471, 2020.
[25] M. Weenk *et al.*, "Continuous monitoring of vital signs using wearable devices on the general ward: pilot study," *JMIR mHealth and uHealth*, vol. 5, no. 7, p. e7208, 2017.
[26] D. Dias and J. Paulo Silva Cunha, "Wearable health devices—vital sign monitoring, systems and technologies," *Sensors*, vol. 18, no. 8, p. 2414, 2018.
[27] N. Mitro *et al.*, "AI-Enabled Smart Wristband Providing Real-Time Vital Signs and Stress Monitoring," *Sensors*, vol. 23, no. 5, p. 2821, 2023.
[28] A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods," *IEEE Access*, vol. 5, pp. 15776-15790, 2017.
[29] L. M. Gartner, "Neonatal jaundice," *Pediatrics in review*, vol. 15, no. 11, pp. 422-432, 1994.
[30] B. Katti and B. S. Babu, "Analytical Review on the Technologies Designed for the Better Detection and Treatment of Neonatal Jaundice," *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 06, pp. 2518-2521, 2021.
[31] J. Rennie, S. Burman-Roy, and M. S. Murphy, "Neonatal jaundice: summary of NICE guidance," *Bmj*, vol. 340, pp.1-7, 2010.
[32] P. Padidar *et al.*, "Detection of neonatal jaundice by using an android OS-based smartphone application," *Iranian Journal of Pediatrics*, vol. 29, no. 2, pp.1-8, 2019.
[33] F. A. Dzulkifli, M. Y. Mashor, and K. Khalid, "Methods for determining bilirubin level in neonatal jaundice screening and monitoring: A literature review," *J. Eng. Res. Educ.*, vol. 10, pp. 1-10, 2018.
[34] Y. Chee, P. Chung, R. Wong, and K. Wong, "Jaundice in infants and children: causes, diagnosis, and management," *Hong Kong Med J*, vol. 24, no. 3, pp. 285-92, 2018.
[35] S. Ullah, K. Rahman, and M. Hedayati, "Hyperbilirubinemia in neonates: types, causes, clinical examinations, preventive measures and treatments: a narrative review article," *Iranian journal of public health*, vol. 45, no. 5, p. 558, 2016.
[36] A. R. Deepthi, C. Singh, and J. Jeeva, "A Device to Detect Hyperbilirubin through Non-Invasive Technique," in *2019 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India*, 2019, pp. 1-4.
[37] M. Aydın, F. Hardalaç, B. Ural, and S. Karap, "Neonatal jaundice detection system," *Journal of medical systems*, vol. 40, no. 7, pp. 1-11, 2016.
[38] R. Singla and S. Singh, "A framework for detection of jaundice in new born babies using homomorphic filtering based image processing," in *2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India*, 2016, pp. 1-5.

[39] M. F. Shafiq, Z. Ahmed, and A. M. Ahmad, "Validity of visual assessment of neonatal jaundice for screening significant hyperbilirubinaemia," *PAFMJ*, vol. 69, no. 1, pp. 212-16, 2019.
[40] M. Mansouri, A. Mahmoodnejad, R. Taghizadeh Sarvestani, and F. Gharibi, "A comparison between transcutaneous bilirubin (TcB) and total serum bilirubin (TSB) measurements in term neonates," *International journal of pediatrics*, vol. 3, no. 3.1, pp. 633-641, 2015.
[41] R. Keren, K. Tremont, X. Luan, and A. Cnaan, "Visual assessment of jaundice in term and late preterm infants," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 94, no. 5, pp. F317-F322, 2009.
[42] A. Riskin, A. Tamir, A. Kugelman, M. Hemo, and D. Bader, "Is visual assessment of jaundice reliable as a screening tool to detect significant neonatal hyperbilirubinemia?," *The Journal of pediatrics*, vol. 152, no. 6, pp. 782-787. e2, 2008.
[43] L. Scalise, "Non contact heart monitoring," in *Advances in Electrocardiograms-Methods and Analysis*: *IntechOpen, London, UK*, 2012.
[44] P. P. Tarjan and R. McFee, "Electrodeless measurements of the effective resistivity of the human torso and head by magnetic induction," *IEEE Transactions on Biomedical Engineering*, no. 4, pp. 266-278, 1968.
[45] R. Guardo, S. Trudelle, A. Adler, C. Boulay, and P. Savard, "Contactless recording of cardiac related thoracic conductivity changes," in *Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society, QC, Canada*, 1995, pp. 1581-1582.
[46] A. Richer and A. Adler, "Eddy current based flexible sensor for contactless measurement of breathing," in *2005 IEEE Instrumentationand Measurement Technology Conference Proceedings, Ottawa, ON, Canada*, 2005, pp. 257-260.
[47] M. Steffen, A. Aleksandrowicz, and S. Leonhardt, "Mobile noncontact monitoring of heart and lung activity," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 1, no. 4, pp. 250-257, 2007.
[48] P. Vetter, L. Leicht, S. Leonhardt, and D. Teichmann, "Integration of an electromagnetic coupled sensor into a driver seat for vital sign monitoring: Initial insight," in *2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES), Vienna, Austria*, 2017, pp. 185-190.
[49] S. Wang *et al.*, "Noninvasive monitoring of vital signs based on highly sensitive fiber optic mattress," *IEEE Sensors Journal*, vol. 20, no. 11, pp. 6182-6190, 2020.
[50] S. Peng, K. Xu, S. Bao, Y. Yuan, C. Dai, and W. Chen, "Flexible electrodes-based smart mattress for monitoring physiological signals of heart and autonomic nerves in a non-contact way," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 6-15, 2020.
[51] H. Dalal, A. Basu, and M. P. Abegaonkar, "Remote sensing of vital sign of human body with radio frequency," *CSI transactions on ICT*, vol. 5, no. 2, pp. 161-166, 2017.
[52] M. S. Rabbani and H. Ghafouri-Shiraz, "Ultra-wide patch antenna array design at 60 GHz band for remote vital sign monitoring with Doppler radar principle," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 38, no. 5, pp. 548-566, 2017.
[53] L. Scalise, P. Marchionni, and I. Ercoli, "Optical method for measurement of respiration rate," in *2010 IEEE International Workshop on Medical Measurements and Applications, Ottawa, ON, Canada*, 2010, pp. 19-22.
[54] B. Lohman, O. Boric-Lubecke, V. Lubecke, P. Ong, and M. Sondhi, "A digital signal processor for Doppler radar sensing of vital signs," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 5, pp. 161-164, 2002.
[55] M. Mercuri, Y.-H. Liu, I. Lorato, T. Torfs, A. Bourdoux, and C. Van Hoof, "Frequency-tracking CW Doppler radar solving small-angle approximation and null point issues in non-contact vital signs monitoring," *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 3, pp. 671-680, 2017.
[56] M. Nosrati, S. Shahsavari, S. Lee, H. Wang, and N. Tavassolian, "A Concurrent Dual-Beam Phased-Array Doppler Radar Using MIMO Beamforming Techniques for Short-

