# POLITECNICO DI TORINO

## Master's Degree in Computer Engineering
## Software

### Master's Degree Thesis

# Contactless estimation of newborn vital parameters using a camera


**Supervisor**

Prof. Gabriella Olmo

**Co-supervisors**

Letizia Bergamasco

Marco Gavelli

**Candidate**

Adriana Margarita Hevia Masbernat


**Academic Year 2022/2023**

# Abstract

This research, conducted in collaboration with the Neonatal Unit of AO Ordine Mauriziano Hospital in Turin and LINKS Foundation, presents non-contact RGB camera techniques for measuring heart rate and respiration rate in newborns as a promising alternative for monitoring vital signs without causing discomfort or increasing the risk of infection. Additionally, these techniques enhance objectivity and convenience in the pain assessment of newborns. While focusing on the Neonatal Intensive Care Unit (NICU) context, the technology has potential applications in remote physiological monitoring beyond clinical settings. A private dataset was elaborated to evaluate vital sign estimation techniques in newborns in the NICU based on different traditional algorithms for remote photoplethysmography. These algorithms were selected based on their essential characteristics for the context, such as robustness to motion and lighting conditions. Ground truth values for heart rate were obtained using character recognition from pulse oximeter values displayed in the videos, while ground truth values for respiration rate were manually obtained by clinical staff for a subset of the dataset. The Virtual Heart Rate python package framework, customized for this research, facilitated the implementation of traditional algorithms and offered efficient computations through Graphics Processing Unit parallelism, enabling real-time processing. Experiments were conducted to determine optimal algorithm parameters. Vital signal estimations were then calculated and compared to the ground truth values using defined error metrics. The results of heart rate estimation were categorized based on different motion levels (motionless, sporadic motion, and motion), and the best-performing algorithms for each category were identified. Projection Plane Orthogonal to the Skin-tone and Independent Component Analysis performed consistently well across different motion categories, indicating their suitability for heart rate estimation in the given context. Notably, the motionless category achieved a Mean Absolute Error of 5.7, which is within the clinically acceptable range, demonstrating the feasibility of this approach for remote heart rate monitoring during rest or sleep. Future research may explore hybrid methodologies to improve performance in categories involving movement. Regarding respiration rate estimation, Chrominance-based method and Principal Component Analysis demonstrated the best performance.

Despite the small sample size of ground truth values of respiration rate to obtain statistically significant results, this part of the work demonstrates the approach's feasibility and opens the doors for future experiments. In conclusion, this study presents a framework for the automatic non-contact camera-based measurement of heart rate and respiration rate, comparing for the first time the performance of different traditional algorithms in the NICU environment. The study acknowledges certain limitations, including homogeneity of skin color for the subjects in the used dataset and challenges related to accurately identifying the Region of Interest to extract the vital signals. These limitations provide opportunities for future exploration and improvements in this field.

# Acknowledgements

Thanks to my guide Professor Gabriella Olmo and supervisors from Links Foundation, Letizia Bergamasco and Marco Gavelli, for monitoring my progress and guiding the procedure of this project. Special thanks to Letizia Bergamasco, for whom I increased my skills and competencies, which will forever be part of my professional future, and who gave me many helpful suggestions and continuous encouragement and support along with Edoardo Pristeri.

Thanks to LINKS Foundation and the Neonatal Unit of AO Ordine Mauriziano Hospital in Turin for the collaboration in this investigation and for trusting in my knowledge and skills to develop it in the best possible way.

Thanks to all of the professors who have formed me as an Engineer. Also, I am grateful to those who have gone beyond professional training and contributed to shaping me.

Thanks to my mother, Karen Mónica Masbernat Muñoz, for the unconditional support she has given me in every decision I have taken, including the one of coming to Italy. Her support has been an engine that has propelled me to get far and chase my dreams.

Furthermore, I thank the Italian Ministry of Foreign Affairs and International Cooperation (MAECI) for the financial support they provided me during my stay in Italy. They were an important contributor to making my studies in Italy possible.

I am also grateful to my friends, family, and boyfriend, people who have continuously stood by my side through the changing events of life and made my chosen path enjoyable.

Lastly, thanks to all the people who believed in me and encouraged me to bring out the best versions of myself.

# Table of Contents

# List of Tables

# List of Figures

# Acronyms

| | |
|---|---|
| AR | Auto-Regressive |
| bpm | Beats per minute |
| B | Blue |
| BCG | Ballistocardiography |
| BKF | Bounded Kelman Filters |
| BSS | Blind Source Separation |
| BVP | Blood volume pulse |
| CEEMDAN | Complete Ensemble Empirical Mode Decomposition with Adaptive Noise |
| CHROM | Chrominance-based method |
| cpm | Cycles per minute |
| CUDA | Compute Unified Device Architecture |
| CPU | Central Processing Unit |
| DAN | Douleur Aiguë du Nouveau-né |
| DANN | Deep Alignment Network |
| DFT | Discrete Fourier Transform |
| ECG | Electrocardiography |
| EEMD | Ensemble Empirical Mode Decomposition |
| EVM | Eulerian Video Magnification |
| FDA | Food and Drug Administration |
| FFT | Fast Fourier Transform |
| fps | Frames per second |
| G | Green |
| GPU | Graphics Processing Unit |
| HR | Heart Rate |

| | |
|---|---|
| HRV | Heart Rate Variability |
| HSV | Hue Saturation Value |
| Hz | Hertz |
| ICA | Independent Component Analysis |
| IMFs | Intrinsic Mode Functions |
| JBSS | Joint Blind Source Separation |
| KLT | Kanade-Lucas-Tomasi |
| LGI | Local Group Invariance |
| MAE | Mean Absolute Error |
| MAECI | Ministry of Foreign Affairs and International Co-operation |
| MAX | Maximum Mean Absolute Error |
| MEMD | Multivariate Empirical Mode Decomposition |
| ML | Machine Learning |
| MSD | Micromotion and Stationarity Detection |
| MTCNN | Multi-Task Convolutional Neural Network |
| NFCS | Neonatal Facial Coding System |
| NICU | Neonatal Intensive Care Unit |
| NN | Neural Network |
| OF | Optical Flow |
| PBV | Blood Volume Pulse Signature |
| PCA | Principal Component Analysis |
| PCC | Pearson Correlation Coefficient |
| PIPP | Premature Infant Pain Profile |
| PLS | Partial Least Squares |
| POS | Projection Plane Orthogonal to the Skin-tone |
| PPG | Photoplethysmography |
| PSD | Power Spectral Density |
| pyVHR | Virtual Heart Rate python package |
| px | Pixels |

| | |
|---|---|
| R | Red |
| RGB | Red-Green-Blue |
| RIP | Respiratory Inductance Plethysmography |
| RMSE | Root Mean Square Error |
| ROI | Region of Interest |
| RR | Respiration Rate |
| rPPG | Remote Photoplethysmography |
| s | Seconds |
| SNR | Signal to Noise Ratio |
| $SpO_2$ | Oxygen saturation |
| SSA | Singular Spectrum Analysis |
| STFT | Short-Time Fourier Transform |
| VJ | Viola-Jones |
| WHO | World Health Organization |
| 2PS | Projection-Plane-Switching |
| 2SR | Spatial Subspace Rotation |

# Chapter 1

# Introduction

## 1.1 Thesis purpose

This thesis research is set in the context of automatic pain and vital parameters assessment in newborns. The objective is to develop automatic non-contact camera-based techniques for continuously measuring vital signs in newborns to improve pain assessment objectiveness and convenience.

## 1.2 Vital signs measurement for a more objective newborn pain assessment

### 1.2.1 Newborns' pain and importance of an objective pain assessment

It is a proven fact that newborns experience pain since neonatal age and the memory of this pain is not only preserved but can even produce other alterations in the newborn, for example, behavioral, hormonal, and cognitive [1], [2], [3]. For this reason, it is fundamental to measure and treat newborns' pain properly. Additionally, the Italian Law 38/2010 guarantees pain therapy in medical procedures in hospitals [4], but as newborns cannot verbally communicate the pain experienced, some scales to evaluate pain (called "algometric pain scales") have been developed and validated [5]. Examples of traditional pain scales are the Neonatal Facial Coding System (NFCS), which relies only on facial expressions; the Premature Infant Pain Profile (PIPP), which also considers contextual and physiological parameters; and the Douleur Aiguë du Nouveau-né (DAN), which evaluates facial expressions, limb movements, and vocal expressions. Recent studies provide evidence to support that noxious stimulation of neonates [6], [7] and toddlers [8] can be differentiated from non-noxious stimulation as the first one produces a significant Heart Rate (HR) increase and behavioral changes such as limb withdrawal and changes in facial expressions. Consequently, in clinical practice, the use of pain-validated scales is strongly recommended.

However, traditional pain assessment methods using these pain scales are highly subjective and time-consuming as they depend on the knowledge and sensitivity of the healthcare staff. [9] showed that discrepancies in pain evaluation can arise even when evaluating objective values like Oxygen saturation ($SpO_2$) using the PIPP scale since clinicians read the value from the pulse oximeter in different moments. Therefore, automation is needed to produce a more objective pain assessment with a reproducible score.

### 1.2.2 Newborn pain assessment using traditional methods for vital signs monitoring

When using some validated pain scales, such as the PIPP scale shown in Figure 1.1 from [10], the measurement of vital signs as the HR and $SpO_2$ are fundamental.

| Infant Indicator | Indicator Score | | | | Infant Indicator Score |
|---|---|---|---|---|---|
| | 0 | +1 | +2 | +3 | |
| Change in Heart Rate (bpm) Baseline: _________ | 0 - 4 | 5 - 14 | 15 - 24 | >24 | |
| Decrease in Oxygen Saturation (%) Baseline: _________ | 0 - 2 | 3 - 5 | 6 – 8 | >8 or Increase in $O_2$ | |
| Brow Bulge (Sec) | None (<3) | Minimal (3 -10) | Moderate (11 – 20) | Maximal (>20) | |
| Eye Squeeze (Sec) | None (<3) | Minimal (3 -10) | Moderate (11 – 20) | Maximal (>20) | |
| Naso-Labial Furrow (Sec) | None (<3) | Minimal (3 -10) | Moderate (11 – 20) | Maximal (>20) | |
| | | | | * Sub-total Score: | |
| **Gestational Age (Wks + Days)** | >36 wks | 32 wks - 35 wks, 6d | 28 wks- 31wks, 6d | <28wks | |
| **Baseline Behavioural State** | Active and Awake | Quiet and Awake | Active and Asleep | Quiet and Asleep | |
| | | | | ** Total Score: | |

**Figure 1.1:** PIPP scale. Source: [10]

Commonly, these physiological signals are estimated using either an Electrocardiography (ECG) or Photoplethysmography (PPG). The first technique consists in recording the electrical signal of the HR by attaching electrodes, which are patches with an adhesive layer, to the chest and limbs of the patient and have wires that connect to a monitor. ECG is used as the standard cardiovascular measurement as mentioned in [11]. The second technique, PPG was first described in the 1930s as an optical technique to identify vital signs [12]. PPG is considered a simple optical measurement that uses light to measure the volumetric variations of blood circulation at the skin's surface. This method is commonly preferred over ECG as it uses a single sensor at a measurement site for PPG signal instead of various electrodes [13] and it provides an equally reliable measurement [11]. A pulse oximeter is an example of an instrument that uses PPG technology.

Therefore, traditional methods for HR monitoring make use of medical equipment that requires constant contact with the newborn's skin and thus can cause discomfort, induce chances of allergy, injury, or epidermal stripping to newborns skin which can cause pain and trauma, and increase the risk of spreading infection in hospitals [2], [3], [14], [15], [16], [17], [18], [19]. Moreover, the humid environment of neonatal incubators and the neonates' thin and underdeveloped skin can cause the adhesive patches or sensors to fail and require frequent changing [20].

Regarding Respiration Rate (RR), the gold standard for measuring this value consists of manually counting breaths while auscultating the patient or palpating for chest rise. This measuring technique is accurate, yet time-consuming and impractical for continuous vital signs monitoring. Other devices, such as Respiratory Inductance Plethysmography (RIP), use a chest belt and require interpretation from a specialist [20].

### 1.2.3 Contactless methods to determine vital signs

Alternative non-contact solutions such as Radar-based systems, Laser Doppler Vibrometers, and Thermal imaging have been explored. On one hand, they have been proven to have penetration capabilities, work unaffected by the color of the subject's skin, and work under different ambient light levels. On the other hand, they are sensitive to motion changes, they require expensive extra specialized hardware, radiation exposure could be unsafe, reflection-based systems require to direct the laser/radar to the target in the subject to monitor, the thermal camera requires calibration and demand high resolution [3], [18], [21], [22], [23].

Whilst, the usage of a Red-Green-Blue (RGB) camera to determine the value of vital signs has proven to be a non-invasive, low-cost, easy-to-use and versatile non-contact alternative to measure vital parameters, ubiquitous and capable of high performance facilitating the intervention of the healthcare staff [15], [17], [19], [20], [24], [22], [25], [23]. Consequently, it was the chosen hardware for the development of this project.

### 1.2.4 Using a RGB camera to determine vital signs

When using a RGB camera to determine these vital signs they can either be obtained by movement detection, a technique called Ballistocardiography (BCG), or by light reflection, using another technique called Remote Photoplethysmography (rPPG).

BCG relies on the mechanical motion of the heart and lungs. For HR estimation, the heart's mechanical motion contributes to a microscopic displacement of the head or facial skin [26] at the cardiac frequency. While for RR, changes in lungs volume generate periodic chest movements [27] at the breathing frequency.

Whilst, rPPG is based on the same principle as PPG, with the difference that this signal is obtained remotely by means of a camera. rPPG estimates vital signals by capturing microscopic color variations of the skin [26]. The basic principle behind this is that blood absorbs more light, specifically the hemoglobin molecule, than surrounding tissues, so changes in blood volume affect transmitted and reflected light [28]. Therefore, in the case of HR measurement, the beating of the heart causes pressure variations in the arteries (even in small vessels), translated in varying amounts of hemoglobin, which consequently produce synchronous varying light absorption [2], [14], [18], [24]. Similarly, for RR measurement, breathing causes changes in pressure at the area of the torso which affects the pressure in the large blood vessels (veins) [18]. This phenomenon produces skin color changes that are invisible to the eye but can be detected using a camera [2].

More in detail, the dichromatic reflection model shown in Figure 1.2 from [29] explains light reflected from the skin surface as a combination of two components, the specular reflection, which does not contain any pulse signal information, and the diffuse reflection which manifests the pulse signal (blood volume changes) as explained by [30]. In this way, rPPG captures the reflected light from the illuminated skin resulting in a waveform that contains the HR and RR information.

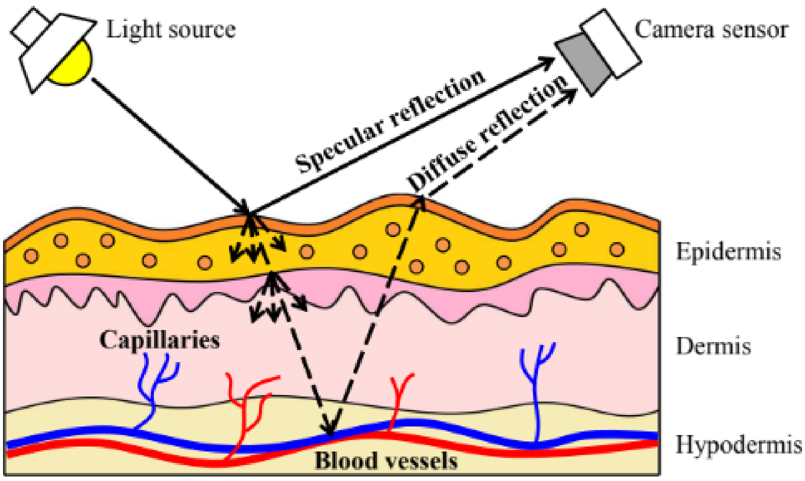


**Figure 1.2:** Dichromatic reflection model. Source: [29]

### 1.2.5 Variables affecting vital signs estimation when using a RGB camera

When estimating vital signs using an RGB camera, several variables need to be considered as they can affect the quality of the measurements obtained.

Firstly, the distance between the subject and the camera has been found to impact the accuracy, with distances less than 1m yielding satisfactory results according to [27]. Increasing the distance from the subject to the camera has been shown to increase errors in vital signal estimations.

Secondly, illumination levels also play a role in the performance of RGB cameras for vital sign estimation. Decreasing illumination has been associated with increased error [26], actually it is recommended to avoid light intensities below 20 lux [27]. Additionally, the type of light source is important, as ambient light has been reported to produce more reliable measurements compared to artificial light, which can introduce noise due to AC current flickering. [2] notes that algorithms for HR estimation exhibit greater robustness during overcast days when diffuse light is present, resulting in less pronounced shadows. Furthermore, [31] mentions that direct light sources produce sharp shadows on the subject's face which can decrease performance.

Thirdly, there is a correlation between skin color and the performance of PPG vital sign measurement. Higher concentrations of melanin, the substance responsible for skin pigmentation, have been associated with lower PPG performance [26], [27].

Facial hair represents a fourth variable that can degrade performance [27], [30], [32].

Fifthly, subject movements significantly impact vital sign measurements. The Region of Interest (ROI), which is the area of the skin of the newborn from which the vital signal estimations are made, has to be tracked through time to be able to have continuous monitoring of the vital signal. However, movement of the individual can cause the tracking algorithms to fail resulting in a ROI loss or drift. This means that the tracking algorithm produces a failure to extract the required signals from the intended region or might start extracting signals from unintended areas, leading to inaccurate measurements. Consequently, movements of the newborn result in varying camera-infant viewing angles which influence performance [17] or can result in ROI-camera blockage. Additionally, infant motion and the presence of healthcare staff passing by can result in pronounced shadows, which decrease the Signal to Noise Ratio (SNR), which measures the desired signal relative to the background noise, of the rPPG signal [2], [23], [26]. In other words, this means that it increases background error.

Lastly, video compression of RGB camera-collected videos can greatly influence video-based vital sign extraction [31]. Consequently, specific algorithms have been developed in the literature to extract the rPPG signal in the presence of video compression artifacts.

### 1.2.6 Importance of newborns' automatic, non-contact vital signs measurement using a camera

Although there is existing research in this field, most of the findings are based on studies conducted with adult subjects under specific lighting and motion conditions, which are not directly applicable to the Neonatal Intensive Care Unit (NICU) context. Studies performed in adults differ from those in newborns for many reasons, one being that they have different physiological characteristics. For instance, normal resting HR and RR ranges for awake and healthy newborns aged 0-1 year are significantly higher, ranging from 90 to 181 bpm for HR and 25 to 68 cpm for RR [33], ranges which are almost double as high as adults. Additionally, newborns exhibit more frequent episodes of rapid movement that cannot be controlled [22]. Although these movements are often gentle and subtle, newborns may occasionally experience spasms that generate artifacts or false measurements [17]. Furthermore, since newborns do not consciously look toward the camera, tracking problems may arise in the presence of high-degree angles relative to the camera [17]. Moreover, newborn monitoring is much more challenging than adults because neonates have obscure facial features [3].