Range Vital-Signs Monitoring," *IEEE Transactions on Antennas and Propagation,* vol. 67, no. 4, pp.2390-2404, 2019.
[57] P. Marchionni, L. Scalise, I. Ercoli, and E. Tomasini, "An optical measurement method for the simultaneous assessment of respiration and heart rates in preterm infants," *Review of Scientific Instruments,* vol. 84, no. 12, p. 121705, 2013.
[58] E. J. Sirevaag, S. Casaccia, E. A. Richter, J. A. O'Sullivan, L. Scalise, and J. W. Rohrbaugh, "Cardiorespiratory interactions: Noncontact assessment using laser Doppler vibrometry," *Psychophysiology,* vol. 53, no. 6, pp. 847-867, 2016.
[59] D. Holdsworth, C. Norley, R. Frayne, D. Steinman, and B. Rutt, "Characterization of common carotid artery blood-flow waveforms in normal human subjects," *Physiological measurement,* vol. 20, no. 3, p. 219, 1999.
[60] P. Arlotto, M. Grimaldi, R. Naeck, and J.-M. Ginoux, "An ultrasonic contactless sensor for breathing monitoring," *Sensors,* vol. 14, no. 8, pp. 15371-15386, 2014.
[61] 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, no. 11, pp. 1113-1119, 2007.
[62] A. K. Abbas, K. Heiman, T. Orlikowsky, and S. Leonhardt, "Non-contact respiratory monitoring based on real-time IR-thermography," in *World Congress on Medical Physics and Biomedical Engineering, Munich, Germany,* 2009, pp. 1306-1309.
[63] S. Que, W. Verkruijsse, M. van Gastel, and S. Stuijk, "Contactless heartbeat measurement using speckle vibrometry," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom,* 2022, pp. 4604-4610.
[64] S. Que *et al.*, "Speckle Vibrometry for Instantaneous Heart Rate Monitoring," *Sensors,* vol. 23, no. 14, p. 6312, 2023.
[65] O. Havakuk *et al.*, "Validation of a novel contact-free heart and respiratory rate monitor," *Journal of Medical Engineering & Technology,* vol. 45, no. 5, pp. 344-350, 2021.
[66] Z. Zalevsky *et al.*, "Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern," *Optics express,* vol. 17, no. 24, pp. 21566-21580, 2009.
[67] M. Yang, Q. Liu, T. Turner, and Y. Wu, "Vital sign estimation from passive thermal video," in *2008 IEEE Conference on Computer Vision and Pattern Recognition, AK, USA,* 2008, pp. 1-8.
[68] S. Bennett, T. N. El Harake, R. Goubran, and F. Knoefel, "Adaptive Eulerian Video Processing of Thermal Video: An Experimental Analysis," *IEEE Transactions on Instrumentation and Measurement,* vol. 66, no. 10, pp. 2516-2524, 2017.
[69] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE,* vol. 86, no. 10, pp. 2009-2025, 1998.
[70] H. E. Elphick, A. H. Alkali, R. K. Kingshott, D. Burke, and R. Saatchi, "Exploratory study to evaluate respiratory rate using a thermal imaging camera," *Respiration,* vol. 97, no. 2, pp. 1-8, 2019.
[71] C. Luca, C. Corciovă, D. Andriţoi, and R. Ciorap, "The Use of Thermal Imaging Techniques as a Method of Monitoring the New Born," in *6th International Conference on Advancements of Medicine and Health Care through Technology,Cluj-Napoca, Romania,* 2018, pp. 35-39.
[72] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography," *Biomedical engineering online,* vol. 10, no. 1, p. 93, 2011.
[73] N. Blanik, A. K. Abbas, B. Venema, V. Blazek, and S. Leonhardt, "Hybrid optical imaging technology for long-term remote monitoring of skin perfusion and temperature behavior," *Journal of biomedical optics,* vol. 19, no. 1, p. 016012, 2014.
[74] R. Chauvin *et al.*, "Contact-free respiration rate monitoring using a pan–tilt thermal camera for stationary bike telerehabilitation sessions," *IEEE Systems Journal,* vol. 10, no. 3, pp. 1046-1055, 2014.

[75] C. B. Pereira, X. Yu, M. Czaplik, R. Rossaint, V. Blazek, and S. Leonhardt, "Remote monitoring of breathing dynamics using infrared thermography," *Biomedical optics express*, vol. 6, no. 11, pp. 4378-4394, 2015.
[76] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery," *IEEE transactions on Biomedical Engineering*, vol. 54, no. 8, pp. 1418-1426, 2007.
[77] I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey, "Interacting with human physiology," *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 150-170, 2007.
[78] N. Sun, M. Garbey, A. Merla, and I. Pavlidis, "Imaging the cardiovascular pulse," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA*, 2005, pp. 416-421.
[79] S. Y. Chekmenev, A. A. Farag, W. M. Miller, E. A. Essock, and A. Bhatnagar, "Multiresolution approach for noncontact measurements of arterial pulse using thermal imaging," in *Augmented vision perception in infrared*: Springer, 2009, pp. 87-112.
[80] K. Pulli, A. Baksheev, K. Kornyakov, and V. Eruhimov, "Real-time computer vision with OpenCV," *Communications of the ACM*, vol. 55, no. 6, pp. 61-69, 2012.
[81] F.-T.-Z. Khanam, A. Al-Naji, A. G. Perera, K. Gibson, and J. Chahl, "Non-contact automatic vital signs monitoring of neonates in NICU using video camera imaging," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 11, no. 2, pp.278-285, 2022.
[82] N. Jasti *et al.*, "Medical Applications of Infrared Thermography: A Narrative Review," *Journal of Stem Cells*, vol. 14, no. 1, pp. 35-53, 2019.
[83] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine vision and applications*, vol. 25, no. 1, pp. 245-262, 2014.
[84] C. LLOYD. "How Does Thermal Imaging Work?" https://www.howtogeek.com/294076/how-does-thermal-imaging-work/ (accessed 22 May, 2020).
[85] N. Bouzida, A. Bendada, and X. P. Maldague, "Visualization of body thermoregulation by infrared imaging," *Journal of Thermal Biology*, vol. 34, no. 3, pp. 120-126, 2009.
[86] "Thermal Imaging for Detecting Elevated Body Temperature. FLIR Systems." https://www.flir.com.au/discover/public-safety/thermal-imaging-for-detecting-elevated-body-temperature (accessed 22 May, 2020).
[87] E. Y. Ng, G. Kawb, and W. Chang, "Analysis of IR thermal imager for mass blind fever screening," *Microvascular research*, vol. 68, no. 2, pp. 104-109, 2004.
[88] G.-A. Bilodeau *et al.*, "Body temperature estimation of a moving subject from thermographic images," *Machine Vision and Applications*, vol. 23, no. 2, pp. 299-311, 2012.
[89] B. Aubakir, B. Nurimbetov, I. Tursynbek, and H. A. Varol, "Vital sign monitoring utilizing Eulerian video magnification and thermography," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA*, 2016, pp. 3527-3530.
[90] A. Sharma and A. R. Yadav, "Image processing based body temperature estimation using thermal video sequence," in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, 2017: IEEE, pp. 846-852.
[91] J.-W. Lin, M.-H. Lu, and Y.-H. Lin, "A Thermal Camera Based Continuous Body Temperature Measurement System," in *Proceedings of the IEEE International Conference on Computer Vision Workshops, Seoul, South Korea*, 2019, pp. 681–687.
[92] S. Sumriddetchkajorn and K. Chaitavon, "Field test studies of our infrared-based human temperature screening system embedded with a parallel measurement approach," *Infrared physics & technology*, vol. 52, no. 4, pp. 119-123, 2009.
[93] N. Silawan, K. Kusukame, K. J. Kek, and W. S. Kuan, "A Novel Environment-Invariant Core Body Temperature Estimation for High Sensitivity and Specificity Fever Screening," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), HI, USA*, 2018, pp. 1612-1615.

[94] J. B. Thomas, L. Pahler, R. Handy, M. S. Thiese, and C. Schaefer, "Pilot study predicting core body temperatures in hot work environments using thermal imagery," *Journal of Chemical Health & Safety*, vol. 26, no. 6, pp. 75-83, 2019.
[95] R. Murthy and I. Pavlidis, "Noncontact measurement of breathing function," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 3, pp. 57-67, 2006.
[96] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis, "Touchless monitoring of breathing function," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *San Francisco, CA, USA*, 2004, pp. 1196-1199.
[97] J. Fei and I. Pavlidis, "Analysis of breathing air flow patterns in thermal imaging," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 946-952.
[98] J. Fei, Z. Zhu, and I. Pavlidis, "Imaging breathing rate in the co 2 absorption band," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, *Shanghai, China*, 2005, pp. 700-705.
[99] S. Y. Chekmenev, H. Rara, and A. A. Farag, "Non-contact, wavelet-based measurement of vital signs using thermal imaging," in *The first international conference on graphics, vision, and image processing (GVIP), Cairo, Egypt*, 2005, pp. 107-112.
[100] J. Fei and I. Pavlidis, "Thermistor at a distance: unobtrusive measurement of breathing," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 988-998, 2009.
[101] M. F. M. Shakhih, A. A. Wahab, and M. I. M. Salim, "Assessment of inspiration and expiration time using infrared thermal imaging modality," *Infrared Physics & Technology*, vol. 99, pp. 129-139, 2019.
[102] C. Barbosa Pereira, M. Czaplik, V. Blazek, S. Leonhardt, and D. Teichmann, "Monitoring of cardiorespiratory signals using thermal imaging: a pilot study on healthy human subjects," *Sensors*, vol. 18, no. 5, p. 1541, 2018.
[103] C. B. Pereira, X. Yu, V. Blazek, B. Venema, and S. Leonhardt, "Multisensor data fusion for enhanced respiratory rate estimation in thermal videos," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA*, 2016, pp. 1381-1384.
[104] C. B. Pereira, K. Heimann, B. Venema, V. Blazek, M. Czaplik, and S. Leonhardt, "Estimation of respiratory rate from thermal videos of preterm infants," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, *Jeju, Korea*, 2017, pp. 3818-3821.
[105] C. B. Pereira *et al.*, "Noncontact monitoring of respiratory rate in newborn infants using thermal imaging," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 4, pp. 1105-1114, 2018.
[106] K. Nakajima, A. Osa, and H. Miike, "A method for measuring respiration and physical activity in bed by optical flow analysis," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136) IL, USA*, 1997, pp. 2054-2057.
[107] M. Frigola, J. Amat, and J. Pagès, "Vision based respiratory monitoring system," in *Proceedings of the 10th Mediterranean Conference on Control and Automation (MED 2002), Lisbon, Portugal*, 2002, pp. 9-13.
[108] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Portland, OR, USA*, 2013, pp. 3430-3437.
[109] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *CVPR (1)*, vol. 1, pp. 511-518, 2001.
[110] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *in Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada*, 1981, pp. 674–679.

[111] L. Shan and M. Yu, "Video-based heart rate measurement using head motion tracking and ICA," in *2013 6th International Congress on Image and Signal Processing (CISP), Hangzhou, China*, 2013, pp. 160-164.
[112] R. Irani, K. Nasrollahi, and T. B. Moeslund, "Improved pulse detection from head motions using DCT," in *2014 international conference on computer vision theory and applications (VISAPP), Lisbon, Portugal*, 2014, pp. 118-124.
[113] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, "Heartbeat rate measurement from facial video," *IEEE Intelligent Systems*, vol. 31, no. 3, pp. 40-48, 2016.
[114] J.-P. Lomaliza and H. Park, "Detecting Pulse from Head Motions Using Smartphone Camera," in *International Conference on Advanced Engineering Theory and Applications, Busan, Vietnnam*, 2016, pp. 243-251.
[115] J.-P. Lomaliza and H. Park, "Improved Heart-Rate Measurement from Mobile Face Videos," *Electronics*, vol. 8, no. 6, p. 663, 2019.
[116] X. He, R. A. Goubran, and X. P. Liu, "Wrist pulse measurement and analysis using Eulerian video magnification," in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, USA*, 2016, pp. 41-44.
[117] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM transactions on graphics (TOG)*, vol. 31, no. 4, pp.1-8, 2012.
[118] A. Al-Naji and J. Chahl, "Contactless cardiac activity detection based on head motion magnification," *International Journal of Image and Graphics*, vol. 17, no. 01, p. 1750001, 2017.
[119] A. Al-Naji and J. Chahl, "Remote respiratory monitoring system based on developing motion magnification technique," *Biomedical Signal Processing and Control*, vol. 29, pp. 1-10, 2016.
[120] 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.
[121] A. Al-Naji and J. Chahl, "Detection of cardiopulmonary activity and related abnormal events using Microsoft Kinect sensor," *Sensors*, vol. 18, no. 3, p. 920, 2018.
[122] A. Al-Naji, S.-H. Lee, and J. Chahl, "An efficient motion magnification system for real-time applications," *Machine Vision and Applications*, vol. 29, no. 4, pp. 585-600, 2018.
[123] A. B. Hertzman, "Observations on the finger volume pulse recorded photoelectrically," *Am. J. Physiol.*, vol. 119, pp. 334-335, 1937.
[124] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479-1491, 2017.
[125] 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.
[126] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics express*, vol. 18, no. 10, pp. 10762-10774, 2010.
[127] 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, 2011.
[128] T. Pursche, J. Krajewski, and R. Moeller, "Video-based heart rate measurement from human faces," in *2012 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA*, 2012, pp. 544-545.
[129] 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), Szczecin, Poland*, 2011, pp. 405-410.
[130] O. Gupta, D. McDuff, and R. Raskar, "Real-time physiological measurement and visualization using a synchronized multi-camera system," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA*, 2016, pp. 46-53.

[131] M.-H. Hu, G.-T. Zhai, D. Li, Y.-Z. Fan, X.-H. Chen, and X.-K. Yang, "Synergetic use of thermal and visible imaging techniques for contactless and unobtrusive breathing measurement," *Journal of biomedical optics*, vol. 22, no. 3, p. 036006, 2017.
[132] M. Hu *et al.*, "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation," *PloS one*, vol. 13, no. 1, p. e0190466, 2018.
[133] S. Yu, S. Hu, V. Azorin-Peris, J. A. Chambers, Y. Zhu, and S. E. Greenwald, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *Journal of biomedical optics*, vol. 16, no. 7, p. 077010, 2011.
[134] L. Feng, L.-M. Po, X. Xu, and Y. Li, "Motion artifacts suppression for remote imaging photoplethysmography," in *2014 19th International Conference on Digital Signal Processing, Hong Kong, China*, 2014, pp. 18-23.
[135] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. J. Wang, "Video-based human heart rate measurement using joint blind source separation," *Biomedical Signal Processing and Control*, vol. 31, pp. 309-320, 2017.
[136] F. Bousefsaf, C. Maaoui, and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 568-574, 2013.
[137] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878-2886, 2013.
[138] G. De Haan and A. Van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiological measurement*, vol. 35, no. 9, p. 1913, 2014.
[139] L. Feng, L.-M. Po, X. Xu, Y. Li, and R. Ma, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 5, pp. 879-891, 2015.
[140] W. Wang, S. Stuijk, and G. De Haan, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," *IEEE transactions on biomedical engineering*, vol. 63, no. 9, pp. 1974-1984, 2016.
[141] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Robust heart rate from fitness videos," *Physiological measurement*, vol. 38, no. 6, p. 1023, 2017.
[142] B.-F. Wu, P.-W. Huang, T.-Y. Tsou, T.-M. Lin, and M.-L. Chung, "Camera-based heart rate measurement using continuous wavelet transform," in *2017 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh City, Vietnam*, 2017, pp. 7-11.
[143] B.-F. Wu, P.-W. Huang, C.-H. Lin, M.-L. Chung, T.-Y. Tsou, and Y.-L. Wu, "Motion resistant image-photoplethysmography based on spectral peak tracking algorithm," *IEEE Access*, vol. 6, pp. 21621-21634, 2018.
[144] K. Xie, C.-H. Fu, H. Liang, H. Hong, and X. Zhu, "Non-contact Heart Rate Monitoring for Intensive Exercise Based on Singular Spectrum Analysis," in *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA*, 2019, pp. 228-233.
[145] D. J. McDuff, E. B. Blackford, J. R. Estepp, and I. Nishidate, "A fast non-contact imaging photoplethysmography method using a tissue-like model," in *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics, San Francisco, CA, USA*, 2018, p. 105010Q.
[146] S. Fallet, Y. Schoenenberger, L. Martin, F. Braun, V. Moser, and J.-M. Vesin, "Imaging photoplethysmography: A real-time signal quality index," in *2017 Computing in Cardiology (CinC), Rennes, France*, 2017, pp. 1-4.
[147] D.-Y. Chen *et al.*, "Image sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 618-627, 2015.
[148] K.-Y. Lin, D.-Y. Chen, and W.-J. Tsai, "Face-based heart rate signal decomposition and evaluation using multiple linear regression," *IEEE Sensors Journal*, vol. 16, no. 5, pp. 1351-1360, 2016.
[149] D. Lee, J. Kim, S. Kwon, and K. Park, "Heart rate estimation from facial photoplethysmography during dynamic illuminance changes," in *2015 37th Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 2015, pp. 2758-2761.

[150] 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.