It is important to note that lighting conditions in the NICU may vary across hospitals [17]. Therefore, specific studies tailored to the unique context of the NICU are necessary to address these challenges effectively.

This thesis aims to fill this gap by developing non-contact techniques for automatic vital parameters continuous calculation, specifically to measure HR and RR, with a real-time application using only a camera in the NICU context. These techniques will be fundamental to developing a more objective and convenient system for automatic pain assessment in newborns. In other words, the idea is to achieve long-term monitoring of the mentioned vital signals by acquiring them continuously, which are important parameters for NICU management as mentioned by [34], in an unobtrusive and comfortable manner.

The project has been developed in LINKS Foundation, a research center actively involved in different technological projects on the frontier of knowledge as it is mentioned on their website [35]. Also, this project has the collaboration of the Neonatal Unit of AO Ordine Mauriziano Hospital in Turin, as this kind of system would provide multiple benefits in clinical practice. This method not only lowers risks for newborns due to the absence of contact with the instrumentation but also provides continuous monitoring of multiple parameters, facilitating the work of healthcare operators, who otherwise would have to observe one pain indicator at a time. Moreover, the system will provide more objectivity in pain assessment, as it will not depend on the observer's knowledge and sensitivity.

### 1.2.7 Other potential applications of this research

In addition to its use in the NICU environment, this technology could also be applied for the remote monitoring of physiological parameters in other contexts, even non-clinical ones. For example, one non-clinical research field that is emerging is the detection of the so-called "deep fakes", i.e., synthetic images or videos where a person is replaced with someone else's likeness; in this case, the study of the skin color changes caused by cardiovascular pulses used to calculate the heart rate may help discriminate real videos from deep fakes [36].

## 1.3 Thesis outline

The following Chapter describes literature techniques used to obtain vital signs and the context in which they were tested, including a Section that shows the state of the art. In Chapter 3, this project's applied methodology and testing context is described. Later in Chapter 4, the experimental results are shown with error metrics. Finally, in Chapter 5, the aforementioned results are discussed, and improvements for future research are proposed.

# Chapter 2

# Literature Review

## 2.1 Image processing and signal processing techniques for HR and RR estimation

To obtain vital signal estimations using rPPG (refer to Subsection 1.2.4 for an explanation of this technology) some steps have been commonly followed in literature. First, a ROI is selected from the subject, this is a region of visible skin from which the rPPG signal is extracted [16], [26]. This region is specific to the vital signal that is estimated because historically some regions have demonstrated to provide a higher quality estimation [37] and therefore, they have been preferred (this will be further discussed in Subsection 2.1.1). Then, the mean color of pixels in the ROI, called raw signal, is tracked over a time window to extract the rPPG signal (tracking and detection of the subject is discussed in Subsection 2.1.2). Later, processing algorithms are applied to this window of data to eliminate noise, mainly coming from motion and lighting variations, consequently getting an rPPG signal that contains representative vital signal information (these algorithms are described in Subsection 2.1.3). Finally, another algorithm is used to get the vital signal estimation from the resulting window of data (commonly used functions are described in Subsection 2.1.4). The process is repeated for multiple windows, so-called moving windows, to output vital signal estimations continuously.

Note that HR is usually estimated in literature using rPPG technology, however the RR is either estimated using rPPG or BCG. The same structure previously described is followed when estimating this vital signal using BCG technology, but different ROI selections (no longer limited to visible skin) and algorithms are used. These will be described in the same Subsections as for rPPG, specifying the base technology used in each case.

### 2.1.1 ROI selection

In [28] rPPG signals were measured from different locations of visible skin in the body (including the wrist, legs, etc.), and the signal obtained from the face was found to be stronger. This can also be an advantage because the face is usually uncovered, unlike other anatomical locations used to obtain this signal. Actually, in the study of [9], newborns were tightly wrapped in a blanket to ease the pain, which left the face uncovered by the blanket but the rest of the body, including the arms, were completely covered.

In 2021 ROI regions of the face were analyzed in [38] to determine the effect of skin thickness on the accuracy of the obtained rPPG signal for HR estimation. It was concluded that some regions perform better due to their anatomical characteristics, specifically the yellow regions of Figure 2.1 have the highest reliability and accuracy among the studied regions, while blue regions are shown as the worst performance.

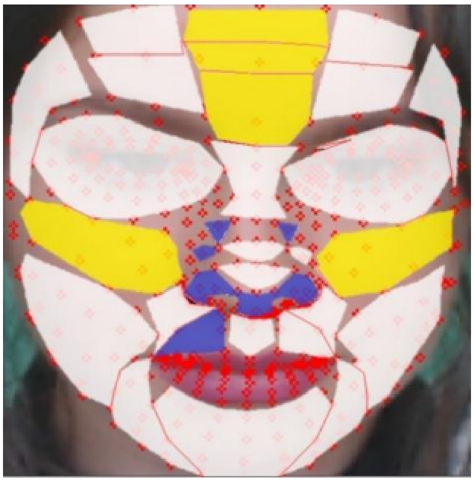

**Figure 2.1:** Anatomical influence of face's ROI selection: regions highlighted in yellow exhibit higher performance, whereas the regions highlighted in blue show lower performance. Source: [38]

The same year, [39] mentioned that non-rigid motions, such as blinking and breathing, which introduce noise in rPPG estimations are likely to occur on the areas around the mouth and eyes, and therefore these areas should be avoided for the extraction of the pulse signal. In compliance with the previously mentioned study [39] states that the forehead and cheeks are areas which contain most of the Blood volume pulse (BVP) information. Note that BVP is the same as PPG, a measure of the HR based on the volume of blood that passes through the tissues in a confined area with each heartbeat. Commonly, HR has been extracted from these regions because they have more capillaries and are unaffected by facial expressions as mentioned by [27], while RR has been measured from the nasal area or torso region.

Moreover, [14] constructed an accurate map of the spatial distribution of HR and RR rPPG information that reaffirms that HR information can be found in the face except for the areas around the eyes and nostrils. It also states that RR information can be found in the face, specifically in the forehead and regions close to the nose. Consequently, HR can be extracted from larger ROIs, while RR is more prominent in smaller ROIs.

Furthermore, [31] studied the influence of part of the ROI being undetected when head rotation of 30° - 45° is present. The symmetry substitution method was used to replace the undetected ROI with the values of a detected one. It was concluded that there was no significant difference in HR measurement between the left and right cheek, and therefore HR accuracy did not decrease compared with the full ROI condition.

In [15] it was demonstrated that it is possible to obtain accurate measurements of the RR from light reflection variation at the level of the collar bones and above the sternum, region called pit of the neck. However, this study was performed in adult subjects in the absence of breathing unrelated movements and in a quiet environment.

In 2022 the Virtual Heart Rate python package (pyVHR) framework (which is further described in Subsection 2.1.3) was used by [40] to assess the effectiveness of popular rPPG processing algorithms in HR estimation using four ROIs: forehead, left cheek, right cheek and the combination of all three. The results suggest preferring large ROIs for challenging scenarios. It was also concluded that the performance of rPPG methods depends on different characteristics of the context of use, such as movement, lighting conditions, and the error metrics applied.

A similar investigation was carried out by [41] the same year. HR was estimated using the forehead and cheeks as ROI, while RR was estimated from the motion of the subject's pit of the neck. Popular rPPG processing methods combined with filters were used to obtain HR. While to obtain RR, Optical Flow (OF) was used to calculate displacement between consecutive images followed by integration of this quantity. OF is a technique that estimates the movement of surfaces between two frames. [32] suggests that it can produce accurate results when tracking small displacements. There are many algorithms for its determination including partial derivatives, phase correlation, and discrete optimization methods. In [41] also, different distances and lightning conditions were evaluated, concluding that for HR estimation a multi-ROI approach has a better performance than a single ROI approach. However, in this study, subjects were motionless and always faced the camera.

### 2.1.2 ROI detection and tracking algorithms

In 2017 an automated pain evaluation framework for newborns based on facial expressions assessment was developed by [9]. The investigation used the Kanade-Lucas-Tomasi (KLT) algorithm for face detection and tracking, which according to [42] is "an algorithm that is used to track face based on trained features". However, its use was discouraged for newborn applications since they do rapid head movements that cause the algorithm to fail.

Another widely used face location detection algorithm as in [31], [38], [41] is the Viola-Jones (VJ) algorithm. According to [42], it "is used to detect the face based on the haar features" that are black and white patterns in pixels defining figures. However, again, the usage of this algorithm is not suitable in motion conditions because VJ classifiers were mostly trained using frontal face images, therefore if the face of the subject is not frontal to the camera VJ algorithm will most likely fail [43].

As mentioned in the systematic review done by [27], in 2022 over 900 articles approximately related to the monitoring of vital signs with camera use. Recently Neural Network (NN) based ROI detection methods gained increased attention mainly because they yield higher accuracy compared to traditional approaches. Although, as mentioned in [31], some of these algorithms are slow and therefore unsuitable for real-time analysis, such as Multi-Task Convolutional Neural Network (MTCNN) and Deep Alignment Network (DANN).

In other investigations, such as in [40], Google Mediapipe was used for ROI detection and tracking. This is a cross-platform, customizable, accelerated, free, and open source Machine Learning (ML) solution that estimates 468 3-D facial points (called landmarks) in real-time using a single camera input or video recording [44], [45].

In 2018 [32] estimated HR under different lightning conditions and motion. It used Bounded Kelman Filters (BKF), which is "a motion estimation model that is employed to track the regions of interest from frame to frame". It identifies blur due to motion in the video recording frames, which can cause discontinued pixel intensity measurements, and denoises it enabling feature points to be identified with higher accuracy than OF. It also addresses illumination changes by using the Hue Saturation Value (HSV), an alternative color space to RGB. This is due to the fact that color in HSV is not sensitive to light changes, successfully capturing the denoised rPPG signal. In this way, ROIs of the forehead and cheeks are used, and better performance is achieved compared to traditional methodologies.

### 2.1.3 rPPG post-processing methods

In 2008 HR and RR were estimated by [28] from the rPPG information from the Green (G) channel as it is said to contain the strongest signal, using the forehead of video recordings as ROI. The latter is explained in the same paper with the fact that the green light absorption capacity of oxyhemoglobin is higher than red light absorption and "penetrates sufficiently deeper into the skin as compared to blue light to probe the vasculature". However, it is suggested that the Red (R) and Blue (B) channels contain complementary rPPG information, and in some cases, the RR signal is more pronounced in them. This method then became the basic comparison method for subsequent ones that were developed and will be described next.

Blind Source Separation (BSS) techniques consist in "recovering a set of signals of which only instantaneous linear mixtures are observed" [46]. In 2010 [47] introduced the first automated and motion-tolerant non-contact HR estimation using Independent Component Analysis (ICA), a BSS technique. ICA is defined by [48] as "a statistical method used to discover hidden factors [...] from a set of measurements or observed data such that the sources are maximally independent". Thus, ICA assumes that the observed signals are linear mixtures of independent sources and that the source signal of interest is the cardiovascular pulse wave. Although the linearity assumption may not be correct, it provides a reasonable approximation of the HR signal when a 30 s window is used. However, the motion artifacts evaluated in this paper were slow and small.

In 2011 Principal Component Analysis (PCA), another BSS technique, was used by [49] to estimate HR in a contactless way. Similarly to ICA, PCA separates source components from the observed signals using linear transforms. These transforms are bidirectional, so no information is lost, and they represent the data with a new coordinate system. According to [50], the separation leads to a number "n" of new source components, where the first one corresponds to the cardiovascular pulse wave information and the last one corresponds to noise in low noise conditions. However, in high noise conditions, the source component separation is more complex because artifacts can have higher energy levels than the cardiovascular pulse wave. In [49], PCA was found to have comparable accuracy to ICA on non-moving subjects using the forehead as ROI, though requiring a lower computational complexity.

BSS techniques find the source component corresponding to the pulse signal, assuming it is the signal that presents the strongest periodicity [51], an assumption that is not necessarily true in particular for scenarios of repetitive movements such as exercising. While Joint Blind Source Separation (JBSS) in [52] solves this problem by extracting "the underlying sources within each dataset and meanwhile keeping a consistent ordering of the extracted sources across multiple datasets". In other words, unlike conventional BSS methods, JBSS-based methods have the advantage that can automate the extraction of the BVP adapted for different scenarios, but this method may lead to performance degradation [39].

Consequently, in 2013 [51] proposed a motion robust Chrominance-based method (CHROM) based on the dichromatic reflection model (see Figure 1.2). This method extracts the pulse signal from a temporally normalized RGB channel plane projection, obtained from the linear combination of chrominance signals that is orthogonal to the specular variation direction. It works regardless of the illumination color, and it assumes skin-tone standardization. The proposed method was compared with ICA and PCA using recordings obtained in controlled environments with daylight-fluorescent light and significant motion using the optimal sliding window for each method; this is 32 picture periods equivalent to 1.6 s in CHROM and 512 picture periods equivalent to 25 s in BSS-based methods. The CHROM algorithm demonstrated its robustness and better performance with shorter latency.

In 2014 [14] presented a novel method using Auto-Regressive (AR) modelling and pole cancellation. The latter is an algorithm for the removal of aliasing caused by strong fluorescent lights. The AR model is a linear predictive modeling technique that predicts the signal based on previous signal samples [53] taking into consideration white noise or residual error with zero mean. Using this method, the values of the HR, RR, and $SpO_2$ were obtained from an rPPG signal from an adult patient's face with minimal motion and using a background reference ROI. However, AR modeling looks for regular frequencies in a signal that is assumed to be stationary, which is not the case in NICU. While this model has the advantage that it is unaffected by quantization errors at typical frame rates, unlike Fast Fourier Transform (FFT)-based methods (FFT is further explained in Subsection 2.1.4). This paper also references the Heart Rate Variability (HRV) method as it is commonly used to obtain RR from HR. As mentioned by [54], the reason for this is that "a person's heart rate tends to increase when he/she breathes in, and fall when he/she exhales". Nonetheless, as it is stated by [14], this method is not likely to have high performance with typical hospital patients.

The same year, a method called Blood Volume Pulse Signature (PBV) was presented by [55] as a method that improves motion robustness. The algorithm determines a unit vector, referred to as the blood volume pulse "signature", that describes the pulse signal independently of skin pigmentation. This method showed results with an accuracy comparable to CHROM. Also, the combination of this new method and CHROM was evaluated together with ICA and PCA. Some of the hybrid methods demonstrate to improve the performance of HR estimation in motion situations.

In 2015 [43] proposed an adaptive color difference operation between the G-R channels because the noise caused by motion and lighting variations are similar in these channels, to reduce motion artifacts in remote HR estimation based on the optical properties of the skin. It showed improved signal quality compared to single-channel approaches nevertheless, it was not robust to all types of movements.

As mentioned by [56] PBV uses a predefined pulse signature and CHROM assumes skin-tone standardization. In consequence, when light variations occur, producing a change in the relative contribution of the blood volume pulse in the RGB channels, the RGB based pulse estimation will also change. Therefore, these fixed assumptions may induce errors. For this reason, in 2016 [56] proposed the Spatial Subspace Rotation (2SR) post-processing algorithm, which is skin tone independent and does not require pulse-related priors. The principal behind this algorithm is to estimate the temporal rotation of a spatial subspace of skin pixels. 2SR requires a "well-defined skin mask measuring the single cluster distribution of skin pixels". The algorithm was compared with ICA, CHROM, and PBV under a variety of subject motions and illumination conditions and showed to improve HR measurement results.

In 2017 [29] introduced the Projection Plane Orthogonal to the Skin-tone (POS) method for pulse signal extraction based on the dichromatic reflection model. This method extracts the pulse signal from a projection plane orthogonal to the skin tone. Later in the paper POS is compared with other commonly used state of the art rPPG techniques including G, G-R, PCA, ICA, CHROM, PBV, 2SR. It is concluded that model-based methods, which are CHROM, PBV, and POS perform significantly better in fitness contexts. Also, CHROM performs better than PBV overall, particularly when the subject is nearly stationary. While 2SR is the best performing in non-fitness scenarios and in the same context, ICA performs better than PCA. Comparatively, [39] states that POS is more robust to illumination variations than CHROM, while this last one is more robust to motion artifacts.

According to [57] Ensemble Empirical Mode Decomposition (EEMD) is a "method to process nonlinear and non-stationary signals. It is a completely adaptive and data-driven algorithmic approach. It decomposes the signal into amplitude and frequency modulated [...] oscillations called Intrinsic Mode Functions (IMFs) without any a priori assumption and defined a basis". While the Multivariate Empirical Mode Decomposition (MEMD) is an extension of this algorithm to analyze multi-channel data. Thus, MEMD is an algorithm for multivariate non-stationary signals analysis. Also the year 2017, [58] used Partial Least Squares (PLS) and MEMD from facial and background ROI to obtain the HR under varying illumination conditions. PLS is "a data analysis technique for testing theoretical relations among a system of variables" [59]. In this method, PLS is used to determine the projection that maximizes the covariance between the ROIs, extracting the illumination variation. Then, MEMD decomposes the information of multiple signal channels from the ROIs into IMFs without considering the dependent information among these channels. The technique assumes that both the facial ROI and background ROI have similar illumination variation sources. The PLS-MEMD was compared with ICA and EEMD, which does the same as MEMD but considering only one channel at a time, on subjects sitting stationary in front of the camera. PLS-MEMD showed better results overall.

In 2018 [60] presented a model robust to nuisance factors called Local Group Invariance (LGI). The algorithm searches for invariant features as a result of local transformations, incorporating uncertainty in the feature distribution. This method was compared with ICA, 2SR, and POS using a self-created database and showed improvements in movement situations, specifically in the category of talking, rotation, and gym.

In 2019 [21] measured the RR on subjects facing the camera with casual walking motion. They applied ICA and then used Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) scheme to decompose the ICA output into its IMFs. CEEMDAN is an improved version of EEMD, which has an addition of adaptive white Gaussian noise. Then an ML algorithm is used to identify the IMF that best represents the RR. A Root Mean Square Error (RMSE) of 2.30 bpm was obtained (RMSE meaning is explained in Section 3.4), yet the algorithm was not used on real-time applications.

In 2020 the Virtual Heart Rate Python package named pyVHR was created by [37]. This is an open framework which implements the following rPPG methods: ICA, PCA, G, CHROM, POS, LGI, PBV. They also investigated which window size to use to yield the highest PCC, which was 10 s for all methodologies. Then, the framework was tested with publicly available databases concluding that CHROM is overall the best methodology with a Mean Absolute Error (MAE) of 2.31 bpm on average (the MAE meaning is explained in Section 3.4).