[151] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1422-1433, 2017.

[152] L. Xu, J. Cheng, and X. Chen, "Illumination variation interference suppression in remote PPG using PLS and MEMD," *Electronics Letters*, vol. 53, no. 4, pp. 216-218, 2017.

[153] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus,OH, USA*, 2014, pp. 4264-4271.

[154] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomedical optics express*, vol. 6, no. 5, pp. 1565-1588, 2015.

[155] A. Al-Naji, A. G. Perera, and J. Chahl, "Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle," *Biomedical engineering online*, vol. 16, no. 1, p. 101, 2017.

[156] A. Al-Naji, S.-H. Lee, and J. Chahl, "Quality index evaluation of videos based on fuzzy interface system," *IET Image Processing*, vol. 11, no. 5, pp. 292-300, 2017.

[157] J.-H. Chen, I.-L. Tang, and C.-H. Chang, "Enhancing the detection rate of inclined faces," in *2015 IEEE Trustcom/BigDataSE/ISPA, Helsinki, Finland*, 2015, pp. 143-146.

[158] S. Kwon, H. Kim, and K. S. Park, "Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA*, 2012, pp. 2174-2177.

[159] A. Al-Naji, A. G. Perera, and J. Chahl, "Remote measurement of cardiopulmonary signal using an unmanned aerial vehicle," in *IOP Conference Series: Materials Science and Engineering, Bristol, UK*, 2018, p. 012001.

[160] D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in remote cardiopulmonary measurement using a five band digital camera," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 10, pp. 2593-2601, 2014.

[161] Y. Sun, V. Azorin-Peris, R. Kalawsky, S. Hu, C. Papin, and S. E. Greenwald, "Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam," *Journal of biomedical optics*, vol. 17, no. 3, p. 037005, 2012.

[162] N. Bernacchia, L. Scalise, L. Casacanditella, I. Ercoli, P. Marchionni, and E. P. Tomasini, "Non contact measurement of heart and respiration rates based on Kinect™," in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Lisboa, Portugal*, 2014, pp. 1-5.

[163] T. Smilkstein, M. Buenrostro, A. Kenyon, M. Lienemann, and G. Larson, "Heart rate monitoring using Kinect and color amplification," in *2014 IEEE Healthcare Innovation Conference (HIC), Seattle,WA,USA*, 2014, pp. 60-62.

[164] E. Gambi *et al.*, "Heart rate detection using microsoft kinect: Validation and comparison to wearable devices," *Sensors*, vol. 17, no. 8, p. 1776, 2017.

[165] L. Scalise, N. Bernacchia, I. Ercoli, and P. Marchionni, "Heart rate measurement in neonatal patients using a webcamera," in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings, Budapest, Hungary*, 2012, pp. 1-4.

[166] J.-C. Cobos-Torres, M. Abderrahim, and J. Martínez-Orgado, "Non-Contact, Simple Neonatal Monitoring by Photoplethysmography," *Sensors*, vol. 18, no. 12, p. 4362, 2018.

[167] K. Gibson *et al.*, "Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: a method comparison study," *Pediatric Research*, vol. 86, no. 6, pp.738-741, 2019.
[168] U. Bal, "Non-contact estimation of heart rate and oxygen saturation using ambient light," *Biomedical optics express*, vol. 6, no. 1, pp. 86-97, 2015.
[169] R. Fouad, O. A. Omer, and M. H. Aly, "Optimizing Remote Photoplethysmography Using Adaptive Skin Segmentation for Real-Time Heart Rate Monitoring," *IEEE Access*, vol. 7, pp. 76513–76528, 2019.
[170] D. Datcu, M. Cidota, S. Lukosch, and L. Rothkrantz, "Noncontact automatic heart rate analysis in visible spectrum by specific face regions," in *Proceedings of the 14th International Conference on Computer Systems and Technologies, Ruse, Bulgaria*, 2013, pp. 120-127.
[171] L. Feng *et al.*, "Dynamic ROI based on K-means for remote photoplethysmography," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), QLD, Australia*, 2015, pp. 1310-1314.
[172] C. Zhao, W. Chen, C.-L. Lin, and X. Wu, "Physiological signal preserving video compression for remote photoplethysmography," *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4537-4548, 2019.
[173] A. Al-Naji, J. Chahl, and S.-H. Lee, "Cardiopulmonary signal acquisition from different regions using video imaging analysis," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 7, no. 2, pp.117-131, 2018.
[174] A. Al-Naji and J. Chahl, "Remote Optical Cardiopulmonary Signal Extraction With Noise Artifact Removal, Multiple Subject Detection & Long-Distance," *IEEE Access*, vol. 6, pp. 11573-11595, 2018.
[175] Y. Hsu, Y.-L. Lin, and W. Hsu, "Learning-based heart rate detection from remote photoplethysmography features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, Australia*, 2014, pp. 4433-4437.
[176] C. Wiede, J. Richter, and G. Hirtz, "Signal fusion based on intensity and motion variations for remote heart rate determination," in *2016 IEEE International Conference on Imaging Systems and Techniques (IST), Chania, Greece*, 2016, pp. 526-531.
[177] 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, San Francisco, CA, USA*, 2014, p. 89351P.
[178] M. Villarroel *et al.*, "Continuous non-contact vital sign monitoring in neonatal intensive care unit," *Healthcare technology letters*, vol. 1, no. 3, pp. 87-91, 2014.
[179] U. Rubins, Z. Marcinkevics, I. Logina, A. Grabovskis, and E. Kviesis-Kipge, "Imaging photoplethysmography for assessment of chronic pain patients," in *Optical Diagnostics and Sensing XIX: Toward Point-of-Care Diagnostics, San Francisco, CA, USA*, 2019, p. 1088508.
[180] N. Zaproudina *et al.*, "Asynchronicity of facial blood perfusion in migraine," *PloS one*, vol. 8, no. 12, p. e80189, 2013.
[181] S. Rasche *et al.*, "Camera-based photoplethysmography in critical care patients," *Clinical hemorheology and microcirculation*, vol. 64, no. 1, pp. 77-90, 2016.
[182] R. Amelard, D. A. Clausi, and A. Wong, "Spectral-spatial fusion model for robust blood pulse waveform extraction in photoplethysmographic imaging," *Biomedical optics express*, vol. 7, no. 12, pp. 4874-4885, 2016.
[183] U. Rubīns, J. Spīgulis, and A. Miščuks, "Photoplethysmography imaging algorithm for continuous monitoring of regional anesthesia," in *2016 14th ACM/IEEE Symposium on Embedded Systems For Real-time Multimedia (ESTIMedia,Pittsburgh, PA, USA*, 2016, pp. 1-5.
[184] A. Trumpp *et al.*, "Camera-based photoplethysmography in an intraoperative setting," *Biomedical engineering online*, vol. 17, no. 1, p. 33, 2018.

[185] 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), Washington, DC, USA*, 2017, pp. 278-285.
[186] J. E. Thatcher *et al.*, "Multispectral and photoplethysmography optical imaging techniques identify important tissue characteristics in an animal model of tangential burn excision," *Journal of Burn Care & Research,* vol. 37, no. 1, pp. 38-52, 2016.
[187] W. Wang, S. Stuijk, and G. De Haan, "Unsupervised subject detection via remote PPG," *IEEE Transactions on Biomedical Engineering,* vol. 62, no. 11, pp. 2629-2637, 2015.
[188] R. M. Seepers, W. Wang, G. de Haan, I. Sourdis, and C. Strydis, "Attacks on heartbeat-based security using remote photoplethysmography," *IEEE journal of biomedical and health informatics,* vol. 22, no. 3, pp. 714-721, 2017.
[189] P. V. Rouast, M. T. Adam, D. J. Cornforth, E. Lux, and C. Weinhardt, "Using contactless heart rate measurements for real-time assessment of affective states," in *Information Systems and Neuroscience*: Springer, 2017, pp. 157-163.
[190] A. M. Unakafov, S. Möller, I. Kagan, A. Gail, S. Treue, and F. Wolf, "Using imaging photoplethysmography for heart rate estimation in non-human primates," *PloS one,* vol. 13, no. 8, p. e0202581, 2018.
[191] P. Puppalwar, K. Goswami, and A. Dhok, "Review on "evolution of methods of bilirubin estimation"," *IOSR Journal of Dental and Medical Sciences,* vol. 1, no. 3, pp. 17-28, 2012.
[192] B. Ansong-Assoku and P. A. Ankola, "Neonatal jaundice," 2018.
[193] M. J. Maisels, "Managing the jaundiced newborn: a persistent challenge," *Cmaj,* vol. 187, no. 5, pp. 335-343, 2015.
[194] K. Mreihil *et al.*, "Uniform national guidelines do not prevent wide variations in the clinical application of phototherapy for neonatal jaundice," *Acta Paediatrica,* vol. 107, no. 4, pp. 620-627, 2018.
[195] N. C. C. f. Women's and C. s. Health, "Neonatal jaundice: Clinical guideline," *London (UK): National Institute for Health and Clinical Excellence (NICE),* 2010.
[196] K. Mreihil *et al.*, "Phototherapy is commonly used for neonatal jaundice but greater control is needed to avoid toxicity in the most vulnerable infants," *Acta Paediatrica,* vol. 107, no. 4, pp. 611-619, 2018.
[197] M. Mussavi, P. Niknafs, and B. Bijari, "Determining the correlation and accuracy of three methods of measuring neonatal bilirubin concentration," *Iranian Journal of Pediatrics,* vol. 23, no. 3, p. 333, 2013.
[198] K. Srividya, K. Renganathan, S. Meha, and U. Yogabhuvaneswari, "Review on Jaundice Detection in Neonates Using Image Processing," in *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT),Chennai, India*, 2022, pp. 1-5.
[199] Z. Osman, A. Ahmad, and A. Muharam, "Rapid prototyping of neonatal jaundice detector using skin optics theory," in *2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, Malaysia*, 2014, pp. 328-331.
[200] W. Hashim, M. Alkhaled, A. Al-Naji, and I. Al-Rayahi, "A Review on Image Processing Based Neonatal Jaundice Detection Techniques," in *2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM), Kuala Lumpur, Malaysia*, 2021, pp. 213-218.
[201] W. Hashim, A. Al-Naji, I. A. Al-Rayahi, and M. Oudah, "Computer Vision for Jaundice Detection in Neonates Using Graphic User Interface," in *IOP Conference Series: Materials Science and Engineering, Mosul, Iraq*, 2021, p. 012076.
[202] A. Laddi, S. Kumar, S. Sharma, and A. Kumar, "Non-invasive jaundice detection using machine vision," *IETE Journal of Research,* vol. 59, no. 5, pp. 591-596, 2013.
[203] A. Gupta, A. Kumar, and P. Khera, "Method and model for Jaundice prediction through non-invasive Bilirubin detection technique," *International Journal of Engineering Research & Technology (IJERT),* vol. 4, no. 8, pp. 34-38, 2015.

[204] R. Angelico *et al.*, "A novel mobile phone application for infant stool color recognition: An easy and effective tool to identify acholic stools in newborns," *Journal of Medical Screening,* vol. 28, no. 3, pp. 230-237, 2021.
[205] M. M. M. Miah *et al.*, "Non-invasive bilirubin level quantification and jaundice detection by sclera image processing," in *2019 IEEE Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA*, 2019, pp. 1-7.
[206] M. Amani, H. Falk, O. D. Jensen, G. Vartdal, A. Aune, and F. Lindseth, "Color calibration on human skin images," in *International Conference on Computer Vision Systems, Thessaloniki, Greece*, 2019, pp. 211-223.
[207] A. Aune, G. Vartdal, H. Bergseng, L. L. Randeberg, and E. Darj, "Bilirubin estimates from smartphone images of newborn infants' skin correlated highly to serum bilirubin levels," *Acta Paediatrica,* vol. 109, no. 12, pp. 2532-2538, 2020.
[208] W. Hashim, A. Al-Naji, I. A. Al-Rayahi, M. Alkhaled, and J. Chahl, "Neonatal Jaundice Detection Using a Computer Vision System," *Designs,* vol. 5, no. 4, p. 63, 2021.
[209] S. Kawano, T. T. Zin, and Y. Kodama, "A Study on Non-contact and Non-invasive Neonatal Jaundice Detection and Bilirubin Value Prediction," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara, Japan*, 2018, pp. 401-402.
[210] Z. Rong *et al.*, "Evaluation of an automatic image-based screening technique for neonatal hyperbilirubinemia," *Zhonghua er ke za zhi= Chinese Journal of Pediatrics,* vol. 54, no. 8, pp. 597-600, 2016.
[211] A. Chakraborty, S. Goud, V. Shetty, and B. Bhattacharyya, "Neonatal Jaundice Detection System using CNN Algorithm and Image Processing," *International Journal of Electrical Engineering and Technology,* vol. 11, no. 3, 2020, pp.248-264.
[212] J. Castro-Ramos, C. Toxqui-Quitl, F. V. Manriquez, E. Orozco-Guillen, A. Padilla-Vivanco, and J. Sánchez-Escobar, "Detecting jaundice by using digital image processing," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXI, San Francisco, California, United States*, 2014, pp. 290-296.
[213] E. Juliastuti, V. Nadhira, Y. W. Satwika, N. A. Aziz, and N. Zahra, "Risk zone estimation of newborn jaundice based on skin color image analysis," in *2019 6th International Conference on Instrumentation, Control, and Automation (ICA), Bandung, Indonesia*, 2019, pp. 176-181.
[214] L. De Greef *et al.*, "Bilicam: using mobile phones to monitor newborn jaundice," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA*, 2014, pp. 331-342.
[215] M. N. Mansor, S. Yaacob, H. Muthusamy, and S. Nisha, "PCA-based feature extraction and k-NN algorithm for early jaundice detection," *System,* vol. 1, no. 1, 2011.
[216] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision,* vol. 57, no. 2, pp. 137-154, 2004.
[217] V. Dantu, J. Vempati, and S. Srivilliputhur, "Non-invasive blood glucose monitor based on spectroscopy using a smartphone," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA*, 2014, pp. 3695-3698.
[218] J. P. Devadhasan, H. Oh, C. S. Choi, and S. Kim, "Whole blood glucose analysis based on smartphone camera module," *Journal of biomedical optics,* vol. 20, no. 11, p. 117001, 2015.
[219] P. Leijdekkers, V. Gay, and E. Lawrence, "Smart homecare system for health tele-monitoring," in *First International Conference on the Digital Society (ICDS'07), Guadeloupe, French Caribbean*, 2007, pp. 3-3.
[220] E. C. Eichenwald, "Apnea of prematurity," *Pediatrics,* vol. 137, no. 1, pp. 1-7, 2016.
[221] "Go Direct® Respiration Belt." https://www.vernier.com/product/go-direct-respiration-belt/ (accessed on April, 2020).
[222] H.-D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern recognition,* vol. 34, no. 12, pp. 2259-2281, 2001.