Also this year, [30] proposed the LAB-EEMD method to obtain illumination variation resistant HR measurements. This algorithm converts RGB color space to LAB color space to separate the luminance signal, and it is followed by EEMD to obtain the pulse signal IMFs. Also, smoothness prior approach and pixel averaging are used to eliminate noise. The technique was tested in scenarios of changing illumination (including slow head rotation) and showed results similar to CHROM.

In 2021 [39] proposed a method combining Projection-Plane-Switching (2PS) and Singular Spectrum Analysis (SSA). This method uses 2PS based on head motion assessment. Specifically, it determines the distance changes by the head between adjacent frames and, for a given threshold, switches between the projection axes of CHROM and POS. Also, it uses SSA, which is "a non parametric procedure based on subspace algorithms for signal extraction. The main task [...] is to extract the underlying signals of a time series like the trend, cycle, seasonal and irregular components" [61]. In this case, SSA is used on the B channel of the face ROI for noise removal. The proposed methodology was evaluated using non-rigid motions such as blinking and strong illumination variations. Results showed 2PS-SSA method had the best performance among the compared methods, including 2PS, CHROM and G.

The same year, [24] proposed an algorithm for HR estimation in which the face gets divided into $25 \times 25$ pixel sub-blocks, logarithmic operations are applied to separate noise from the reflected component of the PPG information. Then EEMD is used to obtain IMFs, and the signal quality of each ROI sub-block is computed to use only high-quality signals for the HR estimation. The study used a 30 s window and obtained a RMSE of 5.62 in stable light conditions and 8.30 in unstable light conditions. Yet, note that subjects were asked to avoid rigid head movement.

In 2022 [25] adapted pyVHR framework into an easy-to-use pipeline. The software exploits NVIDIA Graphics Processing Unit (GPU) to accelerate code execution into real-time inference speed by using parallelism therefore, it has the potential to be used in real-time video processing. Currently it supports input video recordings, not real-time video streaming.

Note that all of the previously mentioned studies have results biased towards age because there are no infants in the tested databases. This issue was also referenced by [62], where public databases were analyzed, and an age bias was confirmed. It was therefore suggested to be cautious with age-undifferentiated products derived from these databases. Moreover, a skin color bias was also confirmed in public databases by [63]. Consequently, products derived from their use could lead to unrecognized health conditions of the under-represented skin pigmentation population. Due to this circumstance, the Food and Drug Administration (FDA) in [64] recommends interpreting trends more meaningfully than one specific pulse oximeter measurement.

### 2.1.4 Algorithms for vital signal estimation

Power Spectral Density (PSD) is the measure of the input signal power over a range of frequencies. It is a commonly used method to track and distinguish signals of interest from the data [15], [25], [37], [41], [65]. From the PSD, the dominant frequency peak in a given frequency interval is evaluated to obtain the HR and RR estimations [65].

As stated in [66], the PSD of a signal is the FFT of the auto correlation of the signal. The FFT is a high processing speed implementation of the Discrete Fourier Transform (DFT), which converts discrete signals from the time domain to the frequency domain [67]. Real-time frequency domain computation of signals sampled at a rate of 16 MHz is currently feasible [68]. Multiple algorithms have been used to estimate PSD. A popular algorithm is the Welch PSD method, nevertheless, this algorithm requires high computational complexity [69]. In [65], an algorithm for PSD estimation with good performance was proposed, yet the algorithm's complexity has prevented real-time applications.

Consequently, [66] proposed a modified version of the Welch PSD method, which computes PSD using the Welch algorithm at a lower computational cost but at the expense of an approximately 8% lower performance.

However, the DFT can become an inadequate technique for signal analysis when the signal has transient and non-periodic components [70]. Hence, Short-Time Fourier Transform (STFT) can be used for time-frequency analysis of non-stationary signals, providing an insight of the time-evolution of each signal component. If a signal is altered in a specific time instant, the entire frequency spectrum can be affected. So, to detect these temporary positions in data, a small window has to be used in the STFT. However, as mentioned by [71], the STFT fails when used for signals with slowly varying components and rapidly changing transient events.

Moreover, as mentioned in [72], using PSD and STFT combined can yield a higher accuracy rate than both techniques separately.

## 2.2 State of the art

In this Section, an overview of literature investigations regarding techniques designed and tested to get contactless estimations of vital signals of newborns are described.

In 2013 [2] estimated HR using a camera by analyzing the G channel with ambient light for the first time at the NICU. The study proved the feasibility of HR estimation using a camera on the NICU environment. However, the motion artifacts and poor illumination conditions were mentioned as improvement sources.

In 2014 [34] obtained continuous estimates of HR, RR and $SpO_2$ for infants nursed in incubators with minimal motion and ambient light, excluding periods of intervention of the clinical staff. They manually selected two ROIs (from the face and the background) and used ICA and pole cancellation in AR models to extract the vital signals without interference from aliased frequencies. This study used band-pass filters of bandwidth 1.3-5 Hz for HR and 0.33-1.67 Hz for RR.

In 2015 [73] used OF, ICA and PCA signal extraction algorithms for RR estimation of neonatal video data. Video frames were manually cropped to show only the chest and abdomen ROI. Then Eulerian Video Magnification (EVM), OF, and STFT were used to obtain the RR estimation. EVM is a technique used to amplify small variations to make them detectable. As mentioned by [73] "the algorithm tracks and amplifies changes in pixel intensity values over time. A constant illumination of the scene is therefore necessary". Further, PCA and ICA were used to improve signal quality, where PCA showed a better performance.

In 2016 [18] proposed an algorithm that finds the linear combination of the color channels with the best SNR to represent the pulse rate and then filtered this signal in the corresponding bandwidth of the RR to obtain its measurement. The algorithm was evaluated using visible light on NICU subjects, yet results show that the existing algorithm CHROM had a better performance.

In 2018 [17] estimated HR and RR using a methodology of low Central Processing Unit (CPU) consumption which was declared to work with varying illumination conditions and motion. It is based on rPPG analysis on the infant's diaphragm, where a $40 \times 40$ pixels ROI was manually selected from this region, and the least squares method was applied to the average pixel intensity of each channel over a moving window to find the linear function that best fits the signal. A HR estimation value was obtained every 1 s, and an estimation of RR was obtained every 3 s. It was concluded that the R-channel has to be analyzed in varying illumination conditions since this channel maintains a high level of pulse signal information independently of light changes.

The same year, [74] proposed a video processing algorithm to obtain RR estimations based on the analysis of local variations and breathing motion magnification. The pixel-wise processing was performed over a single-channel grayscale video leading to preliminary results for steady newborn subjects.

In 2019 [16] analyzed the G channel using two cameras, this time incorporating an improved version of EVM to magnify the signal of the videos to determine HR and RR. For this, ROIs were manually detected on preterm infants. The proposed solution demonstrated to detect apnoea episodes while the reference ECG couldn't. Results were compared with the measurements obtained without the magnification algorithm applied, and it was found that when the infant was moving the magnification increased noise and led to inaccurate results.

The same year, [22] developed a multi-task deep learning algorithm to segment skin areas automatically and to estimate vital signs, specifically HR and RR, of infants in the NICU when no medical procedures where performed. The HR and RR estimations were obtained using multiple algorithms over a window of 8 s and 10 s respectively and then a data fusion technique was applied to combine these estimations. That AR best model used for HR estimation has rules to discard noisy periods, and so "it incorporates a trade-off of high accuracy in exchange for a smaller portion of computed values over time" [22]. However, this study demonstrated that real-time execution is possible with a vital signal estimation value output every second.

In 2020 [20] estimated RR of fully clothed or swaddled infants with a technique called Micromotion and Stationarity Detection (MSD) using various lighting conditions and camera orientations. The MSD is an algorithm that assumes that the standard deviation of the change in pixel intensity over a series of frames with no motion, but noisy, remains relatively small therefore a large change in this standard deviation means a micro-movement associated with RR. In this way, an RR estimation was computed every 5 s. Yet, the algorithm described is not motion robust.

The same year, [3] estimated HR using skin segmentation of the face by transforming RGB color domain to HSV color domain, in which the skin color falls into a particular range. Then EVM was used to detect and magnify changes. An HR estimation was obtained every 1/9 s in real-time. Results got an average MAE of 7.4 and RMSE of 15.2. However, the neonates did not move during the execution of this study.

In 2021 [19] implemented a motion robust algorithm over the G channel consisting in ROI division, EVM and majority voting in order to choose the HR with the highest probability among ROI patches. The proposed algorithm considered head rotation and non-rigid motions, such as blinking and emotion expressing, obtaining a MAE of 4.3 bpm. However, this algorithm uses HSV color space to filter skin cells in each frame, for which it assumes that no objects of similar color are on the background otherwise, these would introduce noise.

The same year, [75] created a publicly available dataset of full-term infants freely available upon request to develop a deep learning method, as there was no public dataset available for this purpose. Also, they proposed a multi-task deep learning method called NBHRnet to estimate the HR from which they obtained a MAE of 3.97 bpm.

# Chapter 3

# Materials and methods

## 3.1 Experimental setup

As shown in Figure 3.1 (obtained from [17] and adapted for the specific setting), newborns in the self-made dataset used in this study (a) were inside an incubator (b) in a supine position. A camera was placed and fixed on the top of the incubator (d) at a distance smaller than 1 m to obtain a video recording of the infant, ensuring that the baby's head and chest were always within the captured recording.

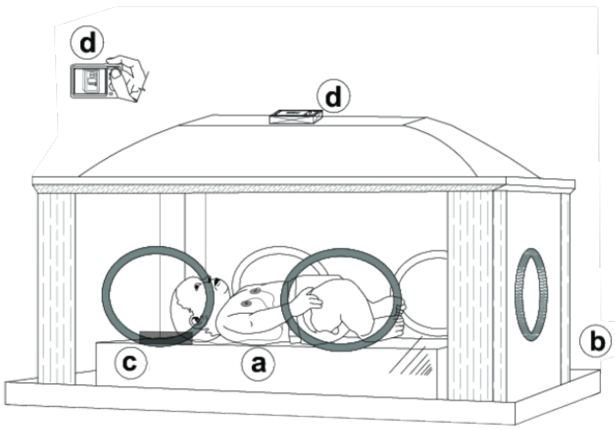


**Figure 3.1:** Newborn incubator experimental setup. Source: Adapted from [17]

These characteristics are important because, as mentioned in Subsection 1.2.5, for the purpose of vital signal estimation using a camera this distance from the subject to the camera has shown to have higher performance compared to greater distances. Also, the face of the baby is used for ROI selection therefore, its presence in the video recording is fundamental for vital signal estimation.

A pulse oximeter with the reference measurement of HR (the ground truth) was connected to the newborn's chest by adhesive patches, and its monitor was positioned next to the infant, as shown in (c). Note that the HR ground truth value is the second number shown on the screen top-down. In this way, the ground truth values were displayed in the video recordings. Additionally, the only source of illumination was artificial light.

An actual frame from a video recording is shown in Figure 3.2 to illustrate the video recording perspective previously described. Note that images exposed in this Chapter corresponding to database video recording frames were blurred during post-processing to preserve the privacy of the newborns.


**Figure 3.2:** Example of a video recording frame.

During video recordings, newborns could freely move, and healthcare personnel could perform procedures normally.

All videos were recorded using a 24-bit RGB camera with 3 channels of 8 bits per channel at 30 fps with a pixel resolution of 480 × 852 and saved in uncompressed MP4 format. Automatic camera parameters were disabled as their changes may affect the intensity of the color channels differently over time [14].

## 3.2 Dataset

The dataset used for this thesis work comprises 23 video recordings obtained from the Neonatal Unit of Mauriziano Hospital in Turin under a study protocol approved by the Local Ethics Committee. Informed consent was obtained from participants' parents prior to the start of video recordings. Video recordings have different durations corresponding to male and female pre-term newborns with less than 40 weeks. Also, according to the color bar in [76] showed in Figure 3.3, the Fitzpatrick skin color classification of all newborns in the videos corresponds to type I. The color bar was used by comparing it with the inside part of the upper arm of the newborn, yet it is important to note that skin sensitivity to sunburn should be taken into consideration to make a more objective skin color classification.


**Figure 3.3:** Fitzpatrick grading color bar tool used by matching these colors with the newborn's skin tone at the inside part of the upper arm. Source: [76]

During the video database analysis, it was found that some videos were duplicated in the dataset therefore, the shorter duplicated videos were excluded from the investigation to avoid redundancy. Also, some videos were excluded due to the absence of ground truth values to compare with the HR or RR estimations (explained in Section 3.3). Finally, video recordings with a shorter duration than the window size selected for the analysis (explained in Section 4.1.2) were also excluded. Thus, videos from which HR estimations were effectively calculated correspond to a total time of 1 hour and 28 minutes, while for RR estimations, a subset of this dataset was used (further explained in Sections 3.3 and 3.5).

These videos were classified into 3 categories of motion: “motionless”, “sporadic motion” and “motion”. The “motionless” category describes completely still newborns, “sporadic motion” corresponds to sleeping newborns occasionally making involuntary movements, and finally “motion” refers to newborns who do not stop consciously moving. Furthermore, it is to be considered that some video recordings had an undergoing medical procedure of blood draw. The respective characteristics of effectively analyzed video recordings are specified in Table 3.1.

| Video number | Sex | Description | | Duration | |
|---|---|---|---|---|---|
| | | Movement | Medical procedure performed | Minutes | Seconds |
| 2 | Female | Motionless | no | 1 | 30 |
| 3 | Male | Sporadic motion | no | 7 | 54 |
| 5 | Male | Motion | yes | 17 | 42 |
| 6 | Male | Motionless | no | 2 | 19 |
| 7 | Male | Sporadic motion | yes | 10 | 5 |
| 9 | Male | Sporadic motion | no | 6 | 17 |
| 11 | Male | Motion | yes | 8 | 03 |
| 14 | Male | Motion | yes | 13 | 43 |
| 17 | Male | Sporadic motion | no | 3 | 14 |
| 21 | Male | Motion | no | 2 | 18 |
| 22 | Male | Sporadic motion | yes | 15 | 28 |

**Table 3.1:** Processed dataset specifications, where the sex of the newborn, movement, and medical procedure characteristics are detailed. Also, the duration of each video from which vital signals were effectively estimated is specified.

## 3.3 Ground Truth

Ground truth values of the HR were obtained from a pulse oximeter used during the recordings, with its corresponding values displayed on the videos. Unfortunately, directly downloading the HR trends was not possible because this software is proprietary software, as opposed to freely distributed software. Therefore, an algorithm was developed to extract the ground truth from the videos using character recognition. This algorithm obtains a value for each video frame corresponding to the pulse oximeter value read in that frame and stored as a CSV file for each recording.

To ensure the accuracy of the extracted values, pre-processing and post-processing is required. Concerning the pre-processing algorithm, each frame is first masked to isolate the region where the number is displayed. Then it is resized to amplify this value. Furthermore, it is grey-scaled, binary thresholded, and blurred to reduce noise. Finally, tesseract, which is a Python library for digit recognition, is used to determine the corresponding number [77], [78].

It is important to note that the frames, in some cases, needed to be rotated to make the values horizontally aligned to facilitate character recognition depending on the instrument position and orientation of the recording. Also, noise-reducing techniques were adapted to the specific light conditions of each video. For these reasons, a strategy design pattern was adopted to code this algorithm meaning different variants of the algorithm are used depending on the specific video characteristics therefore, part of the behavior of the algorithm changes accordingly [79].


**Figure 3.4:** Example of a frame from which to obtain the HR ground truth value, corresponding to the second number top-down on the screen of the pulse oximeter.


**Figure 3.5:** Pre-processed value to be interpreted by tesseract digit recognition.

On Figure 3.4, there is an example of a frame that required a different strategy, compared to that of Figure 3.2. The strategy included rotation and a specific binary threshold because lightning conditions made the pulse oximeter HR value more obscure. Figure 3.5 shows the output of the pre-processing stage, which is interpreted correctly by tesseract as “138”.

Some frames in which there was occlusion of the instrument or where the pulse oximeter value was not fully captured in the recording were discarded. Furthermore, video recordings with light reflection on the instrument screen, which prevented the character recognition algorithm from interpreting its value correctly, were also discarded.

With respect to the post-processing techniques, only non-digit values were discarded to preserve the integrity of the result. This means that, for example, a dot or hyphen, which sometimes were recognized, were discarded but possible digits identified in the same value remained unchanged. Then, blank values, abnormally high values, which are values over 290 bpm as mentioned in [22], and abnormally low values compared to the surrounding frames were checked and properly replaced. Finally, outliers were detected as numbers that increment or decrement by more than 11 bpm from one frame to another. The threshold was set at this number because this is the maximum change of the pulse oximeter instrument values for the HR from one frame to the next one captured in the recordings. Finally, the file containing the ground truth for each video frame was saved as a CSV file.

For the RR, there was no instrumentation to measure this vital signal directly from the newborn. Therefore, in compliance with the gold standard described in Subsection 1.2.2, breaths were manually counted by personnel of the Neonatal Unit of Mauriziano Hospital in Turin by observing the chest movement of the infant for specific instants of a subset of the dataset. Note that obtaining the RR ground truth values is possible since the newborn’s chest is uncovered; thus, its breathing movements are visible. Therefore, the RR for a given instant was obtained by counting the breaths (observing the respiration motion) during a window that starts at that specific instant and spans for one minute from that point in time.

In this manner, the amount of RR ground truth values collected using this methodology included a total of 21 values. However, 4 out of these values were excluded from the analysis because of two different reasons. One of them was that the video was not long enough to account for a 60 s window after the ground truth measuring instant. Therefore it was not possible to make the RR estimation using the same amount of information for comparison (explained in Section 3.5 with Figure 3.11). The second reason was due to the failure of the ROI detection and tracking algorithm, which consequently resulted in a lack of sufficient rPPG data collection to make a reliable RR estimation (explained in Subsection 3.6.2).

## 3.4 Evaluation metrics

The metrics used to evaluate the results are the following: RMSE, MAE, Maximum Mean Absolute Error (MAX) and Pearson Correlation Coefficient (PCC). In Table 3.2, the formulas for each metric are shown, where $x$ is the estimated value, $y$ is the ground truth value, $\{1...T\}$ are the values for each window, $\overline{x}$ and $\overline{y}$ are the mean values of the estimation and the ground truth respectively.

| Metric | Formula |
| --- | --- |
| RMSE | $\sqrt{\frac{1}{T}\Sigma_{t=1}^{T}(x_t - y_t)^2}$ |
| MAE | $\frac{1}{T}\sum_{t=1}^{T}\lvert x_t - y_t\rvert$ |
| MAX | $\max_{t=1}^{T}\lvert x_t - y_t\rvert$ |
| PCC | $\frac{\sum_{t=1}^{T}(x_t-\overline{x})(y_t-\overline{y})}{\sqrt{\sum_{t=1}^{T}(x_t-\overline{x})^2\sum_{t=1}^{T}(y_t-\overline{y})^2}}$ |

**Table 3.2:** Evaluation metrics and formulas, where $x$ is the estimated value, $y$ is the ground truth value, $\{1...T\}$ are the values for each window, $\overline{x}$ and $\overline{y}$ are the mean values of the estimation and the ground truth respectively.