[223] F. Y. Shih and S. Cheng, "Automatic seeded region growing for color image segmentation," *Image and vision computing*, vol. 23, no. 10, pp. 877-886, 2005.
[224] M. Tkalcic and J. F. Tasic, "Colour spaces: perceptual, historical and applicational background", *The IEEE Region 8 EUROCON 2003. Computer as a Tool., Ljubljana, Slovenia*, 2003, pp. 304-308.
[225] X. Wu, "YIQ vector quantization in a new color palette architecture," *IEEE Transactions on Image Processing*, vol. 5, no. 2, pp. 321-329, 1996.
[226] M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, vol. 26, no. 5, pp. 340-350, 2001.
[227] A. Elgammal, C. Muang, and D. Hu, "Skin detection-a short tutorial," *Encyclopedia of Biometrics*, vol. 4, pp. 1218-1224, 2009.
[228] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA*, 2016, pp. 779-788.
[229] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *International Journal of Computer Vision*, vol. 127, pp. 115-142, 2019.
[230] A. Rosebrock. "Facial landmarks with dlib, OpenCV, and Python." https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/ (accessed on June, 2020)
[231] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903-995, 1998.
[232] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Computers in biology and medicine*, vol. 38, no. 1, pp. 1-13, 2008.
[233] X. Zhang and P. Zhou, "Filtering of surface EMG using ensemble empirical mode decomposition," *Medical engineering & physics*, vol. 35, no. 4, pp. 537-542, 2013.
[234] X. Chen, A. Liu, J. Chiang, Z. J. Wang, M. J. McKeown, and R. K. Ward, "Removing muscle artifacts from EEG data: Multichannel or single-channel techniques?," *IEEE Sensors Journal*, vol. 16, no. 7, pp. 1986-1997, 2015.
[235] S. Liu, Q. He, R. X. Gao, and P. Freedson, "Empirical mode decomposition applied to tissue artifact removal from respiratory signal," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada,*, 2008, pp. 3624-3627.
[236] R. Song, J. Li, M. Wang, J. Cheng, C. Li, and X. Chen, "Remote Photoplethysmography with An EEMD-MCCA Method Robust Against Spatially Uneven Illuminations," *IEEE Sensors Journal*, vol. 21, no.12, pp. 13484 - 13494, 2021.
[237] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1-41, 2009.
[238] O. Fialka and M. Cadik, "FFT and convolution performance in image filtering on GPU," *presented at Tenth International Conference on the Information Visualization, London, England*, 2006, pp. 609-614.
[239] L. Asiminoael, F. Blaabjerg, and S. Hansen, "Detection is key-Harmonic detection methods for active power filter applications," *IEEE Industry Applications Magazine*, vol. 13, no. 4, pp. 22-33, 2007.
[240] A. Banerjee, A. S. Dhar, and S. Banerjee, "FPGA realization of a CORDIC based FFT processor for biomedical signal processing," *Microprocessors and Microsystems*, vol. 25, no. 3, pp. 131-142, 2001.

[241] M. M. Butt, U. Akram, and S. A. Khan, "Denoising practices for electrocardiographic (ECG) signals: a survey," in *2015 international conference on computer, communications, and control technology (I4CT), Kuching, Malaysia*, 2015, pp. 264-268.
[242] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, New York: springer, 2009.
[243] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*, 2012.
[244] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning* (no. 4). Springer, 2006.
[245] T. M. Mitchell, *Machine learning*. McGraw-hill New York, 2007.
[246] J. H. Friedman, "Contrast trees and distribution boosting," *Proceedings of the National Academy of Sciences*, vol. 117, no. 35, pp. 21175-21184, 2020.
[247] J. M. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The lancet*, vol. 327, no. 8476, pp. 307-310, 1986.
[248] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Statistical methods in medical research*, vol. 8, no. 2, pp. 135-160, 1999.
[249] "W.H.O. International statistical classification of diseases and related health problems." (accessed on 1 March 2021).
[250] H. Blencowe *et al.*, "National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications," *The lancet*, vol. 379, no. 9832, pp. 2162-2172, 2012.
[251] C. Kenner and J. W. Lott, *Comprehensive neonatal care: an interdisciplinary approach*. Elsevier Health Sciences, 2007.
[252] M. M. Baharestani, "An overview of neonatal and pediatric wound care knowledge and considerations," *Ostomy/wound management*, vol. 53, no. 6, p. 34, 2007.
[253] J. M. Kuller, "Skin breakdown: risk factors, prevention, and treatment," *Newborn and Infant Nursing Reviews*, vol. 1, no. 1, pp. 35-42, 2001.
[254] R. G. Pineda *et al.*, "Alterations in brain structure and neurodevelopmental outcome in preterm infants hospitalized in different neonatal intensive care unit environments," *The Journal of pediatrics*, vol. 164, no. 1, pp. 52-60. e2, 2014.
[255] W. Lv, W. He, X. Lin, and J. Miao, "Non-Contact Monitoring of Human Vital Signs Using FMCW Millimeter Wave Radar in the 120 GHz Band," *Sensors*, vol. 21, no. 8, p. 2732, 2021.
[256] 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.
[257] M. Paul *et al.*, "Non-contact sensing of neonatal pulse rate using camera-based imaging: a clinical feasibility study," *Physiological Measurement*, vol. 41, no. 2, p. 024001, 2020.
[258] Q. Chen *et al.*, "Camera-based heart rate estimation for hospitalized newborns in the presence of motion artifacts," *BioMedical Engineering OnLine*, vol. 20, no. 1, pp. 1-16, 2021.
[259] L. K. Mestha, S. Kyal, B. Xu, L. E. Lewis, and V. Kumar, "Towards continuous monitoring of pulse rate in neonatal intensive care unit with a webcam," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA*, 2014, pp. 3817-3820.
[260] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 211-223, 2015.
[261] T. M. Mahmoud, "A new fast skin color detection technique," *World Academy of Science, Engineering and Technology*, vol. 43, pp. 501-505, 2008.
[262] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern recognition*, vol. 40, no. 3, pp. 1106-1122, 2007.
[263] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon*, 2003, vol. 3: Moscow, Russia, pp. 85-92.

[264] H. V. Lakshmi and S. PatilKulakarni, "Segmentation algorithm for multiple face detection in color images with skin tone regions using color spaces and edge detection techniques," *International journal of computer theory and engineering,* vol. 2, no. 4, p. 552, 2010.
[265] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics,* vol. 9, no. 1, pp. 62-66, 1979.
[266] M. M. Baharestani, "An overview of neonatal and pediatric wound care knowledge and considerations," *Ostomy/wound management,* vol. 53, no. 6, pp. 34-6, 38, 40, passim, 2007.
[267] 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.
[268] J. Jorge *et al.*, "Non-contact monitoring of respiration in the neonatal intensive care unit," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017, Washington, DC, USA*, 2017, pp. 286-293.
[269] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks,* vol. 3361, no. 10, p. 1995, 1995.
[270] S. Chaichulee *et al.*, "Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA*, 2017, pp. 266-272.
[271] J. Jorge, M. Villarroel, S. Chaichulee, K. McCormick, and L. Tarassenko, "Data fusion for improved camera-based detection of respiration in neonates," in *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics San Francisco, California, United States*, 2018, pp. 215-224.
[272] M. Villarroel *et al.*, "Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit," *NPJ digital medicine,* vol. 2, no. 1, pp. 1-18, 2019.
[273] S. Chaichulee *et al.*, "Localised photoplethysmography imaging for heart rate estimation of pre-term infants in the clinic," in *Optical diagnostics and sensing XVIII: toward point-of-care diagnostics, San Francisco, California, United States*, 2018, pp. 146-159.
[274] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767,* pp. 1-7, 2018,.
[275] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision, Zurich, Switzerland*, 2014, pp. 740-755.
[276] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA*, 2014, pp. 580-587.
[277] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision, Santiago, Chile*, 2015, pp. 1440-1448.
[278] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields," *IEEE transactions on pattern analysis and machine intelligence,* vol. 43, no. 1, pp. 172-186, 2019.
[279] D.-Y. Chen *et al.*, "Image sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform," *IEEE Sensors Journal,* vol. 15, no. 1, pp. 618-627, 2014.
[280] F.-T.-Z. Khanam, A. Al-Naji, and J. Chahl, "Remote Monitoring of Vital Signs in Diverse Non-Clinical and Clinical Scenarios Using Computer Vision Systems: A Review," *Applied Sciences,* vol. 9, no. 20, p. 4474, 2019.
[281] Z. Jiang *et al.*, "Detection of respiratory infections using RGB-infrared sensors on portable device," *IEEE Sensors Journal,* vol. 20, no. 22, pp. 13674-13681, 2020.
[282] R. L. Riha, "Diagnostic approaches to respiratory sleep disorders," *Journal of thoracic disease,* vol. 7, no. 8, p. 1373, 2015.
[283] T. Duning *et al.*, "Brainstem involvement as a cause of central sleep apnea: pattern of microstructural cerebral damage in patients with cerebral microangiopathy," *PLoS One,* vol. 8, no. 4, p. e60304, 2013.