The RMSE is the standard deviation of the error thus, it measures the accuracy of the model's prediction. In this way, if it is close to zero, it means that the results are concentrated around the best-fit line and therefore, it is accurate.

The MAE is a similar measurement to RMSE, yet it is more robust because it is less sensitive to outliers. The MAE is the average error expected on the prediction. An absolute error between 3 bpm and 5 bpm is considered clinically acceptable according to [39], however, this error range makes reference to a grown child or an adult, both having a HR range around two times smaller compared to a newborn (as it was explained in Subsection 1.2.6 of Chapter 1). Therefore, for this study, an acceptable absolute error would be approximately between 6 bpm and 10 bpm, considering the differences in the magnitude of newborn HR ranges with respect to adults.

The MAX error is the largest MAE. A small MAX error suggests that the model's error never strays far from the reference values, so the prediction is near the ground truth.

The PCC measures the linear correlation of the ground truth with the prediction. Normalized PCC is between -1 and 1, where 1 means perfect correlation. According to [80], a PCC between 0.3 and 0.6 is a moderate correlation, while a PCC above this range is considered to be a strong correlation. However, note that according to [81], the minimum sample size required for a proper PCC is size 25. Therefore, this error metric cannot be used for the ground truth available for the RR as it has an effective size of 17 ground truth values.

## 3.5 Methodology

### 3.5.1 General software specifications

The software developed uses Python 3.9, and the algorithm for vital signal estimation, which receives as input the dataset of video recordings, is accelerated by Compute Unified Device Architecture (CUDA) on NVIDIA GeForce GTX 1060 6GB. Refer to Subsection 3.5.3 for a detailed description of this architecture.

### 3.5.2 Use of pyVHR framework

To estimate vital signals the pyVHR framework [25], [37] was adapted for the context of use. This framework was used because it implements various popular rPPG processing algorithms to obtain vital signs that have shown good performance in adult studies and some of which declare to have important characteristics for the context of use, such as motion robustness and lightning robustness, as it was mentioned in Subsection 2.1.3. Until now, these methods have not been compared in the NICU context as it was seen in Subsection 2.2 referencing the state of the art.

Another reason to use this framework is that it uses accelerated algorithms by exploiting CUDA NVIDIA GPU parallelism with Python; hence it is suitable for real-time processing, which is a transcendental characteristic for this study.

However, note that as mentioned in Chapter 2, with the current implementation of the framework, video recordings are processed as input; the framework does not receive as input a data stream. So, the processing pipeline receives a complete video yet performs the analysis of it on moving windows as it would be done in an actual real-time implementation of the algorithm. Since the present work has the potential for real-time application, the direction of this study points towards real-time usage. However, some modifications would be needed at the software level to fully support this.

Furthermore, using rPPG technology is convenient because HR and RR vital signal estimations can be extracted using a similar processing analysis with the difference that some parameters have to be adapted depending on the vital signal to be measured, such as the ROI selected, the window size, the bandwidth of the filter, between others. In this sense, using this framework is advantageous as the main pipeline for processing the vital signal has access to these parameters, as it uses a facade-type design pattern making it easier to interact with the main function parameters of the framework without delving into the complexity of the functions [79]. Therefore, it is a flexible framework, as it is possible to change parameters from the processing pipeline.

### 3.5.3 Parallel computing using CUDA Python

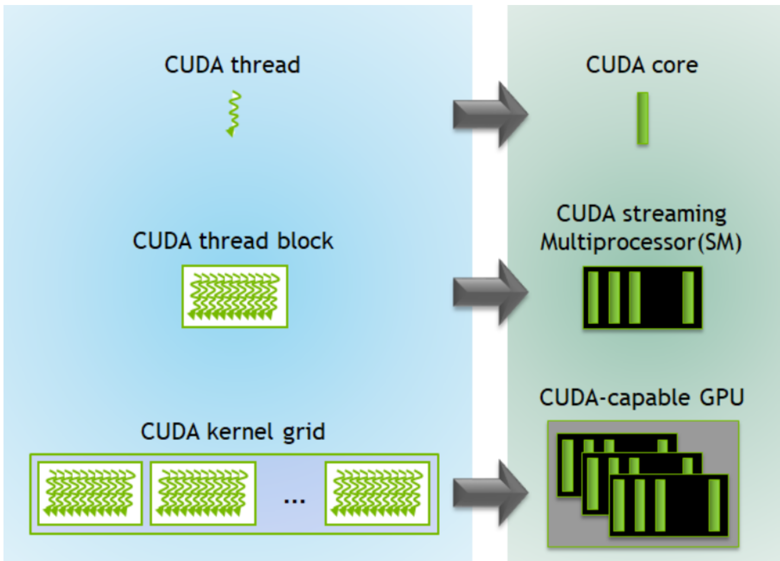


**Figure 3.6:** CUDA components illustration. Source: [82]

The underlying computer architecture used is CUDA, which is a parallel computing platform that allows developers to use the processing power of NVIDIA GPU devices for general-purpose processing. Parallelism in CUDA is achieved through the use of threads, thread blocks, and kernels. As described in [82], a thread is a single unit of execution within a CUDA program, a thread block is a group of threads, while a kernel is a group of threads blocks that are executed synchronously and communicate through shared memory. A visual representation of these aggregation levels is shown in Figure 3.6. By breaking down computations into many threads and organizing those threads into blocks and kernels, CUDA enables developers to achieve massively parallel computations executed independently on the GPU.

In [83] it is mentioned that the benefits of using GPU for parallel computations, apart from increasing computational speed, also include increased power efficiency when using CUDA. The same study describes that accessing the GPU from Python can be as efficient as accessing it from C or C$^{++}$. They concluded that Python could be even faster and provide higher quality code, meaning fewer bugs and crashes. In this way, Python CUDA can lead to a highly productive development environment.

### 3.5.4 Processing pipeline description and modifications

Figure 3.7 provides an outline of the pyVHR framework's stages, obtained from [37]. However, note that there are some inconsistencies between the illustration and the actual implementation of the framework. In the following paragraphs, these stages are described in detail mentioning also the actual implementation used and the changes made to them.

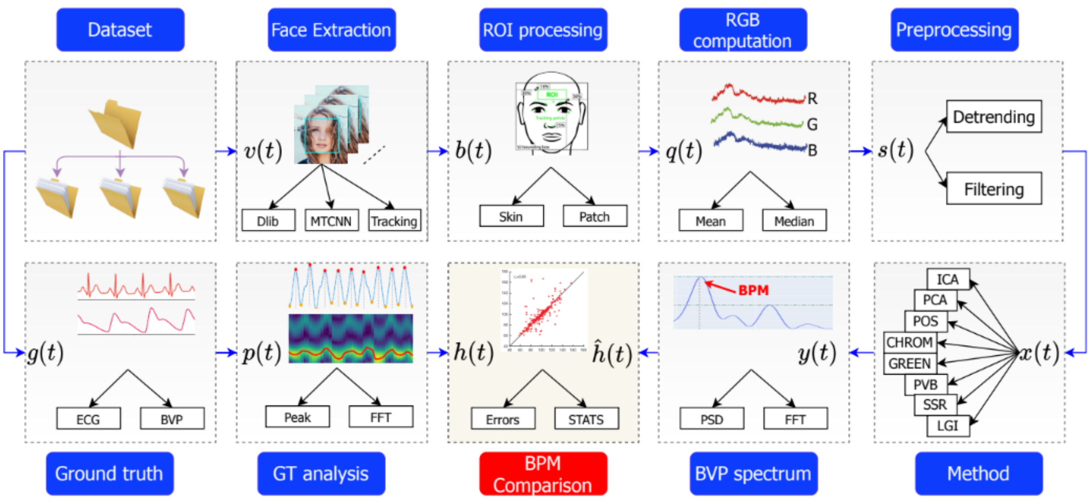


**Figure 3.7:** pyVHR framework stages. Source: [37]

As shown in Figure 3.7, the dataset, composed of video recordings, is loaded into the processing pipeline to obtain the vital signal estimations from them. The framework supports using public databases as the processing pipeline includes a factory design pattern to open and process each video from them. This part was modified to adapt the input dataset settings to the private video recordings described in Section 3.2.

The next stage in the processing pipeline is ROI detection and tracking using Google Mediapipe library, specifically the Google Mediapipe face mesh (see Section 2.1.2 for its generic description and literature applications). It works in the following way: first, Google Mediapipe uses face detection on frames where no figure is recognized as a face until it localizes the face. Then, it tracks this face in consecutive frames.

Google Mediapipe face mesh was used with the following configuration: minimum detection confidence of 0.5, meaning the probability that a recognized figure in a frame was the face of the newborn was at least 50%. Also, the minimum tracking confidence was set at 50%, meaning that the tracking of the figure persisted as long as there was a 50% or higher probability that the recognized figure corresponded to the face of the newborn. In cases where this threshold was not met, automatic facial detection was initiated in the subsequent frame. Increasing these probability thresholds can increase the robustness of the model, but the trade-off is increased latency. Also, static image mode can be set to true to use face detection on every frame instead of face detection and tracking, but at the cost of a higher computational requirement.

Subsequently, pyVHR framework offers two possibilities, either to consider all the skin of the face as ROI or to select pre-defined groups of landmarks in the face as ROI. The second option was chosen because, as it was described in Subsection 2.1.1, literature has shown that ROI region selection impacts the quality of the estimations [37]. This procedure was also further customized to enable specific landmark selection choices. Following, the regions of the eyes and the mouth are cropped from the face of the subject to avoid the contribution of these skin pixels in the vital signal estimation. This is in compliance with the literature discussed in Subsection 2.1.1, which states that these regions should be avoided because they induce non-rigid motion error. Also, a mask is used over the face to avoid the contribution of the background as noise (further details about chosen ROI landmarks can be found in Subsection 3.6.1).

In the RGB computation stage, a square of pixels centered on each selected landmark is used to obtain either the median or mean of all of the pixel color intensities in each square for each color channel. Note that this square has a customized size. In this study, the mean value was used, as previous literature described in Chapter 2 exclusively uses the raw signal, which by definition is the mean RGB color channel value.

When calculating the raw signal in the color intensity extreme values, a threshold is used for two reasons: first, to avoid the contribution of the mask in the computation, and second, because newborns are allowed to move freely, so there are moments when the newborns do not face frontally the camera which reduces the visibility of the selected ROI for the vital signals estimations. Therefore, it is necessary to adaptively select the ROI depending on the newborns' position with respect to the camera. In this way, only RGB pixel values between 5 and 250 are considered in each square when taking the average, and so "a patch may disappear due to subject's movements, hence delivering only partial or none contribution" [37]. In other words, only the pixels that are inside the threshold are taken into consideration when obtaining the mean color intensity of the patches. Refer to Figure 3.8 for an example that illustrates a specific patch that includes pixels that surpass the threshold. In this example, only the blue area is used to calculate the raw signal for this patch.

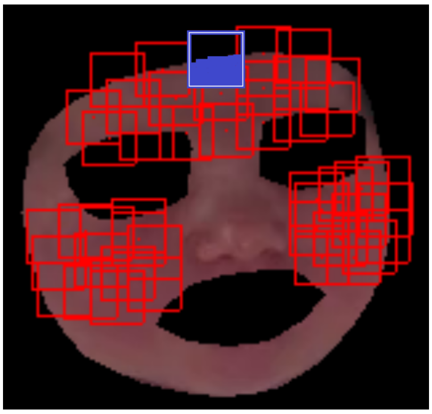

**Figure 3.8:** 45 landmarks from Figure 3.12 with patches of 20 px side size. The partial contribution of a patch from the forehead is shown, where the blue shaded area is considered for the raw signal computation.

Then, the median from all square patches is selected to represent the signal in that frame. According to [84], the median represents the central tendency of a group better than the mean as it is less affected by outliers.

A custom size time window is used to obtain the rPPG signal information. This window moves in a step size which is also customizable (further details about chosen window and step sizes can be found in Subsection 3.6.2). A vital signal value is obtained from each window by analyzing the collected rPPG information within it. In the pyVHR framework's moving window approach, the vital signal value is estimated at the center of the window. However, to simulate real-time application behavior, it is not possible to predict future values. Estimating the vital signal value in the middle of the moving window implies forecasting the future half of that particular window. To address this, a modification was implemented, and now the vital signal value is estimated at the end of the moving window. Subsequently, the window shifts by a step size and the next estimation is computed, as illustrated in Figure 3.9.

Due to the fact that Google Mediapipe can fail to recognize and track the ROI over a moving window, some windows collect rPPG information for a shorter period than the designated window size. To ensure reliable vital signal estimation, an additional step was incorporated into the framework. It verifies that the moving windows contain a minimum quantity of rPPG data information, discussed in Section 3.6.2. This step is applied to each moving window of the video.

More in detail, the presence of artifacts (such as newborn movements and healthcare staff interference) can cause Google Mediapipe ROI detection and tracking to fail. Hence, these frames provide no rPPG data collection. Therefore, it is necessary to ensure that to compute the vital signal value from a moving window, this window should have at least a period of magnitude equivalent to the minimum window size with ROI detection and tracking present from where to extract the vital signal estimation. Otherwise, the framework does not output any estimation for that moving window as it would be considered to be an unreliable value, and so the comparison with the ground truth value is not made, meaning no evaluation metrics are obtained in this case.

Note that this part was programmed in an independent Python script. Its integration required many changes of high complexity due to the design of the pyVHR framework. However, it has the potential to be integrated into the pipeline with a different underlying design. For this reason, this step is shown as part of the processing pipeline in Figure 3.10, which shows the stages of the processing pipeline used in this study.

Then, the rPPG signal in a moving window is processed with an algorithm. Generally, these algorithms, based on mathematician and scientific models, possess additional characteristics such as motion or lighting robustness, which are able

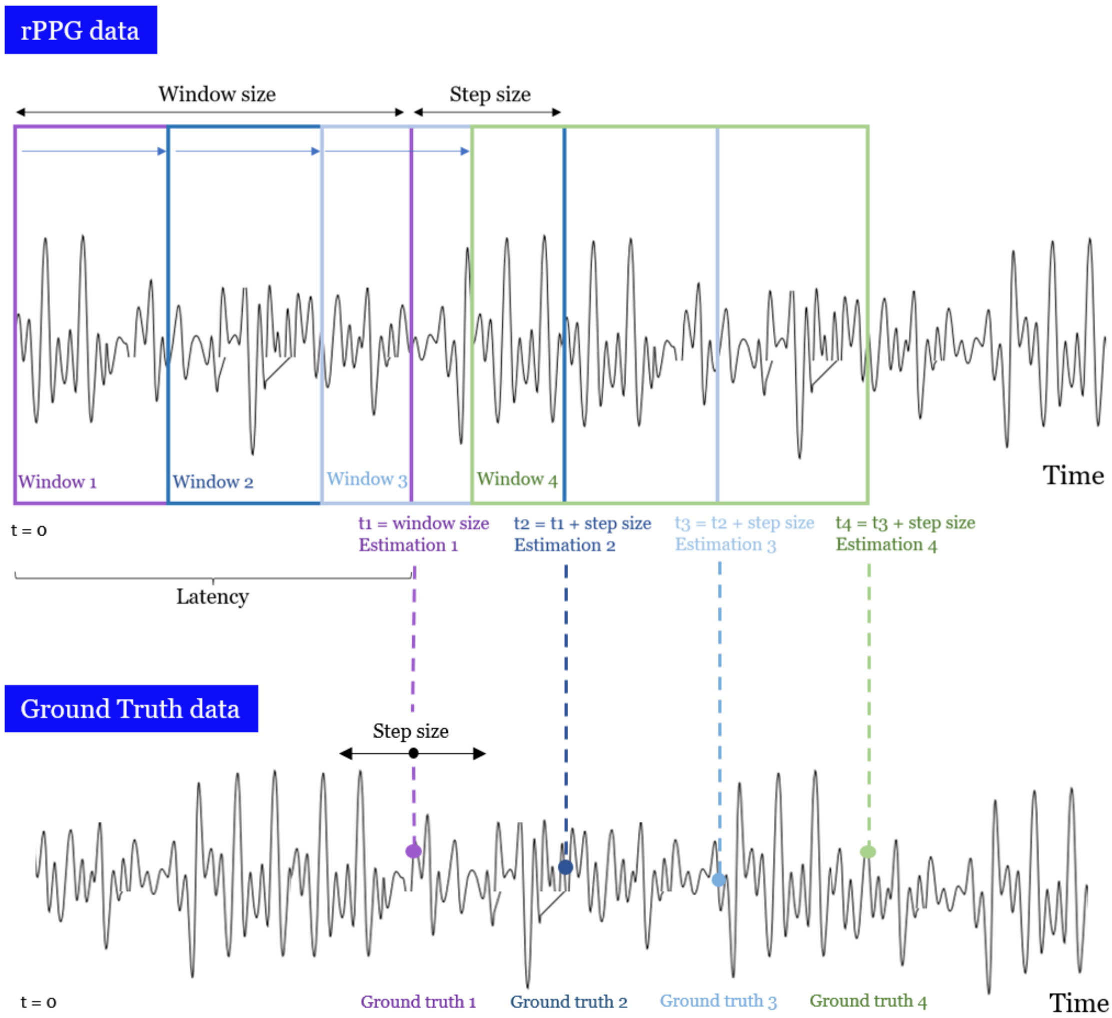


**Figure 3.9:** Moving window and step size illustration used to estimate HR values from rPPG data and ground truth values used for comparison.

to isolate the signal that contributes to the physiological parameter to be determined from the rPPG information in the moving window, thus reducing errors before the final vital signal estimation step. Currently, there is no consensus about a benchmark criterion to be used. The possible algorithms included are: ICA, PCA, G, CHROM, POS, LGI, PBV, and all of them where used in this project. Many of these methodologies have shown promising results in adult studies, as it was outlined in Chapter 2. Table 3.3 shows a description of the rPPG algorithms, which summarize the technologies already described in Subsection 2.1.3.