[284] J. Naranjo, R. Centeno, D. Galiano, and M. Beaus, "A nomogram for assessment of breathing patterns during treadmill exercise," *British journal of sports medicine,* vol. 39, no. 2, pp. 80-83, 2005.
[285] G. Sun *et al.*, "Remote sensing of multiple vital signs using a CMOS camera-equipped infrared thermography system and its clinical application in rapidly screening patients with suspected infectious diseases," *International Journal of Infectious Diseases,* vol. 55, pp. 113-117, 2017.
[286] E. M. Nowara, A. Sabharwal, and A. Veeraraghavan, "PPGSecure: biometric presentation attack detection using photopletysmograms," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA*, 2017, pp. 56-62.
[287] F.-T.-Z. Khanam *et al.*, "Noncontact Sensing of Contagion," *Journal of Imaging,* vol. 7, no. 2, p. 28, 2021.
[288] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Physics & Technology,* vol. 82, pp. 8-17, 2017.
[289] G. Scebba, L. Tüshaus, and W. Karlen, "Multispectral camera fusion increases robustness of ROI detection for biosignal estimation with nearables in real-world scenarios," in *2018 40th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA*, 2018, pp. 5672-5675.
[290] T. Negishi, G. Sun, H. Liu, S. Sato, T. Matsui, and T. Kirimoto, "Stable contactless sensing of vital signs using RGB-thermal image fusion system with facial tracking for infection screening," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA*, 2018, pp. 4371-4374.
[291] G. Scebba, G. Da Poian, and W. Karlen, "Multispectral video fusion for non-contact monitoring of respiratory rate and apnea," *IEEE Transactions on Biomedical Engineering,* vol. 68, no. 1, pp. 350-359, 2020.
[292] F. Yang, S. He, S. Sadanand, A. Yusuf, and M. Bolic, "Contactless measurement of vital signs using thermal and RGB cameras: A study of COVID 19-related health monitoring," *Sensors,* vol. 22, no. 2, p. 627, 2022.
[293] D. E. King, "Dlib-ml: A machine learning toolkit," *The Journal of Machine Learning Research,* vol. 10, pp. 1755-1758, 2009.
[294] "NJN: A Dataset for the Normal and Jaundiced Newborns" https://sites.google.com/view/neonataljaundice.
[295] J. Chen, R. J. Stanley, R. H. Moss, and W. Van Stoecker, "Colour analysis of skin lesion regions for melanoma discrimination in clinical images," *Skin Research and Technology,* vol. 9, no. 2, pp. 94-104, 2003.
[296] M. Tkalcic and F. Tasic, "Project for the Digital signal processing course," ed: University of Ljubljana) Colour spaces, 2002.
[297] R. V. Kumar, K. P. Raju, L. R. Kumar, and M. Kumar, "Gray Level to RGB Using YcbCr Color Space Technique," *Int. J. Comput. Appl,* vol. 147, pp. 25-28, 2016.
[298] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.
[299] A. B. Shaik and S. Srinivasan, "A brief survey on random forest ensembles in classification model," in *International Conference on Innovative Computing and Communications*, 2019, pp. 253-260.
[300] I. Daunhawer *et al.*, "Enhanced early prediction of clinically relevant neonatal hyperbilirubinemia with machine learning," *Pediatric research,* vol. 86, no. 1, pp. 122-127, 2019.
[301] T. I. Trishna, S. U. Emon, R. R. Ema, G. I. H. Sajal, S. Kundu, and T. Islam, "Detection of hepatitis (a, b, c and e) viruses based on random forest, k-nearest and naive Bayes classifier," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India*, 2019, pp. 1-7.
[302] Z. Zhang, N. Li, H. Gao, Z. Cai, S. Si, and Z. Geng, "Preoperative analysis for clinical features of unsuspected gallbladder cancer based on random forest," in *2018 IEEE*

International Conference on Industrial Engineering and Engineering Management (IEEM), Bangkok, Thailand, 2018, pp. 1160-1164.

[303] J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random forests and decision trees," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 5, p. 272, 2012.

Appendix A Camera Sensors

A.1. Nikon D5300 Digital Camera

Specifications

Type of camera	Single-lens reflex (SLR) digital camera
Lens mount	Nikon F mount (with AF contacts)
Image sensor	23.5 × 15.6 mm CMOS sensor
Effective pixels	24.2 Megapixels (MP)
Total pixels	24.78 MP
Max Native Light sensitivity	12800 ISO
Min Native Light sensitivity	100 ISO
Dust-reduction system	Image sensor cleaning, Image Dust Off reference data (optional Capture N× 2 software required)
Magnification	Approx. 0.82× (50 mm f/1.4 lens at infinity, -1.0 m-1)
Shutter Speed	1/4000 to 30 s in steps of 1/3 or 1/2 EV, bulb, time
Image Resolution	6000 × 4000 (24.0 MP, 3:2), 4496 × 3000 (13.5 MP, 3:2), 2992 × 2000 (6.0 MP, 3:2)
Image File Format	JPEG (EXIF 2.3), RAW (12 or 14-bit, NEF), RAW+JPEG
Video Resolution:	1920 × 1080 (60p/50p/30p/25p/24p) 1280 × 720 (60p/50p) 640 × 424 (30p/25p)
Video File Format:	MOV (Video: H.264/MPEG-4 AVC; Audio: Stereo Linear PCM)
Dimensions (W × H × D)	Approx. 125 × 98 × 76 mm
Weight	Approx. 530 g with battery and memory card but without body cap; Approx. 480 g (camera body only)

A.2. Nikon D610 Digital Camera

Specifications

Type of camera	Single-lens reflex (SLR) digital camera
Lens mount	Nikon F mount
Image sensor	35.9 × 24 mm CMOS sensor
Effective pixels	24.3 Megapixels (MP)
Total pixels	24.7 MP
Max Native Light sensitivity	6400 ISO
Min Native Light sensitivity	100 ISO
Magnification	Approx. 0.7×
Shutter Speed	1/4000 to 30 s
Image Resolution	3936 × 2624 2944 × 1968 1968 × 1312
Image File Format	JPEG, RAW
Video Resolution:	1920 ×1080p at 23.98/25 fps [12 to 24 Mb/s] 1280 ×720 at 25/29.97/50/59.94 fps [8 to 24 Mb/s]
Video File Format:	MOV (Video: H.264/MPEG-4 AVC; Audio: Stereo Linear PCM)
Dimensions (W × H × D)	Approx. 141 x 113 x 82 mm
Weight	Approx. 840 g with battery and recording media Approx. 760 g (camera body only)

A.3. FLIR A400

Specifications

Type of camera	Long wave infrared (LWIR) thermal smart sensor
Base mounting	4× M4 on 4 sides
Encoding	H.264, MPEG4, or MJPEG
IR Resolution	320 × 240 Pixels
Image frequency	30Hz
Spectral range	7.5-14 µm
Temperature range	-20°C to 120°C (-4°F to 248°F), 0°C to 650°C (32°F to 1202°F), 300°C to 2000°C (572°F to 3632°F)
Sensitivity	<30 mK, 42° @ +30°C (+86°F) <40 mK, 24° @ +30°C (+86°F) <50 mK, 14° @ +30°C (+86°F)
Accuracy	±2°C (±3.6°F) or ±2% of reading, for ambient temperature 15°C–35°C (59°F–95°F) and object temperature above 0°C (32°F)
Visual Resolution	640 × 480 pixels
Image source	Visual/IR/ MSX® (visual camera is an optional feature)
Dimensions (L × W × H)	Approx. 123 × 77 × 77 mm
Weight	820 gm (camera only)

Appendix B Statement of Authorship

B.1. Remote Monitoring of Vital Signs in Diverse Non-Clinical and Clinical Scenarios Using Computer Vision Systems: A Review

University of South Australia

Statement of Authorship

Title of Paper	Remote Monitoring of Vital Signs in Diverse Non-Clinical and Clinical Scenarios Using Computer Vision Systems: A Review
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam, A. Al-Naji & J. Chahl, "Remote Monitoring of Vital Signs in Diverse Non-Clinical and Clinical Scenarios Using Computer Vision Systems: A Review," Applied Sciences, vol. 9, no. 20, pp. 1-32, 2019.
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	did the literature review and wrote the draft manuscript
Overall Percentage (%)	80%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30-Aug-23
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Ali Al-Naji

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 2	Javaan Chahl	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	Javaan Chahl Digitally signed by Javaan Chahl Date: 2023.09.27 14:57:02 +09'30'	Date 27-Sep-23
Name of Co-Author 3		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 4		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 5		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 7		
% Contribution to the Paper		

B.2. Noncontact Sensing of Contagion

University of South Australia

Statement of Authorship

Title of Paper	Noncontact Sensing of Contagion
Publication Status	■ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam, LA Chahl, JS Chahl, A. Al-Naji, AG Perera, D. Wang, YH Lee, & J. Chahl, "Noncontact sensing of contagion," Journal of Imaging, vol. 7, no. 2, pp. 1-35, 2021.
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	did the literature review and wrote the draft manuscript
Overall Percentage (%)	55%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30.08.2023
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Loris A. Chahl

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 15/9/2023
Name of Co-Author 2	Jaswant S. Chahl	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 15/9/2023
Name of Co-Author 3	Ali Al-Naji	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 4	Asanka G. Perera	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 05 Sep 2023
Name of Co-Author 5	Danyi Wang	
% Contribution to the Paper	2%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 30-Aug-23
Name of Co-Author 6	Y.H. Lee	
% Contribution to the Paper	2%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 12 Sep 2023
Name of Co-Author 7	Titilayo T. Ogunwa	
% Contribution to the Paper	2%	

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

University of South Australia

Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 07/09/2023
Name of Co-Author 8	Samuel Teague	
% Contribution to the Paper	2%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 12/09/2023
Name of Co-Author 9	Tran Xuan Bach Nguyen	
% Contribution to the Paper	2%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 06 Sep 2023
Name of Co-Author 10	Timothy D. McIntyre	
% Contribution to the Paper	2%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 3-9-23

University of South Australia

Name of Co-Author 11	**Simon P. Pegoli**	
% Contribution to the Paper	**2%**	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 13/09/2023
Name of Co-Author 12	**Yiting Tao**	
% Contribution to the Paper	**2%**	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 07/09/2023
Name of Co-Author 13	**John L. McGuire**	
% Contribution to the Paper	**2%**	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date
Name of Co-Author 14	**Jasmine Huynh**	
% Contribution to the Paper	**2%**	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 13/09/2023
Name of Co-Author 15	**Javaan Chahl**	
% Contribution to the Paper	**5%**	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 27/09/2023
Name of Co-Author 16		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date

B.3. Remote Vital Signs Monitoring in Neonatal Intensive Care Unit Using a Digital Camera

University of South Australia

Statement of Authorship

Title of Paper	Remote Vital Signs Monitoring in Neonatal Intensive Care Unit Using a Digital Camera
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam, A. Al-Naji, A. G. Perera, K. Gibson & J. Chahl, "Remote Vital Signs Monitoring in Neonatal Intensive Care Unit Using a Digital Camera," International Journal of Biomedical and Biological Engineering, vol. 16, no. 10, pp. 138–144, 2022.
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	Conceptualization, Methodology, Software, Validation, and Original Draft Preparation
Overall Percentage (%)	70%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30-Aug-23
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Ali Al-Naji

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 2	Asanka G. Perera	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 5-Sep-23
Name of Co-Author 3	Kim Gibson	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 31/8/23
Name of Co-Author 4	Javaan Chahl	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date
Name of Co-Author 5		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 7		
% Contribution to the Paper		

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

B.4. Non-contact Automatic Vital Signs Monitoring of Neonates in NICU Using Video Camera Imaging

University of South Australia

Statement of Authorship

Title of Paper	Non-contact automatic vital signs monitoring of neonates in NICU using video camera imaging
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam, A. Al-Naji, A. G. Perera, K. Gibson & J. Chahl, "Non-contact automatic vital signs monitoring of neonates in NICU using video camera," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 11, no. 2, pp. 278–285, 2023.
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	Conceptualization, Methodology, Software, Validation, and Original Draft Preparation
Overall Percentage (%)	70%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30-Aug-23
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Ali Al-Naji

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 2	Asanka G. Perera	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 5-Sep-23
Name of Co-Author 3	Kim Gibson	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 31.8.23
Name of Co-Author 4	Javaan Chahl	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date
Name of Co-Author 5		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 7		
% Contribution to the Paper		

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

B.5. Non-Contact Automatic Vital Signs Monitoring of Infants in A Neonatal Intensive Care Unit Based on Neural Networks

University of South Australia

Statement of Authorship

Title of Paper	Non-Contact Automatic Vital Signs Monitoring of Infants in A Neonatal Intensive Care Unit Based on Neural Networks
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam, A. G. Perera, A. Al-Naji, K. Gibson & 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, pp.1–19, 2021.
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	Conceptualization, Methodology, Software, Validation, and Original Draft Preparation
Overall Percentage (%)	70%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30-Aug-23
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Asanka G. Perera

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 5-Sep-23
Name of Co-Author 2	Ali Al-Naji	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 3	Kim Gibson	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 31.8.23
Name of Co-Author 4	Javaan Chahl	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date
Name of Co-Author 5		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 7		
% Contribution to the Paper		

B.6. Contactless Vital Signs Monitoring in NICU Based on Machine Vision

University of South Australia

Statement of Authorship

Title of Paper	Contactless Vital Signs Monitoring in NICU based on Machine Vision
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam A. Al-Naji, K. Gibson & J. Chahl, "Contactless Vital Signs Monitoring in NICU based on Machine Vision," The 19th International Conference on Intelligent Unmanned Systems (ICIUS), Adelaide, Australia, 2023.
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	Conceptualization, Methodology, Software, Validation, and Original Draft Preparation
Overall Percentage (%)	70%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30-Aug-23
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Asanka G. Perera

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 5-Sep-23
Name of Co-Author 2	Ali Al-Naji	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 3	Kim Gibson	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 31.8.23
Name of Co-Author 4	Javaan Chahl	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date
Name of Co-Author 5		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 7		
% Contribution to the Paper		

B.7. Integrating RGB-Thermal Image Sensors for Non-contact Automatic Respiration Rate Monitoring

University of South Australia

Statement of Authorship

Title of Paper	Integrating RGB-Thermal Image Sensors for Non-contact Automatic Respiration Rate Monitoring
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam, A. G. Perera, A. Al-Naji, K. Gibson & J. Chahl, "Integrating RGB-Thermal Image Sensors for Non-contact Automatic Respiration Rate Monitoring," Biomedical Signal Processing and Control, 2023. (Submitted)
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	Conceptualization, Methodology, Software, Validation, and Original Draft Preparation
Overall Percentage (%)	70%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30-Aug-23
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Asanka G. Perera

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	5-Sep-2023 Date
Name of Co-Author 2	Ali Al-Naji	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 3	Timothy D. Mcintyre	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 3-9-23
Name of Co-Author 4	Javaan Chahl	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	27/09/2023 Date
Name of Co-Author 5		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 7		
% Contribution to the Paper		

B.8. Non-invasive and Non-contact Automatic Jaundice Detection of Infants Based on Random Forest

University of South Australia

Statement of Authorship

Title of Paper	Non-invasive and non-contact automatic jaundice detection of infants based on random forest
Publication Status	■ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details Relevant reference details	F-T-Z Khanam, A. G. Perera, A. Al-Naji, D. Wang, & J. Chahl, "Non-invasive and Non-contact Automatic Jaundice Detection of Infants Based on Random Forest," Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2023.
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Fatema-Tuz-Zohra Khanam
Contribution to the Paper Brief description of your work in this publication	Conceptualization, Methodology, Software, Validation, and Original Draft Preparation
Overall Percentage (%)	70%
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted] Date 30.08.2023
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Ali Al-Naji

The University of South Australia
www.unisa.edu.au
CRICOS Provider No. 00121B

Last updated: May 2023

University of South Australia

% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date 2/9/2023
Name of Co-Author 2	Asanka G. Perera	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	5-Sep-2023 Date
Name of Co-Author 3	Danyi Wang	
% Contribution to the Paper	5%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	30-Aug-23 Date
Name of Co-Author 4	Javaan Chahl	
% Contribution to the Paper	10%	
Signature *Signatures will be redacted by the Library at the time of publication*	[redacted]	Date
Name of Co-Author 5		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 6		
% Contribution to the Paper		
Signature *Signatures will be redacted by the Library at the time of publication*		Date
Name of Co-Author 7		
% Contribution to the Paper		