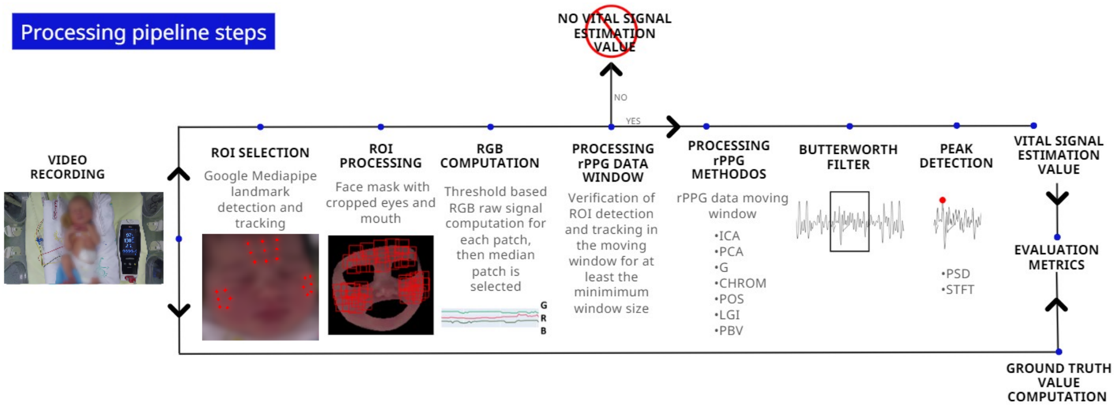


**Figure 3.10:** Processing pipeline steps used.

The purpose of comparing the algorithms described is to gain a comprehensive understanding of their performance within the specific conditions of the NICU, to determine which algorithm demonstrates superior performance under these conditions.

Optionally, filtering or detrending can be added before and/or after using the rPPG signal processing algorithms. In this way, the Butterworth filter was included to remove noise outside the HR and RR frequency bandwidths (for further details about the chosen Butterworth filter order and bandwidth, see Subsection 3.6.4).

Finally, in the BVP spectrum stage, two algorithms are available to estimate the vital signals based on the processed rPPG signal; these are: PSD and STFT (refer to Subsection 2.1.4 for a description of these algorithms, and for further details about the chosen algorithm are available in Subsection 3.6.5).

To summarize, all of the processing pipeline steps used in this research are shown in Figure 3.10.

In parallel, the ground truth was prepared to be able to compare the reference value with the estimated value. To accomplish this, the processing pipeline was modified to dynamically incorporate the ground truth value for comparison and immediately obtain performance evaluation metrics. So, ground truth was prepared to synchronously deliver a value each time an estimation was computed.

The criterion used for HR estimations was the following: i) start the processing of the ground truth starts at the second equivalent to the latency of the window size chosen for the estimation of HR. This is the moment when the first HR estimation

| rPPG algorithm | Description |
|---|---|
| G | Spatial averaging of the G channel. [28] |
| ICA | BSS method, which assumes that the observed signals are linear mixtures of independent sources. [47] |
| PCA | BSS method, which finds source components using linear transforms to represent the data with a new coordinate system. [49] |
| CHROM | Feature transform method that projects the temporally normalized RGB channel signals on a plane orthogonal to the specular variation direction (assuming skin-tone standardization). [51] |
| PBV | Blood volume pulse "signature" unit vector independent of skin-pigmentation. [55] |
| POS | Feature transform method that projects the temporally normalized RGB channel signals on a plane orthogonal to the skin-tone. [29] |
| LGI | Feature transform method that searches for invariant features using local transformations. [60] |

**Table 3.3:** Description of benchmark processing rPPG algorithmns used.

is outputted. ii) Select a window of a size equivalent to the chosen step size for the estimation of HR is selected centered on the position previously determined. iii) Calculate the median of the ground truth values in the frames of this window, and output the value obtained as the ground truth for comparison in this instant. iv) Move a magnitude of step size seconds because a vital sign value will be estimated after each step size seconds and repeat steps ii-iv until finished. Figure 3.9 shows a visualization of the ground truth values used for comparison with the HR estimations.

For RR estimations, a different approach was used. The estimated value was

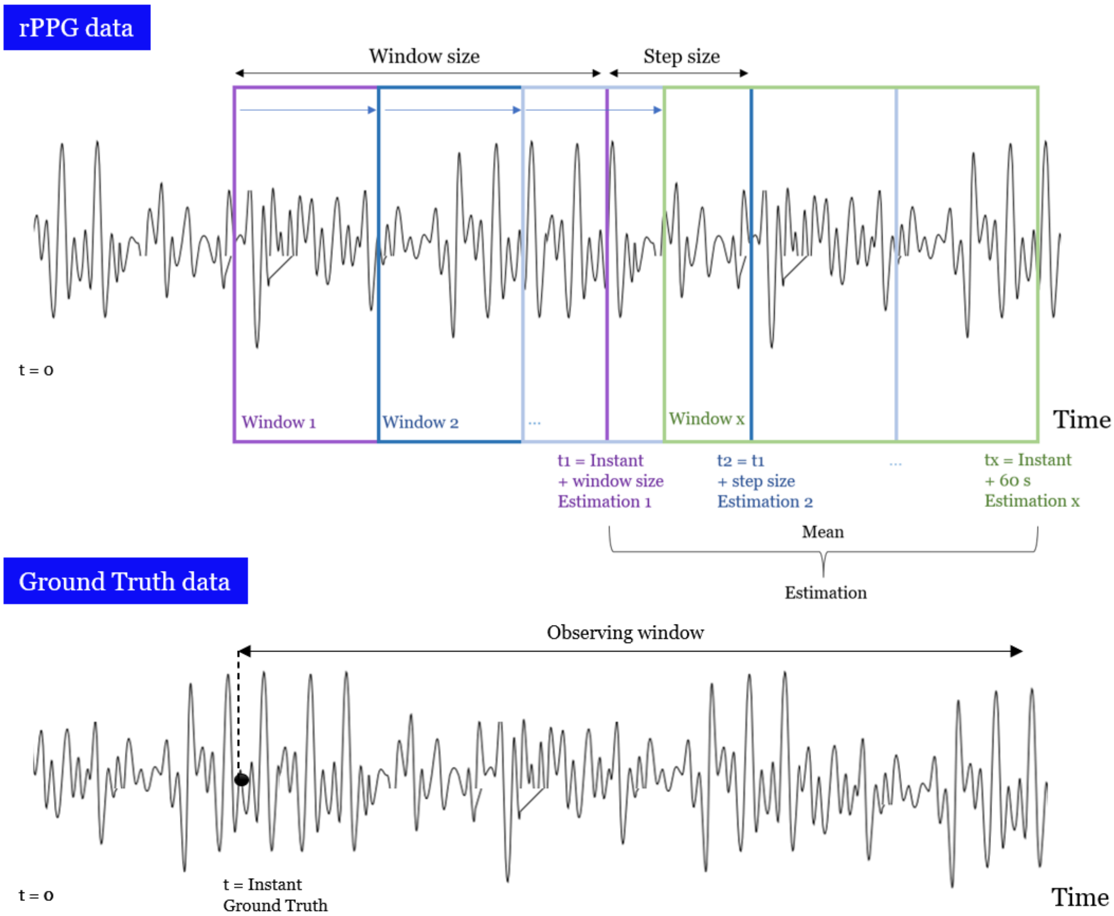


**Figure 3.11:** Moving window and step size illustration used to estimate RR values from rPPG data and ground truth value used for comparison.

computed as the mean value of several RR estimations. The estimations were made starting from the precise time instant at which the ground truth RR was provided and continued for a duration of up to 60 s. The windows were configured with window size and moved in intervals of step size until reaching the previously specified end.

The ground truth was obtained by manually counting breaths by personnel of the Neonatal Unit of Mauriziano Hospital in Turin starting from the recorded instant (as explained in Section 3.3). This means counting the number of times the newborn's chest or abdomen rises over one minute. In this way, the two values are comparable (estimated and ground truth) since the ground truth represents an average of the respiration frequency for an interval equivalent to the minute that follows it. Refer to Figure 3.11 for a visual representation illustrating the utilization of moving windows for RR estimation and the ground truth values used for comparison with these windows.

The reference values can be compared immediately with the estimated one to obtain performance evaluation metrics; see Figure 3.10 to understand this implementation.

## 3.6 Parameter selection and experiments

Before determining HR and RR for all of the dataset videos, some parameters had to be determined. To do this, experiments were carried out, which will be described in this Section.

Note that as a general rule, when carrying out an experiment, only the parameter under analysis is varied while all the other parameters are fixed. Then, at the end of the experiment, this parameter gets fixed at the value determined to yield the best performance based on the evaluation metrics. The G processing rPPG algorithm was used in all experiments because it is considered to be a traditional baseline method. Further, for HR estimation, experiments were performed using video 2 because this video is from a newborn recording whose face can be accurately detected and tracked using the Google Mediapipe library, there is no movement (it belongs to the “motionless” category), and there are reduced lighting variations, so the influence of other variables over the computed vital signal is minimized. Yet, for RR experiments also, a measurement from video 6 was considered, as this video possesses similar characteristics to video 2 and in order to have more than one value of ground truth for comparison.

Results obtained during the process of parameter settings are considered to be the “ideal” case in the NICU environment. As it was mentioned, recordings 2 and 6 correspond to “motionless” newborns; consequently, there is little or no error introduced by movements. Also, the Google Mediapipe face detection and tracking algorithm works without failures and lightning conditions remain relatively stable. Thus, the influence of external variables over the estimated vital signals is reduced. It is expected that the results of the other motion categories, “sporadic motion” and “motion”, will have worse evaluation metrics with respect to this case because they contain more sources of noise.

Finally, when the parameters get set, all the video recordings are tested with all the processing rPPG algorithms for comparison.

In this way, the optimal parameter combination was determined from experimentation, which was used as guidance to determine the parameter selection that yields higher performance. However, note that this is a high-complexity multivariate problem, so the best possible combination of parameters was selected. It is important to note that since changing a parameter can make all the evaluation metrics vary, it is not possible to determine if the chosen configuration is a global optimum (if it even exists). In any case, it corresponds to an optimal solution for the context of the problem.

### 3.6.1 ROI selection experiment

For HR estimation, an experiment is done to evaluate performance changes when different landmarks and region sizes are selected as ROI to understand how changes in these parameters impact the accuracy of HR estimation. Different skin coverage and landmark selections in the regions of the forehead and the cheeks are tested on video 2, and error metrics are used for determining which ROI selection is suitable for the context.

Regions of the forehead and the cheeks are selected because, as explained in Subsection 2.1.1, the face usually contains visible skin, which is required to get the rPPG signal. These specific regions have shown the strongest reliability due to their anatomical characteristics and higher BVP information. Figures 3.12, 3.13, 3.14, 3.15, 3.17 show the tested ROI landmark selections on a newborns face, where the specific landmarks used are shown as red dots.

Similarly, for RR estimation, an experiment is performed using different ROI selections with the same purpose as HR estimation. Different skin coverage and landmark selections in the region of the forehead are tested on video 2 and error metrics are used to determine which ROI selection is suitable for the context.

The region of the forehead is selected because, as explained in Subsection 2.1.1, the face usually contains visible skin, which is required to get the rPPG signal, and this specific region has shown strongest reliability for RR estimation containing higher BVP information. Figures 3.15, 3.16, 3.17 show the tested ROI landmark selections on a newborn's face, where the specific landmarks used are shown as red dots.

**Figure 3.12:** Multiple landmarks in the forehead and the cheeks.

**Figure 3.13:** Forehead and cheeks landmarks selected based on the anatomical regions shown in Figure 2.1.

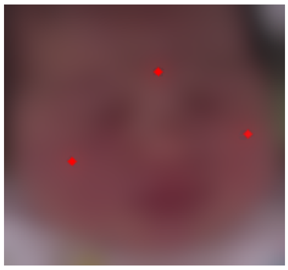

**Figure 3.14:** Forehead and cheeks with one representative landmark.

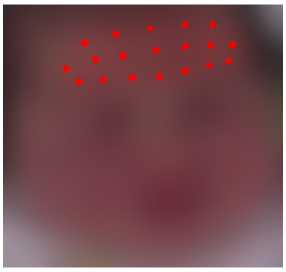

**Figure 3.15:** Multiple landmarks in the forehead.

**Figure 3.16:** Forehead landmarks selected based on the anatomical regions shown in Figure 2.1.

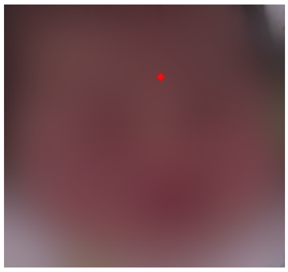

**Figure 3.17:** Forehead with one representative landmark.

### 3.6.2 Window size and step size experiment

For HR estimation an experiment is performed to evaluate performance changes when the window size and the step size parameters are modified. This is important because of the existing trade-off between latency and precision when modifying these parameters in order to find suitable values for them in the context of use. Thus, different combinations of window size and step size are evaluated in video 2.

According to [22], the 99% of HR values for newborns are in the interval 90 - 270 bpm, which corresponds to a frequency range of 1.5 - 4.5 Hz. However, [34] considers a wider bandwidth starting from 1.3 Hz. Both investigations, which were described in the state of the art Subsection 2.2, have a maximum frequency of 4.5 Hz and correspond to ranges used in NICU environment. The widest bandwidth is selected to ensure that the HR information is fully captured.

Equation 3.1, which relates period (T) and frequency (f), is used with the selected bandwidth. It is determined that the signal can be completely captured in a period of 0.77 s in the worst-case scenario.

$$T = \frac{1}{f} \tag{3.1}$$

Nevertheless, the pyVHR framework used does not allow much flexibility as it requires the step size to have a positive integer size greater than the window size. So, the window size cannot take the value of 1 s. Consequently, the smallest window investigated is of size 2 s.

In a similar way, for RR estimation, an experiment is performed using different windows and step sizes for the same purpose as for HR estimation.

RR values for healthy infants from 0 - 1 years old at rest are in the interval 25 - 68 cpm [33], as it was previously mentioned in Subsection 1.2.6, corresponding to a frequency bandwidth of 0.41 - 1.13 Hz. Yet, commonly used bandwidths for the estimation of this vital signal in NICU context are: 0.1 - 3 Hz, 0.5 - 1 Hz, and 0.2 - 2 Hz [27]. Also, ground truth values in the dataset are compressed between 35 cpm and 93 cpm. So, a frequency bandwidth of 0.5 - 2 Hz is considered.

Using Equation 3.1, we obtain that the signal can be completely read in a period of 2 s in the worst-case scenario, and so, the minimum possible tested window size is of the same value.

### 3.6.3 Sampling rate

The Shannon-Nyquist theorem states that a continuous signal can be accurately reconstructed from its samples if the sampling frequency is at least twice the highest frequency present in the signal [85]. In the case of a video recording analysis, the sampling frequency is restricted by the fps of the video.

As mentioned in the previous Subsection, the frequency bandwidth used for HR is 1.3 - 4.5 Hz as it was explained in the previous Subsection. Therefore, according to the Shannon-Nyquist theorem, the minimum sampling frequency required to fully characterize the signal is 9 fps. In compliance with the theorem, a 30 fps sampling frequency is used because recordings from the dataset have 30 fps, and no resource consumption limitation is considered.

Whilst, RR estimations considered a bandwidth of 0.5 - 2 Hz, as mentioned in the previous Subsection. According to the Shannon-Nyquist theorem, the minimum sampling frequency requirement is 4 fps for the RR. So, a 30 fps sampling rate is enough to discretize the breathing movements [15].

### 3.6.4 Butterworth filter range and order experiment

For HR estimation, a Butterworth filter of 1.3 - 4.5 Hz is used. Also, an experiment is performed to determine which order yields better performance in the context of use. As a reference, 3rd order and 5th order have been commonly used in literature [14], [24], [30].

For RR estimation a Butterworth filter of 0.5 - 2 Hz is used with the same order selected from the HR experiment previously conducted. This choice is justified by the fact that the respiratory peak is present in the spectrum of the rPPG waveform, similar to the cardiac peak, but with a lower amplitude [14].

Bandwidth range selections for the vital signals were explained in Subsection 3.6.2.

### 3.6.5 Algorithm for vital signal estimation experiment

Both STFT and Welch PSD algorithms have advantages and disadvantages that were previously discussed in Subsection 2.1.4. So, it is not trivial to determine in advance which algorithm to use. Consequently, for HR estimation, both algorithms are compared using evaluation metrics, and the best performing algorithm is selected.

For RR estimation, the same algorithm for vital signal estimation that was selected from the previous HR experiment is used.

# Chapter 4

# Results

## 4.1 Parameter selection and experiments results

In this Section, the results of the parameter selection experiments described in Section 3.6 of Chapter 3 are shown correspondingly. First, all of the results relative to the HR estimations are shown, followed by the results obtained for the RR estimations.

### 4.1.1 ROI selection experiment results

Table 4.1 shows the different landmarks and patch sizes combinations used in this experiment for HR estimation and the metrics obtained as a result. The landmark column makes reference to the figure showing the landmarks used, and size is the length in pixels of the side of the squared patches over each landmark. The best 5 results for each metric are shown in bold, except for the PCC because this experiment derived no results with a moderate or strong correlation. For this reason, this metric was not considered in the performance analysis. Additionally, note that only the significant part of the table is shown, while the rest of the table can be found in Appendix B.

The experiment considered the following patch sizes (measured in quantity of pixels): 25, 100, 225, 625, 900, and 1225. The impact of larger patches was not explored because results suggested a performance degradation with the increment of the patch size.

The best results were obtained from the landmarks corresponding to Figure 3.12, concentrating the majority of the bold evaluation metrics. These 45 landmarks correspond to the following numbers from Google Mediapipe face mesh: 9, 10, 36,

| Landmarks | Size (px) | RMSE | MAE | MAX | PCC |
|---|---|---|---|---|---|
| 3.12 | 5 | **6.73** | **5.40** | 21.68 | 0.170 |
| 3.12 | 10 | **8.35** | **6.75** | **18.41** | -0.014 |
| 3.12 | 15 | **8.64** | **7.18** | **18.96** | 0.047 |
| 3.12 | 20 | **8.61** | **6.90** | **18.08** | 0.124 |
| 3.12 | 25 | 8.81 | **6.94** | **19.84** | 0.134 |
| 3.12 | 30 | 10.26 | 8.16 | 22.80 | -0.087 |
| 3.12 | 35 | 11.16 | 8.44 | 36.87 | -0.083 |
| 3.15 | 5 | **8.44** | 6.71 | 28.59 | 0.267 |
| 3.15 | 10 | 9.19 | 7.74 | 21.96 | -0.202 |
| 3.15 | 15 | 9.30 | 7.90 | **20.20** | -0.052 |
| 3.15 | 20 | 10.51 | 8.63 | 34.11 | -0.025 |
| 3.15 | 25 | 10.51 | 8.63 | 34.11 | -0.025 |
| 3.15 | 30 | 10.29 | 8.19 | 33.23 | -0.063 |
| 3.15 | 35 | 10.31 | 8.62 | 24.96 | -0.022 |

**Table 4.1:** ROI selection experiment metrics obtained for different landmarks and patch sizes combinations. The landmark column makes reference to the figure showing the landmarks used, and size is the number of pixels on the side of the square patches over each landmark. The best 5 results for each metric are shown in bold, except for the PCC as there is no moderate or strong correlation. The selected configuration of parameters is highlighted in gray.

50, 66, 67, 69, 101, 104, 105, 107, 108, 109, 116, 117, 118, 119, 123, 147, 151, 187, 205, 206, 207, 266, 280, 296, 297, 299, 330, 333, 334, 336, 337, 338, 345, 346, 347, 348, 352, 376, 411, 425, 426, 427. Refer to Appendix A for a visualization of the numbered landmarks in the canonical face model image.

As mentioned in Subsection 2.1.1, HR can be extracted from the larger ROIs, so squares of side size 20 px were selected because it is the biggest patch with best RMSE, MAE and MAX values. The corresponding selected configuration of parameters is highlighted in gray on Table 4.1.

With respect to RR estimations, Figure 3.16 landmark selection showed RR estimations closer to the ground truth values for all estimated instants. The corresponding error metrics were RMSE of 7.99, MAE of 7.95, and MAX of 8.73. Thus, landmark number 9 of Google Mediapipe with a patch size of 15 px proved to be effective.

### 4.1.2 Window size and step size experiment results

Table 4.2 shows the selected window size and step size combinations measured in seconds and the resulting evaluation metrics for HR estimation. The best 5 results for each metric are shown in bold.

From the results Table it is possible to observe that a bigger step size leads to a minimal improvement in the evaluation metrics or no improvement at all. Consequently, the step size of 1 s is more convenient since it provides more frequent vital sign updates, and so this value is chosen for the step size parameter.

Subsequently, it can be observed that the best performance error metrics for RMSE and MAE are found around a window of size 10 - 15 s. While the best PCC is found in larger windows of size between 30 - 60 s, where PCC has a moderate correlation.

To better understand the effects of these metrics on the behavior of the signal, the 12 s window and the 28 s window were graphed and are showed in Figures 4.1 and 4.2 respectively. The green line represents the results obtained using the previously specified parameters configuration and the processing rPPG algorithm G, and the red line is the ground truth.

From these Figures, it is possible to observe that the 12 s window has values closer to the ground truth but is not an accurate representation of its trend. Whilst the 28 s window has underestimated HR values using the processing G method, these estimations seem to follow the ground truth trend more closely than those made with a 12 s window. The situation was discussed with the Neonatal Unit of Mauriziano Hospital in Turin, which stated that it is more important for the healthcare staff to have a representation of the vital signals' trend than its exact measurement. Consequently, the 28 s window was selected for the window size parameter.

The corresponding selected configuration of parameters is highlighted in gray on Table 4.2.

| Window size (s) | Step size (s) | RMSE | MAE | MAX | PCC |
|---|---|---|---|---|---|
| 2 | 1 | 18.13 | 14.95 | 48.93 | -0.148 |
| 4 | 1 | 12.41 | 10.22 | 36.26 | -0.096 |
| 6 | 1 | 9.63 | 7.85 | 28.35 | 0.037 |
| 8 | 1 | 8.82 | 7.16 | 19.08 | 0.118 |
| 10 | 1 | **8.61** | **6.90** | **18.08** | 0.124 |
| 12 | 1 | **8.54** | **6.54** | **18.35** | 0.097 |
| 15 | 1 | 9.12 | **6.97** | 20.35 | 0.053 |
| 20 | 1 | 10.14 | 7.75 | 21.23 | 0.196 |
| 25 | 1 | 10.89 | 9.18 | 21.11 | 0.196 |
| 28 | 1 | 11.61 | 10.37 | **18.59** | **0.567** |
| 30 | 1 | 12.07 | 10.85 | **18.59** | **0.544** |
| 60 | 1 | 13.86 | 13.53 | 24.84 | **0.455** |
| 2 | 2 | 18.04 | 15.27 | 47.29 | -0.227 |
| 4 | 2 | 12.74 | 10.34 | 36.26 | -0.152 |
| 6 | 2 | 9.34 | 7.85 | 19.96 | -0.056 |
| 8 | 2 | **8.71** | 7.09 | 19.08 | 0.115 |
| 10 | 2 | **8.51** | **6.66** | **18.08** | 0.099 |
| 12 | 2 | **8.38** | **6.20** | **18.35** | 0.123 |
| 15 | 2 | 9.30 | 7.08 | 20.23 | -0.021 |
| 30 | 2 | 11.85 | 10.68 | **18.59** | **0.564** |
| 60 | 2 | 13.84 | 13.64 | 19.34 | **0.505** |

**Table 4.2:** Window size and step size (both measured in seconds) experiment results with evaluation metrics. The best 5 results for each metric are shown in bold. The selected configuration of parameters is highlighted in gray.

With respect to RR estimation, a window size of 16 s and a step size of 2 s showed RR estimations closer to the ground truth values for all estimated instants. The corresponding error metrics were: RMSE of 6.92, MAE of 6.92 and MAX of 6.92.

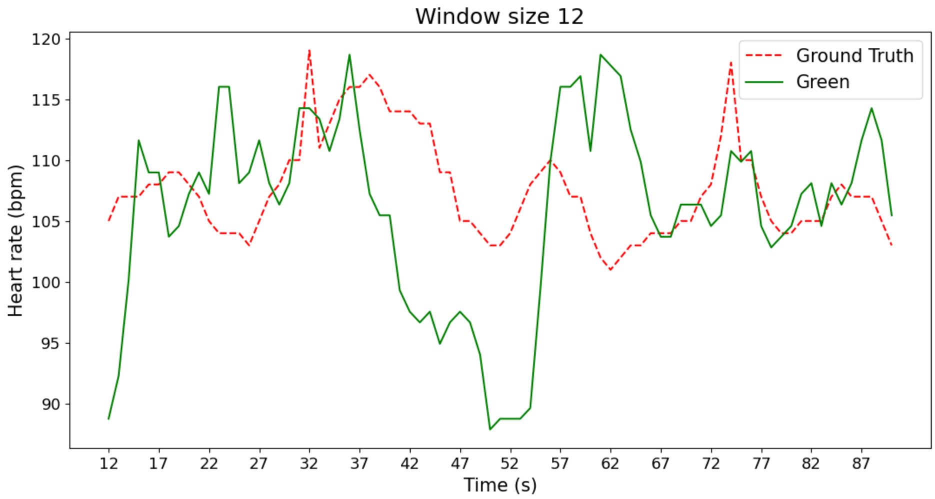


**Figure 4.1:** Visualization of HR values using a 12 s window size and 1 s step size. The green line shows the results obtained with the G rPPG processing algorithm, and the ground truth is shown in red.

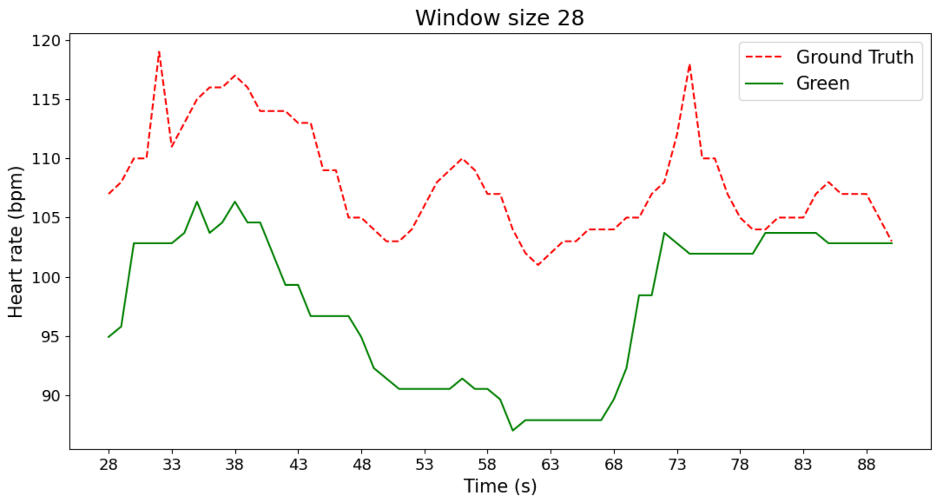


**Figure 4.2:** Visualization of HR values using a 28 s window size and 1 s step size. The green line shows the results obtained with the G rPPG processing algorithm, and the ground truth is shown in red.

### 4.1.3 Butterworth filter order experiment results

For HR estimation Table 4.3 shows the results of the Butterworth filter order experiment. It can be observed that the best results are obtained by the 3rd and 4th-order Butterworth filter.

| Order | RMSE | MAE | MAX | PCC |
|---|---|---|---|---|
| 1 | 13.82 | 11.62 | 28.59 | 0.321 |
| 2 | 8.20 | 6.59 | **17.41** | 0.414 |
| 3 | **8.15** | **6.56** | 17.71 | 0.552 |
| 4 | 9.51 | 8.11 | 17.71 | **0.583** |
| 5 | 10.76 | 9.43 | 18.59 | 0.571 |
| 6 | 11.61 | 10.37 | 18.59 | 0.567 |

**Table 4.3:** Butterworth filter order experiment results with evaluation metrics, where the best result for each metric is shown in bold.

As mentioned in Subsection 3.6.4, the 3rd order is commonly used in literature. Furthermore, the results obtained using the 3rd order filter demonstrate a moderate correlation in the PCC, which is consistent with the best PCC obtained. Consequently, this filter order is established as the configuration for this parameter.

The corresponding parameter selection configuration is highlighted in gray on Table 4.3.

### 4.1.4 Algorithm for vital signal estimation experiment results

Table 4.4 shows the results of the experiment for HR estimation, where the best result for each metric are shown in bold.

Welch PSD algorithm for vital signal estimation performs better than STFT in all of the evaluated error metrics. Consequently, this algorithm is set as a parameter.

The corresponding parameter selection configuration is highlighted in gray on Table 4.4.

| Algorithm | RMSE | MAE | MAX | PCC |
|---|---|---|---|---|
| Welch PSD | **8.15** | **6.56** | **17.71** | **0.552** |
| STFT | 13.72 | 11.10 | 32.75 | 0.292 |

**Table 4.4:** Algorithms for vital signal estimation experiment results with evaluation metrics, where the best result for each metric is shown in bold.

## 4.2 Processing rPPG algorithms comparison

## 4.3 Single video recording example

Values of parameters defined in Section 4.1 were fixed, and a comparison of all of the rPPG processing algorithms for HR estimation was performed. Table 4.5 shows the error metrics obtained from the results, where the best result for each metric is shown in bold. Also, considering the best processing algorithms obtained from the previously mentioned metrics displayed in the Table, Figure 4.3 visually compares these algorithms with the ground truth throughout the duration of the video.

| Algorithm | RMSE | MAE | MAX | PCC |
|---|---|---|---|---|
| G | 8.15 | 6.56 | 17.71 | **0.552** |
| ICA | **4.71** | **3.75** | 13.53 | 0.325 |
| LGI | 5.64 | 4.66 | 17.93 | -0.008 |
| PBV | 5.61 | 4.71 | 14.41 | -0.056 |
| PCA | 6.03 | 4.75 | 14.41 | -0.447 |
| CHROM | 6.19 | 5.25 | 14.41 | -0.220 |
| POS | 5.26 | 4.53 | **10.77** | 0.109 |

**Table 4.5:** Evaluation metrics obtained for estimations with different processing rPPG algorithms, where the best result for each metric is shown in bold.

Note that the grey area in the Figure corresponds to a period where an external entity produces a shadow on the ROI. The impact of this area affects the estimation of the vital sign from the moment it occurs up to a 28 s window. This is due to the fact that the rPPG data collected during this period (containing the shadow information) will be used by the following moving windows. This may explain the

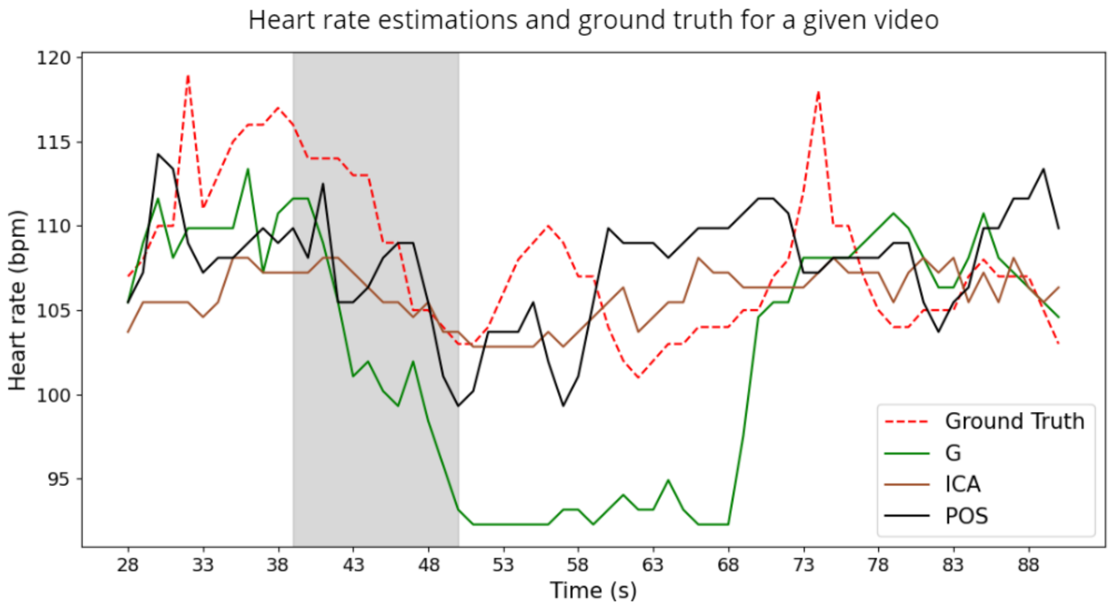


**Figure 4.3:** Graphic comparison of results obtained using ICA, G and POS processing rPPG algorithms and the ground truth, where the grey area corresponds to a period in which an external entity produces a shadow on the ROI.

abrupt drop in the estimations computed with the G processing algorithm after the occurrence of the shadow.

In this example, G and ICA processing algorithms, which have moderate PCC, seem to produce a more representative estimation compared to the ground truth's trend. As mentioned in Subsection 4.1.2 and in Subsection 2.1.3, this characteristic is of the utmost importance.

## 4.4 Aggregated results

Following, the aforementioned algorithms were compared at an aggregated level on the entire dataset. Figure 4.4 displays the aggregate results as an average of the RMSE, MAE and MAX error metrics obtained under 3 motion categories: "motionless", "sporadic motion" and "motion" (refer to Table 3.1 for dataset characterization). The PCC error metric was not considered in the comparison because there were no averages corresponding to a moderate or strong correlation.

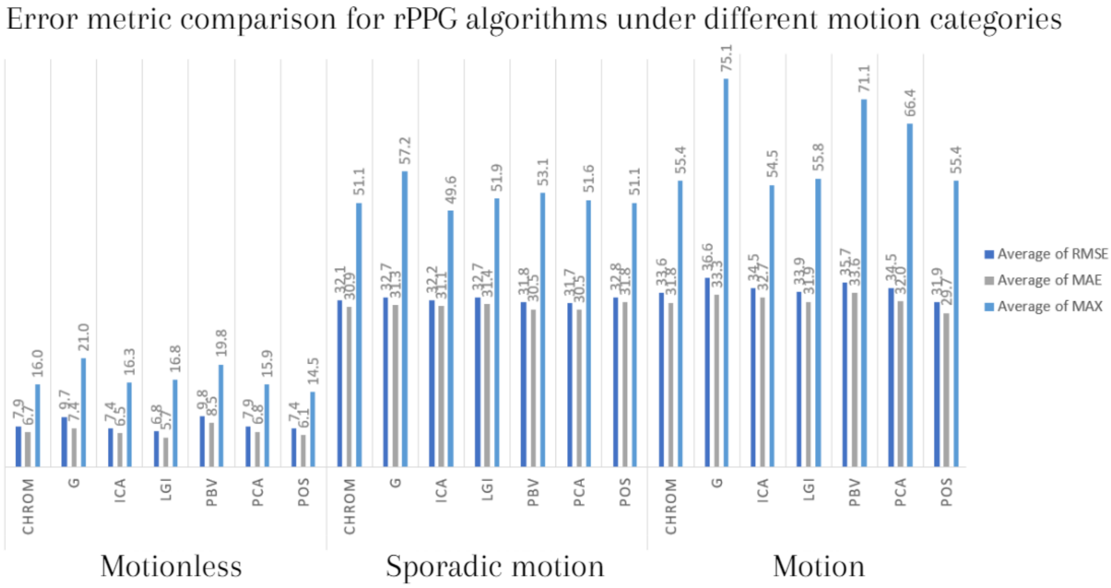


**Figure 4.4:** Comparison graph for RMSE, MAE and MAX average for all dataset HR results obtained with different rPPG processing algorithms distinguished by motion category.

| Motion Category | Best-Performing Algorithms | Error Metrics | | |
|---|---|---|---|---|
| | | RMSE | MAE | MAX |
| Motionless | POS | 7.4 | 6.1 | **14.5** |
| Motionless | LGI | **6.8** | **5.7** | 16.8 |
| Sporadic motion | ICA | 32.2 | 31.1 | **49.6** |
| Sporadic motion | PCA | **31.7** | **30.5** | 51.6 |
| Motion | ICA | 34.5 | 32.7 | **54.4** |
| Motion | POS | **31.9** | **29.7** | 55.4 |

**Table 4.6:** Best performing algorithms for HR estimation in each movement category and the corresponding error metrics, where the best result for each metric is shown in bold.

Table 4.6 summarizes the best-performing algorithms for the different motion categories based on the rPPG processing algorithm that obtained the best error metrics for each motion category. All of the motion categories have two best-performing rPPG processing algorithms because some error metrics were smaller in one of them (for example, MAX), while other error metrics were lower in the others (RMSE and MAE).

Based on the best-performing algorithms POS and ICA are the best processing rPPG methodologies overall because they present the best results transversely through the different motion categories. On the contrary, G and PBV tend to prove to obtain higher error values compared to the other methodologies in all motion categories.

It can also be noted that there are minimal differences in the evaluation metrics obtained by comparing the category “sporadic motion” with the category “motion”. However, there are significant differences when comparing both categories with the “motionless” category.

With respect to the RR results, Table 4.7 shows the error metrics obtained for RR estimations of all of the dataset instants for which there was a ground truth value available using the different rPPG processing algorithms. The best result for each metric is shown in bold.

Note that the results are shown in an aggregate manner and not separated by category due to the reduced sample size of the reference values. For the same reason, the PCC metric was not considered, as it was mentioned in Subsection 3.4.

| Algorithm | RMSE | MAE | MAX |
|---|---|---|---|
| G | 19.73 | 14.97 | 48.67 |
| ICA | 19.21 | 13.48 | 53.53 |
| PCA | **17.99** | 13.70 | 51.18 |
| CHROM | 18.48 | **13.15** | **44.13** |
| POS | 20.24 | 14.58 | 46.95 |
| PBV | 18.98 | 14.25 | 51.46 |
| LGI | 19.76 | 14.23 | 50.12 |

**Table 4.7:** RR estimation error metrics results using different processing rPPG methods for video 2 and 6, where the best result for each metric is shown in bold.

The rPPG processing methodologies of CHROM and PCA show the best error metric results. However, there were no significant differences between the different methodologies used, all of them on average performed similarly.

# Chapter 5

# Discussion and conclusions

## 5.1 Importance of the study and future research directions

Although there is literature investigation to determine HR and RR vital signals using different rPPG processing algorithms (described in Chapter 2), there is no consensus about a benchmark criterion to be used for this purpose. Also, most of the results are obtained from adult subjects with specific lighting and motion conditions and are not suitable for the context of a neonatal unit. Therefore, a private dataset was used to test the traditional methods that have shown good performance in adult studies and possess important characteristics for the context of use, such as motion and lightning robustness. Additionally, until now, these algorithms have not been compared in the NICU context as it was seen in Subsection 2.2 referencing the state of the art.

The pyVHR framework was adapted for the context of use because it has an in-built implementation of these commonly used algorithms. Another reason to use this framework is that it accelerates execution by exploiting CUDA NVIDIA GPU parallelism. Hence it is suitable for real-time processing, which is a transcendental characteristic of this study. The trade-off is that the NVIDIA GPU is a hardware requirement.

Adaptations were made to the framework to make it suitable for the NICU context. For example, the provision of vital signs estimations was moved from the center of a moving window to its end to resemble its real-life use. Also, a processing step was incorporated in the pipeline used to discard the provision of a vital signal estimation from a moving window with insufficient rPPG to make a reliable computation. This insufficiency in the collected data is due to the fact that the

Google Mediapipe algorithm can fail to detect and track the selected ROI (further details were described in Section 3.5).

Results obtained for the HR estimation were classified under three different motion categories (“motionless”, “sporadic motion”, and “motion”) to have a more accurate understanding of the performance of the algorithms. These aggregated results are shown in Figure 4.4, and the best-performing algorithms for each category and the corresponding error metrics are displayed in Table 4.6. According to the results, the algorithms POS and ICA performed consistently well across the different motion categories, indicating their suitability for vital sign estimation in the given context. For future research, a hybrid between the methodologies POS and ICA is to be explored, aiming to improve performance (similarly as it has been explored by [55] between ICA and CHROM or PBV).

On the contrary, G and PBV tend to prove to obtain higher error values compared to the other rPPG processing algorithms in all motion categories.

The motionless category achieved an aggregated MAE within the clinically acceptable range, which was specified in Subsection 3.4 as an error between 6 bpm and 10 bpm. Therefore, monitoring during rest or sleep is viable.

Another possible improvement for the HR estimations would be to add a step at the end of the processing pipeline (refer to Figure 3.10) to compare the estimated value with historical ones. As observed in the ground truth obtained by the pulse oximeter in the dataset, the maximum change from one frame to another was 11 bpm. Therefore, under the same logic, changes from one HR estimation to the following should avoid abrupt changes in consecutive estimated values.

Concerning RR estimation results, the best-performing algorithms, and the corresponding error metrics are displayed in Table 4.7. There was no significant difference between the methodologies used, which obtained good performance on average. Despite the small sample size of ground truth values to obtain statistically significant results, this part of the work demonstrates the feasibility of the approach and opens the doors for future experiments.

It is relevant to mention that the MAX error metric is given by the estimation of the RR of one specific video, which in particular has a higher ground truth compared to the others. It is left for future research to determine if this value is an outlier or if the algorithm has a bias with respect to high RR values.

A possible line of research is in the field of Deep Learning. As it was seen in Chapter

2, several studies have shown good performance results applying this technology. However, in order to use it, a large dataset is needed, which this study did not have.

As mentioned in Subsection 2.1.1, the performance of the different algorithms for obtaining vital signs depends on several factors, including the error metrics used. Therefore, it is important to use different model evaluation metrics in order to understand its behavior, as done in this work.

Part of the error obtained by the evaluation metrics could be explained due to the fact that cyclical movement of blood from the heart to head considerably deteriorates when babies are lying down [19], as they were in the dataset used. Also, as stated by [14] "small proportion of the cardiac-synchronous signal is due to the motion of face landmarks in time with the heart beat".

## 5.2 Limitations and opportunities

The study acknowledges certain limitations, however, these limitations provide opportunities for future exploration and improvements in the field.

This study is limited by the number of infant subjects in the 11 video recordings. Furthermore, the dataset used has a sex bias because it includes only 1 video recording with a female subject and a skin color bias, as all newborns have a Fitzpatrick skin type I (refer to Table 3.1 for dataset specifications). For future research, it is suggested to have a more representative sample of sex and skin color to avoid systematic discrimination of these categories of newborns.

Also, this study is limited by the difficulty of the pulse oximeter instrument to record the HR ground truth accurately. As mentioned in [60], "it can't be expected to obtain clear signals from sensors. This is a major drawback and makes the analysis of processes difficult, enforcing several constraints for real applications". Some of these inaccuracies are due to motion artifacts, which affect most of the traditional instruments for vital signal measurement [22]. Furthermore, values of the ground truth were not obtained directly by the measurement instrument, they were obtained indirectly through character recognition of the instrument values in the video recordings. Although pre-processing and post-processing were performed to reduce the error in obtaining these results, the intermediate step could also be avoided to ensure the accurate obtaining of the ground truth values.

The RR ground truth values were manually obtained. Despite the fact that they were obtained by clinical staff, a measurement instrument could be used instead to increase objectivity and the number of measurements collected from the dataset videos.

Another limitation of the results of this study is that it assumes that the chosen ROI for vital signal calculation has visible skin to be recorded by the camera. However, the dynamism of the NICU environment does not always make it possible to meet this requirement. It is not uncommon that medical personnel passes their hands or implements over the selected ROI. Consequently, interference is generated in the estimation of vital signs. This is due to the fact that the ROI zones are blocked from camera registration; therefore, this produces noise in the estimated vital signals. This aspect is to be considered in future improvements of the algorithm. In addition, this circumstance sometimes results in the failure of the Google Mediapipe algorithm to detect and track the ROI.

Additionally, the use of a pacifier is common in the context of the NICU because it provides advantages for the newborn, as mentioned by [86] it "helps transition from tube to oral feeding, breastfeeding, faster weight gain and earlier discharge from the NICU". Some of the video recordings from the dataset evidenced the use of a pacifier. For the purpose of vital signal estimation, the pacifier can produce the same blockage effect previously described either by itself because it can partially cover the ROI or because the healthcare staff blocks the ROI when putting the pacifier on the newborn's mouth. Moreover, this can hinder Google Mediapipe ROI detection and tracking algorithm because this algorithm was not trained on newborns using a pacifier, which can decrease the precision of the detected landmarks or cause its failure. The use of an algorithm specialized in the recognition and tracking of newborns could be useful to increase accuracy, specifically an algorithm developed with face recognition, even if there is the presence of a pacifier in the mouth of the infant.

It is important to consider that the Google Mediapipe face mesh algorithm used in this study to find the ROI in the video recordings of the dataset finds a 2-D figure in a 3-D world. This approximation may introduce detection and tracking errors. Consequently, using two RGB cameras or a depth camera is suggested to help improve ROI detection and tracking results by getting a more precise understanding of the third dimension.

Moreover, note that in this study, intervals of failure of the Google Mediapipe face mesh algorithm were taken into account in the processing pipeline used (refer to Figure 3.10 for the details of this pipeline), meaning periods of no detection and tracking of the newborn's ROI. However, periods in which the algorithm fails to localize and track the landmarks correctly are harder to discriminate; this phenomenon is called ROI drift. In this case, the collected rPPG data does not represent that of the selected ROI. This issue could be due to newborn movement or to an obstruction of the ROI. A manual analysis could've been made to determine the accuracy of the detected face mesh, but it would be impractical. The idea of this study is to generate an algorithm capable of monitoring vital signs in a continuous and autonomous way without the need for supervision. In line with this objective, the Google Mediapipe face mesh algorithm was selected for this study's use because it is one of the best facial detection and tracking algorithms in state of the art. Yet considering the existence of ROI drifts, which were evidenced within the dataset of this investigation, and to improve results further, an algorithm specialized in newborns could be used or trained (as it was previously suggested). These points and the previously mentioned statements about Google Mediapipe highlight the criticality of having an accurate ROI detection and tracking algorithm.

As previously mentioned in Section 3.6, the parameters selected through experimentation used for posterior vital signal estimations (ROI, window size and step size, Butterworth filter order and algorithm for vital signal estimation) correspond to a local optimum. Determining a global optimum is a high-complexity multivariate problem with no solution guarantee. Consequently, robust experimentation for determining a possible global optimum solution is a subject of future investigation.

The selected window size for the HR vital signal estimation creates an in-built latency of 28 s. This means that initially, there will be no estimates until this period has elapsed. This has to be considered when "turning on" the device because HR estimations will not be provided immediately but after a few seconds. This can be important in the NICU context, as mentioned by [14] "in some clinical scenarios, for example, the detection of apnoea, the effect of this in-built delay will need to be taken into account". In addition, situations that can alter the HR estimation by inducing error, such as shadows over the selected ROI, will affect the vital signal estimation during an equivalent window size from the moment of occurrence.

## 5.3 Advantages and applications

An advantage over most of state of the art investigations mentioned in Subsection 2.2 is that all patients in the dataset were reported. No exclusions were made due to movement artifacts from the newborn or the healthcare staff when performing a blood withdrawal procedure, nor when there were changes in illumination due to shadows produced by external entities or movements of the newborn, nor ROI blockage. Therefore, the obtained results closely resemble the NICU context of the Neonatal Unit of AO Ordine Mauriziano Hospital in Turin. The only video recordings exclusions and frame cuts on the dataset were due to: i) no availability of ground truth values. Therefore, no error metrics could be obtained from them. ii) Duplicated videos in the dataset, which would have introduced redundancy. iii) Videos with a shorter duration than the selected window size to process the rPPG signal, which were not possible to evaluate with a 28 s parameter configuration for HR estimations (the 28 s window size was determined based on experimentation). iv) If there was not enough rPPG data collected inside a moving window to support a valid vital signal measurement estimation, meaning that the amount of rPPG data in that specific window is smaller than the minimum window size mentioned in Subsection 3.6.2. Consequently, based on signal theory, it is not possible to obtain a vital signal estimation from it, and so the processing pipeline outputs no vital signal estimation from this moving window (refer to Figure 3.10 for a visual explanation).

An advantage of estimating vital signs using a RGB camera compared to traditional methods, is that this information can also be integrated with the knowledge it provides of the newborns' behavior. Therefore, apart from vital signs monitoring, the patient's morphology can be understood. In this way, by combining both, this technology has the potential to automatically and objectively obtain the pain assessment of newborns, which is the aim of this study. Moreover, this technology can potentially evaluate multiple subjects simultaneously using only one camera.

In addition, there are several other applications where this technology could be useful, as mentioned in the Subsection 1.2.7, this technology can help discriminate "deepfakes". It could also be used in telemedicine, infant monitoring, and as mentioned in [30] "it has significant advantages in some unconstrained scenarios long-term epilepsy monitoring".

This technology could be critical for newborn monitoring. According to the World Health Organization (WHO), "the majority of all neonatal deaths (75%) occur during the first week of life" [87]. It is expected to continue with the course of development of this technology with the hope that in the near future, it can contribute to reducing this quantity significantly by signaling when professional help is needed based on psychological vital signals.

# Appendix A

# Face model landmark number visualization

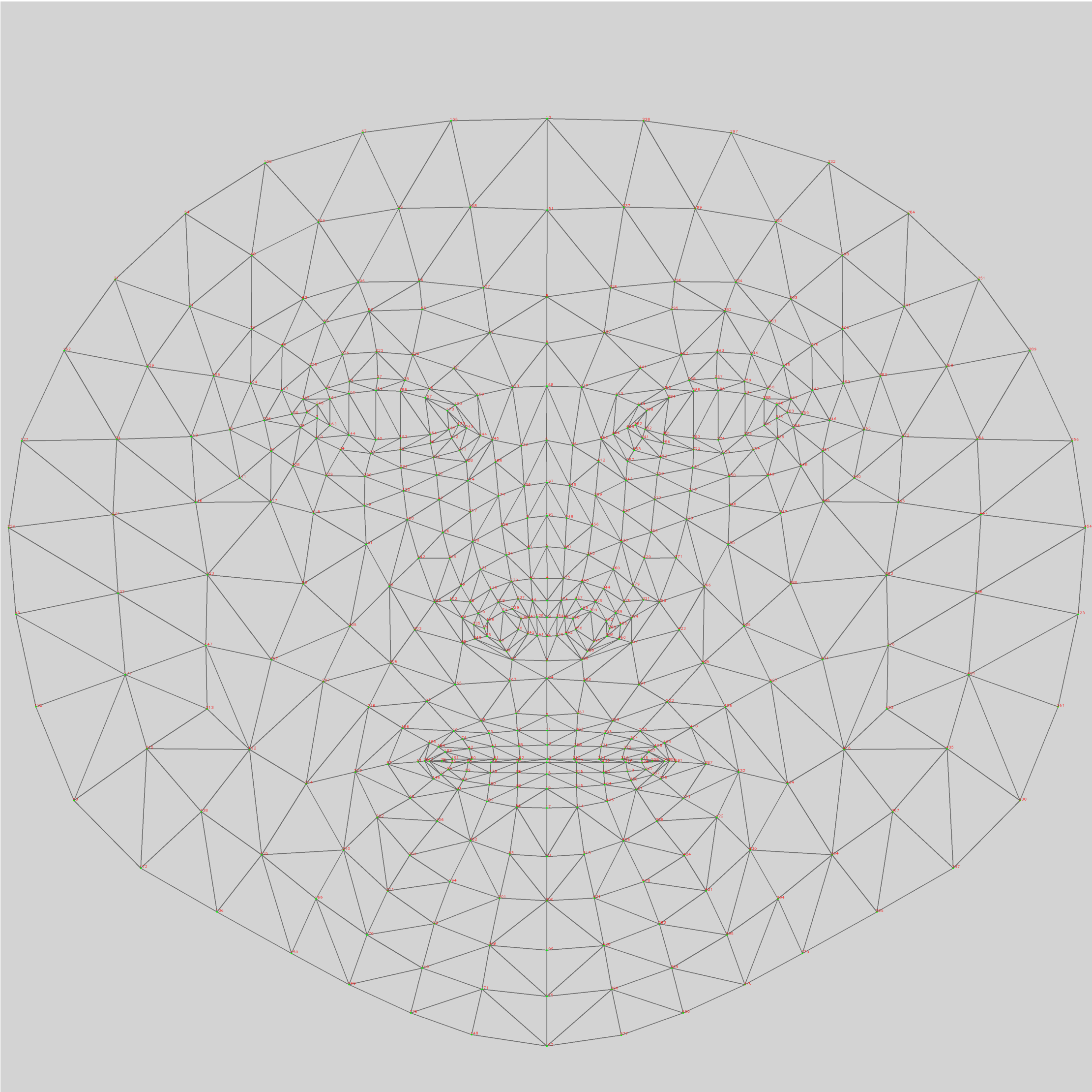

**Figure A.1:** Canonical face model image for landmark number visualization. This image is meant to be seen in the digital version of this thesis in order to zoom in to identify the corresponding numbers of landmarks in the face mesh. However, if there are problems with the display of the numbers, refer to the source file directly. Source: [88].

# Appendix B

# ROI selection experiment result table extension

| Landmarks | Size (px) | RMSE | MAE | MAX | PCC |
|---|---|---|---|---|---|
| 3.13 | 5 | 9.46 | 7.55 | 26.08 | -0.077 |
| 3.13 | 10 | 10.63 | 8.72 | 23.44 | -0.185 |
| 3.13 | 15 | 11.57 | 9.92 | 22.47 | -0.116 |
| 3.13 | 20 | 10.61 | 9.15 | 20.71 | -0.027 |
| 3.13 | 25 | 10.75 | 9.17 | 21.59 | 0.037 |
| 3.13 | 30 | 11.53 | 9.92 | 24.32 | -0.106 |
| 3.13 | 35 | 12.84 | 10.87 | 35.99 | -0.199 |
| 3.14 | 5 | 16.43 | 13.17 | 61.39 | 0.119 |
| 3.14 | 10 | 17.33 | 13.78 | 54.72 | -0.177 |
| 3.14 | 15 | 17.15 | 13.67 | 45.81 | -0.151 |
| 3.14 | 20 | 18.23 | 13.50 | 63.14 | 0.083 |
| 3.14 | 25 | 18.23 | 13.50 | 63.14 | 0.083 |
| 3.14 | 30 | 20.54 | 15.67 | 54.08 | 0.068 |
| 3.14 | 35 | 21.22 | 16.32 | 53.84 | 0.055 |
| 3.17 | 5 | 36.56 | 23.85 | 120.00 | 0.148 |
| 3.17 | 10 | 44.20 | 27.82 | 143.97 | -0.086 |
| 3.17 | 15 | 35.28 | 24.48 | 137.21 | -0.045 |
| 3.17 | 20 | 33.13 | 21.14 | 143.97 | 0.083 |
| 3.17 | 25 | 33.13 | 21.14 | 143.97 | 0.083 |
| 3.17 | 30 | 26.81 | 18.81 | 86.84 | 0.076 |
| 3.17 | 35 | 26.65 | 19.31 | 84.84 | 0.143 |

**Table B.1:** Table 4.1 extension.

# Bibliography


[1] J.A. Lemons et al. «Prevention and management of pain and stress in the neonate». In: *Pediatrics* 105.2 (2000), pp. 454–461 (cit. on p. 1).

[2] L. A. M. Aarts et al. «Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—A pilot study». In: *Early human development* 89.12 (2013), pp. 943–948 (cit. on pp. 1, 3–6, 19).

[3] Q. Chen, X. Jiang, X. Liu, C. Lu, L. Wang, and W. Chen. «Non-contact heart rate monitoring in Neonatal intensive care unit using RGB camera». In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2020, pp. 5822–5825 (cit. on pp. 1, 3, 7, 20).

[4] S. Cevoli and P. Cortelli. «Italian Law "measures to guarantee the access to palliative and pain treatments": rebound on headaches' management». In: *Neurological Sciences* 32.1 (2011), pp. 77–79 (cit. on p. 1).

[5] G.M. Melo, A. L. P. Lélis, A. F. Moura, M. V. L. M. L. Cardoso, and V. M. Silva. «Escalas de avaliação de dor em recém-nascidos: revisão integrativa». In: *Revista Paulista de Pediatria* 32 (2014), pp. 395–402 (cit. on p. 1).

[6] M. van der Vaart et al. «Premature infants display discriminable behavioural, physiological and brain responses to noxious and non-noxious stimuli». In: *medRxiv* (2021) (cit. on p. 1).

[7] N. J. Meesters, S. H. P. Simons, J. van Rosmalen, L. Holsti, I. K. M. Reiss, and M. Van Dijk. «Acute pain assessment in prematurely born infants below 29 weeks: a long way to go». In: *The Clinical Journal of Pain* 35.12 (2019), pp. 975–982 (cit. on p. 1).

[8] J. A. Waxman et al. «An examination of the reciprocal and concurrent relations between behavioral and cardiac indicators of acute pain in toddlerhood». In: *Pain* 161.7 (2020), pp. 1518–1531 (cit. on p. 1).

[9] E. Parodi, D. Melis, L. Boulard, M. Gavelli, and E. Baccaglini. «Automated newborn pain assessment framework using computer vision techniques». In: *Proceedings of the International Conference on Bioinformatics Research and Applications 2017*. 2017, pp. 31–36 (cit. on pp. 2, 10, 12).

[10] N. J. Meesters, S. H. P. Simons, J. van Rosmalen, L. Holsti, I. K. M. Reiss, and M. Van Dijk. «Acute pain assessment in prematurely born infants below 29 weeks: a long way to go». In: *The Clinical Journal of Pain* 35.12 (2019), pp. 975–982 (cit. on p. 2).

[11] S. Beres and L. Hejjel. «The minimal sampling frequency of the photoplethysmogram for accurate pulse rate variability parameters in healthy volunteers». In: *Biomedical Signal Processing and Control* 68 (2021), p. 102589 (cit. on p. 3).

[12] A. B. Hertzmann. «Observations on the finger volume pulse recorded photoelectrically». In: *Am J Physiol* 119 (1937), pp. 334–335 (cit. on p. 3).

[13] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran. «A review on wearable photoplethysmography sensors and their potential future applications in health care». In: *International journal of biosensors & bioelectronics* 4.4 (2018), p. 195 (cit. on p. 3).

[14] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh. «Non-contact video-based vital sign monitoring using ambient light and autoregressive models». In: *Physiological measurement* 35.5 (2014), p. 807 (cit. on pp. 3, 4, 11, 14, 24, 43, 57, 59).

[15] C. Massaroni, D. Simões Lopes, D. Lo Presti, E. Schena, and S. Silvestri. «Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: A single camera approach». In: *Journal of Sensors* 2018 (2018) (cit. on pp. 3, 4, 11, 18, 43).

[16] K. Gibson et al. «Non-contact heart and respiratory rate monitoring of preterm infants based on a computer vision system: A method comparison study». In: *Pediatric research* 86.6 (2019), pp. 738–741 (cit. on pp. 3, 9, 20).

[17] J. C. Cobos-Torres, M. Abderrahim, and J. Martınez-Orgado. «Non-contact, simple neonatal monitoring by photoplethysmography». In: *Sensors* 18.12 (2018), p. 4362 (cit. on pp. 3, 4, 6, 7, 19, 22).

[18] M. Van Gastel, S. Stuijk, and G. de Haan. «Robust respiration detection from remote photoplethysmography». In: *Biomedical optics express* 7.12 (2016), pp. 4941–4957 (cit. on pp. 3, 4, 19).

[19] Q. Chen et al. «Camera-based heart rate estimation for hospitalized newborns in the presence of motion artifacts». In: *BioMedical Engineering OnLine* 20.1 (2021), pp. 1–16 (cit. on pp. 3, 4, 21, 57).

[20] S. L. Rossol et al. «Non-contact video-based neonatal respiratory monitoring». In: *Children* 7.10 (2020), p. 171 (cit. on pp. 3, 4, 20).

[21] M. Ghodratigohar, H. Ghanadian, and H. Al Osman. «A remote respiration rate measurement method for non-stationary subjects using CEEMDAN and machine learning». In: *IEEE Sensors Journal* 20.3 (2019), pp. 1400–1410 (cit. on pp. 3, 16).

[22] M. Villarroel et al. «Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit». In: *NPJ digital medicine* 2.1 (2019), pp. 1–18 (cit. on pp. 3, 4, 6, 20, 27, 41, 57).

[23] L. Maurya, P. Kaur, D. Chawla, and P. Mahapatra. «Non-contact breathing rate monitoring in newborns: A review». In: *Computers in Biology and Medicine* 132 (2021), p. 104321 (cit. on pp. 3, 4, 6).

[24] R.N. Yin et al. «Heart rate estimation based on face video under unstable illumination». In: *Applied Intelligence* 51.8 (2021), pp. 5388–5404 (cit. on pp. 4, 17, 43).

[25] G. Boccignone et al. «pyVHR: a Python framework for remote photoplethysmography». In: *PeerJ Computer Science* 8 (2022), e929 (cit. on pp. 4, 17, 18, 29).

[26] M. A. Hassan, A. S. Malik, D. Fofi, B. Karasfi, and F. Meriaudeau. «Towards health monitoring using remote heart rate measurement using digital camera: A feasibility study». In: *Measurement* 149 (2020), p. 106804 (cit. on pp. 4–6, 9).

[27] V. Selvaraju et al. «Continuous Monitoring of Vital Signs Using Cameras: A Systematic Review». In: *Sensors* 22.11 (2022), p. 4097 (cit. on pp. 4–6, 11, 12, 42).

[28] W. Verkruysse, L. O. Svaasand, and J. S. Nelson. «Remote plethysmographic imaging using ambient light.» In: *Optics express* 16.26 (2008), pp. 21434–21445 (cit. on pp. 4, 10, 13, 37).

[29] W. Wang, A. C. Den Brinker, S. Stuijk, and G. De Haan. «Algorithmic principles of remote PPG». In: *IEEE Transactions on Biomedical Engineering* 64.7 (2016), pp. 1479–1491 (cit. on pp. 4, 5, 15, 37).

[30] Y. Zhang, Z. Dong, K. Zhang, S. Shu, F. Lu, and J. Chen. «Illumination variation-resistant video-based heart rate monitoring using LAB color space». In: *Optics and Lasers in Engineering* 136 (2021), p. 106328 (cit. on pp. 4, 6, 17, 43, 60).

[31] K. Zheng, K. Ci, J. Cui, J. Kong, and J. Zhou. «Non-contact heart rate detection when face information is missing during online learning». In: *Sensors* 20.24 (2020), p. 7021 (cit. on pp. 5, 6, 11, 12).

[32] S. K. A. Prakash and C. S. Tucker. «Bounded Kalman filter method for motion-robust, non-contact heart rate estimation». In: *Biomedical optics express* 9.2 (2018), pp. 873–897 (cit. on pp. 6, 11, 12).

[33] S. Fleming et al. «Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies». In: *The Lancet* 377.9770 (2011), pp. 1011–1018 (cit. on pp. 6, 42).

[34] M. Villarroel et al. «Continuous non-contact vital sign monitoring in neonatal intensive care unit». In: *Healthcare technology letters* 1.3 (2014), pp. 87–91 (cit. on pp. 7, 19, 41).

[35] Fondazione Links. *Get to know Us.* 2021 [Online]. URL: `https://linksfoundation.com/en/get-to-know-us/` (cit. on p. 7).

[36] Intel. *Certified human: How new Intel tech detects deepfakes in real time | Intel.* Dec. 2022. URL: `https://www.youtube.com/watch?v=WYjJM49559I&t=4s` (cit. on p. 7).

[37] G. Boccignone, D. Conte, V. Cuculo, A. d'Amelio, G. Grossi, and R. Lanzarotti. «An open framework for remote-PPG methods and their assessment». In: *IEEE Access* 8 (2020), pp. 216083–216103 (cit. on pp. 9, 16, 18, 29, 31–33).

[38] D. Y. Kim, K. Lee, and C. B. Sohn. «Assessment of ROI Selection for Facial Video-Based rPPG». In: *Sensors* 21.23 (2021), p. 7923 (cit. on pp. 10, 12).

[39] J. Ryu, S. Hong, S. Liang, S. Pak, Q. Chen, and S. Yan. «A New Framework for Robust Heart Rate Measurement Based on the Head Motion State Estimation». In: *IEEE Journal of Biomedical and Health Informatics* 25.9 (2021), pp. 3428–3437 (cit. on pp. 10, 14, 15, 17, 28).

[40] F. Haugg, M. Elgendi, and C. Menon. «Effectiveness of Remote PPG Construction Methods: A Preliminary Analysis». In: *Bioengineering* 9.10 (2022), p. 485 (cit. on pp. 11, 12).

[41] N. Molinaro, E. Schena, S. Silvestri, and C. Massaroni. «Multi-ROI Spectral Approach for the Continuous Remote Cardio-Respiratory Monitoring from Mobile Device Built-In Cameras». In: *Sensors* 22.7 (2022), p. 2539 (cit. on pp. 11, 12, 18).

[42] R. Boda, M. J. Pemeena Priyadarsini, and J. Pemeena. «Face detection and tracking using KLT and Viola Jones». In: *ARPN journal of Engineering and Applied Sciences* 11.23 (2016), pp. 13472–13476 (cit. on p. 12).

[43] L. Feng, L. M. Po, X. Xu, Y. Li, and R. Ma. «Motion-resistant remote imaging photoplethysmography based on the optical properties of skin». In: *IEEE Transactions on Circuits and Systems for Video Technology* 25.5 (2014), pp. 879–891 (cit. on pp. 12, 15).

[44] MediaPipe. *MediaPipe.* 2022 [Online]. URL: `https://google.github.io/mediapipe/` (cit. on p. 12).

[45] MediaPipe. *MediaPipe Face Mesh.* 2022 [Online]. URL: `https://google.github.io/mediapipe/solutions/face_mesh.html` (cit. on p. 12).

[46] A. Belouchrani and M. G. Amin. «Blind source separation based on time-frequency signal representations». In: *IEEE Transactions on Signal Processing* 46.11 (1998), pp. 2888–2897. DOI: `10.1109/78.726803` (cit. on p. 13).

[47] M. Z. Poh, D. J. McDuff, and R. W. Picard. «Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.» In: *Optics express* 18.10 (2010), pp. 10762–10774 (cit. on pp. 13, 37).

[48] V. D. Calhoun, J. Liu, and T. Adalı. «A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data». In: *Neuroimage* 45.1 (2009), S163–S172 (cit. on p. 13).

[49] 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).* IEEE. 2011, pp. 405–410 (cit. on pp. 13, 14, 37).

[50] I. Romero. «PCA-based noise reduction in ambulatory ECGs». In: *2010 Computing in Cardiology.* 2010, pp. 677–680 (cit. on p. 13).

[51] G. De Haan and V. Jeanne. «Robust pulse rate from chrominance-based rPPG». In: *IEEE Transactions on Biomedical Engineering* 60.10 (2013), pp. 2878–2886 (cit. on pp. 14, 37).

[52] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. J. Wang. «Video-based human heart rate measurement using joint blind source separation». In: *Biomedical Signal Processing and Control* 31 (2017), pp. 309–320 (cit. on p. 14).

[53] V. Gandhi. *Brain-computer interfacing for assistive robotics: electroencephalograms, recurrent quantum neural networks, and user-centric graphical interfaces.* academic press, 2014 (cit. on p. 14).

[54] M. Chen, Q. Zhu, H. Zhang, M. Wu, and Q. Wang. «Respiratory rate estimation from face videos». In: *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI).* IEEE. 2019, pp. 1–4 (cit. on p. 14).

[55] G. De Haan and A. Van Leest. «Improved motion robustness of remote-PPG by using the blood volume pulse signature». In: *Physiological measurement* 35.9 (2014), p. 1913 (cit. on pp. 15, 37, 56).

[56] W. Wang, S. Stuijk, and G. De Haan. «A novel algorithm for remote photoplethysmography: Spatial subspace rotation». In: *IEEE transactions on biomedical engineering* 63.9 (2015), pp. 1974–1984 (cit. on p. 15).

[57] A. Mert and A. Akan. «Emotion recognition from EEG signals by using multivariate empirical mode decomposition». In: *Pattern Analysis and Applications* 21.1 (2018), pp. 81–89 (cit. on p. 16).

[58] L. Xu, J. Cheng, and X. Chen. «Illumination variation interference suppression in remote PPG using PLS and MEMD». In: *Electronics Letters* 53.4 (2017), pp. 216–218 (cit. on p. 16).

[59] H. W. Willaby, D. S. J. Costa, B. D. Burns, C. MacCann, and R. D. Roberts. «Testing complex models with small sample sizes: A historical overview and empirical demonstration of what partial least squares (PLS) can offer differential psychology». In: *Personality and Individual Differences* 84 (2015), pp. 73–78 (cit. on p. 16).

[60] C. S. Pilz, S. Zaunseder, J. Krajewski, and V. Blazek. «Local group invariance for heart rate estimation from face videos in the wild». In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018, pp. 1254–1262 (cit. on pp. 16, 37, 57).

[61] J. Bógalo, P. Poncela, and E. Senra. «Circulant singular spectrum analysis: a new automated procedure for signal extraction». In: *Signal Processing* 179 (2021), p. 107824 (cit. on p. 17).

[62] M. Elgendi, R. R. Fletcher, H. Tomar, J. Allen, R. Ward, and C. Menon. «The Striking Need for Age Diverse Pulse Oximeter Databases». In: *Frontiers in Medicine* (2021), p. 2428 (cit. on p. 17).

[63] F. Y. Sinaki et al. «Ethnic disparities in publicly-available pulse oximetry databases». In: *Communications Medicine* 2.1 (2022), pp. 1–5 (cit. on p. 18).

[64] U.S. Food and Drug Administration. *Pulse Oximeter Accuracy and Limitations: FDA Safety Communication*. Nov. 2022 [Online]. URL: `https://www.fda.gov/medical-devices/safety-communications/pulse-oximeter-accuracy-and-limitations-fda-safety-communication` (cit. on p. 18).

[65] A. Fusco, D. Locatelli, F. Onorati, G. C. Durelli, and M. D. Santambrogio. «On how to extract breathing rate from PPG signal using wearable devices». In: *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE. 2015, pp. 1–4 (cit. on p. 18).

[66] K. K. Parhi and M. Ayinala. «Low-complexity Welch power spectral density computation». In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 61.1 (2013), pp. 172–182 (cit. on p. 18).

[67] MathWorks. *Fast Fourier Transform (FFT)*. 2022 [Online]. URL: `https://it.mathworks.com/discovery/fft.html` (cit. on p. 18).

[68] M. J. Corinthios. «A fast Fourier transform for high-speed signal processing». In: *IEEE transactions on computers* 100.8 (1971), pp. 843–846 (cit. on p. 18).

[69] P. Welch. «The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms». In: *IEEE Transactions on audio and electroacoustics* 15.2 (1967), pp. 70–73 (cit. on p. 18).

[70] F. Jurado and J. R. Saenz. «Comparison between discrete STFT and wavelets for the analysis of power quality events». In: *Electric power systems research* 62.3 (2002), pp. 183–190 (cit. on p. 18).

[71] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri. «Heart rate variability: a review». In: *Medical and biological engineering and computing* 44.12 (2006), pp. 1031–1051 (cit. on p. 18).

[72] M. Singh, M. Singh, and N. Singhal. «Emotion recognition along valence axis using Naıve bayes classifier». In: *International Journal of Information Technology & Knowledge Management* 7.1 (2013), pp. 51–55 (cit. on p. 18).

[73] N. Koolen et al. «Automated Respiration Detection from Neonatal Video Data». In: *ICPRAM (2)*. 2015, pp. 164–169 (cit. on p. 19).

[74] D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli. «Respiratory rate monitoring by video processing using local motion magnification». In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE. 2018, pp. 1780–1784 (cit. on p. 20).

[75] B. Huang et al. «A neonatal dataset and benchmark for non-contact neonatal heart rate monitoring based on spatio-temporal neural networks». In: *Engineering Applications of Artificial Intelligence* 106 (2021), p. 104447 (cit. on p. 21).

[76] V. Gupta and V. K. Sharma. «Skin typing: Fitzpatrick grading and others». In: *Clinics in dermatology* 37.5 (2019), pp. 430–436 (cit. on p. 24).

[77] S. Hoffstaetter. *Tesseract OCR*. 2022 [Online]. URL: `https://github.com/tesseract-ocr/tesseract#running-tesseract` (cit. on p. 26).

[78] Python Package Index. *pytesseract 0.3.10*. 2022 [Online]. URL: `https://pypi.org/project/pytesseract/` (cit. on p. 26).

[79] E. Freeman and E. Robson. *Head First Design Patterns*. O'Reilly Media, 2020 (cit. on pp. 26, 30).

[80] V. A. Profillidis and G. N. Botzoris. «Chapter 5 - Statistical Methods for Transport Demand Modeling». In: *Modeling of Transport Demand*. Ed. by V. A. Profillidis and G. N. Botzoris. Elsevier, 2019, pp. 163–224. ISBN: 978-0-12-811513-8. DOI: `https://doi.org/10.1016/B978-0-12-811513-8.00005-4` (cit. on p. 29).

[81] F. N. David. *Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples.* Cambridge University Press, 1938 (cit. on p. 29).

[82] P. Gupta. *CUDA Refresher: The CUDA Programming Model.* 2020 [Online]. URL: `https://developer.nvidia.com/blog/cuda-refresher-cuda-programming-model/` (cit. on p. 30).

[83] H. H. Holm, A. R. Brodtkorb, and M. L. Sætra. «GPU computing with Python: Performance, energy efficiency and usability». In: *Computation* 8.1 (2020), p. 4 (cit. on p. 31).

[84] Central Statistics Office. *Mean vs Median Information Note.* 2023 [Online]. URL: `https://www.cso.ie/en/releasesandpublications/in/rrppi/meanvsmedianinformationnote/` (cit. on p. 33).

[85] H. J. Landau. «Sampling, data transmission, and the Nyquist rate». In: *Proceedings of the IEEE* 55.10 (1967), pp. 1701–1706 (cit. on p. 42).

[86] E. Orovou et al. «Correlation between Pacifier Use in Preterm Neonates and Breastfeeding in Infancy: A Systematic Review». In: *Children* 9.10 (2022), p. 1585 (cit. on p. 58).

[87] World Health Organization. *Newborns: improving survival and well-being.* 2020 [Online]. URL: `https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality` (cit. on p. 60).

[88] Mediapipe. *Canonical Face Model uv Visualization.* 2023 [Online]. URL: `https://github.com/google/mediapipe/blob/master/mediapipe/modules/face_geometry/data/canonical_face_model_uv_visualization.png` (cit. on p. 62).